Bitcoding the brain. Integration and organization of massive parallel neuronal data.

Ljungquist, Bengt

2018

Document Version:
Publisher's PDF, also known as Version of record

Bitcoding the brain

Integration and organization of massive parallel neuronal data

BENGT LJUNGQUIST
EXPERIMENTAL MEDICAL SCIENCE | FACULTY OF MEDICINE | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2018 NORDIC SWAN ECOLABEL 3041 0903

FACULTY OF MEDICINE

Department of
Experimental Medical Science

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2018:97
ISBN 978-91-7619-665-6
ISSN 1652-8220

Bitcoding the brain

Bitcoding the brain

Integration and organization of massive parallel neuronal data

by Bengt Ljungquist

LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy
Thesis advisors: Prof. Martin Garwicz, Prof. Jens Schouenborg, Assoc. Prof. Per Petersson, Assoc. Prof. Anders J. Johansson
Faculty opponent: Assoc. Prof. Rune Berg

To be presented, with the permission of the Faculty of Medicine of Lund University, for public defense in Hörsalen, Medicon Village on Thursday, the 13th of September 2018 at 09:00.

Organization
LUND UNIVERSITY

Department of Experimental Medical Science
Lund University BMC I13
SE–221 84 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2018-09-13

Sponsoring organization

Author(s)
Bengt Ljungquist

Title and subtitle
Bitcoding the brainIntegration and organization of massive parallel neuronal data

Abstract
As a result of the development in the field of Brain Machine Interfaces (BMI), a continuously increasing number of neurons in the awake brain may be recorded from simultaneously over long periods of time. This has led to major challenges in handling the amount of neuronal data from these recordings. In this thesis, we show how to address these challenges by enabling organization and integration of large amounts of data, for both direct analysis in real time and for storing for later analysis online. In paper I, we present a software architecture for organizing and integrating electrophysiological data by using an object model for the data together with a service oriented architecture for data integration. Paper II presents findings providing direct evidence that chronic recordings in primary somatosensory cortex in awake animals can offer a powerful, and much sought for, translational model of the perception of pain magnitude during hyperalgesia, superior to previously used behavioral models for evaluation of pain conditions. Principles from paper I are applied, in order to enable faster comparisons between groups of data. Paper III describes a computationally highly efficient novel data structure using bit encoding for action potentials from single neurons, together with a software architecture for handling sources of data in parallel, for example from different groups of electrodes. We show how this enables collecting and analyzing signals from millions of neurons and then possibly providing direct feedback to the recorded subject's brain in real time.

Key words
electrophysiology, neuroinformatics, scientific data management

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title
1652-8220

ISBN
978-91-7619-665-6

Recipient's notes

Number of pages
109

Price

Security classification

Bitcoding the brain

Integration and organization of massive parallel neuronal data

by Bengt Ljungquist

LUND
UNIVERSITY

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the doctoral student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration front: Bitcoded brain. Credits: Jacopo Annese, The Brain Observatory (`http://thebrainobservatory.org`) + own work.

Funding information: The thesis work was financially supported by the following grants: Swedish Research Council (Linnaeus grant - project number 60012701, project number K2010-62X-21400-01-3, project number 01013), The Knut and Alice Wallenberg Foundation (project number: KAW 2004.0119), the Medical Faculty at Lund University, Olle Engkvist Foundation, Harald Jeansson Foundation, Segerfalk Foundation

Faculty of Medicine, Department of Experimental Medical Science

ISBN: 978-91-7619-665-6 (print)
1652-8220

Printed in Sweden by Media-Tryck, Lund University, Lund 2018

To my family, for their support

Contents

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **A novel framework for storage, analysis and integration through mediation of large-scale electrophysiological data**

B. Ljungquist, P. Petersson, J. Schouenborg, A. J. Johansson, M. Garwicz
Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on (pp. 203-207)

II **Discrepancies between cortical and behavioural long-term readouts of hyperalgesia in awake freely moving rats**

B. Ljungquist[1] , **T. Jensen**[1], L. Etemadi, J. Thelin, G. Lind, M. Garwicz, P. Petersson, F. Tsanakalis, J. Schouenborg
European Journal of Pain, 20(10), 1689-1699

III **A bit-encoding based new data structure for time and memory efficient handling of spike times in an electrophysiological setup**

B. Ljungquist, P. Petersson, A. J. Johansson, J. Schouenborg[2], M. Garwicz[2]
Neuroinformatics, 16(2) pp. 217-229

[1] Equal contribution
[2] Equal contribution

Acknowledgements

The production of this thesis has been great fun and developing, but also quite challenging at times, but this is of course also natural in the process of charting and exploring what has not been done before. Even though I am proud of my work and for achieving the very goal I set out when starting my thesis, to be able to handle large volumes of neuronal data, without the support of others, there would be no thesis.

I would especially like to thank my main supervisor **Martin Garwicz**, your keen interest in and grasp of all science, has fitted the interdisciplinary nature of the thesis and my work very well. The feedback I have received from you has been invaluable, especially concerning the art of writing manuscripts, and written communication in general, which you master fully, and I have learned a lot from you. In addition to your natural scientific skills, you are a true humanist, who have supported me well also during some of the more difficult parts of the work and my life in general, for this you have my warmest thanks. It has been an honor to work with you as my supervisor.

Jens Schouenborg, my co-supervisor, although you have been close to a main supervisor in engaging and providing feedback to my work, especially during the final and critical phase of the thesis. As I stepped into your office the first time discussing to become a graduate student, I felt that you are fitting in every respect the image of the accomplished scientist, in being sharp-minded and... uniquely different, in a positive sense. I also during my thesis learnt that your intellectual skills are matched with a good heart, caring for me and others around you in different respects, which I deeply appreciate and consider myself fortunate to have carried out my thesis at your research center.

Per Petersson, as my co-supervisor, you have provided inspiration in interdisciplinary matters, and also valuable feedback and ideas in development of my research and skills.

Anders J Johansson, as my co-supervisor, I appreciate your sound engineering mind, you have brought the practicalities into consideration in many of my research undertakings.

I thank you all current and former staff at or around the lab, both those whom I got to know at BMC as well as those moving on to Medicon village, for interesting discussions both research related as well as generally in the lunch room and at other occassions: **Lina Petersson**, **Jonas Thelin**, **Cecilia Eriksson Linsmeier**, **Lucas Kumosa**, **Petter Pettersson**, **Ali**

Ghasemi Azar, **Mengliang Zhang**, **Niclas Lindqvist**, **Ulrike Richter**, **Pär Halje**, **Dmitry Suyatin**, **Jörgen Hager**, **Hjalmar Bjartmarz**, **Hiroyuki Watanabe**, **Lars Wallman**, **Marcus Granmo**, **Dan-Anders Jirenhed**, **Henrik Jörntell**, **Germund Hesslow**, **Fredrik Bengtsson**, **Andrea Nord**, **Suzanne Rosander-Jönsson**, **Peter Paulander**, **Alexandra Waldenström**. Special thanks to: **Linda Eliasson** and **Agneta SanMartin**, I deeply appreciate your kind and keen support in all administrative matters and otherwise as well, thank you for your efforts in keeping order in this sometimes unruly world of academia. **Nils Danielsen**, always with the ethical mind set you are in that respect a role model who have supported me well, I appreciate our discussions and your insights into the inner workings and rules of the university. **Palmi Thor Thorbergsson**, for your input and discussions, ailments, ransomware and spike sorting alike. My warmest memories and thoughts also to **Lars-Åke Clementz**, I appreciated our discussions and your practical viewpoints on research as well as administrative issues.

Thank you also to all fellow past and present PhD Students for sharing knowledge, insights and burdens: **Veronica Johansson**, **Gustav Lind**, **Tanja Jensen**, **Per Köhler**, **Lina Gällentoft**, **Leila Etemadi**, **Mohsin Mohammed**, **Johan Agorelius**, **Matilde Forni**, **Joel Sjöbom**, **Nedjeljka Ivica**, **Martin Tamté**, **Alexander Holmqvist**.

My special thanks to **Pomesh Kumar**, for much appreciated discussions on Python and research in general, as well as your interest in and feedback to my work, this thesis included. I wish you the best of progress in years to come.

However, support is not only about research, in order to keep my mind sharp and still its fluctuations, yoga has been of much importance. To my first yoga teachers in Ashtanga Yoga, **Susanna Finocchi** and **Jens Bache**, thank you, I will always cherish the time that I spent building a foundation in my practice with you in Copenhagen, also well before starting the undertaking of this thesis. To **Ulrica Norberg**, my warmest thanks for showing me the path when I needed it. To **Alan Finger**, Guruji and Yogiraj, my gratitude for your teachings, which I look forward to follow and study further. Thank you, **Sarah Finger**, graceful in life as well as in your practice, you are a role model for me and many others. Also thank you all of my friends, yogi or non-yogi alike, with whom I have shared the joys and burdens of research and life in general. Especially, thank you **Andreas Brock**, long-time friend, at all times encouraging me to do my best.

Last, but not least, I wish to thank my family for your support. To my parents **Siv** and **Hans**, thank you for all the ways you have supported me during my

years and encouraged me to endure, also in my research. To my brother **Mats** and your partner **Disa**, thank you for our discussions in all matters, and for believing in my undertakings (at least my research...). To my extended family, **Mikulas**, **Jaroslava**, **Michal** and **Maria**, I am happy to be part of your family, thank you for welcoming me, as well as you are part of mine. To my daughter **Julia**, during the work of this thesis, you have been growing up to a wonderful daughter, although you have endured the ups and downs of your parents undertakings. I am proud to be your father. To my daughter **Emma**, your playful spirit and energy inspires me everyday and provides much happiness to me and and the rest of your family, I am happy to be your father. To my little son **Benjamin**, who already though is fast growing, your light and gentle mind is already bringing us much comfort and bliss, I am joyful to be your father. And my dear **Monika**, love of my life, without you there would be nothing at all, thank you for everything. I am fortunate to be your husband, I look forward to share our lives together.

Popular summary in English

Development in neuroscience places increasing demands on management of collected data, especially when interacting directly with single cells (neurons) in the brains of awake behaving subjects, using so called Brain Machine Interfaces, i.e. electrodes that directly connect the brain with a computer. The number of neurons that may be recorded from simultaneously are continuously increasing, as development in the field has made electrodes smaller, more efficient, and generally better suited to stay longer periods inside the brain. This has led to major challenges in handling the amount of information generated when listening to and communicating with a large number of neurons, which is important both for basic research, clinical diagnostics and treatment. If, for example, controlling a prosthetic device using signals from the brain, or detecting an epileptic seizure for possible intervention, it is important to quickly handle and interpret large amounts of data from many single neurons at the same time. For later analysis, it is also important that data stays organized so that a proper diagnosis or interpretation of the data may be performed after the recording session.

In this thesis, we show how to address these issues by enabling organization and integration of large amounts of data, for both direct analysis in real time and for storing for later analysis offline. In paper I, we present a software architecture for organizing and integrating electrophysiological data by using a data model together with a software architecture for data transfer. Paper II presents findings providing direct evidence that chronic recordings in primary somatosensory cortex in awake animals can offer a powerful, and much sought for, model of the perception of pain magnitude during hyperalgesia, a condition of increased sense of pain. This model could be applied in both research and the clinic. The model is superior to previously used models of animal behavior for evaluation of pain conditions. Principles from paper I are applied, in order to enable faster comparisons between groups of data. Paper III describes a computationally highly efficient novel data structure using binary numbers for encoding action potentials from single neurons, together with a software architecture for handling sources of data in parallel, for example from different groups of electrodes. We show how this enables collecting and analyzing signals from millions of neurons and then possibly providing direct feedback to the recorded subject's brain in real time.

Populärvetenskaplig sammanfattning på svenska

Utvecklingen inom neurovetenskap ställer allt högre krav på hantering av insamlade data, speciellt när man interagerar direkt med enskilda nervceller i den vakna hjärnan under långa tidsperioder med hjälp av så kallade Brain Machine Interfaces, elektroder som direkt kopplar hjärnan till en dator. Antalet nervceller som kan spelas in samtidigt ökar kontinuerligt, eftersom utvecklingen inom forskningsfältet har gjort både elektroderna mindre, effektivare och i allmänhet bättre lämpade att sitta kvar i hjärnan längre perioder. Detta har lett till stora utmaningar vid hanteringen av informationen som genereras när man lyssnar på och kommunicerar med ett stort antal nervceller, vilket är viktigt både i grundforskning, klinisk diagnostik och behandling. Om man till exempel styr en protes med hjälp av signaler från hjärnan, eller upptäcker ett epileptiskt anfall för eventuell intervention, så är det viktigt att snabbt hantera och tolka stora mängder data från enskilda nervceller. För senare analys är det också viktigt att data förblir organiserade så att en korrekt diagnos eller tolkning från data kan utföras efter inspelningstillfället.

I denna avhandling visar vi hur man hanterar dessa problem genom att möjliggöra organisation och integration av stora mängder data, både för direkt analys i realtid och för lagring för senare analys offline. I artikel I presenterar vi en mjukvaruarkitektur för att organisera och överföra elektrofysiologiska data genom att använda en datamodell tillsammans med en programvara för att integrera data. I artikel II presenteras resultat som ger direkta bevis på att kroniska inspelningar i primära somatosensoriska cortex hos vakna försöksdjur kan erbjuda en kraftfull och mycket eftersökt modell av smärtuppfattningen vid hyperalgesi, ett tillstånd av ökad smärtuppfattning. Denna modelll kan appliceras inom forskning och sjukvård för att diagnosticera smärta. Det visas också hur denna är bättre än tidigare använda modeller av djurs beteende för utvärdering av smärtstillstånd. I detta arbete används principer från artikel I för att möjliggöra snabbare jämförelser mellan grupper av data. I artikel III beskrivs en beräkningseffektiv datastruktur med användning av binära tal för kodning av aktionspotentialer, signaler från enskilda nervceller, tillsammans med en mjukvaruarkitektur för parallel hantering av data, t ex från olika grupper av elektroder. Vi visar i artikeln hur dessa ger möjlighet att samla in, analysera och eventuellt ge direkt återkoppling till hjärnan för miljontals nervceller i realtid.

Bitcoding the brain – Integration and organization of massive parallel neuronal data

Pauca, sed matura - few, but ripe
— Carl Friedrich Gauss

1 Introduction

The last decades have seen extensive development of neuroscience in general and particularly in the field of brain machine interfaces. Connecting to, interpreting and interacting with the activity and signaling of the central nervous system (CNS) has led to increasing requirements of handling the data generated by these interfaces, both in terms of data size and complexity. Furthermore, when interpreting or using neural data in order to control other devices, managing data in real time has also becomes highly relevant, in order to e.g. respond in time to make a feedback loop to stimulate the nervous system or to correlate the presented data to an observed behavior. Through an iterative approach, first addressing offline data management and applying it in an experimental electrophysiological setting, and then addressing online data management, this thesis presents a neuroinformatics architecture that can meet current and foreseeable future needs to integrate, organize and interact with large volumes of electrophysiological data, including signals derived from millions of neurons in real time.

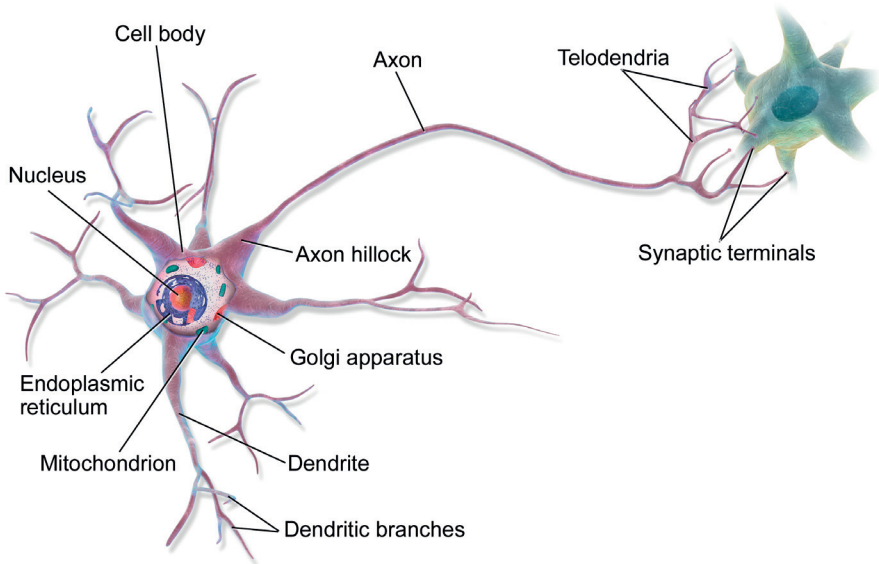

Figure 1: Structure of multipolar neuron. Source: Wikipedia, Author BruceBlaus.

1.1 Neuronal information processing

The human brain is composed of approximately 86 billion neurons, each of which has a computational capacity, and thus forms the basic informational processing unit of the brain. Each neuron is connected to up to 10 000 or more other neurons, and these connections may be formed or removed after learning in a daunting number of ways. If connections are changed just a little bit, either in strength or physical location, the computational results of an input will be different. Thus, the computational complexity of the brain is immense. Yet, using modern investigative methods we may still uncover many aspects of its function, both as a whole, and at the level of single neurons.

1.1.1 The neuron

A neuron is a specialized type of cell within the nervous system that receives chemical signals and converts them to electrical signals. These are then integrated and transmitted as an action potential across the cell membrane, and finally converted back to chemical signals. The neuron is composed of different parts: *dendrites*, thin appendages that receive the majority of input from other cells, *soma* - cell body containing the cell nucleus as well as other organelles like

mitochondria and endoplasmic reticulum, *axon* - a longer projection conducting the electric signal, see Figure 1. The chemical signaling occurs between cells at the synapses, while the electric signaling mostly occurs between different parts of the cell. In addition, some cells also communicate with other cells through electrical synapses using so called gap-junctions.

1.1.2 Chemical and electrical signaling

As a substrate for the chemical signaling, the synaptic axon terminal of one cell forms vesicles, separate spheres of cell membrane containing neurotransmitters. Chemical signaling then takes place when the axon terminal is electrically depolarized. As a result of this depolarization, vesicles dock with the cell membrane at the synaptic cleft and release their neurotransmitter substances. This process is the link between the electrical and chemical signaling. Depending on the neurotransmitter(s) released, it may act either on excitatory or inhibitory receptors on the target neuron, resulting in depolarization or hyperpolarization of the membrane potential. These receptors will open or close ion channels that in turn change the membrane potential. This change then propagates along the membrane and if sufficiently strong at the soma, it may initiate an action potential. Usually many excitatory synapses have to cooperate to elicit an action potential, see Figure 2. The action potential was first described mathematically in the work of Hodgkin and Huxley (Hodgkin and Huxley, 1952), in one of the most influential computational biophysical models of the 20th century. The action potential is also commonly called a “spike”, due to the sharp rise in voltage during the voltage-gated channel caused depolarization. In the inhibitory synapse, the hyper-/repolarization may block other incoming signals as well as decrease the spontaneous firing of the target neuron. Action potential generation, also called ”spiking”, signifies one of the major indicators of activity in the central nervous system, as it directly correlates to the activity of cells.

Other than neurons, the central nervous system also contains glial cells. These may also take part in the signaling to some extent, as they house receptors that can cause membrane potential changes in the glial cells. However, glial cells lack voltage-gated ion channels, which are specific to neurons.

1.2 Brain information processing

The neurons of the central nervous system (CNS) together with the glial cells form the brain and spinal cord. According to the classical view of CNS organiz-

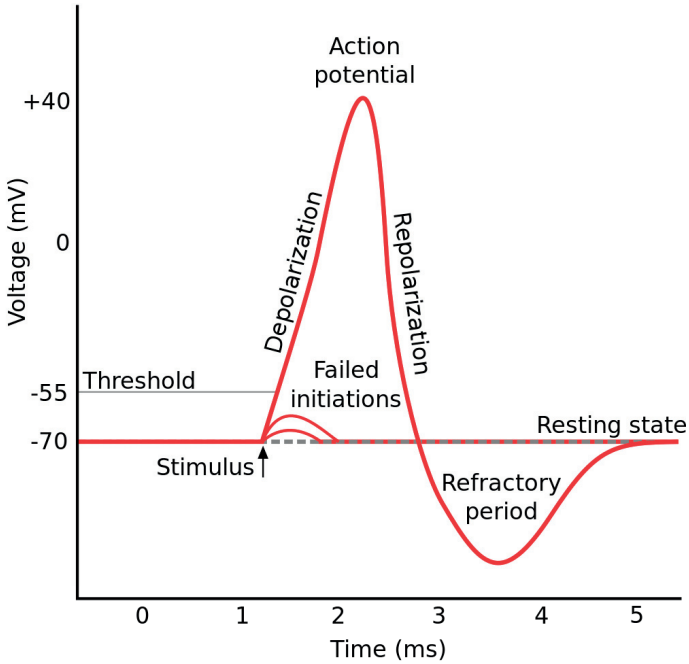

Figure 2: Schematic of an action potential. Activated post-synaptic receptors cause a general dendritic depolarization, which then spreads to the axon hillock at the soma, where it causes voltage-gated sodium channels, which reside in greater numbers there, to open and increase the depolarization further. This in turn results in a sharp rise in potential which then rapidly declines. The depolarization also causes voltage-gated sodium channels in the vicinity along the axon to open, resulting in the action potential propagating along the membrane. Source: (Wikipedia, 2017)

ation and function, the brain and spinal cord typically interact with the environment through the peripheral nervous system to handle the following tasks:

- **Receive information** from the environment through sensory organs and the peripheral nervous system (PNS). Sensory organs are specialized nerve cells, detecting changes in the environment or within the body. These specializations take quite different forms, for example free nerve endings detecting nociception, defined as a sensation that signals harmful or potentially harmful stimuli (Sherrington, 1923). From the sensors, information is conveyed to the CNS by the PNS. In the CNS, this information may be used directly in reflexes, e.g. activating a muscle through a motor neuron or/and be passed on to higher structures in the CNS, such as primary

sensory cortical areas.

- **Integrate** the received **information** by comparing it to current state and memory of past events. Depending on which types of memories that are involved, different parts of the CNS are involved in the storage and retrieval of memory; episodic and spatial memory typically involves hippocampus (Burgess et al., 2002; Moser et al., 2015), while emotional memories (especially fear) involve the amygdala (LeDoux, 2003), and procedural memory involves striatum of the basal ganglia (Barnes et al., 2005). The cerebellum is maybe one of the most striking examples of a feed-forward mechanism within the central nervous system where sensory inputs are used to, among other functions, fine-tune motor function. Another important regulative area of the CNS is the hypothalamus, which regulates autonomic and hormonal response to keep homeostasis, for example temperature regulation. Due to the hypothalamic connections with for example amygdala, this involves memories of past events (LeDoux, 2003). Also, the many different modalities of sensory input, like vision and somatosensory input needs to be integrated into a common experience, which typically take place in association areas like for example parietal association cortex.

- **Coordinate a response**. Many observed functional relationships between the areas and cells initiating and governing higher functions are hierarchical in nature, for example in motor control coding for movement direction of limb (Georgopoulos et al., 1986), or the forward feedback of the cerebellum and the ability of the basal ganglia to select an appropriate response, given a certain motor and emotional state.

In order to understand information processing in the brain, for example processing underlying pain, we need to monitor the neuronal signaling and interact with it through investigative methods such as electrophysiology to elucidate for example the relationship between neuronal activity in the brain or spinal cord and behavior. This interaction may be bidirectional (O'Doherty et al., 2011). The process of coupling the sensory and motor systems and thus bridging the input information with the coordination of the motor response, is often referred to as sensorimotor integration. Its importance is seen when executing a certain motor behavior in response to specific demands of the environment, mediated by the central nervous system at different levels (Machado et al., 2010).

1.3 Brain machine interfaces

The information flow of the brain may be investigated using a Brain Machine Interface (BMI), typically consisting of a set of electrodes that can measure the electrophysiological activity of neurons. The related, but more specific term Brain Computer Interface (BCI), refers to interfaces between nervous tissue and an external device (a computer) allowing for bidirectional communication. Signals can then be extracted from the brain to control a device or the brain may be subjected to stimulation in order to induce neural activity. In therapeutic settings, the latter may be performed by adapting deep brain stimulation (DBS) to simultaneously recorded field potentials in patients with Parkinson's disease (Little et al., 2013). Although BMI is a more general term, including neuroprosthetic devices to compensate for loss of function, the terms are often used interchangeably. BMI in the more general sense refers to various kinds of interfaces that connect to the brain, even chemical interfaces interacting through detection and/or release of transmitter substances or a combination of chemical and electrical interfaces (Bellin et al., 2014).

Electrophysiology concerns the measurement of physiological processes of the body using electrodes, which thus could be used as a BMI. For neurophysiological purposes electrodes may be either non-invasive, placed on the scalp to record electrical activity (EEG, Electro-Encephalogram), or invasive (Lebedev and Nicolelis, 2006), inserting electrodes into brain or spinal cord.

EEG has been available since the 1920s when it was introduced by Hans Berger (Berger, 1931). Since then, EEG has been used in clinical settings for successfully diagnosing different neurological diseases such as epilepsy and sleep disorders. It has also been used for direct neural signal processing in some BMI applications, where it has had limited success (Farwell and Donchin, 1988; Spiga et al., 2006; Sutter, 1992; Wolpaw et al., 1991; Kennedy et al., 2000; Hinterberger et al., 2003). Although its non-invasive nature makes it attractive due to the limited risk of complications, the main problem with EEG is the lack of a fine grade of activity measure, especially for single neurons; the EEG signal typically reflects activity in an underlying volume of a couple of cubic centimeters, in which millions or even billions of neurons may be located.

Microelectrodes placed inside the CNS, on the other hand, are able to capture single neurons, but lack the overview that EEG provides, see Table 1. In order to record action potentials, electrodes must have the suitable properties, for example with regard to impedance for relevant frequencies (usually 1 kHz (Williams et al., 2007).

Table 1: Time and space resolution for various BMIs. (fMRI=functional Magnetic Resonance Imaging, MEG = Magneto EncephaloGram, VSD= Voltage Sensitive Dyes, EEG = Electro Encephalo Gram, ECoG = Electro Cortico Gram, MEA = Multi Electrode Array).

BMI	Maximum sampling rate	Spatial resolution	Scale
fMRI	10 Hz	1 mm3	Whole brain
VSD	500 Hz	Single cells	Cortex area
MEG	1 kHz	2-3 cm	Whole cortex
EEG	100 Hz	2-3 cm	Whole cortex
ECoG	1 kHz	1 cm	Cortex area
Extracellular electrode	32 kHz	0,01-1 mm	Invidual neurons
Patch Clamp	32 kHz	1 μm	Parts of neuron

Correlating single unit recordings to observed behavior started to some extent through the pioneering work of Fetz and coworkers in 1969 (Fetz, 1969), who showed, by using a reinforcement protocol, that a monkey could learn to control the firing of a single neuron in its own brain. Since then, the field has developed, for instance by Georgopolous studies in the 80s (Georgopoulos et al., 1982, 1986, 1989) that demonstrated how a population vector of neurons together encoded the movement directions of an arm. During the last couple of decades the field has grown with many breakthroughs, including for instance reconstruction of more complex movement, such as following a spiral trajectory, from a neuronal population coding (Schwartz, 1994), the general feasibility of real time control of robotic arms using multi-electrode-array (MEA) recordings of cortical neural activity (Wessberg et al., 2000), and, more recently, a robotic device allowing advanced arm and hand movements has been successfully implemented in tetraplegic subjects (Velliste et al., 2008; Hochberg et al., 2012; Gilja et al., 2012; Collinger et al., 2013). This development has depended partly on the identification of important principles of motor control, revealed by neurophysiological investigations of neural activity in awake, behaving animals (Monfils et al., 2005), and partly on advances within the field of robotics (Velliste et al., 2008). The development of BMI has led to many clinical uses; they have successfully been used to alleviate symptoms of Parkinson's disease (Benabid, 2003) and treatment-resistant depression (Mayberg et al., 2005). They have also been used as a tool for paralyzed and tetraplegic (Hochberg et al., 2006) patients to communicate. However, their most fundamental scientific benefit may be the valuable information they provide about how the brain is working, tapping into its live information flow.

1.3.1 Challenges of invasive BMI

When electrodes are implanted invasively, an acute inflammatory tissue response quickly occurs due to the injury caused. This acute reaction often subsides within 1-2 weeks, after which the implants gets encapsulated These responses are mediated to a large extent by different types of glia cells, which are activated by even the slightest pathological change in the CNS, e.g. in inflammation and trauma (Schouenborg, 2016; Kreutzberg, 1996). By interacting with each other through inter-cellular signaling, they form a network of cells within the CNS clearing debris and guiding tissue repair. Their specific response to trauma depends on many factors, e.g. electrode material, insertion method and anchoring method. The Utah Array (Maynard et al., 1997) was one of the first MEAs to be adapted, later also for use in humans (Hochberg et al., 2006), however it affects neural tissue severely (Polikov et al., 2005). The resulting tissue response has been shown to typically increase impedance over time after implantation, peaking at around 7 days (Williams et al., 2007), and then diminishing. Given that the tissue responses not only results in substantial loss of neurons nearby the implanted electrodes but also the isolation of electrodes from the neurons, much efforts are currently focusing on how to reduce them to a minimum. From systematic studies in our and others laboratories it is known that flexible materials elicit less tissue response than rigid (Köhler et al., 2015), that electrode size matters, larger electrodes (200 μm) induce larger tissue reactions than smaller (50 μm) (Thelin et al., 2011) and that the specific weight of the electrodes should be close to that of the tissue (Lind et al., 2013). By using highly biocompatible embedding gelatin materials tissue responses can be further reduced (Lind et al., 2010; Köhler et al., 2015). It can therefore be anticipated that these can be brought to a minimum in the future.

Other than biocompatibility, what is currently holding back the recording of thousands of neurons? In addition to the general complexity of the number of neurons and possible recording sites, there are a number of issues that need to be addressed as they severely limit the information yield. A crucial issue concerns that electrodes may move in the tissue, resulting in changes in the shape of the action potential. A recorded potential will typically have different shapes at different positions in 3D space (Einevoll et al., 2013). For this purpose, methods for tracking spikes over time using various measures have been developed as a workaround (Bar-Hillel et al., 2006; Dickey et al., 2009; Fraser and Schwartz, 2012). In addition, methods to anchor electrodes in the tissue have also been devised with promising results (Agorelius et al., 2015). If perfectly biocompatible electrodes which can be stably positioned in the neural circuits existed, we could find ourselves in a situation where the information yield would increase

dramatically, resulting in an explosion of neuronal data that must be organized properly. For example, we will need to find solutions to store and integrate data from different sources, ranging from individual electrodes to groups of MEAs with many thousands of electrodes.

1.4 Neuroinformatics

As outlined above, MEA recordings provide the possibility of recording the activity in a large number of neurons over long periods of time, resulting in a daunting amount of data that need to be integrated and organized. Moreover, the large amount of data by no means provides an automatic understanding of what the activity signifies or how it relates to different aspects of behavior. To address these two aspects, one needs *neuroinformatics*, an interdisciplinary field employing methods and approaches from computer science, information systems, and integrative biology to identify, analyze, digest, simulate, and compute neuroscience data (Polavaram and Ascoli, 2015). In such a role, neuroinformatics support neurophysiological research by enabling data to be organized and analyzed. If data is to be truly interacted with in real time, i.e. feed it back to the subject for stimulation, functions of the nervous system need to be mirrored in the neuroinformatic system. This means that data is received, stored, compared with previous feedback, and a response to the recording subject is coordinated. There is thus a need to clarify how this may be achieved. Specifically, we need to refine how to efficiently store, integrate, analyze and share electrophysiological data from different sources.

1.4.1 Electrophysiological BMI signal types

As one of the most basic forms of data from a BMI, the activity deemed to come from one single neuron is labeled single unit activity (SUA). In order to separate it from unsorted multi-unit activity (MUA) from many cells, and SUA of other cells, the thresholded waveforms are sorted according to some criteria, e.g. waveform shape and amplitude, in a process referred to as spike sorting. The standard way is to calculate principal components for the waveforms, so called principal component analysis (PCA), project the waveforms onto a "feature space" comprising the first three principal components, and perform clustering in this feature space (Lewicki, 1998). This clustering may be performed manually, using spike sorting tools, or automatically, using different clustering algorithms. The result is discretized as a series of timestamps (each indicating occurrence of one spike) with corresponding waveforms. Once this has been per-

formed, classification is usually a very fast operation (typically involving a linear transformation) and may be performed in real time. A fundamental requirement to be able to identify spikes from a single unit and thereby to be able to monitor the single unit activity (SUA) is that time between consecutive spikes does not violate the refractory period of a single neuron; a neuron is limited by the nature of the action potential generation to fire too often (the sodium channels become inactive for a certain time, typically slightly more than 1 ms). This may be used for judging the quality of a spike sorting; if the percentage of violations is greater than a specific threshold ratio (typically chosen to be 0.1-0.2%), the identified "unit" is in fact probably composed of more than one unit. Multi-Unit Activity, MUA, is collected from the higher spectral range of the recorded signal, typically 600-6000 Hz. The filtered signal is then thresholded, i.e. all waveforms that cross a voltage level, are considered to be the MUA. The threshold is usually chosen to be significantly higher than the channel noise (four standard deviations is usually considered reasonable (Quiroga et al., 2004)). The time where each waveform crosses the threshold (in one direction) is then defined to be the timestamp of that MUA waveform, resulting in a series of timestamps. As a complement to unit activity, local field potentials, LFP may be measured. They are collected from the lower spectral range of the recorded signal, i.e. 1-300 Hz, in contrast to the unit activity. It is widely assumed that they reflect dendritic synaptic - i.e. input - activity to a number of closely situated neurons. This in turn reflects the activity of local neurons in vicinity of the electrode. As the dendritic input is received in one end of the neuron, the neuron becomes a dipole (Nunez and Srinivasan, 2006). The local field potential could be considered the sum of these dipole fields for many neurons (Einevoll et al., 2013). If they are signaling in synchrony, this will thus be reflected as rhythmically occurring LFP. Furthermore, this synchrony may occur between different parts of the brain, reflecting functional connectivity between them (Engel et al., 2001). Although LFPs are considered to contain less information than unit recordings they provide some complementary information (Bansal et al., 2012).

1.4.2 The electrophysiological information flow

These different types of electrophysiological data (SUA, MUA and LFPs) typically have the following information flow:

- **Subject** - The data is recorded as a potential at the electrode. Depending on impedance, it might pick up different qualities of the electrophysiological signals generated by the neurons. Separate stimulation electrodes may also be utilized for providing feedback to the subject.

- **Acquisition system** - integrates data from one or more headstages (which handle amplification and analog filtering of the signal) and performs digital filtering, typically dividing it into the lower frequency LFP content and the higher frequency part containing spiking data. The spiking data is then thresholded and eventually sorted resulting in MUA and SUA (see above for more details). The acquisition system may also receive requests to initiate stimulation protocols from the neuroinformatics system.

- **Local neuroinformatics system** - integrates data from one or more acquisition systems into a data storage. It also analyses data and eventually sends requests to any of the connected acquisition systems for initiation of stimulation protocols for the recording subject.

- **Remote neuroinformatics system** - integrates data from one or more local acquisition systems into a data storage. This typically concerns offline analysis, such as post-experiment data exploration, but could also concern real time integration.

1.4.3 Data integration - transfer and sharing

In science in general, for example astronomy and physics, much of the progress is due to the ability to share recorded data between scientists. Yet, the importance of data sharing in neurophysiological research is less clear. Nevertheless, the past decade has seen increasingly many initiatives to enable sharing of electrophysiology data to a remote neuroinformatics system. However, there are challenges to convey both the data itself and metadata such as the experimental setting to other research groups. Metadata may be different and unique to an experimental setting (Zehl et al., 2016), data may be saved in arbitrary file formats, which are not directly compatible with each other, and a means to transfer the data in a structured way to collaborators may be lacking. This has been addressed in different ways:

- **Central data repositories** - the user submits data to a central data repository, hosted by a server providing upload capability and a reserved physical data space, in which one may choose to share the data with other neuroscientists. To date, the CARMEN project has been one of the more successful projects in this type of integration, with data from over 100 neuroscientists (Smith et al., 2007). There are also other initiatives, but to date they have not attracted as many users to submit data to their repositories (Gardner, 2004)

- **Defining common languages for exchanging data**, for example defining Extensible Markup Language (XML, a markup language for data exchange in a human readable format) schema definitions, such as Genie (Gardner et al., 2002)
- **Defining common storage formats**, such as database schemas, thereby achieving a common base of information (Gardner, 2004; Garcia and Fourcad-Trocmé, 2009)

Because of the challenges involved, most data sharing occurs within research centers, as experimental settings and electrode types differ too much between research centers for data sharing to be successful. At a research center, where there may also be acquisition systems from different vendors, there is a need for an integrated, but localized analysis tools framework, rendering the central data repository approach unfeasible. Furthermore, many research centers develop new analysis tools, requiring access to the data through high-speed local networks. Some of these tools act on real-time data and may in a Brain Machine Interface (BMI) application provide control of an artificial device like a robotic prosthesis based on analysis results of this data; a procedure which further renders the batch oriented federated approach impractical for some applications, even though a portal may serve as a valuable repository for data.

1.4.4 Object models

Electrophysiology data has a certain hierarchical structure; an experiment is composed of one or more recordings, each of which is composed of data from one or more electrodes, forming a structured data model. Object oriented programming (OOP) refers to the concept of structuring data into objects, which may contain data in form of fields as well as methods with code to handle the data. Early efforts include introduction of specific objected oriented programming languages such as Simula (Dahl and Nygaard, 1966) and Smalltalk (Goldberg and Robson, 1983). In the early 90s, the "Three Amigos", Ivar Jacobson, Grady Booch and James Rumbaugh, joined efforts as part of Rational Software in order to develop a unified approach for object oriented analysis and design (OOAD), commonly referred to as the highly influential Rational Unified Process (Rational Software, 2011), which provides a structured approach for constructing an object model from higher level requirements. Design patterns (Gamma, 1995) refers to the concept of reusing specific ways for this construction, i.e. to structure software object models and relationships in order to solve a commonly recurring design problem. To store objects persistently, the most

used technique has been relational database management systems (RDBMS). Yet, this is not without problems, since the object model is not directly translatable to the RDBMS entity relationship (ER) model (Codd and F. 1970), as the latter lacks corresponding ways to model for example inheritance and methods. In electrophysiology, there have been some efforts to design objects models with persistence, such as Neo for Python (Garcia et al., 2014).

1.4.5 Lack of common data formats

Although data may be well structured within a single recording, there may exist many different data formats in a project, for example if using equipment from different manufacturers, which then have to be translated between each other in order to allow comparison of data. Currently, many of the manufacturers have developed their own data formats. These formats fulfill current requirements in electrophysiological research, for which they have been designed. For tomorrow's research and potential clinical applications with as much as millions of neurons simultaneously recorded, mechanisms for integrating this data across different channels have not yet been implemented; the user has to resort to individual analyses per channel.

In order to stimulate data sharing, there have been some initiatives defining a common data format (Teeters et al., 2015; Smith et al., 2007), although none to date have found neither commercial nor academic traction. Some of the more successful and more widely used data models have applied principles of adapting data such as Neuroshare (Bergel et al., 2011) and Neo (Garcia et al., 2014).These can handle the offline translation between different formats, but do not work online.

1.4.6 Organization for faster access

After having been successfully stored in a data structure, data must be analyzed. This includes algorithmic interpretation of the data to detect signal patterns and their correlations between each other and to observed physiological phenomena. However, this task is difficult, especially in the setting with many electrodes. In order to study spectral content, frequency analysis using Fourier transforms has been of great help. Most modern EEG analysis software often house these operations, however not always in real time. For example, EEGlab (Delorme and Makeig, 2004), Chronux (Bokil et al., 2010), Fieldtrip (Oostenveld et al., 2011) and OpenElectrophy (Garcia and Fourcad-Trocmé, 2009) enables such

analysis, of which the three latter deals also specifically with extracellular electrophysiological recordings. Commercially there are many different tools for offline analysis, such as for example Neuroexplorer. Most of these offline tools do quite well with fairly large number of electrodes, but most are not constructed for fast analysis of larger datavolumes ($>$ 10 000 electrodes), and cannot integrate from more than one system at a time.

1.4.7 Closing the data loop - real time interaction

For success in rapid analysis, data has to be structured to facilitate this process both during and after recordings. When considering large scale electrophysiological recordings, some type of compression or data reduction is needed to interact with the data in real time. In computer science, real time computing, RTC, refers to the concept that hardware and/or software is subject to time constraints for the computer responding to an input (Ben-Ari, 2006). The exact time depends on the application domain, but is usually within the range of a fraction of a second.

Electrophysiological real time analysis is often divided into open-loop and closed loop. Open-loop refers to recording from a subject, and either analyzing in real time or offline, but not using the received information for feedback back to the subject. The definition of real time varies, but is generally expected to be within milliseconds of the recording instant. The results from the analysis in an open-loop setting are however not relayed back to the subject but may be presented to the experimenter directly. In closed loop systems, a stimulation pattern is fed back to the subject based on the result of the analysis of the recorded signal. Most of the closed loop setups that have been implemented to date are for single cell patch-clamp recordings such as LCG (Linaro et al., 2015) and RELACS (Benda et al., 2007). However, there is promising progress into real time closed-loop initiatives (Siegle et al., 2017), which currently cover MEA as input and then providing feedback to subject, including claims of hardware in development capable of closed-loop interactions with around 1000 channels or possibly slightly more. However, if recording from a larger number of neurons, this may not be sufficient; development of electrodes will allow for both increased number of simultaneous electrodes (Schwarz et al., 2014) as well as more stable recordings. This will result in better yield in terms of mean number of neurons recorded per electrode (Agorelius et al., 2015).

In line with these developments DARPA (U.S. Defence Advanced Research Projects Agency) issued a program in Jan 2016 called NESD (Neural Engineering System Design, see URL:

https://www.darpa.mil/program/neural-engineering-system-design). This program aims to develop an implantable neural interface able to provide advanced signal resolution and data-transfer bandwidth between the brain and electronics. Specifically, the program aims to develop a BMI with over one million neurons. Apart from the challenges in electrode development, as noted above, the main neuroinformatics challenges of such an undertaking are mainly; 1) the size of the data is difficult to handle, 2) data needs to be integrated so that analysis across all recorded neurons may be performed and 3) current systems may perform the recording task in parallel, but lack a mechanism to integrate to a common memory area when recording from larger number of neurons.

Recording from a single electrode is usually performed with a 32 kHz sampling rate, resulting in a data acquisition rate of 32kB/s per electrode. With two neurons in mean per electrode assumed (a rather optimistic assumption, but with quality electrodes, it may be possible), this results in 16 kB/s per neuron. The amount of data required to record one million neurons would be 16 GB/s. Recording for 24 hours would result in 3600 * 24 * 16 GB = 1 382 400 GB or 1 382 TB if storing the raw data as it is. With today's storage technologies, and also for a foreseeable near future, the storage of a single 24 h recording would thus be unfeasible unless direct local access to large data centers is available. Clearly, some type of compression and carefully considered data organization is needed.

2 Purpose

The purpose of this thesis is to develop a neuroinformatics architecture and data encoding in order to meet current and foreseeable future needs to integrate and organize data from millions of neurons in a way that enables real time interaction.

3 Aims

- Accomplish a concept for integration and organization of data from different data sources
- Validate the concept in an authentic neurophysiological setting
- Accomplish efficient storing of data from one million neurons or more in real time (<25ms)
- Enable real time analysis of data which in turn enables for instance feedback, based on feature extraction and pattern matching

4 Method

4.1 Data sources

The data in this thesis comes from chronic recordings with invasive electrodes implanted in awake, freely moving rats and from computer simulations.

4.1.1 Chronic recordings

In paper II, a microwire array electrode was built in house. The array consisted of 29, 12 μm platinum-irridium wires insulated with parylene C (Paratech, Järfälla, Sweden) and embedded in gelatine type A (2% Sigma-Aldrich Co, Saint Louis, MO, USA) for optimal stiffness during implantation into the cortex (Lind et al., 2010). Seven rats were anaesthetized i.p. with 6.3 mL/kg solution of 1 mg/mL Domitor vet (medetomidin hydrochloride; Orion pharma, Turku, Finland) and 50 mg/mL fentanyl (Braun, Aschaffenburg, Germany), mounted in a stereotaxic frame and implanted with a microwire array electrode in primary somatosensory cortex (S1).

Some of the data that was later published in paper II was also used in Paper I.

4.1.2 Simulations

In paper III, we developed a software test suite emulating an acquisition system in Python `https://www.python.org/` (with critical parts precompiled in Cython (Behnel et al., 2011)). This allowed us to vary the number of neurons, run parallel sessions and simulate additional systems (in our tests with a fixed number of 320 000 neurons per system), which were all simulated on a single computer with standard hardware (see specifications below) and sent over a local area network, or simulated on the same computer as the data was stored. The test suite did not limit the sampling rate to the typically expected maximal rate of 1000 Hz per neuron, but rather sent data as quickly as possible. If the system was able to handle at least 1000 Hz per neuron, it was considered to be successfully able to store the data that may be obtained in a real physiological setting. No "new" neurons were created during testing, although the data format supports it (see Results). Data integrity was ensured by using TCP (Transmission Control Protocol) for transmission as well as checking for data size of each transmitted frame. Spike trains were generated to correspond to an encoded array of integers in sequence (e.g. array of 1001 at time T, array

of 1002 at time T+1, array of 1003 at time T+2, etc.), interrupted by events corresponding to firing simultaneously (array of integer (2^{32}-1)=4294967295). Although this exact pattern is unlikely to observe in a live recording, it allows a demonstration of the capability of the system to detect a given firing pattern.

Ordinary off the shelf hardware was used both for storing and analyzing the data, as well as for generating the simulated data. A dedicated HP Proliant HL350E Server with Ubuntu Linux 15.04, 16 GB RAM memory, dual Intel XEON Quad Core 2.4 GHz processors and a single HP Solid State Drive was used as the server integrating, analyzing and storing the data. A 10 GBit/s Ethernet network using copper cabling was used for network transfers. The simulated data was generated by a HP z400 workstation with 12 GB Ram memory and dual Intel XEON Quad Core 2.4 Ghz processors. The data from longitudinal tests was stored on a 2TB Samsung NVMe SSD 960 PRO Solid State Drive.

4.2 Data information flow - management of electrophysiological data

After recordings or simulations, data was stored as part of a data information flow, as outlined previously in the Introduction. The focus of the different papers was as follows:

- Paper I addresses local and remote storage of data, which allows for offline data sharing and structuring.
- In paper II, the storing and structuring of data is applied to a neurophysiological setting and is used to facilitate offline data analysis.
- In paper III, an architecture and data format is presented for both storing and structuring data offline as well as in real time. Our specific solution allows for extremely fast data analysis as well as feedback – based on this analysis – to the recording subject.

4.3 Readouts of pain (II)

Most common animal models of pain and analgesia are based on indirect measures such as nocifensive behaviours, which cast doubt if they provide valid measures of pain perception as such. To address this issue, authentic neuronal data was obtained in paper II from a new animal model of pain, comprising a more direct readout, including both evoked potentials as well as spiking

data from individual neurons, via chronically (> 1 month) implanted MEA in rat primary somatosensory cortex (S1, which is known to be involved in pain perception in humans). Spike sorting was performed using a fully automatic k-means based approach supplied by the Fieldtrip toolbox (Oostenveld et al., 2011). The readouts was compared to commonly used behavioral measures of pain during development of hyperalgesia. A translational method to induce hyperalgesia, UVB irradiation of the skin, was used. Localized CO2 laser stimulation was made of twenty skin sites (20 stimulations/site/observation day) on the plantar hind paw, before and during the time period when enhanced pain perception is reported in humans after UVB irradiation. In this work, the database structure developed in paper I was used in order to store the recorded data for later analysis offline. MySQL (`https://www.mysql.com`), a standard database software freely available as open source, was used as database. MATLAB (`https://www.mathworks.com`), a proprietary programming language and development environment focused on data analysis, was used as programming language for the data processing.

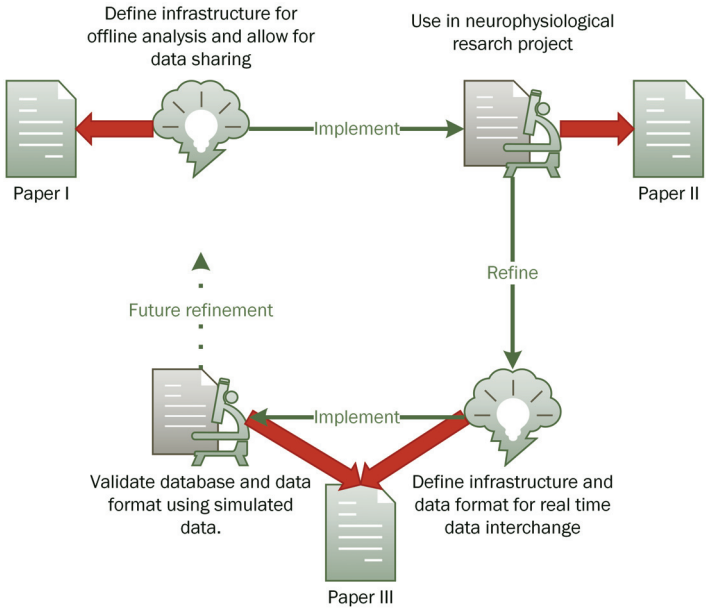

Figure 3: Milestones and deliverables (articles)

5 Results

In order to achieve the aims of this thesis and further progress, the following milestones were defined (see Figure 3):

1. Define a database infrastructure for offline analysis and allow for data sharing.
2. Implement database for offline analysis in a neurophysiological research project in awake freely behaving recording subjects.
3. Refine data structure and define a data format for real time data interchange, based on experiences in the implementation.
4. Validate database and data format using simulated data.

5.1 Data sharing infrastructure (I)

Using a database to store data is often sufficient in order to organize data, but to increase the feasibility for meaningful data sharing, between scientists at the

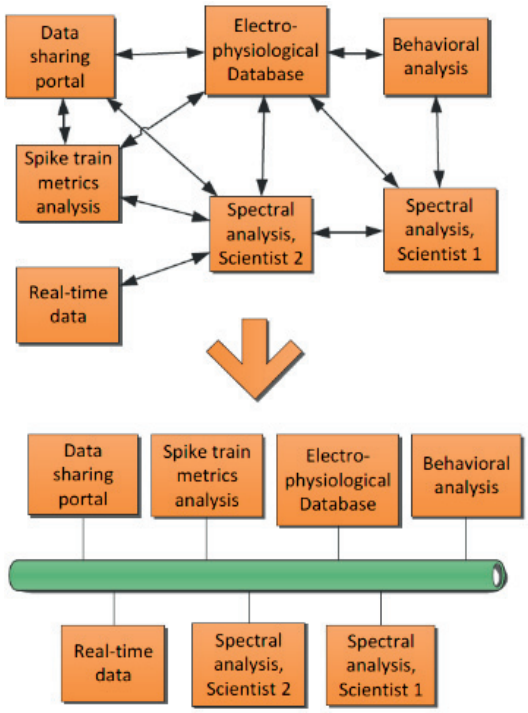

Figure 4: Point-to-point connection architecture vs. service bus architecture of an analysis application. Instead of constructing translators between all different applications, just one common language has to be defined to which the different applications transform their data, thereby reducing the integration effort, as translation only needs to be done once for each entry of data into the service bus. In the point-to-point case, data would have needed to be translated at every time of communication between different systems. Figure from paper I

same research center or between research centers, some additional mechanism for data integration is required as it is often impractical to transfer an entire database. Furthermore, while a common language is likely to be of importance for a total integration system, it does not automatically provide a full solution, since each application would have to use an interpreter for the common language. In commercial systems, such as supply chain and customer relationship management, the challenge of integrating various systems, many by different vendors, is well-recognized. Some of the principles and common solution patterns in "Enterprise integration patterns" (Hohpe and Woolf, 2004), may potentially be applied also to electrophysiology.

In paper I, it is proposed that such an Enterprise integration pattern, a "Service Bus" architecture, may be used for integration of different types of data(see

Figure 4), using a service oriented architecture (SOA) interchange of a common object model through web services. This is referred to in paper I as Core Classes, the language of integration, as data is translated to this intermediary format. Such an architecture forms an electrophysiological service bus, in which the data is translated upon each entry to the neuroinformatic system, see Figure 4. This provides a means for local integration at the research center level, but also the possibility to add interfaces to global data repositories. Data is recorded and then shared after the recording has been finished, resulting in sharing of *offline* data. This is different to *real time* or *online* which, in an electrophysiological context, means the ability to process data within 25 ms. In paper III data sharing was extended to be performed within the limit of this definition of real time, as the bit-encoded data format permitted more efficient interchange of data.

5.2 Analysis infrastructure - objects models - (I,II)

The object core data structure of paper I was implemented both as MATLAB objects and as different entities of the database, a relational database in MySQL. They are a part of a implementation of *abstract factory* and *decorator* design patterns (Gamma, 1995) in order to cope with integration of data from different sources into one data model; the abstract factory pattern provides a way to encapsulate (restricting access to) a creator class (a class that creates other classes), a *factory*, without specifying concrete classes, while the decorator allows for adding behaviors to an individual object without affecting other objects of the same class, see Figure 5. The base classes of this core data structure form a tree composed of the following entities from root to leaves :

- **Subject** - the subject for electrophysiological investigation
- **Experiment** - a set of recordings, from the same or different animals, with the purpose of answering a research question
- **Recording session** - the different recording sessions performed on the subject in time
- **Electrode** - the individual electrode recording an electrophysiological signal
- **Event** - (at the same level as **Electrode**) a discrete event of any kind in time, e.g. onset of a behavior or application of a stimulus
- **Unit** - The neuronal unit of a recording, as identified by spike sorting.

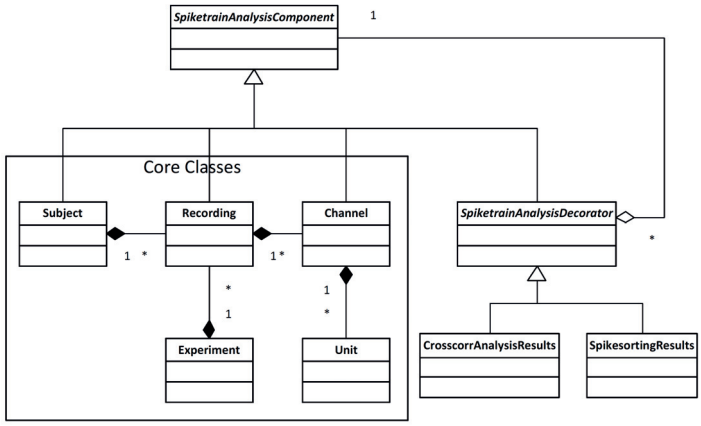

Figure 5: The decorator design pattern UML (Unified Modeling Language, a graphical representation of a software structure) class diagram. The core classes at the bottom left are the classes that represent fundamental data structure of electrophysiological entities. Black diamonds denote a composition relationship, while open diamonds denotes aggregation. They implement the abstract super-class SpiketrainAnalysisComponent which is connected to a SpiketrainAnalysisDecorator abstract super-class. This class is inherited by different concrete subclasses, of which two examples are provided here, corresponding to results of the analysis. The result is a decoupling of the analysis data from the Core Classes.

This core data structure was applied in an authentic neurophysiological setting in paper II, enabling grouping and selection of data, for example finding all recordings made at the first day after UV-stimulation.

5.3 Data format for neuroscientific data (III)

When recording with many electrodes and from many units simultaneously, there are several reasons why it is important that data is integrated in the same memory area, i.e. so that all of the data may be accessed and used in one computation simultaneously. In this way it is possible to process data from all electrodes at the same time, so as to look, for instance, for complex spatiotemporal signaling patterns involving all different structures recorded from. However, such a multi-structure integration is usually not possible, since the computer system that is collecting the signals usually can handle only a limited number of operations in a certain time, for example online spike sorting. Instead, data is recorded to be analyzed at a later point in time in separate files. However, in order to store all the neural data in a common database

one has to overcome the challenge that many data formats and databases do not support true parallelism. In other words, data cannot be written to a data structure by multiple processes in real time, especially if considering the amounts of data that is yielded by parallel MEA recordings. A software architecture adequate for this context should be able to take advantage of a multiprocessor or multi-core CPU (Central Processing Unit of a computer), where each processor or core may run its own piece of code separately, but cooperating across a common memory area. An MPI, a message passing interface, addresses this problem by establishing a standard for communicating among different processes, both point-to-point as well as collective communication is supported. To meet this challenge of these processing requirements, during the development of this thesis, many different formats were systematically tried out, including relational databases, flat files and different acquisition system vendor data formats. After considerable and time consuming evaluations and testing, it became evident that in order to handle the large amounts of data required or expected, most of them would not suffice, as they could not handle the parallel I/O (Input/Output) required to support distributed recordings. The final successful candidate, HDF5 (`https://www.hdfgroup.org`), is a data format that has gotten increasing attention in various biomedical applications (Dougherty et al., 2009; Mouček et al., 2014; Hanuma Chaitanya et al., 2014; Teeters et al., 2015), including in Neuroscience. HDF5 supports a large variety of data types, and is designed for flexible, efficient and parallel (with special MPI support) I/O also for complex data, including high dimension data. Furthermore, the support for a large number of high-level languages, such as Java and Python, is quite appealing, since it enables the data format to be used also in more complex projects. The complexity of the software would be difficult to build using a lower level language; alternatives such as pure C and C ++ were discarded as they were deemed to take too much time to implement the desired features in often needing more extensive development efforts to result in a stable program, although attractive due to performance as fairly low level languages. Python is a script-language that has powerful expressional possibilities. Combined with precompilation using Cython, it also has the performance capabilities required to process large amounts of data, as shown in paper III. It also has excellent support for HDF5, due to the h5py library (Collette, 2013), which also supports parallel Python programs writing to the same HDF5 file. During discussions in the Electrophysiology task force of the INCF (International Neuroinformatics Coordinating Facility, `https://www.incf.org/`) during 2011-2013, which the author of this thesis was part of, HDF5 was suggested as a candidate for storing specifically offline electrophysiology data. As a result of that discussion, but also since the author realized the potential for online data management, considering the solid parallel processing support, we decided to try HDF5 out as a suitable

container format (with other more specific data-structures contained within it, as described in the next section) for storing online data, this undertaking turned out to be successful, as HDF5 fulfilled (and surpassed) all expected performance requirements.

5.4 Encoding of spike data to bit code (III)

As data rate transfers in various neurophysiological applications often will be limited due to hardware and software constraints, it is crucial to use data size as efficiently as possible, without trading speed. For this purpose, we developed an encoding algorithm and minimalistic data storage format, in order to allow for rapid management of real time spiking data, as follows: Each spike is considered as a binary event in discrete time, either spiking during that period (the length of the time bin) of time or not, encoded as a single bit during a specific time as part of a 32-bit integer. Regarding time resolution, due to the refractory period of a single neuron, it is assumed that no neuron will fire with a frequency greater than 1000 Hz. This yields an array of 32-bit integers, of length N/32, where N is the number of neurons recorded in total at the time point (ms), see Figure 6. The array is then saved in the HDF5-file, which supports multiple parallel encoding for different acquisition systems at different locations in the brain or from entirely different subjects. These parallel sessions are handled by different threads of a multiprocessor environment, which are synchronized through MPI.

The input to our system is spike train data, that is waveforms, and time of the threshold crossing (or other time of detection, if for example a wavelet-based detection of spikes was used). If recording a large numbers of neurons,say up to 1 million, different systems are encoding a smaller part of the total number of neurons, although they must share a common timebase. The algorithm is then as follows (as provided in Ljungquist et al. (2018), paper III):

1. Divide the neurons in segments of 32, each corresponding to a bit in the 32-bit integer to be encoded. If enough neurons do not exist for a segment, the reserve bits are set to 0. If new neurons are detected, the reserve bits are allocated in order. A finite number of segments (N/32) are pre-allocated in memory and file system cache, covering the maximum number of neurons expected to be recorded during the session in each acquisition system. If a segment is filled, a new bit segment from that acquisition systems segments is allocated. Neurons which are not recognized to have been identified before will always get “new” IDs. We did not reuse IDs, but rather created a new bit group to store this information.

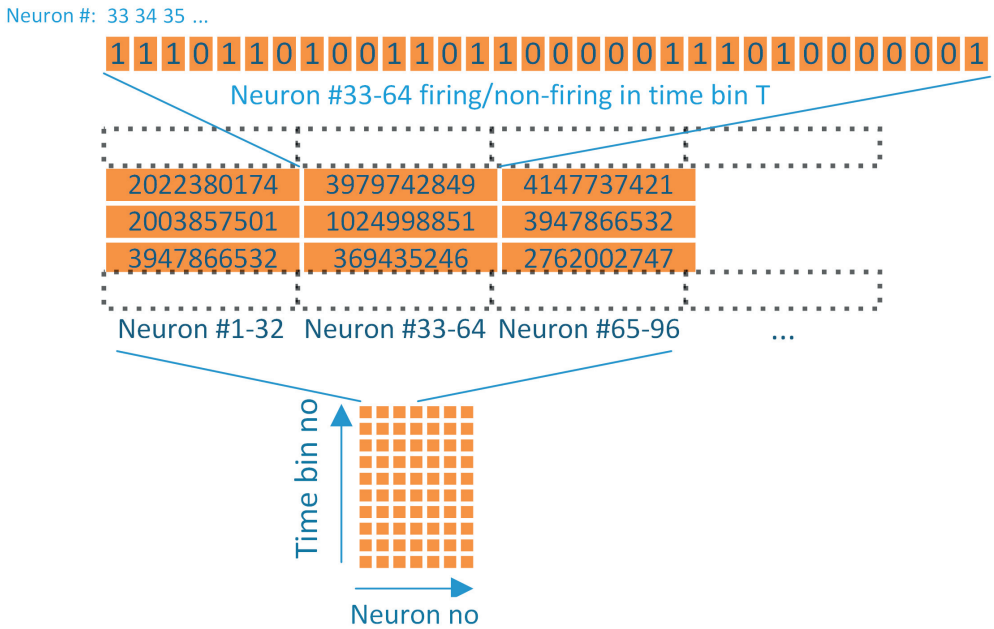

Figure 6: Proposed encoding of spike data in 32-bit integer data grids. At any given time, each neuron spike/non-spike is encoded as a binary bit (1 or 0) in the corresponding time bin. The binary bits of groups of 32 neurons are then expressed as a 32-bit integer. The 32-bit integers are then grouped together in a data grid, where each row corresponds to a specific time bin and each column to the activity in a group of 32 neurons.

2. For each neuron, if a spike was detected since last transfer, the bit corresponding to that neuron is set to 1, otherwise it is set to 0. The result then is a 32-bit integer.

3. Each integer in its position of the array to be sent is set to the resulting value, yielding a N/32 length array.

4. The array, which now encodes whether the neurons recorded by the acquisition system fired or not during the latest millisecond, is sent to data storage (HDF5 file in our system) for concatenation with previously recorded arrays and arrays from the other acquisition systems.

The actual encoding into a bit format is a transformation that may also be described by the matrix[3] operation $A * B = C$, where A is j x k x l matrix with

[3] A matrix is the mathematical concept of a rectangular array of numbers arranged in rows and columns

j = number ofneurons, k = window size and $l = 32$, B:

$$\begin{bmatrix} 1 & 2^1 & 2^2 & \dots & 2^{i-1} & \dots & 2^{31} \end{bmatrix}$$

and C is a m x n matrix with m = number of neurons and n = window size. Also, this encoding could be implemented by both CPUs and GPUs (Graphical Processing Unit) efficiently. In paper III, the encoding is performed in less than a millisecond for one million neurons.

5.5 Enabling storage of large scale recordings (III)

Even if data is stored in a structured way such as object orientation, any refinement of the object structure has to be traded for speed; inserting objects for every single data point might result in a slowly responding system due to the complex structure. One way to go around this is to store binary items of time series data in so called BLOBs, binary large objects, a common way of storing data in larger pieces. However, in a BLOB, performance drops with data sizes over a couple of megabytes (Stancu-Mara et al., 2014), which was also our experience during the implementation of paper I and II. Using relational databases in data intense real time applications is thus not feasible, although it might be fine for offline applications. In paper III we developed, therefore, a multithreaded software architecture using precompiled code for increasing speed, which was capable of storing data from millions of neurons simultaneously, integrating data from different acquisition systems and performing analysis of data across channels. The results yielded a capability of storing data corresponds to storing data of 3.2 million neurons in real time, see Figure 7. As a part of the performance investigation, the most powerful configuration was found to be integrating from different remote systems over a 10 GBit Ethernet, in contrast to integrating from a local system (handled by one computer), thus confirming the benefits of a distributed architecture for a BMI neuroinformatics system, comprised of many subsystems collaborating to produce data in a common data structure.

The data encoding and format vas validated through a process in which the before transmission and encoding was randomly pregenerated, with same random data for each message of 20 bins, except for every 40th message, starting with the 21st, where in bin 0 (out of 20) of the message, all bits of the integer are set to spiking. In the HDF5 file, which integrates from different sources, this may be observed in bin 420 and every 800th bin thereafter.

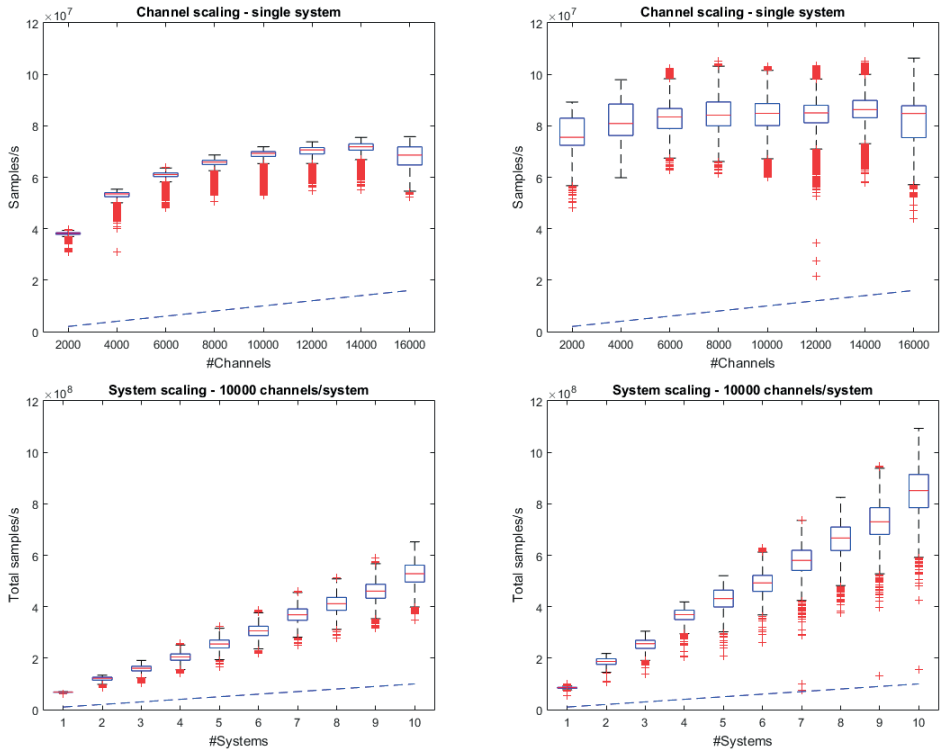

Figure 7: Boxplots of throughput in samples/s as a function of number of neurons for a simulation with data A) from a single system generating data locally, B) from a single system sending data over 10 GbE LAN. Throughput in samples/s as a function of number of systems C) from an increasing number of systems generating data locally, D) from an increasing number of systems sending data over 10 GbE LAN. For C and D the number of neurons were fixed at 320 000 /system. Boxplots are constructed with a 95% confidence interval. The blue dashed lines show the samples per second corresponding to the theoretical limit from the test success criteria with the sampling rate of 1000 Hz per neuron, which, as may be seen, the tested architecture is well above.

5.6 Comparison between cortical readouts and behavioral tests (II)

Paper III provides means to immediately process and access data, for example in visualization, plotting and pattern recognition of data, by providing the data structure. During the analysis of the evoked potential and spiking data from paper II, the analysis process was sometimes hindered by slow access to data, for example in grouping and plotting of data, that a data structure providing faster access to data would be more beneficial to the analysis process. As

some of the recordings were also discovered to have relatively low quality, online analysis directly during the acquisition of the neural data would have been beneficial to allow corrections of the recording, for example directly during electrode implantation, such as with the software architecture developed in paper III.

In paper II, we demonstrate a 2–10 fold significant enhancement of cortical activity evoked from both irradiated and adjacent skin. The time course corresponded to previously reported enhancement of pain magnitude during development of primary and secondary hyperalgesia in humans (Gustorff et al., 2013). In contrast, withdrawal reflexes were only significantly increased from the irradiated skin area and this increase was significantly delayed as compared to activity in S1, which had an earlier peak in both potential and neuronal data. The present findings thus give direct evidence that, provided an efficient data handling system is used, chronic recordings in S1 in awake animals can offer a powerful translational model of the perception of pain magnitude during hyperalgesia, with a potential also for real time evaluation, especially if combined with the data format and system architecture from paper III.

6 Discussion

6.1 Highlights

In this thesis, a novel data format is presented that allows, for the first time, integration and organization of electrophysiological data, not only offline, but also in real time, making possible extremely fast online analysis and closed loop feedback. Due to its highly efficient encoding, this format enables fast storing, transferring and analysis of data from millions of neurons in parallel, thus meeting present and future neuroinformatics challenges in the use of BMI. Although other challenges related to recording quality and stability over time remain, the novel data format and encoding principles provide powerful research tools in neuroscience and will also open up novel clinical applications by making possible for instance complex pattern recognition and real time interaction with neuronal readouts, both in animal models and in clinical diagnostic or therapeutic applications.

6.2 Data format for real time neuronal data

The data format enables real time integration and organization of neuronal data from a very large number of neurons. In fact, in trials on a standard computer, we show in the present thesis that data from recordings of a couple of millions of neurons can be handled. We also show that the data format is both memory and performance efficient and hence the analysis capabilities are extensive. This enables complex operations on the data, such as identifying specific, but complex patterns of neuronal activity, either over time, in the same set of neurons, or across different sets of neurons. Importantly, this feature is crucial for advanced analyses of causal relationships between, for instance, neuronal activity patterns and different aspects of motor (and other) behavior. The performance efficiency of the data format also makes it a powerful tool for investigating interactions in real time. For instance, if a certain part of a specified set of neurons is active at any point in or period of time, excitatory or inhibitory electrical stimulation may be sent back to the subject in real time, as shown in Figure 8. In the present thesis, this has been shown to be possible, again using a standard computer, within a physiologically relevant time, i.e. in less than 25 milliseconds. Importantly, complex pattern recognition may in principle be extended to any arbitrary activity pattern, such as an activity pattern over time or activity with a specific frequency. Alternatively, specific populations of neurons with assumed functional roles may be stimulated and

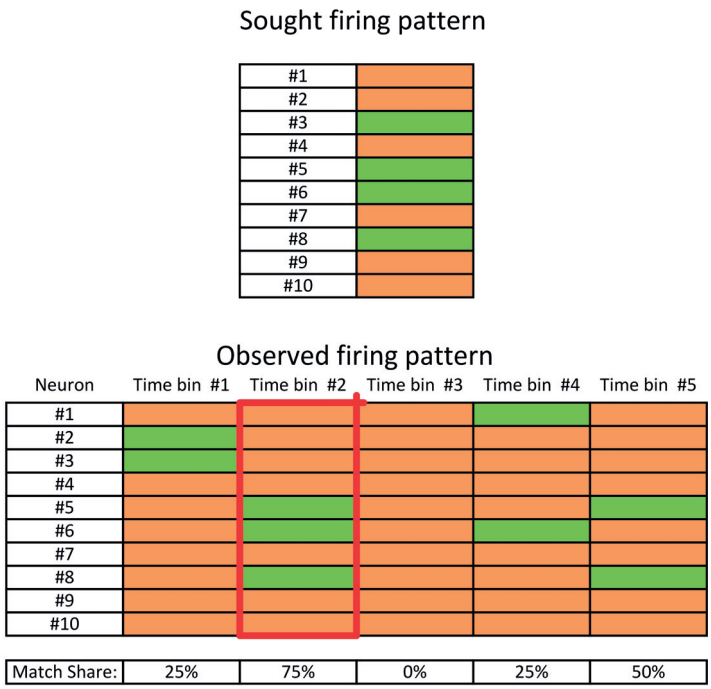

Figure 8: Example of pattern matching, as tested in paper III, using the bit code format. A pattern is sought, in which 4 neurons are firing, indicated by green boxes. In the last row, it is indicated the share of how many neurons that are matching. In the example, an observed firing pattern is deemed matching if 75% of the sought neurons are firing in the observed pattern, indicated by a red line around the time bin. This may then act as a trigger for a stimulation or inhibition sent back to the subject. In paper III this type of matching was able to be performed in less than one millisecond for one million neurons, and a signal indicating triggering stimulation sent back within 25 ms.

changes in subsequent neuronal activity patterns or behaviors may be analyzed.

6.3 Spike train data encoding

6.3.1 Performance

When recording neuronal activity using a BMI, a spike (action potential) is typically represented by 1) *time* – when the spike occurred, i.e. the time for threshold crossing, 2) *waveform shape* – the potential change measured by the electrode around the time of threshold crossing and 3) *neuronal identity* – which

neuron was the origin of the measured action potential. In practice, time and waveform shape are used to determine neuronal identity by a computational process called spike sorting (Lewicki, 1998). A method often used for spike sorting is principal component analysis (PCA). This allows all spikes to be projected onto an orthogonal base of, typically, the first two or three principal components (PC). In previous work from our laboratory (Thorbergsson et al., 2010), it has been shown that the first six PCs account for 99% of the variance in waveform shape. However, once the identities of the neurons have been determined, the waveform shapes are not necessary for the analysis to be carried out, as a given single neuron tends to have the same waveform shape over time, unless the recording electrode is moving. Thus, with highly stable electrodes, such as those being developed by our laboratory, time and neuronal ID are sufficient for representing neuronal action potentials in the form of spikes.

Spike trains are usually represented with either of the following data structures:

1. A 1-dimensional (1D) array of neuronal unit IDs (as integers) for all of the n data samples together with a 1D array of n spike times (usually represented with float data type in microseconds since system start).

2. A binary 2-dimensional (2D) array, i.e. a matrix, of elements c_{ij} denoting if unit j in column j is spiking in time bin i, where the time has been binned at a specific binning frequency, usually at 1000 Hz.

Acquisition system vendors almost exclusively use the first alternative to represent spike train data. It is more efficient in memory for the total amount of data used, especially for neurons that are spiking at relatively low frequencies. By contrast, the second alternative allows for simultaneous computation across units using matrix operations, albeit with some loss of time resolution, since spikes are binned. Furthermore, since the only information required to be stored in a single bin is if a neuron is spiking or not, a binary digit (0 or 1; also referred to as *bit*) is sufficient to encode a spike. If using the bit encoding proposed in the present thesis, in which the bits of an integer are used to represent individual neurons, a single 32-bit integer may represent 32 neurons. For a single spike, the first alternative above is using 64 bits (32 bits for each integer of time and 32 bits for ID), while the second alternative is using only one bit, thus a 64-fold performance difference in terms of storage efficiency. When it comes to analysis, for instance to determine if a set of neurons is spiking or not, the performance difference between the first and the second alternatives is even greater. All that is required to determine if 32 neurons are spiking in a specific pattern is a bitwise comparison of two integers, which in most processor

architectures is a very fast computational operation. These factors explain the performance efficiency of the new encoding format (see paper III), and also why the format is so suitable for computational purposes – it simply does not need to be converted further for computations to be carried out. This applies and extends also to modern processor architectures such as GPUs, which due to their suitability for large array operations may be readily combined with our new data format to achieve unprecedentedly efficient processing of signals from millions of neurons simultaneously (or even billions, if more powerful computers than ordinary laptops would be used).

Considering a sampling rate for the binary 2D array with 1000 bit/s, this corresponds to, for the ID and timestamp 1D format, a maximum spiking rate per neuron of $1000/64 = 16$ Hz. Over this rate of firing, the 1D format is using more memory per spike. However, during for example burst mode firing of thalamic neurons (Sherman and Guillery, 1996), or during high-frequency discharges in a large number of neurons, so called neuronal avalanches, seen in cortical synchronization in awake subjects (Petermann et al., 2009), the individual neuronal firing rate may increase to many times this value. Thus, a system based on timestamp and unit ID for spike data would not be able to cope with the information. On the other hand, the bit format is currently more memory intensive as compared to representing spikes with unit ID and timestamp for firing rates below 16 Hz. It is important to point out in the present context that during a given time bin most neurons are not spiking. This is crucial since storing, for a large number of neurons, the instances (or information about) when an individual neuron is non-spiking, would make the suggested bit code format take quite a lot of space, even given its compression (each spike being represented only by a single bit). Therefore, when storing spike trains in a binary 2D structure, as described above, it is highly advantageous to represent the result as a sparse matrix, i.e. a matrix where most elements have the value of zero. Sparse matrices may be stored in different ways, which take advantage of the sparseness in order to save space. Most frequently used in arithmetic such as matrix multiplication is the Compressed Sparse Column format (CSC), in which values are read first by column, a row index is stored for each value, and column pointers are stored. Despite the fact that a sparse matrix representation was used during the encoding of the firing pattern to a bit pattern in the present thesis (paper III), a standard matrix format was used for transmission of data over Ethernet, as the Numpy (Oliphant, 2007) internal buffer very conveniently and efficiently serializes the data for transfer, which then may be efficiently de-serialized back into standard matrix format. Since HDF5 does not have a standard format for storing data into sparse matrix format, data is stored as non-compressed integer arrays in HDF5 data files. Although there is a data type

in HDF5 for a string of bits, a so called *bitfield* (see documentation of HDF5 at `https://www.hdfgroup.org`, *HDF5 Predefined Datatypes*), we were not able to use it as it is not represented in the h5py interface. In future work, a specific sparse matrix format for bit arrays may be investigated and implemented for further increased performance, possibly also using a supplemental data format as base in addition to HDF5.

6.3.2 Information theoretical efficiency

The proposed encoding of spike data is saving space, but the question is whether it is also reducing information. Importantly, unnecessary information, for example large amounts of non-spiking data, should not be transmitted. The data format may, as a complement to the above analysis, be evaluated based on principles of information theory, which concerns transmission, processing, extraction and utilization of information, as first formulated by Shannon in his seminal paper "A Mathematical Theory of Communication" (Shannon, 1948).

A central concept in information theory is *entropy*, also called Shannon (or Information) entropy, in order to separate it from the thermodynamical concept of entropy. Shannon entropy measures the information *degree of content* (which is not the same as size) of a signal. For a symbol composed of n number of binary digits, also commonly referred to as *bits*, Shannon entropy H is defined as:

$$H = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

where p_i is the probability for the i-th (of the total number of n) possible binary value to occur. If the receiver has prior knowledge of all the values (they all have a probability of either 0 or 1 to occur), no information is considered to have been transmitted and the Shannon entropy is 0. If, on the other hand, the receiver cannot predict the value of a transmitted bit to be either a 0 or a 1, p_i will be 0.5 and the Shannon entropy will have the value of 1. An efficient coding is one where the Shannon entropy is close to 1.

Information theory has been used in neuroscience for analyzing the nature of neural coding, for example the information contents in LFPs and neuronal spike trains respectively (Quiroga and Panzeri, 2009), but also for investigating the importance of timing of individual spikes in relation to each other (Borst and Theunissen, 1999). The Shannon entropy for the bit encoding, if represented

with a standard format is quite low, since neurons more often are not spiking than spiking (at least if assuming a 1000 Hz sampling rate). However, if representing the firing pattern matrix as a sparse matrix, fewer bits are needed to transfer the message and the encoding becomes more efficient with higher entropy for each transferred symbol, as only the spiking activity is transferred. It is then assumed that if spike is not received, the neuron is non-spiking. With this increased efficiency, sparse matrix representation of the proposed bit code format is more efficient from an information theory perspective than if transferring also non-spiking data as well as if transferring timestamps and neuronal unit IDs.

6.3.3 Robustness and vulnerabilities

As the neuronal identity is central to the prosed encoding format, it is important that this identity is valid for units over time. The encoding handles spike train data and it is required that all units are identified properly, so that all spikes sent for encoding represent single unit and not multi-unit activity. This assumes that a correct online spike sorting and tracking is performed by each recording system, a process that is not trivial. Despite recent progress (Rossant et al., 2016), spike sorting still remains a challenge. The tracking also is problematic as recordings obtained separately in time are often sorted separately and the result is a set of single units for each recording (Rey et al., 2015). The units in these different sets may then either: 1) correctly be mapped onto a unit of the same cell 2) erroneously be mapped onto a unit of another cell or 3) lack a corresponding cell. The causes for the two latter outcomes are mainly biocompatibility issues, such as variations in distance between electrode and neuron due to tissue movements, tissue reactions and reorganization of tissue around electrode and loss of neurons nearby electrode immune response, as discussed in the Introduction. In order to ameliorate these problems, any features that are (relatively) stable over a certain time may be used for tracking a specific neuron. To date, the following features indicating stability of single unit activity (SUA) over time have been used for tracking:

1. **Waveform shape** - Due to the properties of a neuron, the shape of the action potential waveforms usually is the same over time, a feature which also is the basis for spikesorting. It is therefore natural to make use of this feature also for keeping track of cells, and waveform shape thus has become the most commonly used feature for SUA stability, by calculating the cross correlation coefficient for the averaged sorted waveforms (Jackson and Fetz, 2007; Dickey et al., 2009) or to calculate the Pearson's correlation

coefficient computed between the two averages of the sorted waveforms (Dickey et al., 2009). A recent study (Eleryan et al., 2014) uses a set of four waveform-based measures, correlation coefficient, normalized peak-to-peak height difference, normalized time-to-time difference, and peak matching. Another study by Fraser and Schwartz (Fraser and Schwartz, 2012) also utilized waveform shapes.

2. **Interspike interval (ISI) histogram shape** - one pioneering study in following neurons over time (Dickey et al., 2009) used the ISI shape together with waveform shape for tracking neurons. Neurons may however have different preferred firing modes and thus ISI characteristics (Chen and Fetz, 2005) that differ during different brain states (indicated by for example different EEG spectra) (Wörgötter et al., 1998). Algorithms have now been developed that may accurately assess unit stability across days with minimal human expert intervention (Eleryan et al., 2014), using ISI characteristics.

3. **Physiological context** - neuron-stimuli/behavior relationship - correlation between neuronal activity and behavioral states, in which the neurons also adapt their behavior, have been shown to be stable over longer time periods, for example in their different roles in idle or active motor states (Velliste et al., 2014),

4. **Population dynamics** - from the relation with the activity of other neurons (correlation in time domain) or by relating it to LFPs (phase locking or Spike Field Coherence) (Berke et al., 2004).

A tool that visualizes all the features and provides decision support for letting an experimenter decide if what appears to be two different neurons as indicated by spike sorting, in fact are one and the same neuron would be highly valuable. We have developed software that utilizes a subset of features that is relatively stable over time for tracking single units during discontinuous recordings from cortical or striatal neurons. We have also implemented classification software using these features, which allows for tracking of neurons, (Ljungquist et al., 2010). This work is not covered further in this thesis, but could be used as an evaluation if recordings are suitable to be encoded into the proposed bit coding format; if units from a specific multielectrode array (MEA, one out of possibly many from a single subject) in general are stable over time, the data from that MEA could be considered ready for encoding. For example, the ratio of stable neurons (neurons present during two disjunctive recordings) to the total number of identified neurons, could be used as a measure of recording quality. Other

groups have since our tool was first published developed similar tools (Eleryan et al., 2014; Fraser and Schwartz, 2012), which use subsets of these features.

During recordings, these tools need data to operate on and evaluate stability for that data. Therefore, they may readily be combined with the software architecture proposed in the present thesis (paper III). For spike tracking purposes, a secondary storage for spike waveforms was proposed as a feature of this architecture. This secondary spike waveform storage would only record a subset of the neurons at a given time and cycle through all the neurons periodically in order to track the waveforms shape progression of all neurons. The waveform storage may reside on the same computer as the spike data storage, since it does not require extensive bandwidth. This would also facilitate practical rapid access if using waveforms in analysis calculations. A further development in future work for storing the waveform data in an efficient way could be to make use of principal components (as described in (Thorbergsson et al., 2010)) for reducing the waveform data as well as providing a feature space for tracking changes in waveform shape over time.

6.4 Clinical applications

As outlined in the Introduction, BMI are gradually finding increased clinical use and this development strongly emphasizes the importance of neuroinformatics. For example in prosthetics applications guided by neuronal readouts, data needs to be integrated, analyzed and organized in real time in order to be able to control the prosthetic device, possibly also using sensory feedback. With the developed bit code, implementation of such systems, involving data from millions of neurons is not only possible, but clearly feasible.

A recent comprehensive review of the literature on neural prosthetics reveals that it is often not specified how data is stored (Lebedev and Nicolelis, 2017), except to indicate which specific acquisition system was used. Most likely, the format specified by the system manufacturer is used. For offline data, this is sufficient for most projects, as long as the files are well organized. However, during multi-trial experiments, such as presented in the present thesis (paper II), keeping track of all recordings quickly becomes very challenging, especially if data is integrated for analysis. For online real time data, lacking a specific format leads to development of experiment specific protocols and applications. For example, in a study of neuroprosthetic control by an individual with tetraplegia (Collinger et al., 2013), it is specified that unit activity was converted to firing rate in 30 ms bins and low pass-filtered using an exponential smoothing function with a 450 ms window. Although successful in this particular application, a finer level

of detail may be needed for finer control applications, requiring for example information about complex collaborative firing patterns with millisecond onset and offset. The bit code data format provides such possibilities.

Clinical applications are, however, not limited to neuroprosthetics, but could be used also for example to reliably detect epileptic seizures, where it could be combined with an interactive stimulation device to alleviate or counteract upcoming seizures. Epilepsy is a chronic disorder of the brain that affects people of all ages. According to WHO (World Health Organization), approximately 50 million people worldwide have epilepsy, making it one of the most common neurological diseases globally. For the individual neuron, firing patterns reflecting seizure onset and termination may appear evident (Truccolo et al., 2011). However, these patterns are in fact different in different neurons; some neurons are increasing their firing rate, while others decrease their firing. It is thus crucial to be able to discern the patterns of individual neurons. To this end, a BMI with neuroinformatic backend using the bit-code and software architecture, as specified in paper III, could easily find these individual patterns. One recent study showed a promising application of dorsal column stimulation (DCS) for decreasing seizure frequency and duration (Pais-Vieira et al., 2016), which, for example, could be used to alleviate the detected upcoming seizure. Other foreseeable applications are Deep Brain Stimulation for treating movement disorders, depression or persistent or intermittent pain guided by detected neuronal patterns of activity.

6.5 Cortical readouts as a translational model for pain

We showed in paper II that chronic cortical measurements in the awake rat can replicate many features of pain perception in humans (Gustorff et al., 2013), in particular pain magnitude, whereas motor responses do not. This suggests that cortical recordings, using chronically implanted BMI and appropriate information analysis, can provide a more valid measure of pain magnitude than commonly used models based on motor responses (Woolf, 2012).

The finding that motor responses in rodents do not correlate well to human perception of pain is not surprising since it is known that motor responses are organized by other neuronal systems than those directly involved in pain perception. Moreover, the control of and processing of information in motor systems such as those underlying spinal reflex responses or coordination and adaptation of movements (e.g. cerebellum) are known to be strikingly different from systems devoted mainly to sensory aspects, such as the spinothalamocortical pathways (Schouenborg, 2008). Rats, monkeys and humans have similar

spinothalamic pain termination targets in the thalamus, the VPL (Ventral posterolateral) nucleus, CL (Central Lateral) nucleus and the posterior complex (Mehler, 1962, 1969; Lund and Webster, 1967; Zemlan et al., 1978; Peschanski et al., 1983; Willis et al., 2001, 1979). The main target of the VPL is the S1 (primary sensory) cortex (Kenshalo and Isensee, 1983; Andersson et al., 1997). Thus, the S1 cortex of the rat, as used in paper II, appears to be a suitable and translational target for the study of pain perception. The concept of deciphering pain related signals directly from cortical recordings with chronically implanted BMI in awake freely moving animals suggests that, it is possible to gain profound insights into the pain experience of animals, something that has been missing previously. It should be kept in mind, though, that pain is a complex phenomenon, and other dimensions of pain than pain magnitude, such as affective dimensions will most likely require recordings from other brain areas than the S1 cortex. It may also be speculated that similar recordings can provide powerful translational models also of, for instance, itch.

6.6 Conclusion and perspective

The novel data format presented in this thesis allows an unprecedented integration and organization of electrophysiological data offline, but perhaps most importantly also directly in real time. The key innovative step was to develop a bit code representation of spiking data, which dramatically facilitates all subsequent computational processes. The capabilities for fast storing and transferring of data allows massive parallel recordings including millions of neurons simultaneously, but also complex analyses in order to provide direct feedback to the subject recorded from, using for example machine learning. The principles of database organization proposed in the thesis, not only the bit code data format, but also an object model with design patterns and service bus architecture, may be readily used to organize and integrate data for a wide range of research projects, thus providing a much-needed infrastructure for present and future electrophysiological research.

The main practical limitation of the suggested bit code format is the need for stable and high-quality input from the neurophysiological recordings (as also discussed above). The tissue reaction to the recording electrode is one of the main limiting factors. During the efforts to address this issue, it has been found previously at the NRC that flexible electrode materials elicit less tissue response than rigid (Köhler et al., 2015), that electrode size matters, with small electrodes provoking less reaction (Thelin et al., 2011), that the specific weight of the electrodes should be close to that of the tissue (Lind et al., 2013), and that

by using embedding gelatin materials tissue responses can be further reduced (Lind et al., 2010; Köhler et al., 2015). Another factor limiting recording quality is electrode stability in the tissue, as moving electrodes will be unable to follow signals from single neurons. To this end, methods to anchor electrodes in the tissue have also been devised at the NRC with promising results (Agorelius et al., 2015). Further work is however needed and is currently in progress.

Further work also involves a more complete implementation of all the proposed architecture of paper III for more extensive electrophysiological data analysis. Such a full implementation would include for example the waveform storage, which could be combined with a dimensionality reduction using principal components as described in (Thorbergsson et al., 2010) for minimizing the waveform data to relevant features, as well as providing a base for tracking unit waveform stability over time. Also, the integration of multiple data modalities, such as imaging, with electrophysiological data remains to be carried out. This could be used together with existing methods for analysis of binned representations of massive parallel spike train data such as detection of synfire chain (a feed-forward network of neurons with many layers) activity and sequences of synchronous neuronal events (Schrader et al., 2008; Torre et al., 2016).

Data from large amounts of neurons will be managed well by the bit code format, but the subsequent analysis will be daunting if performed manually in order to, for example, identify signaling patterns by visual inspection of spike train data. Although visualization tools exist to make this task easier, complex patterns, such as phase shifted spike trains or intermittent firing patterns across many neurons will still be difficult to detect, especially without a priori knowledge of what to look for. Here, modern computational power with AI and machine learning may be of great use. Machine learning refers to the capability of computers to learn without being explicitly programmed. It is part of the greater concept of artificial intelligence (AI), which also includes machines carrying out tasks in an intelligent manner. There are many different algorithms for machine learning, for example decision trees, support vector machines, clustering and deep learning. In particular, deep learning, which is intended to mimic the computations of the human brain, has come to recent attention as prices of specific computational hardware adapted and designed for solving these types of problems have fallen. All commercially available speech recognition systems (for example Amazon Alexa, Google Now and Apple Siri) are built based on principles of deep learning. When analyzing neuroscientific data, often some a priori knowledge about the problem domain is assumed. For example, if analyzing an observed motor pattern, not only which neurons to record from must often be known, but also which spatiotemporal computational correlation that are sought, for example a

specific relationship in firing phase of physically close neurons. The alternative is to discover these relations in a more a posteriori manner, by observing if a pattern exists after data has been collected e.g. in correlation to a behavior or stimulus. For larger number of neurons, this has to date been difficult due to the sheer computational power needed. With the capabilities given by the data format of paper III with large amounts of data resulting from millions of cells recording at the same time, this alternative is within reach. The binary bit code data format of paper III enables fast processing and analysis of the recorded data. In particular, it would enable binarized neural network classifiers, which are deep neural networks constrained to operate on data and have weights that are binary (Courbariaux et al., 2016), to operate on the recorded data. These types of classifiers have recently been highlighted due to their significantly lower computational and power cost in implementations. They have been shown to have both very high throughput (in the range of Tera operations per second, TOPS) and very low latency (for most datasets in the millisecond range) in classifier implementations on FPGAs (Field-programmable-gate-array) (Umuroglu et al., 2016; Fraser et al., 2017) when benchmarking pattern recognition on established test data sets . In addition to impressive computational results, power consumption needed for calculations is comparably low (Umuroglu et al., 2016). If successfully implemented in the domain of neuroinformatics, machine learning would enable complex pattern classification in real time, thus providing many new neurophysiological insights by being able to classify and detect various signaling patterns of millions of neurons and determine how they correlate with observed physiological processes in real time, at a time scale comparable and relevant to real world neural computations. However, this remains future work.

7 References

Agorelius, J., Tsanakalis, F., Friberg, A., Thorbergsson, P. T., Pettersson, L. M., and Schouenborg, J. (2015). An array of highly flexible electrodes with a tailored configuration locked by gelatin during implantation-initial evaluation in cortex cerebri of awake rats. *Frontiers in Neuroscience*, 9(SEP):331.

Andersson, J. L., Lilja, A., Hartvig, P., Långström, B., Gordh, T., Handwerker, H., and Torebjörk, E. (1997). Somatotopic organization along the central sulcus, for pain localization in humans, as revealed by positron emission tomography. *Experimental brain research*, 117(2):192–9.

Bansal, A. K., Truccolo, W., Vargas-Irwin, C. E., and Donoghue, J. P. (2012). Decoding 3D reach and grasp from hybrid signals in motor and premotor cortices: spikes, multiunit activity, and local field potentials. *Journal of Neurophysiology*, 107(5):1337–1355.

Bar-Hillel, A., Spiro, A., and Stark, E. (2006). Spike sorting: Bayesian clustering of non-stationary data. *Journal of neuroscience methods*, 157(2):303–316.

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., and Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, 437(7062):1158–1161.

Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D. S., and Smith, K. (2011). Cython: The Best of Both Worlds. *Computing in Science & Engineering*, 13(2):31–39.

Bellin, D. L., Sakhtah, H., Rosenstein, J. K., Levine, P. M., Thimot, J., Emmett, K., Dietrich, L. E. P., and Shepard, K. L. (2014). Integrated circuit-based electrochemical sensor for spatially resolved detection of redox-active metabolites in biofilms. *Nature Communications*, 5:595–600.

Ben-Ari, M. (2006). *Principles of concurrent and distributed programming*. Pearson Education.

Benabid, A. L. (2003). Deep brain stimulation for Parkinson's disease. *Current Opinion in Neurobiology*, 13(6):696–706.

Benda, J., Gollisch, T., Machens, C. K., and Herz, A. V. (2007). From response to stimulus: adaptive sampling in sensory physiology. *Current Opinion in Neurobiology*, 17(4):430–436.

Bergel, T., Gruner, C., Guilory, S., Loffler, H., Plummer, T., Reina, T., Stengel, C., Korver, K., Wiggins, H., and Wilson, W. (2011). Neuroshare - Open data specifications and software for neurophysiology.

Berger, H. (1931). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 94(1):16–60.

Berke, J. D., Okatan, M., Skurski, J., and Eichenbaum, H. B. (2004). Oscillatory Entrainment of Striatal Neurons in Freely Moving Rats. *Neuron*, 43(6):883–896.

Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S., and Mitra, P. P. (2010). Chronux: A platform for analyzing neural signals. *Journal of Neuroscience Methods*, 192(1):146–151.

Borst, A. and Theunissen, F. E. (1999). Information theory and neural coding. *Nature neuroscience*, 2(11):947–57.

Burgess, N., Maguire, E. A., and Keefe, J. O. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, 35:625–641.

Chen, D. and Fetz, E. E. (2005). Characteristic Membrane Potential Trajectories in Primate Sensorimotor Cortex Neurons Recorded In Vivo. *Journal of Neurophysiology*, 94(4):2713–2725.

Collette, A. (2013). *Python and HDF5: Unlocking Scientific Data*. " O'Reilly Media, Inc.".

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., McMorland, A. J., Velliste, M., Boninger, M. L., and Schwartz, A. B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*, 381(9866):557–564.

Courbariaux, M., Hubara, I., Soudry, D., El-Yaniv, R., and Bengio, Y. (2016). Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1. In *arXiv*, page 9.

Dahl, O.-J. and Nygaard, K. (1966). SIMULA: an ALGOL-based simulation language. *Communications of the ACM*, 9(9):671–678.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

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

Dougherty, M. T., Folk, M. J., Zadok, E., Bernstein, H. J., Bernstein, F. C., Eliceiri, K. W., Benger, W., and Best, C. (2009). Unifying Biological Image Formats with HDF5. *Queue*, 7(9):20.

Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, 14(11):770–785.

Eleryan, A., Vaidya, M., Southerland, J., Badreldin, I. S., Balasubramanian, K., Fagg, A. H., Hatsopoulos, N., and Oweiss, K. (2014). Tracking single units in chronic, large scale, neural recordings for brain machine interface applications. *Frontiers in Neuroengineering*, 7:23.

Engel, a. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature reviews. Neuroscience*, 2(10):704–16.

Farwell, L. A. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, 70(6):510–23.

Fetz, E. E. (1969). Operant Conditioning of Cortical Unit Activity. *Science*, 163(3870).

Fraser, G. W. and Schwartz, A. B. (2012). Recording from the same neurons chronically in motor cortex. *Journal of Neurophysiology*, 107(7):1970–1978.

Fraser, N. J., Umuroglu, Y., Gambardella, G., Blott, M., Leong, P., Jahre, M., and Vissers, K. (2017). Scaling Binarized Neural Networks on Reconfigurable Logic. In *Proceedings of the 8th Workshop and 6th Workshop on Parallel Programming and Run-Time Management Techniques for Many-core Architectures and Design Tools and Architectures for Multicore Embedded Computing Platforms - PARMA-DITAM '17*, pages 25–30, New York, New York, USA. ACM Press.

Gamma, E. (1995). *Design patterns: elements of reusable object-oriented software*. Pearson Education India.

Garcia, S. and Fourcad-Trocmé, N. (2009). OpenElectrophy: an electrophysiological data and analysis sharing framework. *Frontiers in Neuroinformatics*, 3.

Garcia, S., Guarino, D., Jaillet, F., Jennings, T., Pröpper, R., Rautenberg, P. L., Rodgers, C. C., Sobolev, A., Wachtler, T., Yger, P., and Davison, A. P. (2014). Neo: an object model for handling electrophysiology data in multiple formats. *Frontiers in Neuroinformatics*, 8(February):10.

Gardner, D. (2004). Neurodatabase.org: networking the microelectrode. *Nature Neuroscience*, 7(5):486–487.

Gardner, D., Xiao, Y., Abato, M., Knuth, K., and Gardner, E. (2002). BrainML and GENIE: Neuroinformatics schemas for neuroscience data sharing. In *Society for Neuroscience Abstracts*, volume 28.

Georgopoulos, A., Kalaska, J., Caminiti, R., and Massey, J. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.*, 2(11):1527–1537.

Georgopoulos, A., Lurito, J., Petrides, M., Schwartz, A., and Massey, J. (1989). Mental rotation of the neuronal population vector. *Science*, 243(4888).

Georgopoulos, A., Schwartz, A., and Kettner, R. (1986). Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419.

Gilja, V., Nuyujukian, P., Chestek, C. A., Cunningham, J. P., Byron, M. Y., Fan, J. M., Churchland, M. M., Kaufman, M. T., Kao, J. C., Ryu, S. I., and Others (2012). A high-performance neural prosthesis enabled by control algorithm design. *Nature neuroscience*, 15(12):1752–1757.

Goldberg, A. and Robson, D. (1983). *Smalltalk-80 : the language and its implementation*. Addison-Wesley.

Gustorff, B., Sycha, T., Lieba-Samal, D., Rolke, R., Treede, R.-D., and Magerl, W. (2013). The pattern and time course of somatosensory changes in the human UVB sunburn model reveal the presence of peripheral and central sensitization. *Pain*, 154(4):586–597.

Hanuma Chaitanya, C., Subhasis, R., Upinder, B., and Daniel, W. (2014). Neuroscience Simulation Data Format (NSDF) : HDF-based format for large simulation datasets. *Frontiers in Neuroinformatics*, 8(26).

Hinterberger, T., Kübler, A., Kaiser, J., Neumann, N., and Birbaumer, N. (2003). A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 114(3):416–25.

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., and Others (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375.

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–71.

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

Hohpe, G. and Woolf, B. (2004). *Enterprise integration patterns : designing, building, and deploying messaging solutions*. Addison-Wesley.

Jackson, A. and Fetz, E. E. (2007). Compact movable microwire array for long-term chronic unit recording in cerebral cortex of primates. *Journal of neurophysiology*, 98(5):3109–18.

Kennedy, P. R., Bakay, R. A., Moore, M. M., Adams, K., and Goldwaithe, J. (2000). Direct control of a computer from the human central nervous system. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 8(2):198–202.

Kenshalo, D. R. and Isensee, O. (1983). Responses of primate SI cortical neurons to noxious stimuli. *Journal of neurophysiology*, 50(6):1479–96.

Köhler, P., Wolff, A., Ejserholm, F., Wallman, L., Schouenborg, J., and Linsmeier, C. E. (2015). Influence of probe flexibility and gelatin embedding on neuronal density and glial responses to brain implants. *PloS one*, 10(3):e0119340.

Kreutzberg, G. W. (1996). Microglia: A sensor for pathological events in the CNS. *Trends in Neurosciences*, 19(8):312–318.

Lebedev, M. a. and Nicolelis, M. a. L. (2006). Brain-machine interfaces: past, present and future. *Trends in neurosciences*, 29(9):536–46.

Lebedev, M. A. and Nicolelis, M. A. L. (2017). Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. *Physiological Reviews*, 97(2):767–837.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology*, 23(4-5):727–738.

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

Linaro, D., Couto, J., and Giugliano, M. (2015). Real-time Electrophysiology: Using Closed-loop Protocols to Probe Neuronal Dynamics and Beyond. *Journal of visualized experiments : JoVE*, (100):e52320.

Lind, G., Linsmeier, C. E., and Schouenborg, J. (2013). The density difference between tissue and neural probes is a key factor for glial scarring. *Scientific Reports*, 3:1–7.

Lind, G., Linsmeier, C. E., Thelin, J., and Schouenborg, J. (2010). Gelatine-embedded electrodes–a novel biocompatible vehicle allowing implantation of highly flexible microelectrodes. *Journal of neural engineering*, 7(4):046005.

Little, S., Pogosyan, A., Neal, S., Zavala, B., Zrinzo, L., Hariz, M., Foltynie, T., Limousin, P., Ashkan, K., FitzGerald, J., Green, A. L., Aziz, T. Z., and Brown, P. (2013). Adaptive deep brain stimulation in advanced Parkinson disease. *Annals of Neurology*, 74(3):449–457.

Ljungquist, B., Petersson, P., Johansson, A. J., Schouenborg, J., and Garwicz, M. (2018). A Bit-Encoding Based New Data Structure for Time and Memory Efficient Handling of Spike Times in an Electrophysiological Setup. *Neuroinformatics*.

Ljungquist, B., Petersson, P., Thelin, J., Schouenborg, J., Johansson, A., and Garwicz, M. (2010). Spike feature stability during long-term recordings. In *Neuroscience Meeting Planner*, page 918.12/OOO7, San Diego, CA. Society for Neuroscience.

Lund, R. D. and Webster, K. E. (1967). Thalamic afferents from the spinal cord and trigeminal nuclei. An experimental anatomical study in the rat. *Journal of Comparative Neurology*, 130(4):313–327.

Machado, S., Cunha, M., Velasques, B., Minc, D., Teixeira, S., Domingues, C. a., Silva, J. G., Bastos, V. H., Budde, H., Cagy, M., Basile, L., Piedade, R., and Ribeiro, P. (2010). Sensorimotor integration: basic concepts, abnormalities related to movement disorders and sensorimotor training-induced cortical reorganization. *Revista de neurologia*, 51(7):427–436.

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., Schwalb, J. M., and Kennedy, S. H. (2005). Deep Brain Stimulation for Treatment-Resistant Depression. *Neuron*, 45(5):651–660.

Maynard, E. M., Nordhausen, C. T., and Normann, R. A. (1997). The Utah Intracortical Electrode Array: A recording structure for potential brain-computer interfaces. *Electroencephalography and Clinical Neurophysiology*, 102(3):228–239.

Mehler, W. R. (1962). The anatomy of the so-called "pain tract" in man: an analysis of the course and distribution of the ascending fibers of the fasciculus anterolateralis. *Basic research in paraplegia*, 26:55.

Mehler, W. R. (1969). Some neurological species differences-a posteriori. *Annals of the New York Academy of Sciences*, 167(1):424–468.

Monfils, M.-H., Plautz, E. J., and Kleim, J. A. (2005). In search of the motor engram: motor map plasticity as a mechanism for encoding motor experience. *The Neuroscientist*, 11(5):471–483.

Moser, M.-B., Rowland, D. C., and Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology*, 7(2):a021808.

Mouček, R., Ježek, P., Vařeka, L., Rondík, T., Brůha, P., Papež, V., Mautner, P., Novotný, J., Prokop, T., and Stěbeták, J. (2014). Software and hardware infrastructure for research in electrophysiology. *Frontiers in neuroinformatics*, 8(March):20.

Nunez, P. L. and Srinivasan, R. (2006). *Electric fields of the brain : the neurophysics of EEG*. Oxford University Press.

O'Doherty, J. E., Lebedev, M. a., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., and Nicolelis, M. a. L. (2011). Active tactile exploration using a brain–machine–brain interface. *Nature*, 479(7372):228–231.

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, 9(3):10–20.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011:156869.

Pais-Vieira, M., Yadav, A. P., Moreira, D., Guggenmos, D., Santos, A., Lebedev, M., and Nicolelis, M. A. L. (2016). A Closed Loop Brain-machine Interface

for Epilepsy Control Using Dorsal Column Electrical Stimulation. *Scientific Reports*, 6(1):32814.

Peschanski, M., Mantyh, P. W., and Besson, J. M. (1983). Spinal afferents to the ventrobasal thalamic complex in the rat: an anatomical study using wheat-germ agglutinin conjugated to horseradish peroxidase. *Brain research*, 278(1):240–244.

Petermann, T., Thiagarajan, T. C., Lebedev, M. A., Nicolelis, M. A. L., Chialvo, D. R., and Plenz, D. (2009). Spontaneous cortical activity in awake monkeys composed of neuronal avalanches. *Proceedings of the National Academy of Sciences*, 106(37):15921–15926.

Polavaram, S. and Ascoli, G. (2015). Neuroinformatics. *Scholarpedia*, 10(11):1312.

Polikov, V. S., Tresco, P. A., and Reichert, W. M. (2005). Response of brain tissue to chronically implanted neural electrodes. *Journal of Neuroscience Methods*, 148(1):1–18.

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

Quiroga, R. Q. and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(March):173–185.

Rational Software (2011). Rational Unified Process Best Practices for Software Development Teams.

Rey, H. G., Pedreira, C., and Quian Quiroga, R. (2015). Past, present and future of spike sorting techniques. *Brain Research Bulletin*, 119:106–117.

Rossant, C., Kadir, S. N., Goodman, D. F. M., Schulman, J., Hunter, M. L. D., Saleem, A. B., Grosmark, A., Belluscio, M., Denfield, G. H., Ecker, A. S., and Others (2016). Spike sorting for large, dense electrode arrays. Technical report, Nature Publishing Group.

Schouenborg, J. (2008). Action-based sensory encoding in spinal sensorimotor circuits. *Brain research reviews*, 57(1):111–7.

Schouenborg, J. (2016). *Development of Stable Long-term Electrode Tissue Interfaces for Recording and Stimulation*, chapter 2, pages 38–54. Wiley-Blackwell.

Schrader, S., Grün, S., Diesmann, M., and Gerstein, G. L. (2008). Detecting synfire chain activity using massively parallel spike train recording. *Journal of neurophysiology*, 100(4):2165–2176.

Schwartz, A. B. (1994). Direct Cortical Representation of Drawing. *Source: Science, New Series*, 265(5171):540–542.

Schwarz, D. A., Lebedev, M. A., Hanson, T. L., Dimitrov, D. F., Lehew, G., Meloy, J., Rajangam, S., Subramanian, V., Ifft, P. J., Li, Z., Ramakrishnan, A., Tate, A., Zhuang, K. Z., and Nicolelis, M. A. L. (2014). Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nature Methods*, 11(6):670–676.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423.

Sherman, S. M. and Guillery, R. W. (1996). Functional organization of thalamocortical relays. *Journal of Neurophysiology*, 76(3):1367–1395.

Sherrington, C. S. (1923). (1). THE INTEGRATIVE ACTION OF THE NERVOUS SYSTEM. *The Journal of Nervous and Mental Disease*, 57(6):589.

Siegle, J. H., Lopez, A. C., Patel, Y. A., Abramov, K., Ohayon, S., and Voigts, J. (2017). Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. *Journal of Neural Engineering*, 14(4):045003.

Smith, L. S., Austin, J., Baker, S., Borisyuk, R., Eglen, S., Feng, J., Gurney, K., Jackson, T., Kaiser, M., Overton, P., Panzeri, S., Quiroga, R. Q., Schultz, S. R., and Sernagor, E. (2007). The CARMEN e-Science pilot project : Neuroinformatics work packages . *PLoS Computational Biology*, (September):591–598.

Spiga, F., Lightman, S. L., Shekhar, A., and Lowry, C. A. (2006). Injections of urocortin 1 into the basolateral amygdala induce anxiety-like behavior and c-Fos expression in brainstem serotonergic neurons. *Neuroscience*, 138(4):1265–76.

Stancu-Mara, S., Baumann, P., and Marinov, V. (2014). A Comparative Benchmark of Large Objects in Relational Databases. Technical report, Jacobs University.

Sutter, E. E. (1992). The brain response interface: communication through visually-induced electrical brain responses. *Journal of Microcomputer Applications*, 15(1):31–45.

Teeters, J., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., Denisov, G., Siegle, J., Olsen, S., Martin, C., Chun, M., Tripathy, S., Blanche, T., Harris, K., Buzsáki, G., Koch, C., Meister, M., Svoboda, K., and Sommer, F. (2015). Neurodata Without Borders: Creating a Common Data Format for Neurophysiology. *Neuron*, 88(4):629–634.

Thelin, J., Jörntell, H., Psouni, E., Garwicz, M., Schouenborg, J., Danielsen, N., and Linsmeier, C. E. (2011). Implant size and fixation mode strongly influence tissue reactions in the CNS. *PLoS ONE*, 6(1).

Thorbergsson, P. T., Garwicz, M., Schouenborg, J., and Johansson, A. J. (2010). Statistical modelling of spike libraries for simulation of extracellular recordings in the cerebellum. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4250–4253. IEEE.

Torre, E., Canova, C., Denker, M., Gerstein, G., Helias, M., and Gruen, S. (2016). ASSET: analysis of sequences of synchronous events in massively parallel spike trains. *PLoS Comput Biol*, 12(7):e1004939.

Truccolo, W., Donoghue, J. A., Hochberg, L. R., Eskandar, E. N., Madsen, J. R., Anderson, W. S., Brown, E. N., Halgren, E., and Cash, S. S. (2011). Single-neuron dynamics in human focal epilepsy. *Nature Neuroscience*, 14(5):635–643.

Umuroglu, Y., Fraser, N. J., Gambardella, G., Blott, M., Leong, P., Jahre, M., and Vissers, K. (2016). FINN: A Framework for Fast, Scalable Binarized Neural Network Inference. In *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays - FPGA '17*, pages 65–74, New York, New York, USA. ACM Press.

Velliste, M., Kennedy, S. D., Schwartz, A. B., Whitford, A. S., Sohn, J.-W., and McMorland, A. J. C. (2014). Motor cortical correlates of arm resting in the context of a reaching task and implications for prosthetic control. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34(17):6011–22.

Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature*, 453(7198):1098–101.

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., and Nicolelis, M. A.

(2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810):361–5.

Wikipedia, O. b. e. . (2017). Schematic of an action potential.

Williams, J. C., Hippensteel, J. A., Dilgen, J., Shain, W., and Kipke, D. R. (2007). Complex impedance spectroscopy for monitoring tissue responses to inserted neural implants. *Journal of Neural Engineering*, 4(4):410–423.

Willis, W. D., Kenshalo, D. R., and Leonard, R. B. (1979). The cells of origin of the primate spinothalamic tract. *The Journal of Comparative Neurology*, 188(4):543–573.

Willis, W. D., Zhang, X., Honda, C. N., and Giesler, G. J. (2001). Projections from the marginal zone and deep dorsal horn to the ventrobasal nuclei of the primate thalamus. *Pain*, 92(1):267–276.

Wolpaw, J. R., McFarland, D. J., Neat, G. W., and Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography and clinical neurophysiology*, 78(3):252–9.

Woolf, C. J. (2012). Central sensitization: implications for the diagnosis and treatment of pain. *Pain*, 152(3 Suppl):S2–15.

Wörgötter, F., Suder, K., Zhao, Y., Kerscher, N., Eysel, U. T., and Funke, K. (1998). State-dependent receptive-field restructuring in the visual cortex. *Nature*, 396(6707):165–8.

Zehl, L., Jaillet, F., Stoewer, A., Grewe, J., Sobolev, A., Wachtler, T., Brochier, T. G., Riehle, A., Denker, M., and Grün, S. (2016). Handling Metadata in a Neurophysiology Laboratory. *Frontiers in neuroinformatics*, 10(July):26.

Zemlan, F. P., Leonard, C. M., Kow, L.-M., and Pfaff, D. W. (1978). Ascending tracts of the lateral columns of the rat spinal cord: a study using the silver impregnation and horseradish peroxidase techniques. *Experimental neurology*, 62(2):298–334.

Scientific publications

Author contributions

Co-authors (in-alphabetical order) are abbreviated as follows:
Leila Ehtemadi (LE), Martin Garwicz (MG), Tanja Jensen (TJ), Anders J Johansson (AJ), Gustav Lind (GL), Per Petersson (PP), Jens Schouenborg (JS), Jonas Thelin (JT), Fotios Tsanakalis (FT).

Paper i: A novel framework for storage, analysis and integration through mediation of large-scale electrophysiological data

I constructed the software architecture and object model and wrote the majority of the paper. JS, MG, AJ and PP provided feedback during the development of the software and co-authored the paper.

Paper ii: Discrepancies between cortical and behavioural long-term readouts of hyperalgesia in awake freely moving rats

I wrote all software, structured and performed the statistical analysis of the electrophysiological data, constructed the electrophysiology setup prototype together with PP, performed laser mapping experiments together with LE, FT and TJ. TJ structured and performed the behavioral tests and performed the statistical analysis of the behavioral data. JT and GL performed surgery. JS wrote the majority of the paper, with TJ, JT, MG, PP and myself as co-authors.

Paper iii: A bit-encoding based new data structure for time and memory efficient handling of spike times in an electrophysiological setup

I developed the bit encoding format as well as the software architecture. MG and JS gave feedback during the development of the software architecture and data format. I wrote the majority of the paper with MG, JS, PP and AJ as co-authors.