Brain Research Bulletin 119 (2015) 118–126

Contents lists available at ScienceDirect

Brain Research Bulletin

journal homepage: www.elsevier.com/locate/brainresbull

Review

Microelectronics, bioinformatics and neurocomputation for massive neuronal recordings in brain circuits with large scale multielectrode array probes

Alessandro Maccione[a,*], Mauro Gandolfo[b], Stefano Zordan[a], Hayder Amin[a], Stefano Di Marco[c], Thierry Nieus[a], Gian Nicola Angotzi[a], Luca Berdondini[a]

[a] NetS3 Laboratory, Neuroscience and Brain Technologies, Fondazione Istituto Italiano di Tecnologia, Genova, Italy
[b] 3Brain GmbH, Landquart, Switzerland
[c] Università degli Studi dell'Aquila, Dipartimento di Scienze cliniche applicate e biotecnologiche, Aquila, Italy

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form 22 July 2015
Accepted 24 July 2015
Available online 29 July 2015

Keywords:
CMOS-MEAs
Electrophysiology
Data analysis
Neurocomputation
Big data

ABSTRACT

Deciphering neural network function in health and disease requires recording from many active neurons simultaneously. Developing approaches to increase their numbers is a major neurotechnological challenge. Parallel to recent advances in optical Ca^{2+} imaging, an emerging approach consists in adopting complementary-metal-oxide-semiconductor (CMOS) technology to realize MultiElectrode Array (MEA) devices. By implementing signal conditioning and multiplexing circuits, these devices allow nowadays to record from several thousands of single neurons at sub-millisecond temporal resolution. At the same time, these recordings generate very large data streams which become challenging to analyze.

Here, at first we shortly review the major approaches developed for data management and analysis for conventional, low-resolution MEAs. We highlight how conventional computational tools cannot be easily up-scaled to very large electrode array recordings, and custom bioinformatics tools are an emerging need in this field. We then introduce a novel approach adapted for the acquisition, compression and analysis of extracellular signals acquired simultaneously from 4096 electrodes with CMOS MEAs. Finally, as a case study, we describe how this novel large scale recording platform was used to record and analyze extracellular spikes from the ganglion cell layer in the wholemount retina at pan-retinal scale following patterned light stimulation.

© 2015 Elsevier Inc. All rights reserved.

Contents

* Corresponding author.
E-mail address: alessandro.maccione@iit.it (A. Maccione).

http://dx.doi.org/10.1016/j.brainresbull.2015.07.008
0361-9230/© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Neurotechnologies allowing to increase the number of simultaneously recorded spiking neurons in large brain circuits are essential to advance our understanding of functioning and dysfunctioning networks. Developing such technologies is challenging (Marblestone et al., 2013) and requires to introduce novel interdisciplinary approaches to improve existing methodologies for recording and data analysis of unprecedented large-volumes of electrophysiological signals.

This challenging goal has been developed since the 1960's, either by gradually improving extracellular electrode array recordings or functional Ca^{2+} imaging technologies. These efforts have resulted in approximately doubling the number of simultaneously recorded single units in-vivo every other year (Stevenson and Kording 2011), allowing a few laboratories to record from up to several hundreds of neurons. More recently, implantable microendoscopes were developed to record Ca^{2+} signals from more than a thousand of neurons located over less than 1 mm^2 in a single cortical region (Ziv et al., 2013), or from a few hundreds of neurons in two closely spaced regions, i.e., V1 and LM (Lecoq et al., 2014). In contrast to optical approaches, electrodes remain a valuable solution for label-free, high signal-to- noise, sub-millisecond resolution extracellular recordings of spiking activity and field potentials. Remarkably, dense microwire electrode arrays have recently made it possible to perform five years long chronic recordings from up to ~500 cortical neurons distributed across multiple cortical areas (Schwarz et al., 2014). In addition, over the last decade cutting edge developments in engineering have provided an alternative technological substrate for the realization of very large electrode arrays based on microelectronic monolithic chips. These new devices now make it possible to substantially increase the current number of simultaneously recorded neurons, while taking advantage of the quality of electrode sensing. Here, we focus on such emerging microelectronic electrode arrays and in particular on the computational challenges that need to be addressed for managing the ensuing volume of recorded data.

Briefly, active microelectronic MultiElectrode Arrays (MEAs) integrate on-chip circuits for addressing the electrodes, for amplification and for signal multiplexing. Conventional (passive) electrode arrays rely on the individual wiring of each electrode to off-chip amplifiers, thus constraining the electrode density and the number of locally recording electrodes (i.e. typically 64-512). Consequently, the capability to record neuronal activity on passive devices is limited to few hundreds of neurons, resulting in strong undersampling of the neuronal activity within a network that is generally composed by several thousands of neurons.

More than 15 years ago we, as well as other laboratories, have started investigating MEA devices with several thousands of simultaneously recording electrodes that do not rely on wiring of individual electrodes. To that aim, monolithic active microelectronic circuits realized with complementary-metal-oxide-semiconductor (CMOS) technology and post-processing micro-/nano-technologies were implemented and tested in-vitro, in particular on cell cultures (Berdondini et al., 2014a,b; Maccione et al., 2010; Lambacher et al., 2011; Hierlemann et al., 2011), brain slices (Ferrea et al., 2012; Frey et al., 2009) and retinal whole mounts (Fiscella et al., 2012; Maccione et al., 2014). Addressing and multiplexing circuit solutions, i.e., to sequentially scan the electrode array at high-frequency and read-out extracellular signals sensed from thousands of electrodes on a few output channels at sub-millisecond resolution has respectively allowed to reduce the electrode separations to a few micrometers and to record neuronal activity spanning large areas extending over several mm^2. This novel generation of MEAs have recently become commercially available for in-vitro electrophysiology (i.e. from 3Brain GmbH www.3Brain.com and MultiChannelSystems www.multichannelsystems.com) and can nowadays provide whole array recordings from more than 4000 closely spaced electrodes at sampling rates ranging from 7 to 25 kHz/electrode. More recently, CMOS approaches were also proposed for in-vivo probes (Lopez et al., 2014; Ruther and Paul, 2014).

Undoubtedly, such an upscaling in the number of simultaneously recording electrodes requires adapted computational approaches to manage the massive volume of signals and their analysis. As already developed for genomic data with centralized servers for storage and ad hoc compression algorithms (Brandon et al., 2009), bioinformatic approaches can be adopted. Here, we first briefly summarize the main software tools available for conventional MEAs consisting of few tens of electrodes. Given the relatively small volume of data acquired with these systems, except for a few pioneering studies, this field still suffers from a lack of advanced data formats, data compression and optimized analysis algorithms. In addition, many laboratories rely on custom tools designed for specific analyses. Then, based on our experience gained in the development and experimental application of large-scale CMOS-MEAs, we describe a chain for the acquisition and analysis of large electrophysiological datasets based on a custom SW/HW platform providing simultaneous recordings from 4096 electrodes at ~7 kHz. Considering that ten minutes of continuous recording from our system generates ~30 GBytes of raw data, hard drives of 1-2 TeraBytes, that are typically mounted on workstations are inadequate for storing multiple experimental sessions. Finally, as a case-study, we present the workflow used to analyze responses to patterned light recorded with that device from a large population of retinal ganglion cells in the whole mount mouse retina.

2. Available data formats, data sharing and analysis tools for multielectrode arrays

Commercial MEA systems based on passive devices (i.e. 64–512 electrodes) usually provide only a few software tools, mainly for handling data and for enabling the users to develop their own custom analysis algorithms. These systems implement their own proprietary data formats (e.g. .mcd for MultiChannelSystems, or .nex for Plexon Neuroexplorer www.plexon.com) and ad-hoc libraries and wrappers are typically provided as third party applications to access the data for analysis. In some cases, tools for exporting data to Matlab files (or to text files) are available directly in the acquisition software. Common open source data formats, such as those developed in the Neuroshare project (www.neuroshare.sourceforge.net/) or NEO (www.pythonhosted.org/neo/), were proposed to provide hierarchical data models and libraries to support interoperability between different data formats. For the analysis, different open source analysis libraries are available (e.g. Bologna et al., 2010; Huang et al., 2008; Egert et al., 2002), but they are generally not optimized in terms of computational costs. Even though the analysis of signals from only a few electrodes is not very computationally demanding, once multiple sessions of the same experiment are acquired and repetitive analysis is required, it can become a serious bottleneck. In fact, conventional approaches are generally not optimized in terms of read and write routines and typically involve many intermediate steps that require direct conversions or external libraries to access the data. Approaches to solve the bottleneck of the computational overload were proposed (Mahmud et al., 2012; Sobolev et al., 2014), but their use typically requires a deep knowledge in software programming. Alternatively, in the last years there has been an increasing

effort to identify more appropriate data formats and structures to handle the analysis and to share data and results. A remarkable example is the web portal Carmen (https://portal.carmen.org.uk), a repository for data storage and exchange between collaborators. It also allows to perform online data analysis (Eglen et al., 2014). Additionally, the Neurophysiology Data Translation Format (NDF) was developed within the CARMEN project. It consists of a configuration file in XML format that contains metadata and references to the associated data files. This file format presents the advantage of accessing metadata without the need of loading the entire file and it is well suited for web applications to provide the end user with analysis reports that exploit high-performance computational infrastructures. Another interesting approach is the NIX project (previously called Pandora) which was developed by the Electrophysiology Task Force in the context of the International Neuroinformatics Coordination Facility (INCF) Datasharing Program (https://github.com/G-Node/nix/wiki). This consortium proposed a highly generic structure of data and metadata based on a "open metaData Markup Language" (odML) model that provides a flexible architecture for storing any information as extended key–value pairs in a hierarchical organization (Grewe et al., 2011). Interestingly, to support the model structure, the Hierarchical Data File (HDF) technology (https://www.hdfgroup.org/) was used. This presents several advantages, by allowing to store large datasets (even in the order of TeraBytes) with theoretically no limit and enabling the access to subsets of data without the need of loading or reading the entire data stream and by enabling parallel multiple input/output operations. All these aspects make the HDF file format a perfect candidate for implementing a flexible and computationally optimized structure for handling massive neuronal recording from thousands of electrodes.

3. Implementation of an adapted analysis framework for a 4096CMOS-MEA platform

3.1. Platform architecture

A schematic representation of the 4096CMOS-MEA platform (for a review see Berdondini et al., 2014) and of the distributed processing tasks is illustrated in Fig. 1. The platform consists of a CMOS-MEA chip, a real-time interface hardware (i.e., FPGA board) and a Host PC running a dedicated software application. Here, we briefly describe these different components and their role in sampling and processing extracellular signals.

The CMOS-MEA chip integrates a 64×64 electrode-pixel array with amplification and low-pass filtering circuits as well as addressing and multiplexing on-chip functionalities for the read-out of electrode signals on 16 output analogue lines. In particular, the CMOS-MEA chips were designed with a scalable circuit architecture (Imfeld et al., 2008) that can be up-scaled in the number of simultaneously recording electrodes and achieving sub-millisecond temporal resolution when performing whole array recordings (7 kHz up to 18 kHz on the novel platform of 3Brain GmbH). To do so, the circuit is based on the Active Pixel Sensor (APS) concept that was originally developed for high-speed light imaging sensors (Berdondini et al., 2014). Active electrode-pixel circuits ($42 \times 42\ \mu m^2$) integrating a microelectrode ($21 \times 21\ \mu m^2$) and a local amplifier with a gain of ~40 dB are integrated in a regular 64×64 array. Originally, to meet the small area constraints imposed by the pixel size and in order to avoid saturating the in-pixel amplifier with large amplitude and drifting DC-offsets caused by charge accumulation when the chip is directly exposed to light, the in-pixel amplifier was not designed with an AC-coupling (Harrison and Charles, 2003; Mollazadeh et al., 2009) that requires a large capacitor, rather with an auto-zeroing circuit (Fig. 1, in-pixel circuit inset). This circuit solution allows to manage drifting DC-offsets under normal experimental conditions as well as under exposure to direct light (up to 100–300 $\mu W/mm^2$, unpublished). In this case, light exposures result in faster DC drifts that can be managed by increasing the calibration frequency to maintain the amplifiers at their working point. In addition to the recording capabilities from the 4096 electrode array, a more recent generation of chips also provides an array of 16 electrodes, interleaved within the recording array, for electrical stimulation ($21 \times 21\ \mu m^2$ in size, pitch of 81 μm).

The real-time hardware for data acquisition is based on an FPGA board. It controls the CMOS-MEA operation and sends at high-frequency the addresses of the electrode-pixels that have to be read-out on the output multiplexed analogue lines. On the same board, these analogue signals are then converted to digital signals at 12 bits resolution. In addition, the FPGA implements a programmable high-pass digital filter and sends the data to a frame grabber installed on the PC via a high-speed Camera-link interface. The same platform includes input ports for externally generated signals (e.g. time-stamps of light projected pattern stimuli to stimulate the retina), or for externally generated electrical stimulation signals (e.g. using the 16 channels PlexStim Electrical stimulator, Plexon, USA). Additionally, it is possible to output a single electrode signal for implementing closed-loop experimental setups.

The host PC controls the full platform and implements a software (BrainWave, 3Brain GmbH) that provides on-line and off-line tools to process, visualize, store and export the recorded data. The same software tool provides off-line tools for data mining, event detection and basic analysis (e.g. Mean Firing Rate and Inter Spike Interval). In order to perform online processing, visualization and data storage, the acquisition core of the software is organized as a circular memory buffer, with one producer and multiple concurrent consumers of data. In addition, ad-hoc routines were developed for accessing the same area of the buffer simultaneously and for optimizing the writing operations on the hard drives (up to 140 MB/sec at 18 kHz in the novel platform). The acquired data are stored on the local hard drive and are in our case transferred to a server for further analysis.

3.2. Data compression and structured file format of the acquisition platform

In order to manage large-scale electrophysiological experimental datasets that can be composed of multiple recording sessions, structured files with metadata and online data compression strategies are needed.

In our platform, the software embeds online event detection algorithms and waveform extraction tools that allow to store the acquired signals using different modalities, either as compressed raw data or by using different levels of compression to reduce the file size and to facilitate data sharing (Fig. 2a). Two compression modalities are currently available, namely the "spike-mode" (i.e., time-stamps of the detected events) or the "blanking mode" (i.e., only an interval of time of raw data around the detected events, see Fig. 2b). The "blanking mode" allows not only to store the electrodes where the events are detected, but also to store raw data of neighboring sites (Fig. 2c). This allows to apply emerging spike sorting algorithms that exploit the redundant signals of the same neuron recorded from nearby closely spaced electrodes as has already been described in another type of high density MEA system in (Litke et al., 2004). It also defines spatially distributed spike templates for optimizing unit discrimination (Marre et al., 2012), saving space into the hard disk without losing any spatiotemporal information of the signals.

For the event detection, three different online algorithms are provided in the software, i.e., a simple Hard Threshold, a Differen-

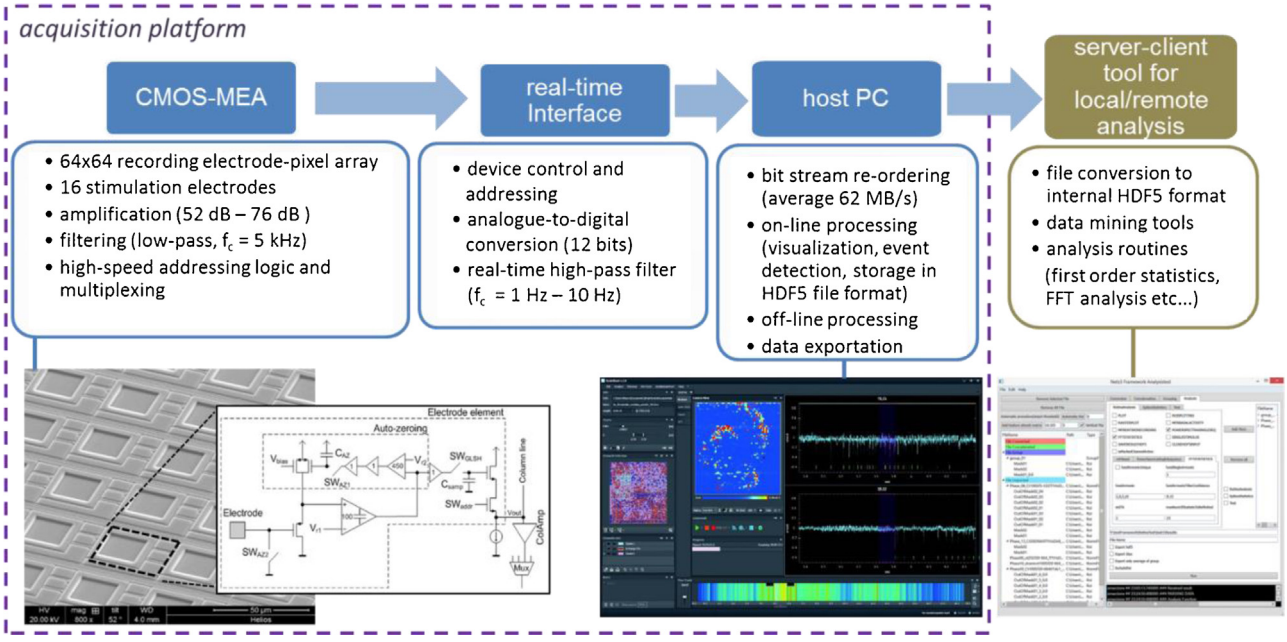

Fig. 1. Overview of the 4096CMOS-MEA platform adapted to large-scale electrophysiological data acquisition and analysis. The platform is composed of (left) an acquisition platform based on CMOS-MEAs and enabling simultaneous recordings at sub-millisecond resolution from 4096 electrodes (real-time hardware and optimized software tools from 3Brain GmbH, Switzerland); (right) and of a custom server-client tool for local or remote data analysis with minimized data transfer.

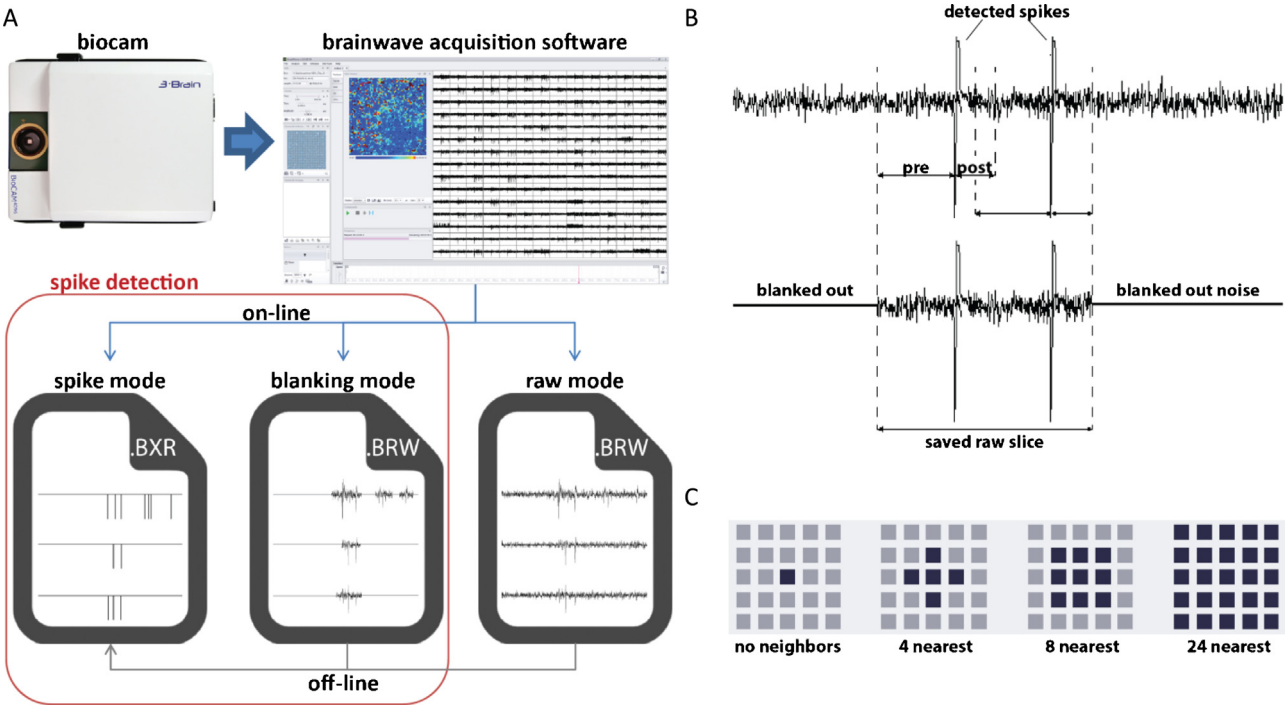

Fig. 2. Acquisition and compression modalities implemented on the CMOS-MEA platform. (A) The instrument software acquires data from the hardware and, through online algorithms, it can store extracellular signals using three different modalities. These are characterized by an increasing rate of compression (from right to left): raw mode – the full raw data is stored; blanking mode – only chunks of data around a detected event are stored; spike mode – only the time stamps of the detected spikes are stored. (B) Working principle of the *blanking mode*: pre- and post-time intervals with respect to a detected spike are used to define the chunk of signal that is stored. (C) The blanking mode can also store neighboring electrodes for spike detection and analysis that exploit signals recorded by adjacent electrodes.

tial Threshold (Novellino et al., 2009) and the Precise Time Spike Detection (Maccione et al., 2009). The simultaneous online spike detection on 4096 electrodes is enabled by the optimized multi-threading 64bit architecture of the software and by the use of fast, yet affordable, workstations (e.g., Intel quad-core 3 GHz with 8GByte RAM 16,000 MHz complemented with 2HD SATA3 in raid 0 configuration).

On this platform, the acquired data (rate of ~60 MB/s at 7 kHz/el.) are directly stored as HDF5 files, organized in a structure including metadata information as the array layout, the software and hardware version and also specific fields dedicated to the recording modalities. Briefly, two types of files can be generated. These are obtained either when storing raw data (i.e., BRW file), or as a result file after an online/offline analysis step (i.e., BXR file), thus

Table 1
Comparison of the file sizes obtained with the different compression modalities available on the platform. The quantifications are referred to an acquisition of 10 min (~850000 detected spikes) from 4096 electrodes of a spontaneously active hippocampal neuronal culture (24 DIV).

	Raw mode	Blanking mode	Spike + wave	Spike
File extension	.brw	.brw	.bxr	.bxr
Dimension	32 GByte	0,98 GByte	27,4 MByte	2,5 MByte
Notes	Sampling rate 7 kHz/channel	Data chunk 10msec with 8 nearest neighbors	Waveform 4 msec	–

allowing the user to tune the amount of stored information and to manage the physical dimension of the dataset.

For the first, the "raw mode" fills a Raw matrix, where each column is an electrode and each row is a given frame (i.e., extracellular potentials recorded by the array of 64 × 64 electrodes), while the "blanking mode" fills two separated arrays named RawEncoded and RawEncodedTOC. The former stores chunks of data of the extracted waveforms, while the latter stores the positions in the RawEncoded array at fixed time intervals, respectively. A chunk of data in the RawEndoded array contains data from the electrodes where portions (ranges) of relevant data were identified. Each electrode data then consists of a tag identifying the electrode, the size in bytes of the electrode data and a series of one or more saved ranges. A range contains the raw data associated to one or more, but sufficiently adjacent to be merged together, detected events (see Fig. 2b). An electrode data chunk can have more ranges since there can be consecutive events on the same channel. Each range is then composed by the starting and the ending frames of the range and by the raw data between them. Since the position of a certain time interval in the RawEncoded array depends on the spiking rate, the length positions in the array are stored in the RawEncodedTOC array at fixed time intervals. In this way, when data are requested, it is possible to jump immediately at the correct position of the RawEncoded array without the need to read the full array.

For the second file format, only the occurrence of the detected events is stored. This file does not contain any raw data and can also be obtained offline by executing algorithms of event detection and waveform extraction on one of the raw data files. This file structure still contains metadata in common with the .brw files, but data are stored into three arrays in a section field named 3BChEvents. Furthermore, complementary data as for the timestamps of detected artifacts can be added, as needed for the case of electrical stimulation.

In order to illustrate the effectiveness of the compression methods implemented in this platform, Table 1 compares the different file sizes obtained by analyzing 10 min of recording of spontaneous spiking activity from a hippocampal culture aged of 24 Days In Vitro (DIV). This dataset provides >3000 active electrodes (i.e. spiking activity > 0.01 spike/sec) and a total of about 850000 detected spikes. As it can be noticed, depending on the level of compression, the file size can scale down from tens of GBytes to a few MBytes.

3.3. A Python™ based client-server environment for analysis and data exchange

The emerging CMOS-MEA technology not only needs new strategies for data acquisition and compression, but also introduces challenging issues with respect to computational cost in data analysis and data sharing. Here, we propose a client-server architecture designed for the analysis of large-electrode array data, with a scalable data format that can include further information to be correlated with the electrical recordings (e.g., visual stimuli time events in a light-evoked retina responses experimental paradigm). This implementation is fully interoperable and provides management tools and a graphical-user-interface (GUI) either for grouping datasets of different experiments or for selecting specific subsets of electrodes within a dataset. We chose the open source Python (www.python.org) language as development environment, since it is fully integrated with the h5 format and it is becoming a standard in neuroscience.

3.3.1. Platform architecture

The Python language (2.7 release), complemented by HDF5 and PyTable modules for data handling, and QT libraries for the GUI (http://qt-project.org/), was adopted to guarantee fully portability among different operating systems. Socket, lock and threading modules have been used to design a client-server application. As shown in the diagram of Fig. 3a, the platform is organized by distinguishing two working modalities, i.e., "local" mode (green paths) and "remote" mode (red paths). Classes are grouped into four main layers: a GUI/client Layer (GUIL), the Kernel Layer (KL), the Remote Communication Layer (RCL) and a Server Layer (SL). The first two layers represent the core of the platform and are used both for the local (green dashed lines) and remote modalities (red dashed lines). These layers include the GUIs and all classes and functions for the manipulation, organization and analysis of the data. The RCL layer (installed on both the server and the client) and the SL layer are needed only for the remote modality. While in local mode the entire architecture is on the user PC, in the remote mode, the RCL layers communicate with a server class via a TCP/IP socket protocol. However, large-volume data exchanges between the client and the server would be limiting the analysis performances of the client-server mode. For this reason, the system was designed to minimize such data transfer. In addition, experimental data can be stored in a separated remote server, thus leaving the researcher to choose his own data repository strategy and to minimize the transfer of large experimental files. To achieve this, the client only transmits to the server the name and parameters of the function to be executed and the data file paths (that are automatically converted to the relative paths of the server) and retrieves the results.

Finally, in order to guarantee parallelization and increased computational performances the platform was designed as a multitasking architecture while dictionary objects exchanged between different threads running on the server and the client are managed by a Sender/Receiver queue. When the Receiver has accomplished a requested task, it informs the Sender who made the request, thus guaranteeing an optimized asynchronous parallel management of multiple tasks.

3.3.2. Graphical-user-interface (GUI) and functionalities

The server-client analysis tool provides several organized GUIs to perform the following tasks:

1. *Conversion to internal HDF5 file format*: the internal file format allows to integrate complementary information coming from different instruments (e.g., electrical stimulator, light projecting systems) that work in parallel with the CMOS-MEA platform.
2. *Concatenation of multiple files*: different recordings can be merged and analyzed as a unique file.
3. *Clustering of electrode channels within a file*: subsets of electrodes can be selected and stored to the internal HDF5 file to be analyzed separately. The selection of the electrodes can be either manual or assisted by the data integrated in the internal file format.

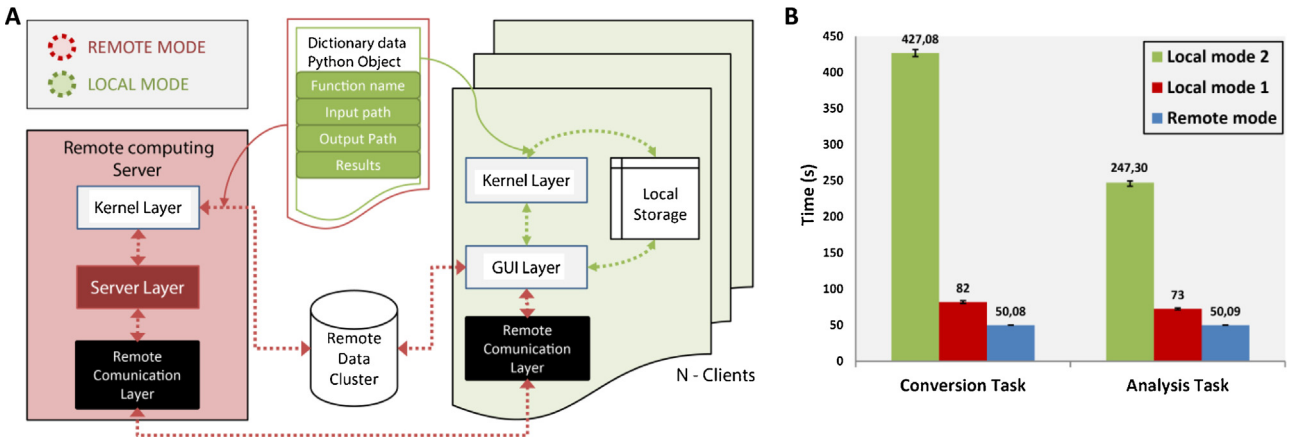

Fig. 3. Schematic representation of the software organization and computational performances. (A) Architecture of the client-server analysis tool. The same tool allows for both local (green) and remote (red) analysis by minimizing data transfer and by exploiting data storage located on a remote network. (B) Comparison of the computing performances obtained with local and remote modalities. A task of data conversion and a task of data analysis (computation of the mean firing rate, mean bursting rate, mean intra fire burst, mean burst duration, bursting network and averaging over 5 trials per task) were performed on 28 dataset, each one consisting of 10 min of recorded spontaneous activity from a hippocampal cultures. Each dataset provides an average of 2600 active electrodes (>0.01 spike/sec) for a total of about 1000000 spikes. The performances of three computational hardware are compared and include the use of the tool either locally or remotely with the server-client configuration.

The selections of electrode subsets can be stored as templates and then applied to other files for successive analysis.

4. *Grouping of files*: files can be collected in groups. During the analysis each file will be at first analyzed separately and the results obtained from different experimental data can be averaged together.

5. *Analysis of datasets*: first order statistical parameters, e.g., Mean Firing Rate, Mean Bursting Rate, Mean Intra Fire Burst, Mean Burst Duration, Bursting Network and Power Spectrum analysis, are included in the platform. New analysis tools can be easily added without deep *a-priori* knowledge of the platform structure. A researcher can create his own python script with a custom analysis function, locate it in a defined directory of the software, and include it in the software by editing an XML configuration file. In this way, the analysis file will be loaded by dynamically creating an analysis tab on the GUI to set the new function parameters. A similar procedure can be executed in order to include different input data formats.

3.3.3. Computational performances

The high configurability of the platform allows to exploit different hardware architectures to optimize the computational execution time of a given analysis. As an example, Fig. 3b shows the comparison in performance obtained on different machines and configurations, when executing a data conversions and an analysis task (i.e. computation of different first order statistical parameters) on 27 datasets, counting approximatively 1000000 detected spikes each. More specifically, these tests were performed using the "Remote mode" on a Server machine (Ubuntu Server SO) that mounts 4Hexa core AMD Opteron 6238 s 1400 MHz CPU (six cores per processor), 120 GB of RAM; the "Local mode 1" an Apple MacBook pro 2,8 GHz Intel Core i7 (4core) with 16 GByte of DDR3 RAM at 16,000 MHz with an SSD 1 TByte hard disk; and using the "Local mode 2" on a workstation "dell precision T7400" with an Intel® Xeon® X5472 Quad-core 3 GHz and 6 GB of RAM. As shown, task parallelization on the server clearly provides better performances up to eight times in conversion time and five times in analysis time with respect to the configuration used for "Local mode 2". Interestingly, with respect to "Local mode 1" that takes advantage of an SSD disk for read/write operations, the gain of the client server configuration for this analysis task is of about 25%.

4. Adoption of the proposed platform for pan-retinal electrophysiological recordings of light evoked RGCs responses on CMOS-MEAs

As a case study aimed at illustrating the application of the proposed platform, we report here an experimental paradigm designed for studying the effects induced by the extent of a projected image on the spatial tuning properties of Retinal Ganglion Cells (RGCs) receptive fields. The physical dimensions of the chip (2.67×2.67 mm^2) allow simultaneous recordings of RGCs light-evoked activity from almost the entire retina of a mouse. This is an ideal scenario to study how an image is encoded in the retina: light-stimuli are projected on the photoreceptors layer and responses from RGCs are recorded by the CMOS-MEA.

The receptive field of a RGC was defined as the area in which light-stimulation evokes a response in that neuron (Levine and Shefner, 1991). Retinal receptive fields are organized in a center-surround antagonistic fashion (e.g. central excitation and surround inhibition). The consequence of this organization is that for a RGC that has an excitatory center and inhibitory surround, stimulating the receptive field with a bar that is as large as the central part of the receptive field will evoke the largest spiking response that can be obtained by that RGC. When the size of the bar increases beyond the limit of the receptive field center and begins to invade the surround, the response of the RGC will decrease as a function of the increase in the bar size. This property of the receptive field is called "spatial tuning" and is a function of the amount of suppression generated by the RF. The larger the suppression, the sharper the spatial tuning of a specific RGC. The logical consequence of this is that RGCs may differently encode two identical images that only differ in their projection sizes on the retina (in our case, one quarter and one eighth of the retina, respectively).

To test this hypothesis, we presented moving bar-stimuli with different spatial frequencies (from 0.045 to 3 cycle/degree at 1 Hz), moving in the four cardinal directions, within two concentric areas of the retina. In addition, to investigate that the observed changes involve the synaptic inhibition in the surround, the same set of stimuli was applied while perfusing the retina with Bicuculline (10 μM), a competitive antagonist of $GABA_A$ receptors that are known to contribute to the inhibitory surround of the RF.

We acquired and analyzed light-evoked electrophysiological responses from RGCs from mouse retinal wholemounts on CMOS-MEAs. As schematically represented in Fig. 4a, ex-vivo light evoked

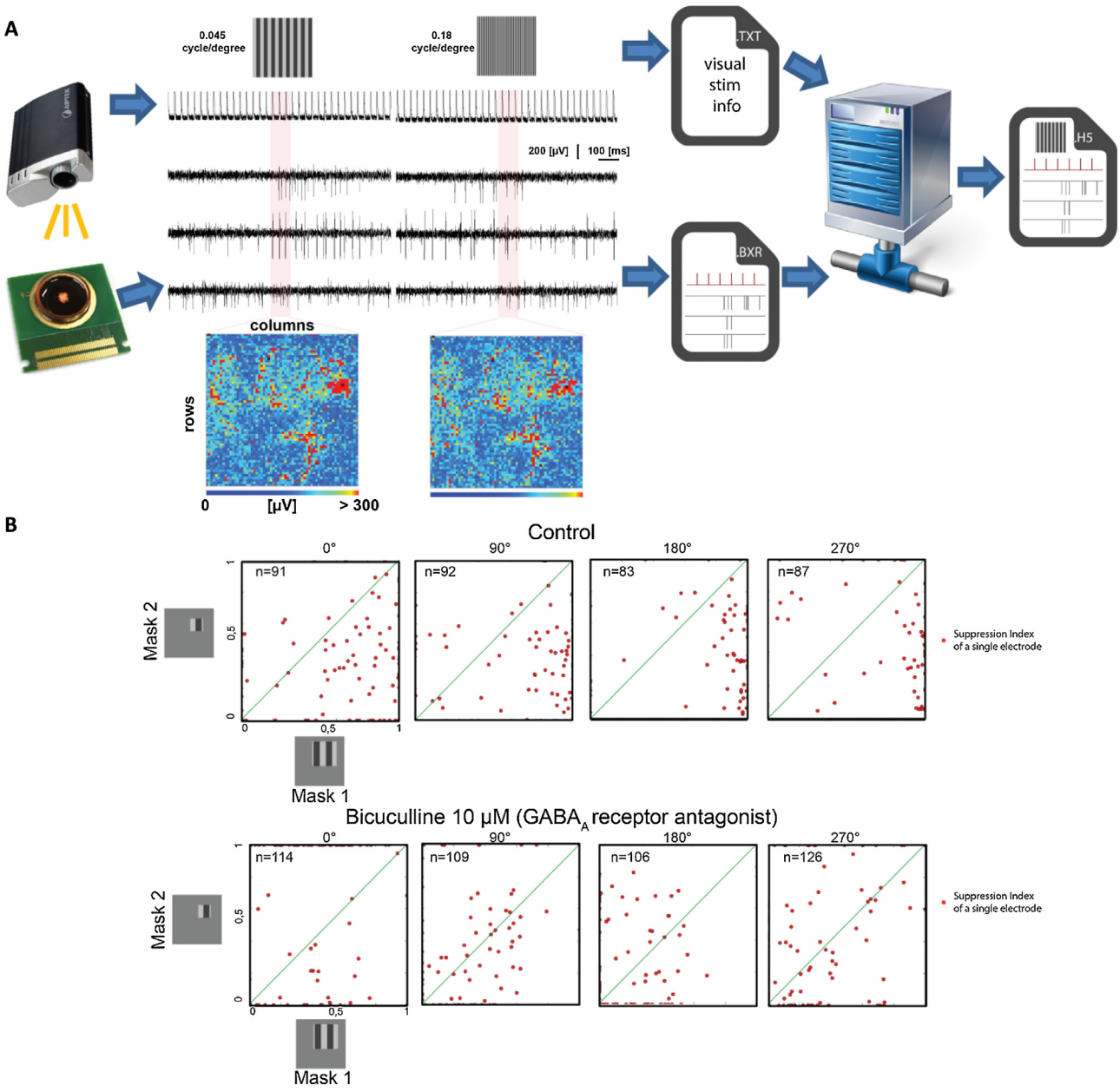

Fig. 4. Example of the platform application for electrophysiological studies on mouse explanted retinas subjected to patterned light stimulations. (A) Overview of the experimental and analysis flow. (from left to right) retinas, with the ganglion cell layer contacting the electrode array, are subjected to light stimuli generated with a Digital Light Projector. Illustrative example showing light evoked responses to moving bars with different spatial frequencies as well as the trigger signals, corresponding to each projected frame, that are aligned and simultaneously recorded with the electrophysiological signals. Below, false colour maps representing the signal variation (100 ms interval) recorded during stimulation from the 64 by 64 electrode array are reported and show extended light evoked RGCs responses. Files containing the detected spike events and the time stamps of the trigger signals are combined with a text file and converted into the internal HDF5 file. (**B**) Comparison of the computed Suppression Index (SI) for the recorded RGCs responses to moving bars following four main directions, both (up) under control conditions and (down) under pharmacological manipulation (Bicuculline, 10 µM). Each dot represents an electrode and the X and Y coordinates are the SI values calculated when the stimulation extent covers a quarter or one eighth of the retina, respectively. Not surprisingly, the spatial tuning is lost under bicucculine, thus indicating the relevant role of lateral inhibition.

responses of RGCs are recorded at pan retinal scale by taking advantage of the 4096 electrode array that is here combined with a Digital Light Projector (DLP) (cf. *Platform architecture*). For this type of experiments, the spiking activity sensed by the electrode array and the trigger signals corresponding to the time-stamp of all visual stimuli, were both recorded and stored as .bxr files after executing the Precise Time Spike Detection algorithm implemented on BrainWave acquisition software (cf. *Data compression and structured file format of the acquisition platform*). The detected events were then converted into the internal .HDF5 format of the server-client analysis platform and completed with text files containing information on the associated light-stimuli.

The subset of electrodes that was always exposed to each visual stimulus was selected and considered for the analysis of all the recorded phases. Recordings under control/pharmacological blockade conditions and for both projection conditions were grouped separately. All these procedures were easily executed by using our client-server analysis platform as described in Section 3.3.2.

For this analysis, spike sorting was not performed. Indeed, from previous unpublished data we observed >70% of the total recor-

ding electrodes are typically acquiring from a single unit (Offline Sorter from Plexon, www.plexon.com). The purpose of the experiment reported here is mainly to show the ability of the proposed platform to handle large and complex biological data and not to perform a rigorous study on the RF properties of RGCs. However, the effects observed by blocking $GABA_A$ receptors, strongly support the rationale of the experiment, thus confirming that the described phenomena are purely biological effects and not artifacts. By considering the high probability of recording from a single RGC from individual electrode, the response of RGCs was evaluated by Fourier analysis of neural responses at the fundamental input frequency of stimulation (i.e. the inverse of the temporal period of the moving bars). Finally, following an existing methodology (Di Marco et al., 2013), we computed the Suppression Index (SI) by using the expression:

$$SI = 1 - \left(\frac{R_{large}}{R_{peak}}\right)$$

where, R_{large} is the response generated by the largest bars size and R_{peak} is the peak response. As a result, the SI index is close to 1 when a RGC does not respond to large uniform fields (i.e., is sensitive to high cycle/degree spatial frequency), and close to 0 when there is no spatial tuning. Results of this analysis are shown in Fig. 4b and illustrate the variation of the SI for the four directions of the moving bars and for the two projection conditions (i.e., stimulated areas on the retina). Under control conditions (upper graphs) the receptive fields of the RGCs appear well tuned spatially for large stimuli. However, when Bicuculline (lower graphs) is applied, no significant trend is observed anymore. Thus, even though outside of the scope of this paper, these results indicate that the spatial tuning of the receptive fields is most probably due to lateral inhibitory mechanisms that can be finely investigated with the proposed large-scale recording and analysis platform.

5. Conclusions

Active electrode array technologies based on CMOS microelectronic circuits are rapidly emerging as a valuable methodology for large-scale neuronal recordings in brain circuits with several thousands of simultaneously recording electrodes. However, the development of these devices challenges the implementation of adapted computational infrastructures and optimized data structures for the acquisition and analysis of such a massive recording of electrophysiological signals. Indeed, conventional approaches used for MEAs with only a few tens of electrodes are not optimized to handle signals from a much higher number of electrodes and their application introduces several bottlenecks for the data storage, data sharing and computational analysis costs.

Here, we have described a complete and expandable platform integrating specific functionalities for data compression, event detection, data sharing and analysis. This was implemented both on the acquisition instrument as well as on a custom server-client software application. In particular, online data compression capabilities were introduced on the acquisition instrument and a structured HDF5 file format structure was presented. These tools allow to manage a data rate of ~60 MB/sec acquired from our CMOS-MEA devices (4096 electrodes, 7 kHz sampling rate/electrode, 12 bits resolution) and to achieve different levels of data compression rates that can be selected depending on the level of information that needs to be preserved.

In addition, the use of an accessible and structured file format enables to easily integrate this acquisition instrument with custom analysis tools. In this respect, here we have presented a Python based client-server platform for analysis and data exchange. Originally, this allows to be operated either locally or remotely with a server-client architecture. Interestingly, this tool was designed with the aim of minimizing the flow of exchanged data among different machines through (slow) TCP/IP protocols and by leaving the freedom to store the acquired and analyzed data either on local or server networks.

Finally, we have shortly described, as a case study, how this platform can be applied to specific experimental studies on retinal whole mounts. These results show how the large-scale recording capabilities of CMOS-MEAs from dense electrode arrays on large active areas are enabling the study of light evoked responses at pan-retinal scale from thousands of RGCs simultaneously. In addition, when integrated with the proposed platform, time-stamps of light stimuli as well as the recorded spikes from specific electrode areas can be easily analyzed over multiple experimental phases.

The proposed framework is a first step toward the challenge of managing and analysis massive electrode array recordings. More advanced approaches, as already developed in bioinformatics for genomic data, might also be applied to this type of electrophysiological data. Additionally, progresses in event detection and spike sorting algorithms that exploit the signal redundancy available from densely integrated electrode arrays are needed and can effectively improve the application of these devices for several experimental studies in neuroscience and electrophysiology. Finally, in our opinion these developments need to be complemented with the development of adapted computational infrastructures designed for facilitating sharing and analysis of large-scale experimental data.

Conflict of interest

Alessandro Maccione, Stefano Zordan, Hayder Amin, Stefano Di Marco, Thierry Nieus, Gian Nicola Angotzi and Luca Berdondini declare that there are no conflicts of interest.

Mauro Gandolfo has a fellowship position at 3Brain GmbH, a private company commercializing part of the technology presented in this paper. His fellowship is sponsored within the Seventh Framework Program for the NAMASEN (FP7-264872) Marie-Curie Initial Training Network.

Acknowledgments

We acknowledge the financial support of the SI-CODE project of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: FP7-284553.

References

Berdondini, L., Bosca, A., Nieus, T., Maccione, A., 2014a. Active pixel sensor multielectrode array for high spatiotemporal resolution. In: De Vittorio, Massimo, Martiradonna, Luigi, Assad, John (Eds.), Nanotechnology and Neuroscience: Nano-Electronic, Photonic and Mechanical Neuronal Interfacing SE - 7. Springer, New York, pp. 207–238, http://dx.doi.org/10.1007/978-1-4899-8038-0_7, http://dx.doi.org/10.1007/978-1-4899-8038-0_7.

Berdondini, Luca, Bosca, Alessandro, Nieus, Thierry, Maccione, Alessandro, 2014b. Active pixel sensor multielectrode array for high spatiotemporal resolution. In: De Vittorio, Massimo, Martiradonna, Luigi, Assad, John (Eds.), Nanotechnology and Neuroscience: Nano-Electronic, Photonic and Mechanical Neuronal Interfacing. Springer New York, New York, NY, http://dx.doi.org/10.1007/978-1-4899-8038-0 http://link.springer.com/10.1007/978-1-4899-8038-0.

Bolgna, Luca Leonardo, Pasquale, Valentina, Garofalo, Matteo, Gandolfo, Mauro, Laurens, Pieter Baljon, Maccione, Alessandro, Martinoia, Sergio, Chiappalone, Michela, 2010. Investigating neuronal activity by SPYCODE multi-channel data analyzer. Neural Networks 23 (6), 685–697, http://dx.doi.org/10.1016/j.neunet.2010.05.002 http://www.ncbi.nlm.nih.gov/pubmed/20554151.

Brandon, Marty, C., Douglas Wallace, C., Baldi, Pierre, 2009. Data structures and compression algorithms for genomic sequence data. Bioinformatics (Oxford, England) 25 (14), 1731–1738 http://bioinformatics.oxfordjournals.org/content/25/14/1731.short.

Di Marco, S., Protti, D.A., Solomon, S.G., 2013. Excitatory and inhibitory contributions to receptive fields of alpha-like retinal ganglion cells in mouse. J.

Neurophysiol. 110 (6), 1426–1440, http://dx.doi.org/10.1152/jn.01097.2012 http://jn.physiology.org/content/110/6/1426.
Egert, U., Knott, T., Schwarz, C., Nawrot, M., Brandt, A., Rotter, S., Diesmann, M., 2002. MEA-Tools: an open source toolbox for the analysis of multi-electrode data with MATLAB. J. Neurosci. Method 117 (1), 33–42 http://www.ncbi.nlm.nih.gov/pubmed/12084562.
Eglen, John, Stephen, Weeks, Michael, Jessop, Mark, Simonotto, Jennifer, Jackson, Tom, Sernagor, Evelyne, 2014. A data repository and analysis framework for spontaneous neural activity recordings in developing retina. GigaScience, 3, http://dx.doi.org/10.1186/2047-217X-3-3 http://www.gigasciencejournal.com/content/3/1/3.
Ferrea, E., Maccione, A., Medrihan, L., Nieus, T., Ghezzi, D., Baldelli, P., Benfenati, F., Berdondini, L., 2012. Large-scale, high-resolution electrophysiological imaging of field potentials in brain slices with microelectronic multielectrode arrays. Front. Neural Circuits, 80, http://dx.doi.org/10.3389/fncir.2012.00080 http://www.scopus.com/inward/record.url?eid=2-s2.0-84869452450&partnerID=tZOtx3y1.
Fiscella, Michele, Farrow, Karl, Jones, Ian L., Jäckel, David, Müller, Jan, Frey, Urs, Bakkum, Douglas J., Hantz, Péter, Roska, B., Hierlemann, A., 2012. Recording from defined populations of retinal ganglion cells using a high-density cmos-integrated microelectrode array with real-time switchable electrode selection. J. Neurosci. Methods 211 (1), http://dx.doi.org/10.1016/j.jneumeth.2012.08.017, Elsevier B.V.: 103–13 http://dx.doi.org/10.1016/j.jneumeth.2012.08.017.
Frey, U., Egert, U., Heer, F., Hafizovic, S., Hierlemann, A., 2009. Microelectronic system for high-resolution mapping of extracellular electric fields applied to brain slices. Biosens. Bioelectron., http://dx.doi.org/10.1016/j.bios.2008.11.028, Elsevier B.V. http://dx.doi.org/10.1016/j.bios.2008.11.028.
Grewe, Jan, Wachtler, Thomas, Benda, Jan, 2011. A bottom-up approach to data annotation in neurophysiology. Front. Neuroinf. (January 5), 16 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3171061&tool=pmcentrez&rendertype=abstract.
Harrison, R.R., Charles, C., 2003. A low-power low-noise cmos for amplifier neural recording applications. IEEE J. Solid-State Circuits 38 (6), 958–965, http://dx.doi.org/10.1109/JSSC.2003.811979 http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1201998.
Hierlemann, By Andreas, Urs Frey, Sadik Hafizovic, and Flavio Heer. 2011. "Growing Cells Atop Microelectronic Chips : Interfacing Electrogenic Cells In Vitro With CMOS-Based Microelectrode Arrays" 99 (2).
Huang, Yu, Xiangning, Li, Yanling, Li, Qingwei, Xu, Qiang, Lu, Qian, Liu, 2008. An integrative analysis platform for multiple neural spike train data. J. Neurosci. Methods 172 (2), 303–311, http://dx.doi.org/10.1016/j.jneumeth.2008.04.026 http://www.ncbi.nlm.nih.gov/pubmed/18538855.
Imfeld, Kilian, Neukom, Simon, Maccione, Alessandro, Bornat, Yannick, Martinoia, Sergio, Farine, Pierre-André, Koudelka-Hep, Milena, Berdondini, Luca, 2008. Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity. IEEE Trans. Bio-Med. Eng. 55 (8), 2064–2073, http://dx.doi.org/10.1109/TBME.2008.919139 http://www.scopus.com/inward/record.url?eid=2-s2.0-47649113214&partnerID=tZOtx3y1.
Lambacher, A., Vitzthum, V., Zeitler, R., Eickenscheidt, M., Eversmann, B., Thewes, R., Fromherz, P. 2011. Identifying firing mammalian neurons in networks with high-resolution multi-transistor Array (MTA), 1–11. doi:10.1007/s00339-010-6046-9.
Lecoq, Jérôme, Joan Savall, Dejan Vučinić, Benjamin F Grewe, Hyun Kim, Jin Zhong Li, Lacey J Kitch, and Mark J Schnitzer. 2014. Visualizing mammalian brain area interactions by dual-axis two-photon calcium imaging, Nat. Neurosci., 17 (12): 1825–29. doi:10.1038/nn.3867. http://www.ncbi.nlm.nih.gov/pubmed/25402858.
Levine, Michael W., and Jeremy M. Shefner. 1991. "Fundamentals of Sensation and Perception." http://philpapers.org/rec/LEVFOS.
Litke, A.M., N. Bezayiff, E.J. Chichilnisky, W. Cunningham, W. Dabrowski, A.A. Grillo, M. Grivich, et al. 2004. What does the eye tell the brain?: development of a system for the large-scale recording of retinal output activity, IEEE Trans. Nucl. Sci. 51 (4): 1434–40. doi:10.1109/TNS.2004.832706. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1323709.
Lopez, Carolina Mora, Alexandru Andrei, Srinjoy Mitra, Marleen Welkenhuysen, Wolfgang Eberle, Carmen Bartic, Robert Puers, Refet Firat Yazicioglu, and Georges G. E. Gielen. 2014. An implantable 455-active-electrode 52-Channel CMOS neural probe, IEEE J. Solid-State Circuits, 49 (1): 248–61. doi:10.1109/JSSC. 2013.2284347. http://www.mendeley.com/catalog/implantable-455activeelectrode-52channel-cmos-neural-probe/.
Maccione Alessandro, Mauro Gandolfo, Paolo Massobrio, Antonio Novellino, Sergio Martinoia, and Michela Chiappalone. 2009. A novel algorithm for precise identification of spikes in extracellularly recorded neuronal signals, J. Neurosci. Methods, 177 (1): 241–49. doi:10.1016/j.jneumeth.2008.09.026. http://www.scopus.com/inward/record.url?eid=2-s2.0-58049217315&partnerID=tZOtx3y1.
Maccione Alessandro, Mauro Gandolfo, Mariateresa Tedesco, Thierry Nieus, Kilian Imfeld, Sergio Martinoia, and Luca Berdondini. 2010. Experimental investigation on spontaneously active hippocampal cultures recorded by means of high-density MEAs: analysis of the spatial resolution effects, Front. Neuroeng., 3 (May): 4. doi:10.3389/fneng.2010.00004. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2871691&tool=pmcentrez&rendertype=abstract.
Maccione Alessandro, Matthias H Hennig, Mauro Gandolfo, Oliver Muthmann, James van Coppenhagen, Stephen J Eglen, Luca Berdondini, and Evelyne Sernagor. 2014. Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse, J. Physiol., 592 (Pt 7). Blackwell Publishing Ltd: 1545–63. doi:10.1113/jphysiol.2013.262840. http://www.scopus.com/inward/record.url?eid=2-s2.0-84897418805&partnerID=tZOtx3y1.
Mahmud, Mufti, Alessandra Bertoldo, Stefano Girardi, Marta Maschietto, and Stefano Vassanelli. 2012. Sigmate: a matlab-based automated tool for extracellular neuronal signal processing and analysis, J. Neurosci. Methods, 207 (1). Elsevier B.V.: 97–112. doi:10.1016/j.jneumeth.2012.03.009. http://www.ncbi.nlm.nih.gov/pubmed/22513383.
Marblestone, Adam H, Bradley M Zamft, Yael G Maguire, Mikhail G Shapiro, Thaddeus R Cybulski, Joshua I Glaser, Dario Amodei, et al. 2013. Physical principles for scalable neural recording, Front. Comp. Neurosci., 7 (January). Frontiers: 137. doi:10.3389/fncom.2013.00137. http://journal.frontiersin.org/article/10.3389/fncom.2013.00137/abstract.
Marre, Olivier, Dario Amodei, Nikhil Deshmukh, Kolia Sadeghi, Frederick Soo, Timothy E Holy, and Michael J Berry. 2012. Mapping a complete neural population in the retina, J. Neurosci., 32 (43): 14859–73. doi:10.1523/JNEUROSCI.0723-12.2012. http://www.jneurosci.org/content/32/43/14859.abstract.
Mollazadeh, M, K Murari, G Cauwenberghs, and N Thakor. 2009. Micropower CMOS integrated low-noise amplification, filtering, and digitization of multimodal neuropotentials, IEEE Trans. Biomed. Circuits Syst., 3 (1): 1–10. doi:10.1109/TBCAS.2008.2005297. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2747318&tool=pmcentrez&rendertype=abstract.
Antonio, Novellino, Chiappalone, Michela, Maccione, Alessandro, Martinoia, Sergio, 2009. Neural signal manager: a collection of classical and innovative tools for multi-channel spike train analysis. Int. J. Adapt. Control Signal Process. 23 (11), 999–1013, http://dx.doi.org/10.1002/acs.1076 http://www.scopus.com/inward/record.url?eid=2-s2.0-70350490059&partnerID=tZOtx3.
Ruther Patrick, and Oliver Paul. 2014. New approaches for cmos-based devices for large-scale neural recording, Curr. Opin. Neurobiol., http://www.sparrho.com/article/UmF3QXJ0aWNsZQ==/1007038/?title=new-approaches-for-cmos-based-devices-for-large-scale-neural-recording.
Schwarz, David A, Mikhail A Lebedev, Timothy L Hanson, Dragan F Dimitrov, Gary Lehew, Jim Meloy, Sankaranarayani Rajangam, et al. 2014. Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys, Nat. Methods, 11 (6). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 670–76. doi:10.1038/nmeth.2936. http://dx.doi.org/10.1038/nmeth.2936.
Sobolev, Andrey, Adrian Stoewer, Michael Pereira, Christian J Kellner, Christian Garbers, Philipp L Rautenberg, and Thomas Wachtler. 2014. Data management routines for reproducible research using the g-node python client library, Front. Neuroinf., 8 (March): 15. doi:10.3389/fninf.2014.00015. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3942789&tool=pmcentrez&rendertype=abstract.
Stevenson Ian H, and Konrad P Kording. 2011. How advances in neural recording affect data analysis, Nat. Neurosci., 14 (2): 139–42. doi:10.1038/nn.2731. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3410539&tool=pmcentrez&rendertype=abstract.
Ziv Yaniv, Laurie D Burns, Eric D Cocker, Elizabeth O Hamel, Kunal K Ghosh, Lacey J Kitch, Abbas El Gamal, and Mark J Schnitzer. 2013. Long-term dynamics of ca1 hippocampal place codes, Nat. Neurosci., 16 (3): 264–66. doi:10.1038/nn.3329. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3784308&tool=pmcentrez&rendertype=abstract.