Denker et al. BMC Neuroscience 2011, 12(Suppl 1):P131
http://www.biomedcentral.com/1471-2202/12/S1/P131

BMC Neuroscience

POSTER PRESENTATION Open Access

Towards guiding principles in workflow design to facilitate collaborative projects involving massively parallel electrophysiological data

Michael Denker[1*], Andrew Davison[2], Markus Diesmann[3], Sonja Grün[3]

From Twentieth Annual Computational Neuroscience Meeting: CNS*2011
Stockholm, Sweden. 23-28 July 2011

The recent years have seen a rapid increase of interest in simultaneously analyzing the activity recorded from large numbers of channels in order to investigate the role of concerted neural activity in brain function. These efforts have led to advances in data analysis methods [1] that exploit the parallel properties of such data sets [2]. However, an often neglected aspect is that massively parallel data streams place new demands on handling their complexity during all stages of the project [3]: from the initial recording, throughout the analysis process, to the final publication. Three factors contribute these new demands: First, the sheer quantity of data complicates the organization of data sources, and the resulting automatization of analysis steps renders the validation of interim and final results difficult. Second, modern analysis methods often require intricate, multi-layered implementations, leading to sophisticated analysis toolchains [4]. Third, a growing number of projects needs to be carried out in teams, within a laboratory or in collaborative efforts, requiring transparent workflows that guarantee smooth interaction. Taken together, the increase in complexity calls for a reevaluation of the ad-hoc traditional approaches to such projects. Can we derive general guiding principles that may be adopted for designs of efficient workflows? How could these improve our confidence in handling the data by providing better cross-validation of findings, reliably managing provenance data, and enabling tighter collaborative research, while at the same time leaving the scientist with the flexibility required for creative research?

Although several projects are devoted to finding solutions for specific aspects of a workflow design (e.g., [5-7]), on a more general level there is lack of a thorough discussion on what goals are expected from a workflow, and which of these can be realistically addressed. Here, we summarize feedback received from experimenters and theoreticians that pinpoints the fundamental problems typically encountered in the analysis of high-dimensional electrophysiological data. Illustrated by examples from our own experience, we further show obstacles that prevent us from harmonizing workflows to common guidelines. For selected issues we draw parallels to other communities that are faced with similar problems (e.g., neuronal network modeling [8,9]; neuroimaging [10]). Lastly, we propose how existing concepts and software [9,11] could assist in practically implementing workflows that are tailored to the needs of a specific project, yet guarantee high standards by adhering to general guidelines of accepted best-practice.

Acknowledgements
This project was supported by the European Union (FP7-ICT-2009-6, BrainScales).

Author details
[1]Laboratory for Statistical Neuroscience, RIKEN BSI, Wako-shi, 351-0198 Saitama, Japan. [2]Unité de Neurosciences, Information et Complexité (UNIC), CNRS UPR-3293, 91198 Gif sur Yvette, France. [3]Institute of Neuroscience and Medicine (INM-6), Research Center Jülich, 52428 Jülich, Germany.

Published: 18 July 2011

References

1. Brown EN, Kass RE, Mitra PP: **Multiple neural spike train data analysis: state-of-the-art and future challenges.** *Nat Neurosci* 2004, **7**:456-461.
2. Stevenson IH, Kording KP: **How advances in neural recording affect data analysis.** *Nat Neurosci* 2011, **14**:139-142.
3. Buzsáki G: **Large-scale recording of neuronal ensembles.** *Nat Neurosci* 2004, **7**:446-451.

* Correspondence: mdenker@brain.riken.jp
[1]Laboratory for Statistical Neuroscience, RIKEN BSI, Wako-shi, 351-0198 Saitama, Japan
Full list of author information is available at the end of the article

4. Denker M, Wiebelt B, Fliegner D, Diesmann M, Morrison A: **Practically trivial parallel data processing in a neuroscience laboratory.** *Analysis of parallel spike trains* New York: Springer-Verlag; 2010.
5. CARMEN: **Code analysis, repository & modeling for e-neuroscience.**[http://www.carmen.org.uk].
6. Herz AVM, Meier R, Nawrot MP, Schiegel W, Zito T: **G-Node: An integrated tool-sharing platform to support cellular and systems neurophysiology in the age of global neuroinformatics.** *Neural Networks* 2008, **21**:1070-1075.
7. CRCNS: **Collaborative research in computational neuroscience.**[http://crcns.org/].
8. Nordlie E, Gewaltig MO, Plesser HE: **Towards reproducible descriptions of neuronal network models.** *PLoS Comput Biol* 2009, **5**:e1000456.
9. Sumatra: automated electronic lab book:[http://neuralensemble.org/trac/sumatra].
10. **LONI Pipeline.** [http://pipeline.loni.ucla.edu/].
11. **VisTrails.** , [http://www.vistrails.org/]; Taverna [http://www.taverna.org.uk/]; Kepler [https://kepler-project.org/].

doi:10.1186/1471-2202-12-S1-P131
Cite this article as: Denker *et al.*: Towards guiding principles in workflow design to facilitate collaborative projects involving massively parallel electrophysiological data. *BMC Neuroscience* 2011 **12**(Suppl 1): P131.