Biologically informed cortical models predict optogenetic perturbations

Christos Sourmpis (christos.sourmpis@epfl.ch)
Laboratory of Computational Neuroscience and
Laboratory of Sensory Processing, Brain Mind Institute
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

Carl C.H. Petersen (carl.petersen@epfl.ch)
Laboratory of Sensory Processing, Brain Mind Institute
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

Wulfram Gerstner (wulfram.gerstner@epfl.ch)
Laboratory of Computational Neuroscience, Brain Mind Institute
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

Guillaume Bellec (guillaume.bellec@tuwien.ac.at)
Machine Learning Research Unit
Technical University of Vienna (TU Wien)
Vienna, Austria

Abstract

A recurrent neural network fitted to large electrophysiological datasets may help us understand the chain of cortical information transmission. In particular, successful network reconstruction methods should enable a model to predict the response to optogenetic perturbations. We test recurrent neural networks (RNNs) fitted to electrophysiological datasets on unseen optogenetic interventions, and measure that generic RNNs used predominantly in the field generalize poorly on these perturbations. Our alternative RNN model adds biologically informed inductive biases like structured connectivity of excitatory and inhibitory neurons, and spiking neuron dynamics. We measure that some biological inductive biases improve the model prediction on perturbed trials in a simulated dataset, and a dataset recorded in mice in vivo. Furthermore, we show in theory and simulations that gradients of the fitted RNN can predict the effect of micro-perturbations in the recorded circuits, and discuss potentials for measuring brain gradients or using gradient-targeted stimulation to bias an animal behavior.

Keywords: Perturbation testing, Machine Learning, Optogenetics, Electrophysiology, Data-constrained RNN, Dale's law, Spiking neurons, Mechanistic modeling

Introduction and state-of-the-art

We need both (1) data modeling approaches that scale well with large-scale electrophysiology datasets (Siegle et al., 2021; Esmaeili et al., 2021; Urai, Doiron, Leifer, & Churchland, 2022; International Brain Laboratory et al., 2023), and (2) metrics to quantify when the models provide a plausible mechanism for the observed phenomena.

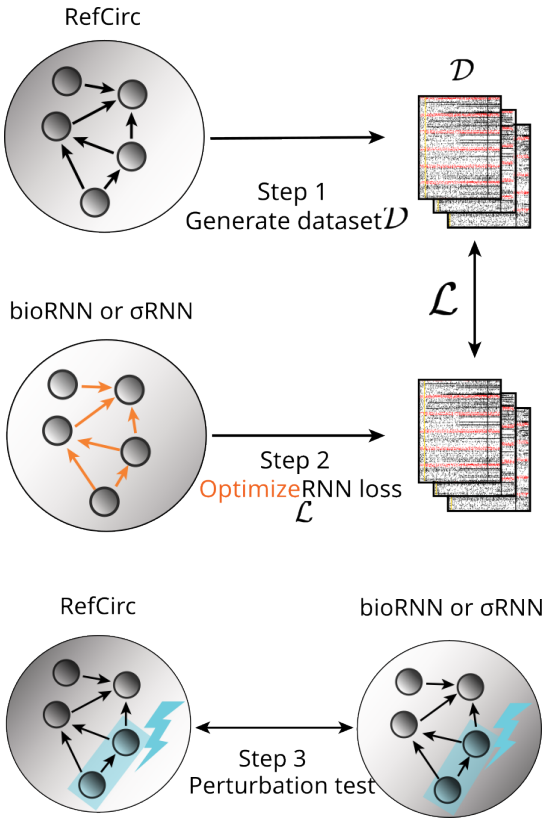

Figure 1: **Method overview.** The three steps to reconstruct the reference circuit (RefCirc) using a biologically informed RNN (bioRNN) or a simgoidal RNN (σRNN) and evaluate the reconstruction based on perturbation tests.

A promising approach to constrain models to electrophysiological data lies in the optimization of the simulation parameters by gradient descent. These methods were successful in quantitatively classifying functional cell types (Pozzorini et al., 2015; Teeter et al., 2018), and modeling micro-circuit interactions (Pillow et al., 2008; Deny et al., 2017; Mahuas, Isacchini, Marre, Ferrari, & Mora, 2020). To bridge the gap from single neurons or small retinal networks to cortical recordings in vivo, recent studies made substantial progress towards data-constrained recurrent neural network (RNN) models (Perich et al., 2020; Bellec, Wang, Modirshanechi, Brea, & Gerstner, 2021; Arthur, Kim, Chen, Preibisch, & Darshan, 2023; Kim, Finkelstein, Chow, Svoboda, & Darshan, 2023; Dinc, Shai, Schnitzer, & Tanaka, 2023; Sourmpis, Petersen, Gerstner, & Bellec, 2023). In this line of work, neurons in the RNN are mapped one-to-one to recorded cells and optimized by gradient descent to predict recorded activity at large.

An important question is whether these data-constrained RNNs can reveal a truthful mechanism of neuronal activity and behavior. By construction, the RNNs can generate brain-like network activity, but how can we measure whether the reconstructed network faithfully represents the biophysical mechanism? To answer this question, we submit a range of RNN reconstruction methods to a difficult *perturbation test*: we measure the similarity of the network response to unseen perturbations in the RNN and the recorded biological circuit.

Methods summary

Optogenetics is a powerful tool to induce precise causal perturbations in vivo (Esmaeili et al., 2021; Guo et al., 2014). It involves the expression of light-sensitive ion channels (Aravanis et al., 2007), such as channelrhodopsins, in specific populations of neurons (e.g., excitatory/pyramidal or inhibitory/parvalbumin-expressing). In this work, we use datasets combining dense electrophysiological recordings with optogenetic perturbations to evaluate RNN reconstruction methods. Since the neurons in our RNNs are mapped one-to-one to the recorded cells, we can model optogenetic perturbations targeting the same cell-types and areas as done in vivo. Yet, we observe that the similarity between the simulated and recorded perturbations varies greatly depending on the RNN reconstruction methods.

Results summary

Most prominently, we study two opposite types of RNN specifications. First, as a control model, we consider a traditional sigmoidal RNN (σRNN) which is arguably the most common choice for contemporary data-constrained RNNs (Perich et al., 2020; Arthur et al., 2023; Pals, Sağtekin, Pei, Gloeckler, & Macke, 2024); and second, we develop a model with biologically informed inductive biases (bioRNN): (1) neuronal dynamics follow a simplified spiking neuron model, and (2) neurons associated with fast-spiking inhibitory cells have short-distance inhibitory projections (other neurons are excitatory with both local and long-range interareal connectivity). Following (Neftci, Mostafa, & Zenke, 2019; Bellec, Salaj, Subramoney, Legenstein, & Maass, 2018; Bellec et al., 2021; Sourmpis et al., 2023), we adapt gradient descent techniques to optimize the bioRNN parameters of neurons and synapses to explain the recorded neural activity and behavior.

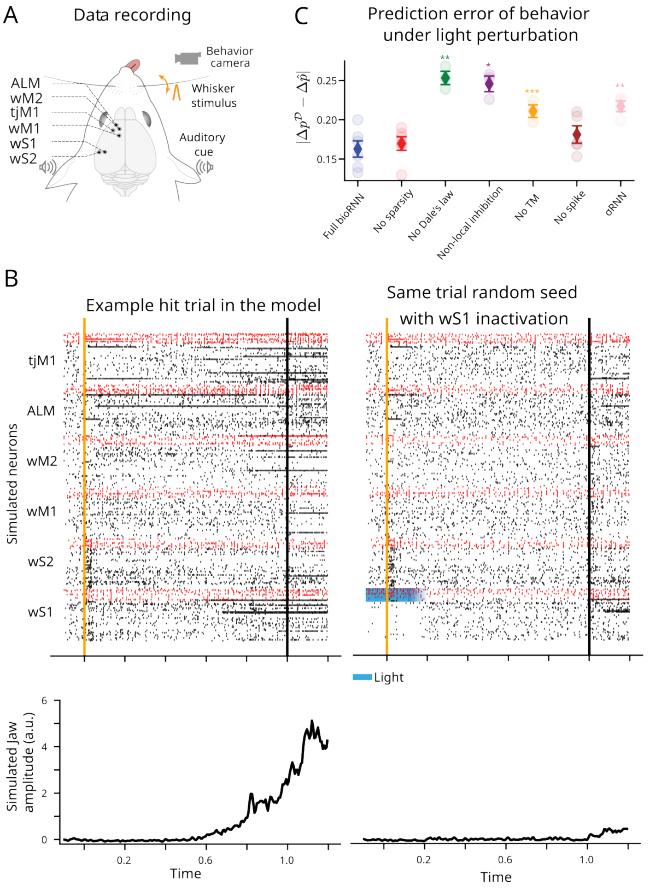

Figure 2: **Predicting optogenetic perturbations for in vivo electrophysiology data A.** During a delayed whisker detection task, the mouse reports a whisker stimulation by licking to obtain a water reward. Jaw movements are recorded by a camera. Our model simulates the jaw movements and the neural activity from six areas. **B.** Example hit trial of a reconstructed network (left). Using the same random seed, the trial turns into a miss trial if we inactivate area wS1 (right, light stimulus indicated by blue shading) during the whisker period by stimulation of inhibitory neurons (red dots). **C.** The experimentalists performed optogenetic inactivations of cortical areas (one area at a time) in three temporal windows. Error of the change in lick frequency caused by the perturbation, $\Delta\hat{p}$ is predicted by the model, and $\Delta p^{\mathcal{D}}$ is recorded in mice. Light-shaded circles show individual reconstructed networks with different initializations. The whiskers are the standard error of means. No TM means "No Trial-matching loss" (Sourmpis et al., 2023), the single trial variability is not fitted.

Strikingly, we find that the bioRNN is more robust to perturbations than the σRNN. This is nontrivial because it is in direct contradiction with other metrics often used in the field:

the σRNN simulation achieves higher similarity with unseen recorded trials before perturbation, but lower than the bioRNN on perturbed trials. This contradiction is confirmed both on synthetic and in vivo datasets. To analyze this result, we submit a spectrum of intermediate bioRNN models to the same *perturbation tests*, and identify two bioRNN model features that are most important to improve robustness to perturbation: (1) Dale's law (the cell type constrains the sign of the connections (Eccles, 1976)), and (2) local-only inhibition (inhibitory neurons do not project to other cortical areas). Other biological inductive biases, like spiking neuron dynamics and a sparsity prior, may improve the robustness to perturbations, but we measure that their effect is smaller in our case. We speculate that spikes might be more crucial for predicting other types of timed perturbations, but the recorded opto-genetic perturbations target one cell type broadly in an entire area and for a sustained duration, so modeling excitatory/inhibitory connection properties is more crucial for correctly modeling and predicting the effect of opto-genetic perturbations.

Furthermore (data not shown in this short abstract), a perturbation-robust bioRNN enables the prediction of the causal effect of perturbations in the recorded circuit. This becomes particularly interesting with micro-perturbations (μ-perturbation) targeting dozens of neurons in a small time window. We show in silico that the causal effect of μ-perturbations can be well approximated by the RNN gradients, which has two important implications for experimental neuroscience: (1) In a close-loop experimental setup, we can use RNN gradients to target a μ-perturbation which optimally increases (or decreases) movement in a simulated mouse (this is demonstrated in silico); (2) our RNN reconstruction methodology enables the estimation of gradients of the recorded circuit. So fitting RNNs becomes a tool to measure "brain gradients" and potentially relate contemporary in vivo measurement to decades of theoretical results from machine learning, where the gradient is a foundational concept (LeCun, Bengio, & Hinton, 2015; Richards & Kording, 2023).

Data availability statement

More details are available in the long preprint `biorxiv.org/content/10.1101/2024.09.27.615361`. and code repo `github.com/Sourmpis/BiologicallyInformed`. The in vivo dataset was published in (Esmaeili et al., 2021) and available at: `zenodo.org/records/4720013`.

Acknowledgment

We thank Alireza Modirshanechi, Shuqi Wang, and Tâm Nguyen for their valuable feedback on the manuscript. We are grateful to Vahid Esmaeili for collecting the dataset and ongoing support throughout this project. This research is supported by the Sinergia project CRSII5_198612, the Swiss National Science Foundation (SNSF) project 200020_207426 awarded to WG, SNSF projects TMAG-3_209271 and 31003A_182010 awarded to CP, and the Vienna Science and Technology Fund (WWTF) project VRG24-018 awarded to GB.

References

Aravanis, A. M., Wang, L.-P., Zhang, F., Meltzer, L. A., Mogri, M. Z., Schneider, M. B., & Deisseroth, K. (2007). An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *Journal of Neural Engineering*, *4*(3), S143.

Arthur, B. J., Kim, C. M., Chen, S., Preibisch, S., & Darshan, R. (2023). A scalable implementation of the recursive least-squares algorithm for training spiking neural networks. *Frontiers in Neuroinformatics*, *17*, 1099510.

Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., & Maass, W. (2018). Long short-term memory and learning-to-learn in networks of spiking neurons. *Advances in Neural Information Processing Systems*, *31*.

Bellec, G., Wang, S., Modirshanechi, A., Brea, J., & Gerstner, W. (2021). Fitting summary statistics of neural data with a differentiable spiking network simulator. *Advances in Neural Information Processing Systems*, *34*.

Deny, S., Ferrari, U., Mace, E., Yger, P., Caplette, R., Picaud, S., ... Marre, O. (2017). Multiplexed computations in retinal ganglion cells of a single type. *Nature Communications*, *8*(1), 1964.

Dinc, F., Shai, A., Schnitzer, M., & Tanaka, H. (2023). Cornn: Convex optimization of recurrent neural networks for rapid inference of neural dynamics. *Advances in Neural Information Processing Systems*, *36*.

Eccles, J. C. (1976). *From electrical to chemical transmission in the central nervous system: The closing address of the sir henry dale centennial symposium*. Springer.

Esmaeili, V., Tamura, K., Muscinelli, S. P., Modirshanechi, A., Boscaglia, M., Lee, A. B., ... others (2021). Rapid suppression and sustained activation of distinct cortical regions for a delayed sensory-triggered motor response. *Neuron*, *109*(13), 2183–2201.

Guo, Z. V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J. T., ... Svoboda, K. (2014). Flow of cortical activity underlying a tactile decision in mice. *Neuron*, *81*(1), 179–194.

International Brain Laboratory, Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., ... others (2023). A brain-wide map of neural activity during complex behaviour. *bioRxiv*, 2023–07.

Kim, C. M., Finkelstein, A., Chow, C. C., Svoboda, K., & Darshan, R. (2023). Distributing task-related neural activity across a cortical network through task-independent connections. *Nature Communications*, *14*(1), 2851.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Mahuas, G., Isacchini, G., Marre, O., Ferrari, U., & Mora, T. (2020). A new inference approach for training shallow and deep generalized linear models of noisy interacting neurons. *Advances in Neural Information Processing Systems*, *33*.

Neftci, E. O., Mostafa, H., & Zenke, F. (2019). Surrogate gradient learning in spiking neural networks: Bringing the power

of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, *36*(6), 51–63.

Pals, M., Sağtekin, A. E., Pei, F., Gloeckler, M., & Macke, J. H. (2024). Inferring stochastic low-rank recurrent neural networks from neural data. *arXiv:2406.16749*.

Perich, M. G., Arlt, C., Soares, S., Young, M. E., Mosher, C. P., Minxha, J., ... others (2020). Inferring brain-wide interactions using data-constrained recurrent neural network models. *bioRxiv*, 2020–12.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999.

Pozzorini, C., Mensi, S., Hagens, O., Naud, R., Koch, C., & Gerstner, W. (2015). Automated high-throughput characterization of single neurons by means of simplified spiking models. *PLoS Computational Biology*, *11*(6), e1004275.

Richards, B. A., & Kording, K. P. (2023). The study of plasticity has always been about gradients. *The Journal of Physiology*, *601*(15), 3141–3149.

Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., ... others (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, *592*(7852), 86–92.

Sourmpis, C., Petersen, C., Gerstner, W., & Bellec, G. (2023). Trial matching: capturing variability with data-constrained spiking neural networks. *Advances in Neural Information Processing Systems*, *36*.

Teeter, C., Iyer, R., Menon, V., Gouwens, N., Feng, D., Berg, J., ... others (2018). Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature Communications*, *9*(1), 709.

Urai, A. E., Doiron, B., Leifer, A. M., & Churchland, A. K. (2022). Large-scale neural recordings call for new insights to link brain and behavior. *Nature Neuroscience*, *25*(1), 11–19.