PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Multitask learning of a biophysically-detailed neuron model

Jonas Verhellen[1☯]*, Kosio Beshkov[1☯]*, Sebastian Amundsen[1], Torbjørn V. Ness[2], Gaute T. Einevoll[1,2,3]*

1 Department of Biosciences, University of Oslo, Oslo, Norway, **2** Department of Physics, Norwegian University of Life Sciences, Ås, Norway, **3** Department of Physics, University of Oslo, Oslo, Norway

☯ These authors contributed equally to this work.
* jverhell@gmail.com (JV); kosio.neuro@gmail.com (KB); gaute.einevoll@nmbu.no (GTE)

Check for updates

OPEN ACCESS

Citation: Verhellen J, Beshkov K, Amundsen S, Ness TV, Einevoll GT (2024) Multitask learning of a biophysically-detailed neuron model. PLoS Comput Biol 20(7): e1011728. https://doi.org/10.1371/journal.pcbi.1011728

Editor: Hugues Berry, Inria, FRANCE

Received: December 4, 2023

Accepted: June 28, 2024

Published: July 31, 2024

Copyright: © 2024 Verhellen et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The LFDeep: Multitask Learning of Biophysically-Detailed Neuron Models project code is hosted on GitHub. The repository includes the implementation of the paper's three multitask learning architectures for the distillation of biophysically-detailed neuron models. To access the latest version of the code and contribute to the project, please visit the official GitHub repository at https://github.com/Jonas-Verhellen/LFDeep. All data along with checkpoints with the optimal parameters of the MMoE model are available at https://www.kaggle.com/datasets/kosiobeshkov/data-for-mmoe-prediction.

Abstract

The human brain operates at multiple levels, from molecules to circuits, and understanding these complex processes requires integrated research efforts. Simulating biophysically-detailed neuron models is a computationally expensive but effective method for studying local neural circuits. Recent innovations have shown that artificial neural networks (ANNs) can accurately predict the behavior of these detailed models in terms of spikes, electrical potentials, and optical readouts. While these methods have the potential to accelerate large network simulations by several orders of magnitude compared to conventional differential equation based modelling, they currently only predict voltage outputs for the soma or a select few neuron compartments. Our novel approach, based on enhanced state-of-the-art architectures for multitask learning (MTL), allows for the simultaneous prediction of membrane potentials in each compartment of a neuron model, at a speed of up to two orders of magnitude faster than classical simulation methods. By predicting all membrane potentials together, our approach not only allows for comparison of model output with a wider range of experimental recordings (patch-electrode, voltage-sensitive dye imaging), it also provides the first stepping stone towards predicting local field potentials (LFPs), electroencephalogram (EEG) signals, and magnetoencephalography (MEG) signals from ANN-based simulations. While LFP and EEG are an important downstream application, the main focus of this paper lies in predicting dendritic voltages within each compartment to capture the entire electrophysiology of a biophysically-detailed neuron model. It further presents a challenging benchmark for MTL architectures due to the large amount of data involved, the presence of correlations between neighbouring compartments, and the non-Gaussian distribution of membrane potentials.

Author summary

Our research focuses on cutting-edge techniques in computational neuroscience. We specifically make use of simulations of biophysically detailed neuron models. Traditionally these methods are computationally intensive, but recent advancements using artificial

Funding: This research was funded by UiO:Life Science through the 4MENT convergence environment to JV, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 945371 to KB and the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement No. 945539 Human Brain Project (HBP) SGA3 to JV. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

neural networks (ANNs) have shown promise in predicting neural behavior with remarkable accuracy. However, existing ANNs fall short in providing comprehensive predictions across all compartments of a neuron model and only provide information on the activity of a limited number of locations along the extent of a neuron. In our study, we introduce a novel approach leveraging state-of-the-art multitask learning architectures. This approach allows us to simultaneously predict membrane potentials in every compartment of a neuron model. By distilling the underlying electrophysiology into an ANN, we significantly outpace conventional simulation methods. By accurately capturing voltage outputs across the neuron's structure, our method invites comparisons with experimental data and paves the way for predicting complex aggregate signals such as local field potentials and EEG signals. Our findings not only advance our understanding of neural dynamics but also present a significant benchmark for future research in computational neuroscience.

1 Introduction

In the seven decades since Hodgkin and Huxley first described the action potential in terms of ion channel gating [1–7], while the scientific community is gaining a comprehensive understanding of how individual neurons process information, the behavior of large networks of neurons remains comparatively poorly understood. Experimental studies provide qualitative insights through statistical correlations between recorded neural activity and sensory stimulation or animal behavior [8–11], but statistical modelling offers little information on how networks perform neural computation or give rise to neural representations. Mechanistic modelling, in which detailed neuron models or networks of detailed neuron models are simulated on a computer, offers an alternative approach to studying the network dynamics of neural circuits [12, 13].

Thanks to recent pioneering efforts facilitated by large supercomputers, we are now able to construct simulations containing tens of thousands of model neurons that mimic specific cortical columns in mammalian sensory cortices [14–18]. Even more recent advances have reduced the need for supercomputers by distilling the output of biophysically-detailed neuron models into easier-to-evaluate artificial neural networks (ANN) [19, 20]. Note that in this context distillation means deep learning the dynamics of a complex physics-based model based on observations of the data generated by that model, without direct access to the underlying equations.

In the present paper we explore accelerated techniques for biophysically-detailed neuron models with full electrophysiological detail, a task that was previously deemed impractical due to the high computational cost involved. Our approach builds on previous work that focused on predicting outgoing action potentials or other experimental variables in a limited number of compartments. However, our approach is unique in that it allows for the simultaneous prediction of the membrane potentials (and membrane currents) for all compartments, see Fig 1. A sufficiently accurate deep learning model in which errors do not accumulate allows for future comparison of model output with a wider range of experimental recordings of membrane potentials (dendritic patch electrodes, voltage-sensitive dye imaging), and also the calculation of extracellular signals such as local field potentials (LFPs), electroencephalogram (EEG) signals, and magnetoencephalography (MEG) signals [21]. This approach is an important step forward in our ability to simulate and eventually better understand the functioning of the

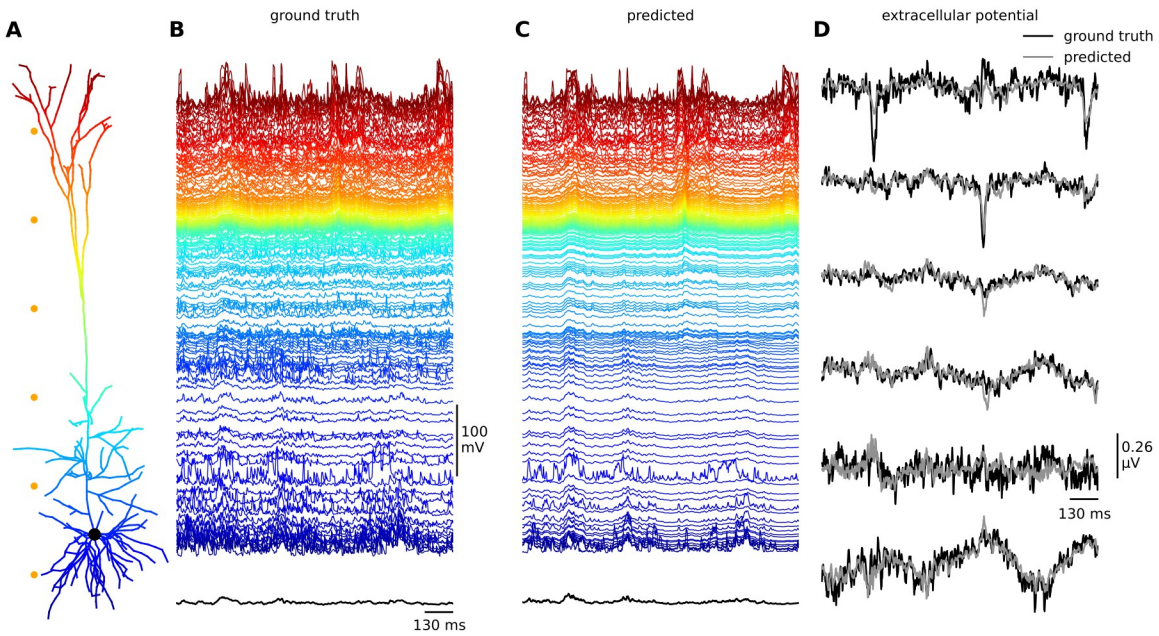

Fig 1. A) Illustration of the biophysically-detailed model with 639 compartments of a cortical layer V pyramidal cell model [29], which is the main object of study in this paper, color-coded by the compartment depth (consistent across all panels). B) Membrane voltages as calculated by a biophysically-detailed simulation of the multi-compartment model, used as the ground truth throughout this paper. C) Membrane voltages as predicted by our best-performing multi-task learning architecture, one time step at a time with a time resolution of 1 ms for the prediction. D) Comparison between the ground truth and predicted extracellular potentials as calculated at six points outside the neuron (orange dots in panel A).

https://doi.org/10.1371/journal.pcbi.1011728.g001

brain in health and disease. Note that the model does not include the intricacies which arise from complex synapse dynamics like plasticity.

By predicting all membrane potentials across a biophysically-detailed neuron model simultaneously, rather than in the soma only as recently done by [19], we have made the transition from single-task learning to multi-task learning (MTL). In contrast to single-task learning that requires a separate model to be trained for each target, MTL optimises a single artificial neural architecture to predict multiple (heterogeneous) targets simultaneously. MTL approaches aim to improve generalisation and efficiency across tasks by leveraging statistical relationships between multiple targets [22–26].

In our work, statistical relationships between tasks, i.e. membrane potential predictions, arise from the correlations between compartments due to the biophysical mechanisms of the ion currents running through a neuron. We note that obtaining good results in each compartment individually does not strictly guarantee that downstream errors will not accumulate due to correlations. To capture shared patterns that could be missed in single-task learning, multi-task learning generally relies on either one of two categories of neural architectures, respectively known as hard parameter [27] or soft parameter sharing models [28].

Hard parameter sharing models have a shared bottom layer in the neural network, while the output branches are task-specific. This means that the lower layers of the neural network are identical for all tasks, and only the final layers are tailored to each individual task. In contrast, soft parameter sharing models use dedicated sets of learning parameters and feature mixing mechanisms, allowing each task to have its own set of features and learning parameters while still efficiently sharing information. The choice between soft and hard parameter sharing models often depends on the nature of the tasks being learned. For example, if there is a

substantial amount of overlap between the features necessary for each task, a hard parameter sharing model may be more appropriate, while if each (or any) task requires its own unique set of features for effective learning, a soft parameter sharing model may be more effective.

In this paper, we explore the capacity of various MTL architectures for distilling the full electrophysiology of a multi-compartment, biophysically-detailed layer 5 pyramidal neuron model [29]. Specifically, we compare a single type of hard parameter sharing model with two novel versions of state-of-the-art soft parameter sharing models, known as Multi-gate Mixture-of-Experts (MMoE) [30] and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEx) [31]. Notably, in our computational experiments, we observed that the soft parameter sharing models substantially outperformed the hard parameter sharing model in the knowledge distillation of the electrophysiology of a biophysically-detailed neuron model. These findings highlight the potential benefits of soft-parameter sharing models for distilling complex electrophysiological data from biophysically-detailed neuron models, which we expect to have broad implications for computational neuroscience research.

For a comprehensive understanding of the MTL architectures utilised in this study, along with corresponding network diagrams, please refer to the detailed information provided in the Methods and Materials section. To make the following sections more accessible however, we briefly highlight that MMoE and MMoEEx, rely on a learnable feature mixing mechanism that controls the contribution of each learned data representation to the prediction of each task. The amount of learnable parameters in the feature mixing mechanisms grows linearly with the flattened input data size (batch size excluded). For the neural input data considered in this paper, the memory requirements regarding the amount of learnable parameters becomes prohibitively large. As a solution, we extend the MMoE and MMoEEx architectures to contain a compressor module that fixes the amount of learnable parameters in the feature mixing mechanisms to a predetermined number.

2 Results

To test the capability of different MTL architectures in accurately representing the full dynamic membrane potential of each of the 639 compartments of the large biophysically-detailed neuron model, we trained the models on a balanced dataset of simulated neural activity. For each target—a collection of membrane potentials (and the presence or absence of an outgoing action potential)—in the dataset, we provide a 100 ms history of neural activity and synaptic inputs from all compartments of the biophysically-detailed neuron model to the MTL models. Further details regarding the dataset and pre-processing steps can be found in the Methods and Materials section. To facilitate a fair comparison between MTL methods, we constructed a hard parameter sharing model (14 million trainable parameters) that is close in size to the soft parameter sharing models (12 million trainable parameters). Similarly, we used the same training procedures for each of the models—Adam [32] with or without task-balancing [33]—and evaluated the inference speed in a single session on publicly available hardware.

2.1 Multitask prediction of membrane potential dynamics in biophysically-detailed neuron models

After conducting comprehensive training on all three Multi-Task Learning (MTL) models, namely Hard Parameter Sharing (MH), Multi-gate Mixture-of-Experts (MMoE), and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEx), we observed that the soft parameter sharing models exhibited a superior performance compared to the hard parameter sharing model, not only in terms of training but also in terms of generalization loss, see Fig 2 which showcases the training progress and the respective losses for each model. During the training

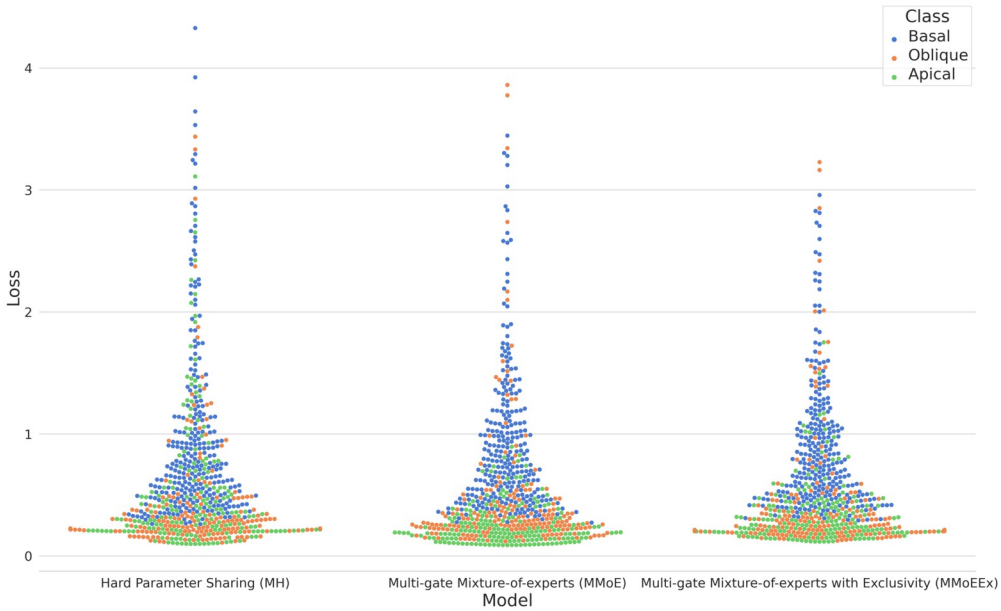

Fig 2. Swarm plots of the generalisation loss for each of the compartments (basal, oblique, apical) of the neuron model for each of the three models: Hard Parameter Sharing (MH), Multi-gate Mixture-of-experts (MMoE), and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEx). Note that basal dendrites of a neuron receive incoming signals from other neurons and convey them towards the cell body, while the apical dendrite extends from the cell body to integrate signals from distant regions, and oblique dendrites play a role in the integration of synaptic inputs at various angles away from the other dendrites. The vertical axis represents the numeric loss values while the horizontal axis, the different ANN models are indicated. The density of the data points in a specific region indicates how many compartments have a similar loss value for each model.

https://doi.org/10.1371/journal.pcbi.1011728.g002

process, we closely monitored the performance of each model through Tensorboard [34]. The hard parameter sharing model achieved a minimal training loss of 0.439 and a minimal validation loss of 0.680. The MMoE model, on the other hand, demonstrated better progress, reaching a minimal training loss of 0.241 and a minimal validation loss of 0.533. Similarly, the MMoEEx model exhibited better training results than the hard parameter sharing model, with a minimal recorded training loss of 0.262 and a minimal validation loss of 0.551.

For further evaluation, we focused on the MMoE model, as it showcased the best performance based on the validation results. By utilising the model weights saved at its optimal validation performance, we were able to predict the membrane potential of a compartment within the biophysically-detailed neuron model with a root mean squared error of 3.78 mV compared to a standard deviation of 13.20 mV across the compartments and batches of the validation data. As explained in more detail in the Discussion section, the electrophysiological data upon which the biophysically-detailed neuron model is built, has an experimental standard deviation of around 5 mV, depending on the type of neural activity. It is worth noting that the network models we explored primarily prioritised learning the membrane potentials of the neuron model's compartments rather than accurately predicting spike generation at the axon initial segment. As a consequence, the presented results omit the performance analysis related to spike generation at the axon initial segment. However, given that our model does not distinguish in its predictions between sub and supra-threshold membrane potentials, one can obtain the spikes within each compartment directly from the predictions.

To further illustrate the complex dynamics of the Hay neuron and how they are captured by our multi-task deep learning model, we present a back-propagating action potential-

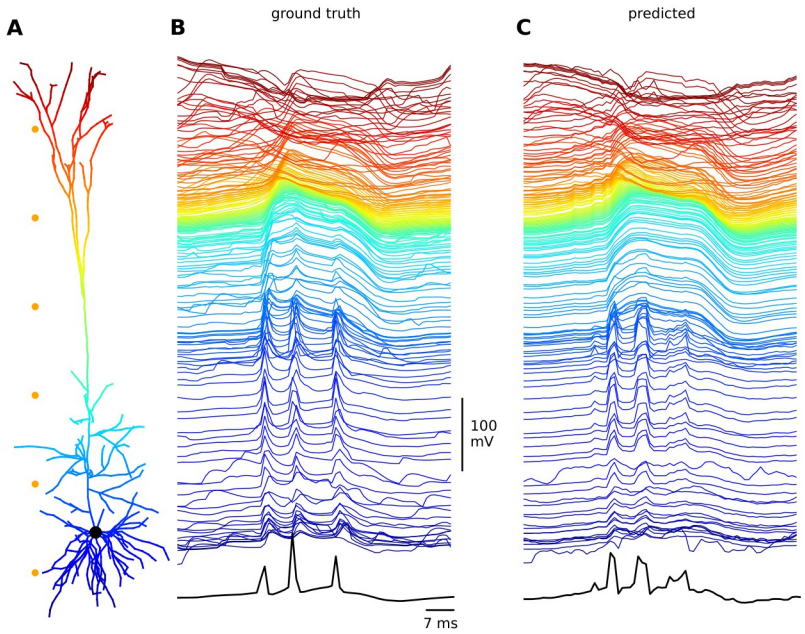

Fig 3. A) Illustration of a biophysically-detailed model of a cortical layer V pyramidal neuron [29], color-coded by the compartment depth (consistent across all panels). B) Membrane voltages of a back-propagating action potential activated Ca^{2+} spike as calculated by a biophysically-detailed simulation of the multi-compartment model. Note the presence of the characterizing features such as the long depolarization of around the main bifurcation point of the apical dendrite (yellow) and the two somatic (blue) action potentials that follow. C) Membrane voltages of a back-propagating action potential activated Ca^{2+} as predicted by our best-performing multi-task learning architecture, one time step at a time with a time resolution of 1 ms for the prediction.

https://doi.org/10.1371/journal.pcbi.1011728.g003

activated Ca^{2+} spike event, a key active dendritic property of layer 5 pyramidal cells. These types of events have been observed in vitro and are the consequence of the convergence of back-propagating action potentials from the soma and the local excitatory post-synaptic potential at the distal dendrites. This convergence triggers a Ca^{2+} spike at the main bifurcation point of the apical dendrite, which in turn triggers several somatic Na^{+} spikes. As shown in Fig 3, the MMoE multi-task deep learning model has the ability to faithfully reproduce these complex dynamical behaviors, including the characteristic long-lasting depolarization that coincides with several somatic action potentials. Note that this also shows that the somatic Na-spikes are clearly visible from the predicted membrane potential, even if the binary spike prediction task itself was unsuccessful. It therefore seems likely that somatic Na-spikes can be extracted from the predicted membrane potential through simple thresholding procedures although this was outside the scope of the current work.

2.2 The importance of expert diversity in MMoE and MMoEEx trained on neural data

Previous studies have suggested that higher diversity among experts in terms of the data representation they generate, see the Methods and materials section, could potentially improve training and generalisation outcomes in MTL [30, 31], particularly for soft parameter sharing models. Motivated by these findings, we conducted a thorough investigation into the effect of expert diversity within the training process of the MMoE and MMoEEx models. To quantify and analyse expert diversity, we employed various diversity metrics. In addition to the diversity score, previously introduced by [31], which is based on the standardised distance matrix

between experts, we extended our analysis to also include its determinant and permanent, see Methods. By considering these metrics, we were hoping to gain deeper insight into the level of diversity present among the experts throughout the training procedure. However, all three metrics exhibited the same temporal pattern. For a visual representation of these correlations and their implications on expert diversity, refer to S1 Fig.

Interestingly, despite MMoEEx originally having been proposed in the literature as a means to enhance expert diversity, we made the intriguing observation that expert diversity exhibited significant variation across different training runs of the same model (MMoE or MMoEEx) with identical training data. This observation suggests that a soft parameter sharing model's architecture alone does not guarantee any consistent trend in expert diversity, see the appendices for more details, and that multiple training runs are necessary for complex deep learning models to learn about their properties. Furthermore, we studied the strength of the expert weights used to predict each compartment in both MMoE and MMoEEx, see Fig 4. While larger average weight does not necessarily imply a better prediction, it corresponds to a larger contribution to the prediction coming from a particular expert. As one can see in both models, different experts focus on different parts of the neuron, with relatively higher mean weights in higher apical and basal regions. One noticeable difference is that in MMoEEx, the oblique parts of the neuron are more strongly represented, whereas somatic compartments are visibly different from zero in one of the experts.

2.3 Fast simulation of full membrane potential dynamics of multiple neurons

Traditional simulation environments such as NEURON [35] rely on numerical integration of compartment-specific differential equations that represent the active and passive biophysical mechanisms of the modeled neuron. This approach requires a significant amount of computational resources. One of the main attractive features of distilling biophysically-detailed neuron models into ANNs is the significant speed-up deep learning architectures can provide. Previous work from the literature, where only the output of the soma compartment was predicted, has shown that NEURON simulations of biophysically-detailed neuron model networks, consisting of up to 5000 model neuron instances, are up to five orders of magnitude slower than their ANN counterparts [20]. Note that, in network simulations, accelerators like GPUs can be used to compute independent timesteps—for different instances of the same model neuron—of the ANN model in parallel. In practice, the batch-size of the MTL models during inference can directly be interpreted as the number of model neurons being evaluated at the same time. Given the size of the training dataset (61 GB) and the typical time needed to train one of these models (approximately a week on four RTX2080Ti GPUs), training is highly non-trivial during model development. However, for verified biophysically detailed neuron models training only has to be performed once, after which the distilled model can be used readily.

We recorded the mean and standard deviation runtimes of seven independent predictions for each of the three MTL models for different amounts of neurons and simulation times. In Table 1, we present the results of 100 neurons simulated for 1 ms, 10 ms and 100 ms, and for a 100ms simulation of 1000 neurons. To provide context, we also recorded the mean and standard deviation runtimes of seven independent 100 ms simulations of a single neuron in NEURON. Note that in the NEURON simulation, each timestep is passed to the next internally, while for the runtime tests of the deep learning models we are limited to running each timestep from Python causing an overhead for the MTL models. Clearly, predictions by MTL models are significantly faster than single-core simulations in NEURON. Note that support for the use of NEURON on GPUs is actively being developed [36] but is not considered in this paper. For

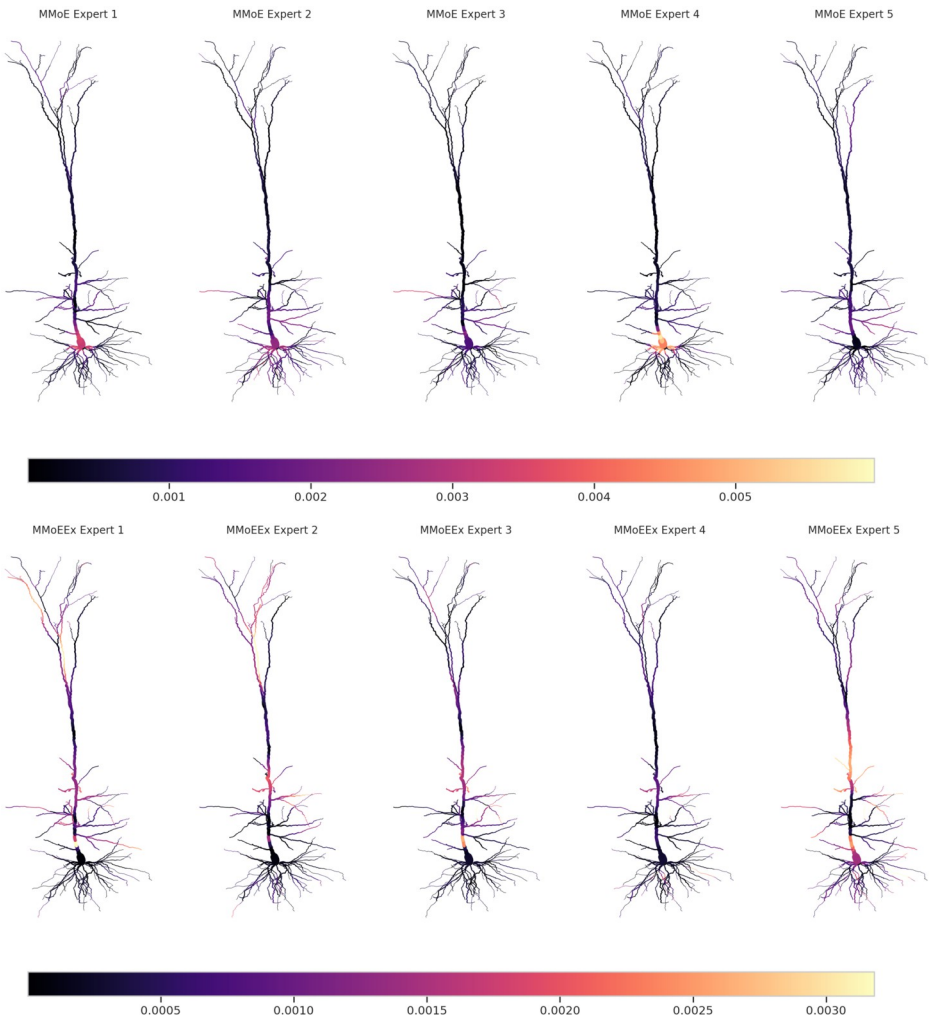

Fig 4. Mean weights from the experts in both Multi-gate Mixture-of-experts (MMoE; top row) and Multi-gate Mixture-of-experts with Exclusivity (MMoEEx; bottom row) after training the models on neural data, projected onto the different compartments of the cortical layer V biophysically-detailed neuron model. Each row consists of five subplots, representing different experts within the models. The color scale (normalised to the mean weights of the first expert) is indicated by the horizontal color bar located below each row.

https://doi.org/10.1371/journal.pcbi.1011728.g004

the set-up with 1000 neurons, the hard parameter sharing model reaches an acceleration of 2 orders of magnitude over the NEURON simulation. Being able to predict voltages in parallel has a profound effect on efficiency, as a 1ms simulation of 100 neurons can be predicted in 0.45 seconds which is nearly half the time necessary for NEURON to simulate a single neuron for 100 ms, while requiring the same amount of update steps.

Also note that while prediction times for the MTL models scale roughly linearly with simulation times, they do not scale linearly with batch size. Within the maximum batch size that can be accommodated by the GPU memory, the number of simulated neurons should only affect the evaluation time of a neural network through the previously discussed Python overhead, as is seen in the recorded prediction times for the hard parameter sharing model. However, for the soft parameter sharing models which contain matrix multiplications in the gates we observe a steep increase in runtime for large batch sizes. In general, it is worth remarking

Table 1. Inference speed of all three MTL architectures—Hard parameter sharing (MH), Multi-gate Mixture-of-experts (MMoE), and Multi-gate Mixture-of-experts with Exclusivity (MMoEEx)—After training on neural data generated by a biophysically-detailed model of a cortical layer 5 pyramidal neuron model compared to the classical NEURON simulation.

Runtime (s)	Std. Dev. (ms)	Model	Hardware	Simulation (ms)	Batch Size
1.01	29.7	Neuron	CPU	100	1
0.07	0.85	MH	GPU	1	100
0.45	1.03	MMoE	GPU	1	100
0.44	1.43	MMoEEx	GPU	1	100
0.70	25.4	MH	GPU	10	100
4.36	35.7	MMoE	GPU	10	100
4.35	15.0	MMoEEx	GPU	10	100
7.19	145	MH	GPU	100	100
43.6	58.6	MMoE	GPU	100	100
43.5	82.4	MMoEEx	GPU	100	100
10.9	153	MH	GPU	100	1000
365.0	34.2	MMoE	GPU	100	1000
365.0	33.5	MMoEEx	GPU	100	1000

https://doi.org/10.1371/journal.pcbi.1011728.t001

that MMoE and MMoEEx demonstrate inference runtimes which are significantly slower compared to the hard parameter sharing model. This discrepancy can be attributed to the hard parameter sharing model having a lower number of convolution filters and the soft parameter sharing models containing large and computationally expensive matrix multiplications.

3 Discussion

3.1 Multitask prediction of membrane potential dynamics in biophysically-detailed neuron models

In our investigation, we evaluated three advanced multi-task learning (MTL) neural network architectures [27, 30, 31], including MMoE and MMoEx which were augmented with a novel compressor module, to tackle the challenging task of capturing intricate membrane potential dynamics in a complex, multi-compartment, biophysically-detailed model of a layer 5 pyramidal neuron [29]. Given the considerable complexity of the model comprising 639 compartments, each generating distinct timeseries data potentially containing spikes, this MTL problem presented an exceptionally difficult distillation task. Nevertheless, our findings are encouraging, with the MMoE model achieving the lowest validation loss and demonstrating a root mean squared error of 3.78 mV in predicting membrane potential across all compartments. In comparison, the root mean squared error for the MMoEEx and MH models was 3.90 mV and 4.81 mV respectively. Notably, these errors are considerably smaller than both the standard deviation of our test set (13.20 mV) and the experimentally measured standard deviations of the peak membrane voltages during perisomatic step current firing (4.97 mV—6.93 mV) or back-propagating action potential Ca^{2+} firing (5 mV) in layer 5 pyramidal neurons [29, 37]. While our model is capable of reproducing the membrane potential of the trained multicompartment model, it is not a direct replacement of an implementation in NEURON. For example, the modification of new compartments will require retraining of the multi-task-learning architecture in PyTorch.

Experimental recordings of the after-hyperpolarization depth of the membrane potential in the soma of layer 5 pyramidal neurons have a slightly lower standard deviation (3.58 mV—5.82 mV) than electrophysiological measurements of action potentials [29, 38, 39], indicating

that a more sensitive evaluation of distilled neuron models could be based on their performance in specific neuronal scenario's. Notably, thus far, all trained models did not manage to learn the binary somatic spike prediction task, a challenge that could potentially be mitigated through a computationally intensive hyperparameter search for γ (explained in the Methods and materials section) to assign higher importance to this specific task during the learning process. Although task balancing methods for soft parameter sharing models were explored in accordance with the procedures detailed in the Methods and Materials section, our findings indicate no improvement in performance, as outlined in S3 Fig. Furthermore, it is essential to highlight that these MTL models are anticipated to find utility in accelerating LFP calculations, where the significance of subthreshold components in membrane potential dynamics is well-established [40–42]. For further insights into calculating extracellular potentials based on the deep learning model's membrane potential predictions and an example of such a downstream prediction, please refer to S4 Fig.

3.2 Measuring experts diversity in MMoE and MMoEEx trained on neural data

To test the conjecture that in soft parameter-sharing MTL models, high diversity between experts can be beneficial in training and generalisation, we computed several diversity measures. Indeed, initially it looked like the model with highest diversity (MMoE) performed the best, despite the fact that unlike its counterpart (MMoEEx), it has no explicit inductive bias towards expert diversity. However, a subsequent retraining of the models showed that the known diversity metrics are neither robust nor correlated with training or generalisation performance. Further exploration of the exclusivity hyperparameter α might lead to more desirable results for the MMoEEx model. Additionally, increasing the number of experts should only lead to improvements in training and generalisation if a substantial amount of independence between the tasks had not yet been incorporated. Future work can make these statements more precise by studying the contribution of individual experts to the prediction of membrane potentials in morphologically distinct neuronal compartments.

3.3 Fast simulation of full membrane potential dynamics of multiple neurons

Biophysically-detailed neuron models distilled into ANNs can be evaluated at significantly higher speeds than their classical counterparts. Because a single instance of a deep learning model can be used to predict outputs for multiple instances of the same neuron model in parallel, on accelerators such as GPUs, these models are particularly suited to accelerating large networks of model neurons. Previous results from the literature [20] have shown that using deep learning models that only predict the output voltage of the soma, could result in a five order magnitude speed-up for a network model of 5000 neurons. In this paper, we have shown that a speed-up of two orders of magnitude can be obtained for a network of 1000 neurons by making use of MTL deep learning models, when the voltage traces for each compartment of the underlying biophysically-detailed neuron model need to be predicted. The highest speedup was achieved with the MH model, which has a slightly higher root mean squared error (4.81 mV) compared to the other two models, but still performs within experimentally observed variability [29].

The presented MTL models did not succeed in the binary somatic spike prediction task. As such, the models are not particularly suited for running recurrently connected neural network simulations. They can, however, be used for investigating, for example, LFP, EEG, or MEG signals from different types of predetermined synaptic input to neural populations, which has

commonly been used in the literature to study the origin and information content of these brain signals [43–57]. Furthermore, in cases where the synaptic input is predetermined, the individual timesteps can be treated independently and in parallel. For example, if we assume a batch size of 1000, this can either correspond to simultaneously simulating 1000 timesteps of a single neuron, or one timestep for a population of 1000 neurons.

An important application of the biophysically-detailed neuron model is the calculation of LFPs and downstream EEG and MEG signals. Recent work has shown the importance of LFP recordings in validating large computational models of brain tissue [58], such as the mouse V1 cortical area, developed by the Allen Institute [18]. The Allen Institute mouse V1 model contains 114 distinct model neuron types, and could be accelerated (by several orders of magnitude) by 114 separate deep learning models each representing one such neuron type. The current biophysically-detailed Allen V1 model makes use of neuron models with passive dendrites which should be significantly easier to distil into an MTL architecture than the layer 5 pyramidal neuron model represented here. Further acceleration of the MTL models discussed in this paper could be achieved by running model inference on multiple GPUs or more advanced accelerators such as TPUs and IPUs.

4 Methods and materials

4.1 Multicompartmental NEURON simulations and data balancing

As a baseline for training and testing, we used an existing dataset of electrophysiological data generated in NEURON [19] based on a well-known biophysical-detailed and multi-compartment model of cortical layer V pyramidal cells [29]. This model contains a wide range of dendritic (Ca^{2+}-driven) and perisomatic (Na^{+} and K^{+}-driven) active properties which are represented by ten key active ionic currents that are unevenly distributed over different dendritic compartments. The data was generated in response to presynaptic spike trains sampled from a Poisson process, with a firing rate of 1.4Hz for excitatory and 1.3Hz for inhibitory synapses. For the purposes of this paper, it is important to note that the biophysically-detailed model contains 639 compartments and 1278 synapses and that 128 simulations of the complete model for 6 seconds of biological time each were included before data balancing. In accordance with previous work from the literature [19] the synapses were distributed uniformly across the span of the neuron with one inhibitory and one excitatory synapse per compartment.

The subthreshold dynamics of the membrane potential in a compartment have small variations and as a result one would expect them to be easy to predict. However, suprathreshold deviations generated by action potentials propagating through the dendrites can be more problematic to predict. To address this issue, we implemented a form of data balancing. We first identified the time points at which somatic spikes occurred and afterwards we used them to create a dataset in which one third of the targets included a spiking event. Additionally, we standardised the membrane potential through z-scoring. We used input data consisting of membrane potentials and incoming synaptic events across all compartments during a 100 ms time window, and target data consisting of 1 ms of membrane potentials and a binary value for the presence or absence of a somatic spike.

4.2 Hard and soft parameter sharing architectures for multitask learning

We implemented a temporal convolutional network (TCN) architecture [59] to learn spatio-temporal relationships between inputs (synaptic events and membrane potentials) and outputs (membrane potentials), for both hard and soft parameter-sharing architectures for multi-task learning. While recurrent neural networks [60, 61] such as GRU's and LSTM's are commonly

used for sequence tasks, research has shown that convolutional architectures can perform just as well or better on tasks like audio synthesis, language modelling, and machine translation [59]. A TCN is a one-dimensional convolutional network with a causal structure, which guarantees causality layer by layer through the use of causal convolutions, padding and dilation. TCNs preserve the temporal ordering of the input data, making it impossible for the network to use information from the future to make predictions about the past. This is important because it ensures that the network can be applied in real-world scenarios where only past information is available.

In early studies of modern multi-task learning, hard parameter sharing was used to share the initial layers ("the bottom") and task-specific top layers ("the towers") of the neural network architecture. This approach has the advantage of being scalable with increasing numbers of tasks, but it can result in a biased shared representation that favours tasks with dominant loss signals. To address this issue, soft parameter sharing architectures have been developed, which utilise dedicated representations for each task. In this study, we employ two soft parameter sharing models, the multi-gate mixture-of-experts (MMoE) [30] and the multi-gate mixture-of-experts with exclusivity (MMoEEx) [31]. These models combine representations learned by multiple shared bottoms ("the experts") through gating functions that apply linear combinations using learnable and data-dependent weights. MMoEEx is an extension of MMoE that encourages diversity among expert representations by randomly setting a subsection of weights to zero.

In our study, we employed a three-layered TCN with varying channel sizes of 32, 16, and 8 and a kernel size of 10. We also implemented a dropout of 0.2 for both the bottom architecture (hard parameter sharing) and the expert architectures (soft parameter sharing). For each tower, we used a three-layered feed-forward network with ELU activation functions and 10 or 25 hidden nodes, depending on whether it was a soft or hard parameter sharing model, respectively. To ensure increased stability, we replaced the final RELU activation function in the original TCN implementation with a sigmoid activation function. In addition, we included an extra TCN expert as a compressor module in the soft parameter sharing models to reduce the size of the data before feeding it into the gating functions, see Fig 5, which helped to avoid quadratic growth of the trainable parameters with respect to the input data size. To improve efficiency, we implemented mixed precision and multi-GPU training for all three models. Our models were trained on up to six parallel RTX2080Ti GPUs, and we performed inference experiments on publicly available Tesla P100 GPUs via Kaggle.

4.3 Task balancing and expert diversity

Effective multi-task learning sometimes requires some form of task balancing to reduce negative transfer or to prevent one or more tasks from dominating the optimisation procedure. To avoid these issues we make use of loss-balanced task weighting (LBTW) [33] which dynamically updates task weights in the loss function during training. For each batch, LBTW calculates loss weights based on the ratio between the current loss and the initial loss for each task, and a hyperparameter α. As α goes to 0, LBTW approaches standard multitask learning training. All taken together, the loss function can be summarised as

$$L(\hat{y}, y) = L_{spike}(\hat{y}_{spike}, y_{spike}) + \sum_{i=0}^{N} L_i(\hat{y}_i, y_i) = w_{spike} \cdot \mathrm{BCE}_{logits}(\hat{y}_{spike}, y_{spike}) + \sum_{i=0}^{N} w_i \cdot \mathrm{MSE}(\hat{y}_i, y_i)\,, \quad (1)$$

where $\hat{y}$ and y are the target and the predicted data respectively, w_{spike} and w_i are the task weight, BCE_{logits} is the binary cross entropy loss combined with a sigmoid activation function, and MSE is the mean square error. The summation index i runs over all N compartments of

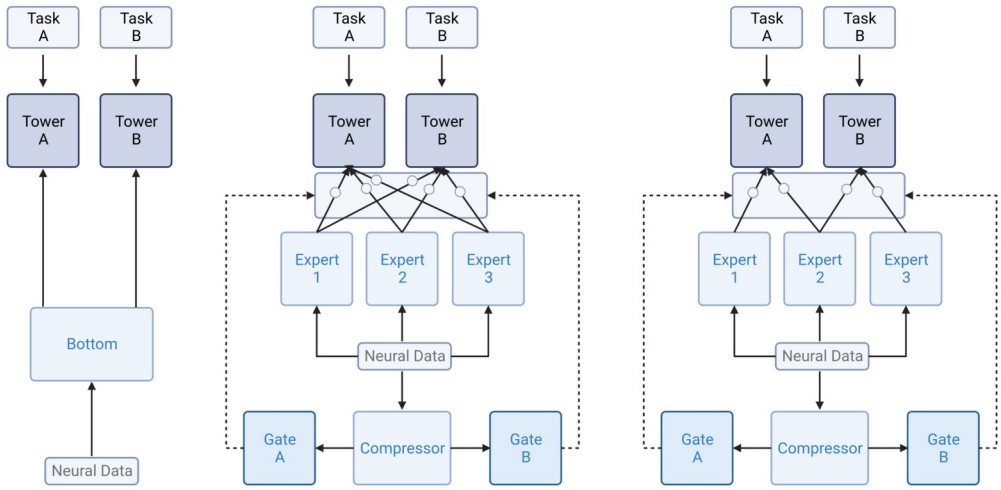

Fig 5. Schematic representation of the architectures of the hard parameter model (left), the MMoE model (middle), and the MMoEEx model (right).

https://doi.org/10.1371/journal.pcbi.1011728.g005

the biophysically-detailed neuron model, and the LBTW task weights are recalculated every epoch E, for each batch B, according to

$$w_{spike} = \left(\frac{L_{spike,E,B}(\hat{y}_{spike}, y_{spike})}{L_{spike,E,0}(\hat{y}_{spike}, y_{spike})}\right)^{\alpha} \quad \text{and} \quad w_i = \left(\frac{L_{i,E,B}(\hat{y}_i, y_i)}{L_{i,E,0}(\hat{y}_i, y_i)}\right)^{\alpha}. \tag{2}$$

To measure the diversity in expert representations in MMoE and MMoEEx, we made use of diversity measurement proposed in the original MMoEEx paper. In that paper, the diversity between two experts n and m is calculated as a (real-valued) distance $d(n, m)$ between the learned representations f_n and f_m, as defined by

$$d(n, m) = \sqrt{\sum_{i=0}^{N} (f_n(x_i) - f_m(x_i))^2}, \tag{3}$$

where N is the number of samples x_i in the validation set, which is used to probe the diversity of the experts. The diversity matrix D of a trained MMoE or MMoEEx model is defined by calculating all pairwise distances between expert representations as described above and normalising the matrix. In this normalised matrix, a pair of experts with distance close to 0 are considered near-identical, and experts with a distance close to 1 are considered to be highly diverse. To compare two different models in terms of overall diversity, we respectively define the first, second, and third diversity scores of a model as the mean entry ($\bar{d}_1$) and the determinant ($\bar{d}_2$), and the permanent ($\bar{d}_3$) of its diversity matrix D.

Supporting information

S1 File. Text, containing all supporting figures.
(PDF)

S1 Fig. Smoothed time series plot of the evolution of the three diversity metrics (determinant, permanent and score) for two different training runs for both the MMoE (left) and

the MMoEEx models (right).
(TIFF)

S2 Fig. Smoothed time series plot of the evolution of the training and validation loss for all three models.
(TIFF)

S3 Fig. Swarm plots of the validation loss for each of the compartments (basal, oblique, apical) of the three LBTW-trained models: Hard Parameter Sharing (MH), Multi-gate Mixture-of-experts (MMoE), and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEx).
(TIFF)

S4 Fig. From left to right. Illustration of a biophysically-detailed model of a multi-compartment cortical layer V pyramidal cell model. Membrane voltages as calculated by a biophysically-detailed simulations of the multi-compartment model, used as the ground truth throughout this paper. Membrane voltages as predicted by our best-performing multi-task learning architecture, one time step at a time (1 ms). Comparison between the ground truth and predicted extracellular potentials calculated at eight points representing the position of the electrodes.
(TIFF)

Acknowledgments

Fig 3 has been made with Biorender.com through an academic license.

Author Contributions

Conceptualization: Jonas Verhellen, Torbjørn V. Ness.

Data curation: Jonas Verhellen, Kosio Beshkov, Torbjørn V. Ness.

Formal analysis: Jonas Verhellen, Kosio Beshkov, Sebastian Amundsen, Torbjørn V. Ness.

Funding acquisition: Gaute T. Einevoll.

Investigation: Jonas Verhellen, Kosio Beshkov, Torbjørn V. Ness.

Methodology: Jonas Verhellen, Kosio Beshkov, Sebastian Amundsen, Torbjørn V. Ness.

Project administration: Jonas Verhellen, Gaute T. Einevoll.

Resources: Gaute T. Einevoll.

Software: Jonas Verhellen, Kosio Beshkov, Sebastian Amundsen, Torbjørn V. Ness.

Supervision: Gaute T. Einevoll.

Validation: Jonas Verhellen, Kosio Beshkov, Sebastian Amundsen, Torbjørn V. Ness.

Visualization: Jonas Verhellen, Kosio Beshkov, Sebastian Amundsen, Torbjørn V. Ness.

Writing – original draft: Jonas Verhellen, Kosio Beshkov, Torbjørn V. Ness.

Writing – review & editing: Jonas Verhellen, Kosio Beshkov, Sebastian Amundsen, Torbjørn V. Ness, Gaute T. Einevoll.

References

1. Hodgkin AL, Huxley AF, Katz B. Measurement of current-voltage relations in the membrane of the giant axon of Loligo. The Journal of physiology. 1952; 116(4):424. https://doi.org/10.1113/jphysiol.1952.sp004716 PMID: 14946712
2. Hodgkin AL, Huxley AF. Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. The Journal of physiology. 1952; 116(4):449. https://doi.org/10.1113/jphysiol.1952.sp004717 PMID: 14946713
3. Hodgkin AL, Huxley AF. The components of membrane conductance in the giant axon of Loligo. The Journal of physiology. 1952; 116(4):473. https://doi.org/10.1113/jphysiol.1952.sp004718 PMID: 14946714
4. Hodgkin AL, Huxley AF. The dual effect of membrane potential on sodium conductance in the giant axon of Loligo. The Journal of physiology. 1952; 116(4):497. https://doi.org/10.1113/jphysiol.1952.sp004719 PMID: 14946715
5. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of physiology. 1952; 117(4):500. https://doi.org/10.1113/jphysiol.1952.sp004764 PMID: 12991237
6. Häusser M. The Hodgkin-Huxley theory of the action potential. Nature neuroscience. 2000; 3 (11):1165–1165. PMID: 11127828
7. Brown A. The Hodgkin and Huxley papers: still inspiring after all these years; 2022.
8. Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology. 1959; 148(3):574. https://doi.org/10.1113/jphysiol.1959.sp006308 PMID: 14403679
9. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology. 1962; 160(1):106. https://doi.org/10.1113/jphysiol.1962.sp006837 PMID: 14449617
10. Hubel DH, Wiesel TN. Ferrier lecture-Functional architecture of macaque monkey visual cortex. Proceedings of the Royal Society of London Series B Biological Sciences. 1977; 198(1130):1–59. PMID: 20635
11. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology. 1968; 195(1):215–243. https://doi.org/10.1113/jphysiol.1968.sp008455 PMID: 4966457
12. Dayan P, Abbott LF. Theoretical neuroscience: computational and mathematical modeling of neural systems. MIT press; 2005.
13. Einevoll GT, Destexhe A, Diesmann M, Grün S, Jirsa V, de Kamps M, et al. The scientific case for brain simulations. Neuron. 2019; 102(4):735–744. https://doi.org/10.1016/j.neuron.2019.03.027 PMID: 31121126
14. Traub RD, Contreras D, Cunningham MO, Murray H, LeBeau FE, Roopun A, et al. Single-column thalamocortical network model exhibiting gamma oscillations, sleep spindles, and epileptogenic bursts. Journal of neurophysiology. 2005; 93(4):2194–2232. https://doi.org/10.1152/jn.00983.2004 PMID: 15525801
15. Potjans TC, Diesmann M. The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. Cerebral cortex. 2014; 24(3):785–806. https://doi.org/10.1093/cercor/bhs358 PMID: 23203991
16. Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and simulation of neocortical microcircuitry. Cell. 2015; 163(2):456–492. https://doi.org/10.1016/j.cell.2015.09.029 PMID: 26451489
17. Schmidt M, Bakker R, Hilgetag CC, Diesmann M, van Albada SJ. Multi-scale account of the network structure of macaque visual cortex. Brain Structure and Function. 2018; 223(3):1409–1435. https://doi.org/10.1007/s00429-017-1554-4 PMID: 29143946
18. Billeh YN, Cai B, Gratiy SL, Dai K, Iyer R, Gouwens NW, et al. Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. Neuron. 2020; 106(3):388–403. https://doi.org/10.1016/j.neuron.2020.01.040 PMID: 32142648
19. Beniaguev D, Segev I, London M. Single cortical neurons as deep artificial neural networks. Neuron. 2021; 109(17):2727–2739. https://doi.org/10.1016/j.neuron.2021.07.002 PMID: 34380016
20. Olah VJ, Pedersen NP, Rowan MJ. Ultrafast simulation of large-scale neocortical microcircuitry with biophysically realistic neurons. Elife. 2022; 11:e79535. https://doi.org/10.7554/eLife.79535 PMID: 36341568
21. Hagen E, Næss S, Ness TV, Einevoll GT. Multimodal modeling of neural network activity: computing LFP, ECoG, EEG, and MEG signals with LFPy 2.0. Frontiers in neuroinformatics. 2018; 12:92. https://doi.org/10.3389/fninf.2018.00092 PMID: 30618697

22. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. PMLR; 2017. p. 1126–1135.

23. Zhang Y, Yang Q. An overview of multi-task learning. National Science Review. 2018; 5(1):30–43. https://doi.org/10.1093/nsr/nwx105

24. Ruder S. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:170605098. 2017;.

25. Lee S, Son Y. Multitask learning with single gradient step update for task balancing. Neurocomputing. 2022; 467:442–453. https://doi.org/10.1016/j.neucom.2021.10.025

26. Mallya A, Davis D, Lazebnik S. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 67–82.

27. Caruana R. Multitask learning: A knowledge-based source of inductive bias1. In: Proceedings of the Tenth International Conference on Machine Learning. Citeseer; 1993. p. 41–48.

28. Duong L, Cohn T, Bird S, Cook P. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In: Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers); 2015. p. 845–850.

29. Hay E, Hill S, Schürmann F, Markram H, Segev I. Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. PLoS computational biology. 2011; 7(7): e1002107. https://doi.org/10.1371/journal.pcbi.1002107 PMID: 21829333

30. Ma J, Zhao Z, Yi X, Chen J, Hong L, Chi EH. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining; 2018. p. 1930–1939.

31. Aoki R, Tung F, Oliveira GL. Heterogeneous multi-task learning with expert diversity. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2022; 19(6):3093–3102. PMID: 35576418

32. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014;.

33. Liu S, Liang Y, Gitter A. Loss-balanced task weighting to reduce negative transfer in multi-task learning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33; 2019. p. 9977–9978.

34. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. Available from: https://www.tensorflow.org/.

35. Hines ML, Carnevale NT. The NEURON simulation environment. Neural computation. 1997; 9 (6):1179–1209. https://doi.org/10.1162/neco.1997.9.6.1179 PMID: 9248061

36. Awile O, Kumbhar P, Cornu N, Dura-Bernal S, King JG, Lupton O, et al. Modernizing the NEURON simulator for sustainability, portability, and performance. Frontiers in Neuroinformatics. 2022; 16:884046. https://doi.org/10.3389/fninf.2022.884046 PMID: 35832575

37. Le Bé JV, Silberberg G, Wang Y, Markram H. Morphological, electrophysiological, and synaptic properties of corticocallosal pyramidal cells in the neonatal rat neocortex. Cerebral cortex. 2007; 17(9):2204–2213. https://doi.org/10.1093/cercor/bhl127 PMID: 17124287

38. Larkum ME, Zhu JJ, Sakmann B. A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature. 1999; 398(6725):338–341. https://doi.org/10.1038/18686 PMID: 10192334

39. Larkum ME, Zhu JJ, Sakmann B. Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons. The Journal of physiology. 2001; 533(2):447–466. https://doi.org/10.1111/j.1469-7793.2001.0447a.x PMID: 11389204

40. Pettersen KH, Einevoll GT. Amplitude variability and extracellular low-pass filtering of neuronal spikes. Biophysical journal. 2008; 94(3):784–802. https://doi.org/10.1529/biophysj.107.111179 PMID: 17921225

41. Pettersen KH, Lindén H, Dale AM, Einevoll GT. Extracellular spikes and CSD. Handbook of neural activity measurement. 2012; 1:92–135. https://doi.org/10.1017/CBO9780511979958.004

42. Lindén H, Pettersen KH, Einevoll GT. Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. Journal of computational neuroscience. 2010; 29:423–444. https://doi.org/10.1007/s10827-010-0245-4 PMID: 20502952

43. Pettersen KH, Hagen E, Einevoll GT. Estimation of population firing rates and current source densities from laminar electrode recordings. Journal of computational neuroscience. 2008; 24:291–313. https://doi.org/10.1007/s10827-007-0056-4 PMID: 17926125

44. Lindén H, Pettersen KH, Einevoll GT. Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. Journal of computational neuroscience. 2010; 29:423–44. https://doi.org/10.1007/s10827-010-0245-4 PMID: 20502952

45. Lindén H, Tetzlaff T, Potjans TC, Pettersen KH, Grün S, Diesmann M, et al. Modeling the Spatial Reach of the LFP. Neuron. 2011; 72:859–72. https://doi.org/10.1016/j.neuron.2011.11.006 PMID: 22153380

46. Schomburg EW, Anastassiou CA, Buzsaki G, Koch C. The Spiking Component of Oscillatory Extracellular Potentials in the Rat Hippocampus. Journal of Neuroscience. 2012; 32:11798–11811. https://doi.org/10.1523/JNEUROSCI.0656-12.2012 PMID: 22915121

47. Lęski S, Lindén H, Tetzlaff T, Pettersen KH, Einevoll GT. Frequency dependence of signal power and spatial reach of the local field potential. PLoS computational biology. 2013; 9:1–23. https://doi.org/10.1371/journal.pcbi.1003137 PMID: 23874180

48. Mazzoni A, Lindèn H, Cuntz H, Lansner A, Panzeri S, Einevoll GT. Computing the Local Field Potential (LFP) from Integrate-and-Fire Network Models. PLOS Computational Biology. 2015; 11:e1004584. https://doi.org/10.1371/journal.pcbi.1004584 PMID: 26657024

49. Hagen E, Dahmen D, Stavrinou ML, Lindén H, Tetzlaff T, Albada SJV, et al. Hybrid scheme for modeling local field potentials from point-neuron networks. Cerebral Cortex. 2016; 26:4461–4496. https://doi.org/10.1093/cercor/bhw237 PMID: 27797828

50. Ness TV, Remme MWH, Einevoll GT. Active subthreshold dendritic conductances shape the local field potential. Journal of Physiology. 2016; 594:3809–3825. https://doi.org/10.1113/JP272022 PMID: 27079755

51. Hagen E, Fossum JC, Pettersen KH, Alonso JM, Swadlow HA, Einevoll GT. Focal local field potential signature of the single-axon monosynaptic thalamocortical connection. Journal of Neuroscience. 2017; 37:5123–5143. https://doi.org/10.1523/JNEUROSCI.2715-16.2017 PMID: 28432143

52. Luo J, Macias S, Ness TV, Einevoll GT, Zhang K, Moss CF. Neural timing of stimulus events with microsecond precision. PLoS biology. 2018; 16:1–22. https://doi.org/10.1371/journal.pbio.2006422 PMID: 30365484

53. Ness TV, Remme MWH, Einevoll GT. h-type membrane current shapes the local field potential from populations of pyramidal neurons. Journal of Neuroscience. 2018; 38:6011–6024. https://doi.org/10.1523/JNEUROSCI.3278-17.2018 PMID: 29875266

54. Skaar JEW, Stasik AJ, Hagen E, Ness TV, Einevoll GT. Estimation of neural network model parameters from local field potentials (LFPs). PLoS Computational Biology. 2020; 16:e1007725. https://doi.org/10.1371/journal.pcbi.1007725 PMID: 32155141

55. Næss S, Halnes G, Hagen E, Hagler DJ, Dale AM, Einevoll GT, et al. Biophysically detailed forward modeling of the neural origin of EEG and MEG signals. NeuroImage. 2021; 225:2020.07.01.181875. PMID: 33075556

56. Martínez-Cañada P, Ness TV, Einevoll GT, Fellin T, Panzeri S. Computation of the electroencephalogram (EEG) from network models of point neurons. PLOS Computational Biology. 2021; 17:e1008893. https://doi.org/10.1371/journal.pcbi.1008893 PMID: 33798190

57. Hagen E, Magnusson SH, Ness TV, Halnes G, Babu PN, Linssen C, et al. Brain signal predictions from multi-scale networks using a linearized framework. PLOS Computational Biology. 2022; 18:2022.02.28.482256. https://doi.org/10.1371/journal.pcbi.1010353 PMID: 35960767

58. Rimehaug AE, Stasik AJ, Hagen E, Billeh YN, Siegle JH, Dai K, et al. Uncovering circuit mechanisms of current sinks and sources with biophysical simulations of primary visual cortex. eLife. 2023; 12:e87169. https://doi.org/10.7554/eLife.87169 PMID: 37486105

59. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:180301271. 2018;.

60. Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997; 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 PMID: 9377276

61. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:14123555. 2014;.