

# NOBLE - Neural Operator with Biologically-informed Latent Embeddings to Capture Experimental Variability in Biological Neuron Models

Luca Ghafourpour<sup>1,2</sup> Valentin Duruisseaux<sup>2,\*</sup> Bahareh Tolooshams<sup>2,3,4,\*</sup> Philip H. Wong<sup>5</sup>  
 Costas A. Anastassiou<sup>5,6,7,8</sup> Anima Anandkumar<sup>2</sup>

<sup>1</sup>Department of Mathematics, ETH Zürich, Zürich, Switzerland

<sup>2</sup>Department of Computing and Mathematical Sciences, California Institute of Technology,  
 Pasadena, CA, USA

<sup>3</sup>Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada

<sup>4</sup>Neuroscience and Mental Health Institute, University of Alberta, Alberta, Canada

<sup>5</sup>Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA

<sup>6</sup>Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA

<sup>7</sup>Center for Biomedical Science, Cedars-Sinai Medical Center, Los Angeles, CA, USA

<sup>8</sup>Archimedes AI, Athena Research Center, Marousi, Greece

## Abstract

Characterizing the diverse computational properties of human neurons via multimodal electrophysiological, transcriptomic, and morphological data provides the foundation for constructing and validating bio-realistic neuron models that can advance our understanding of fundamental mechanisms underlying brain function. However, current modeling approaches remain constrained by the limited availability and intrinsic variability of experimental neuronal data. To capture variability, ensembles of deterministic models are often used, but are difficult to scale as model generation requires repeating computationally expensive optimization for each neuron. While deep learning is becoming increasingly relevant in this space, it fails to capture the full biophysical complexity of neurons, their nonlinear voltage dynamics, and variability. To address these shortcomings, we introduce **NOBLE**, a neural operator framework that learns a mapping from a continuous frequency-modulated embedding of interpretable neuron features to the somatic voltage response induced by current injection. Trained on data generated from biophysically realistic neuron models, NOBLE predicts distributions of neural dynamics accounting for the intrinsic experimental variability. Unlike conventional bio-realistic neuron models, interpolating within the embedding space offers models whose dynamics are consistent with experimentally observed responses. NOBLE is the first scaled-up deep learning framework validated on real experimental data, enabling efficient generation of synthetic neurons that exhibit trial-to-trial variability and achieve a  $4200\times$  speedup over numerical solvers. To this end, NOBLE captures fundamental neural properties, opening the door to a better understanding of cellular composition and computations, neuromorphic architectures, large-scale brain circuits, and general neuroAI applications.

---

\*These authors contributed equally to this work.

# 1 Introduction

Hundreds of distinct neuronal cell types co-exist and compute within neural circuits, yet how they shape cognitive functions remains essentially unanswered [1–5]. This is particularly true in the human brain, where access and monitoring capabilities are severely limited compared to animal models. Over the past decade, multimodal cellular datasets that integrate electrophysiology, morphology, and transcriptomics have emerged for human cell types [6–10]. While integrating across the different modalities remains a challenge, clear differences in gene expression, morphology, and electrophysiology are evident across cell types. However, how such diversity scales up to circuit-level function is still unknown. Understanding how these differences affect overall brain function is crucial in many contexts, for instance, to uncover genes or phenotypes related to neurological diseases.

Cellular models representing multiple data modalities are invaluable as they offer a degree of control and perturbations that are experimentally infeasible (e.g., [11–13]). Recently, evolutionary multi-objective optimization algorithms [14] have been used to generate and validate realistic models of neurons in the form of 3D multi-compartment partial differential equation (PDE) models that mirror both their shape and ion channel expression, shaping their electrical properties [12, 15, 13, 16]. The downside is that such models are deterministic and fail to capture the variability observed experimentally, where identical input to the same neuron often results in different electrophysiological responses. One approach is to generate families of models, sometimes referred to as "hall-of-fame" (HoF) models [12, 13, 17] to represent a single cell.

While each HoF model is distinct and reproduces the electrophysiological features of parts of an experiment, the ensemble of deterministic HoF models is used as a collective representation that captures both the main features as well as their variability in an experiment [12, 13]. Typically, neurons exhibit highly nonlinear behavior, necessitating equally complex models with a large number of parameters, which makes multi-objective evolutionary optimization computationally demanding, requiring about 600k CPU core hours per single-neuron model [12, 13]. Yet, even tiny perturbations of the model parameters lead to large deviations from experimental data [18]. In summary, capturing the nature and variability of neurons is a challenge with existing computational techniques.

The challenges of scalability and computational cost associated with traditional modeling approaches, such as numerical integrators and evolutionary optimization algorithms, have led the scientific community to explore the use of machine learning to accelerate simulations by learning underlying relationships between variables directly from data. While neural networks have been used successfully for many applications, they learn mappings between finite-dimensional vectors, which can limit their ability to model physical phenomena that are better described using functions in infinite-dimensional spaces and functional relationships between them [19]. As a result, neural networks can overfit to the training discretization and suffer from limited out-of-distribution capabilities. Neural operators [19, 20] are a principled way to generalize neural networks to learn operators mapping functions to functions, with a universal operator approximation property [21]. A variety of neural operators have been proposed, such as the Fourier Neural Operator (FNO) [22], and successfully applied to a wide range of problems [23–25]. Prior knowledge of the relevant physics laws and differential equations can also be incorporated as additional loss terms during training, to supplement or replace reference data, as done with physics-informed neural operators [26, 27].

Earlier applications of machine learning to single-cell electrophysiology focused on directly learning the dynamics of canonical point-neuron models like FitzHugh-Nagumo (FHN) [28, 29] and Hodgkin-Huxley (HH) [30]. Fully connected and convolutional neural networks were trained to reproduce FHN dynamics [31], while ResNet-based multilayer perceptrons showed promise in learning HH

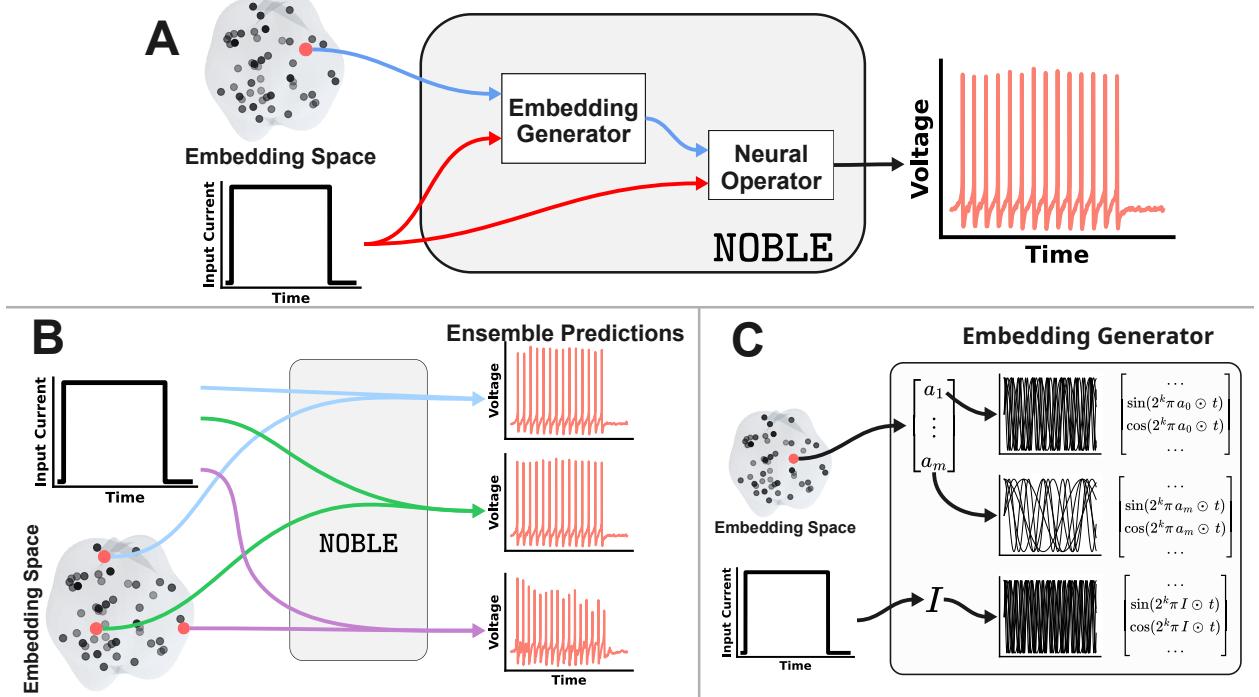


Figure 1: The Neural Operator with Biologically-informed Latent Embeddings (NOBLE) framework. **A)** In NOBLE, a current injection and neuron model features are first encoded using the proposed embedding strategy, before passing through a neural operator to produce a prediction for the somatic voltage response. **B)** NOBLE can be queried in parallel with different model latent representations to produce ensemble predictions. **C)** The proposed embedding in NOBLE encodes specified neuron features and the input current as a stack of trigonometric time-series, as described in Section 3.3.

dynamics [32, 33]. Physics-informed neural networks (PINNs) extend conventional neural networks by introducing prior knowledge about the underlying dynamical system [34–36]. This formulation was used to predict FHN dynamics [37] and to learn the HH model ionic conductances from simulated voltage recordings [38]. Further refinements with PINNs incorporated wavelet bases to capture localized multiscale dynamics and compute derivatives analytically, improving both accuracy and convergence speed when training on the FHN model [39]. While these methods demonstrate capabilities in capturing the core spiking dynamics of these simplified models, their ability to accurately represent the full spectrum of electrophysiological behavior, particularly the highly nonlinear onset of firing, remains largely untested. Moreover, as function approximators, they necessitate retraining for each new input stimulus, significantly limiting their practical utility.

To address these limitations, Centofanti et al. [40] explored using operator learning approaches for forward simulations of the HH model. Among other approaches, FNOs showed promising results by demonstrating a strong capacity for learning the governing operators of this biophysical system. However, this work still exhibits key limitations and a limited scope: (1) it relies on relatively simple simulated data from a point-neuron model, (2) it does not explicitly attempt to capture the full spectrum of electrophysiological dynamics, particularly the highly nonlinear onset of firing, and (3) its formulation on a single operator inherently lacks the capacity to represent the trial-to-trial variability observed in biological recordings. This paper builds on the promise of operator learning for neuroscience demonstrated in [40], extending its scope and addressing its limitations to enable deeper insights into brain function and its applications in neuroAI.

**Contributions.** We introduce **NOBLE** (Neural Operator with Biologically-informed Latent Embeddings), a neural operator framework for learning the nonlinear somatic dynamics across a population of HoF models for a single neuron (Figure 1). **NOBLE** is the first scaled-up deep learning framework whose performance is validated with experimental data. Rather than training a separate independent surrogate for each HoF model, **NOBLE** learns a single neural operator that maps from a continuous latent space of user-defined, interpretable neuron characteristics to an ensemble of somatic voltage responses induced by current injection. This latent space is constructed using an embedding strategy informed by the specified characteristics of the neuron models.

As an example application, we instantiate **NOBLE** with an FNO, and train and evaluate it on a parvalbumin-positive (PVALB) neuron dataset generated using 50 HoF PVALB models (Figure 2). We show that a single **NOBLE** model accurately captures both subthreshold and spiking dynamics across all 50 HoF models (in-distribution) as well as unseen HoF (out-of-distribution) while achieving a significant speedup of  $4200\times$  over the numerical solver used to generate the dataset (Figure 4). In addition, the **NOBLE** predictions across 15 electrophysiological features of interest (including spike count, amplitude, and width) remain within the variability observed in experimental data (Figure 7). We further show that **NOBLE** can successfully generate novel bio-realistic neuron models by sampling and interpolating within the latent space of models. The dynamics of novel neuron models generated by **NOBLE** align both with previously unseen HoF PVALB models and experimentally observed somatic responses. In contrast, direct interpolation between the parameters of bio-realistic PDE-based neuron models fails due to the sensitivity and nonlinearity of the underlying PDEs [41, 42] (Figure 5).

Owing to **NOBLE**'s ability to generate novel bio-realistic neuron models, ensemble predictions are no longer constrained to the original 50 HoF models used during training. **NOBLE** can produce somatic voltage responses for an arbitrary number of biologically plausible neurons. We demonstrate this enhanced flexibility by predicting responses to input stimuli across a larger set of models (Figure 6). The results showcase **NOBLE**'s ability to accurately capture a broad range of biophysically realistic neurons while enabling dense interpolation across model space. By unlocking the efficient generation of infinitely many diverse yet biophysically plausible neurons from a continuous embedding, **NOBLE** offers a scalable alternative to computationally intensive, scale-limited evolutionary approaches, laying the foundation for brain-scale neural circuit modeling.

Finally, the biologically-informed latent representation of the neuron models together with the capability of **NOBLE** to generate arbitrarily many new biophysically-realistic neuron models also offers further insight into the behavior of neural dynamics. We can use **NOBLE** to obtain the somatic responses to current injections on a fine grid in the model latent representation space, and consequently construct heat maps and surface plots to better understand how neuron features used for the model latent representation affect any electrophysiological feature of interest.

## 2 Background

### 2.1 Bio-realistic Neuron Modeling

We create bio-realistic PDE-based neuron models using actual reconstructions of neuron morphologies from human cortical data [6, 8, 9] (Figure 2). We instantiate these models using a framework [41] built on the NEURON simulation environment [42]. We place ion channels in an "all-active" configuration [12, 13], where active ion channels are distributed along both somatic and dendritic compartments along the neuron morphology. For each experimental neuron, models are generated

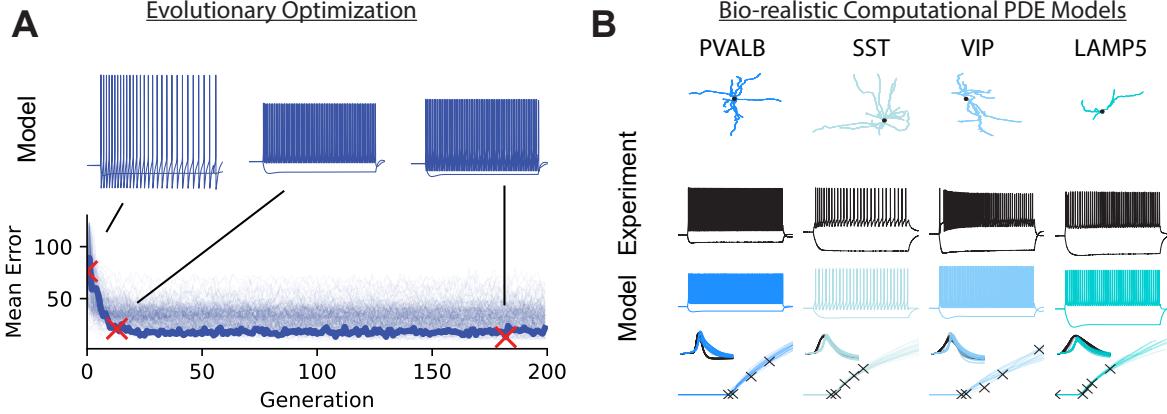


Figure 2: Creation of Bio-realistic PDE-based Neuron Models. **A)** Evolutionary optimization process for a neuron of interest, with voltage responses sampled at different generations (top) and the error history with other experimental neurons overlaid in the background (bottom). **B)** Sample HoF models of various inhibitory cell-types, showing morphology (top), experimental voltage traces (2nd row), simulated voltage traces (3rd row), and spike waveform and frequency-current curves (bottom).

using a multi-objective evolutionary optimization framework [13] to find ion conductance parameters replicating a standard set of electrophysiological features from patch clamp recordings (Figure 2A). We adopt a two-stage optimization strategy, first fitting passive subthreshold responses followed by capturing the active dynamics above the spiking threshold and the full frequency-current curve of each neuron. After 250 generations of evolutionary optimization, the models that best minimize the mean z-score error between simulated and actual experimental electrophysiological features are selected as hall-of-fame (HoF) models (Figure 2B). More details about the electrophysiological features of interest are provided in Appendix E.

To illustrate the proposed approach, we consider in this paper a randomly selected PVALB human cortical neuron, for which we have created 60 HoF models. PVALB neurons are fast-spiking inhibitory interneurons regulating high frequency gamma oscillations (30-80Hz) [43] and their dysfunction has been associated with cognitive impairments such as schizophrenia and Alzheimer’s disease [44, 45].

## 2.2 Neural Operators

Neural operators compose linear integral operators  $\mathcal{K}$  with pointwise nonlinear activation functions  $\sigma$  to approximate highly nonlinear operators. More precisely, we define the neural operator

$$\mathcal{G}_\theta = \mathcal{Q} \circ \sigma(W_L + \mathcal{K}_L + b_L) \circ \dots \circ \sigma(W_1 + \mathcal{K}_1 + b_1) \circ \mathcal{P} \quad (1)$$

where  $\mathcal{P}, \mathcal{Q}$  are the pointwise neural networks that encode the lower dimension function into a higher dimensional space and vice versa. The model stacks  $L$  layers of  $\sigma(W_l + \mathcal{K}_l + b_l)$  where  $W_l$  are pointwise linear operators (matrices),  $\mathcal{K}_l$  are integral kernel operators,  $b_l$  are bias terms, and  $\sigma$  are fixed activation functions. The parameters  $\theta$  consists of all the parameters in  $\mathcal{P}, \mathcal{Q}, W_l, \mathcal{K}_l, b_l$ . Kossaifi et al. [46] maintain a comprehensive open-source PyTorch library for learning neural operators, which serves as the foundation for our implementation.

A Fourier neural operator (FNO) [22] is a neural operator which uses Fourier integral operator layers  $(\mathcal{K}(\phi)v_t)(x) = \mathcal{F}^{-1}(R_\phi \cdot (\mathcal{F}v_t))(x)$  where  $R_\phi$  is the Fourier transform of a periodic function  $\kappa$  parameterized by  $\phi$ . On a uniform mesh, the Fourier transform  $\mathcal{F}$  can be implemented using the fast Fourier transform (FFT). A depiction of the FNO architecture is provided in Figure 8.

### 3 Method

We present the proposed **N**O**B**LE framework, along with the additional data generation and processing mechanisms we developed to enhance training efficiency and reduce computational cost. Specifically, we begin in Section 3.1 by introducing a subsampling strategy for somatic voltage responses that preserves the key neuronal features of interest. We then propose a non-uniform training distribution of current injection amplitudes in Section 3.2 to effectively capture the transition from non-spiking to spiking behavior. Next, we present the neuron model embedding strategy at the core of **N**O**B**LE in Section 3.3, and conclude by detailing the full architecture of **N**O**B**LE in Section 3.4.

#### 3.1 Subsampling

**N**O**B**LE employs a notable property of neural operators on training on low-resolution data while reserving the capability to generate dynamics at higher resolution. In this regard, we subsample the reference HoF simulations in time. To avoid discarding high-resolution information necessary for capturing neuron features of interest, we analyze how these features are affected by different subsampling factors and strategies. In particular, we compare their effect on the discrepancy between HoF simulations and experimental data. We consider (i) low-pass filtering followed by decimation in time, (ii) low-pass filtering followed by truncation in the frequency domain, (iii) direct truncation in the frequency domain, and (iv) decimation in time without filtering. Across neuron features, we observe no consistent differences in performance between these strategies and thus opt for low-pass filtering followed by decimation in time. Our analysis (see Appendix B) reveals that a subsampling factor of 3 preserves the fidelity of extracted features, introducing no more error than the intrinsic discrepancy between HoF simulations and experimental data, and does not lead to significant aliasing. This reduces the time series length from 25,750 to 8,583, substantially decreasing the computational load without compromising biological realism.

#### 3.2 Current Amplitude Sampling

We sample square-pulse amplitudes from a skew-normal distribution whose support matches the experimentally validated range of our HoF models,  $I \in [-0.11, 0.28]\text{nA}$ . To effectively capture the highly nonlinear dynamics around the spiking threshold (0 to 0.05nA) where neural responses transition abruptly from being non-spiking to spiking, the mode of our sampling distribution is strategically located within this peri-threshold window. To address the greater learning challenge posed by the high-frequency components of depolarizing, spiking responses (characterised by features such as spike width, latency to first spike, and spike count), we deliberately use a heavier positive tail in our sampling distribution. This bias ensures the model encounters numerous examples of spike onset and complex spiking patterns during training while still covering the full validated input range. The distribution of square-pulse amplitudes considered is shown in the appendix in Figure 11.

#### 3.3 Embedding Strategy for Neuron-Model Variability

The proposed **N**O**B**LE framework learns a single neural operator that maps from a continuous latent space of user-defined, interpretable neuron characteristics to an ensemble of somatic voltage responses induced by current injection. For the PVALB neuron considered in our numerical experiments, the firing rate (F-I) curve is a useful electrophysiological descriptor that summarises cellular excitability

by relating injected current amplitude to the neuron’s firing rate [47]. Differences between HoF parameterisations manifest as shifts in key features of this curve: the threshold current  $I_{thr}$  (the minimum amplitude that elicits spiking) and the local slope  $s_{thr}$  at  $I_{thr}$ , which measures the rate of increase in the firing rate upon spiking. Figure 5A displays examples of F-I curves for different HoF PVALB models, showing the variability in  $I_{thr}$  and  $s_{thr}$  across models.

Using this observation, we propose representing a given PVALB neuron model by its threshold current  $I_{thr}$  and local slope  $s_{thr}$ , that is, we represent the model as  $(I_{thr}, s_{thr})$ . We propose to use this representation as part of a NeRF-style (Neural Radiance Field) embedding [48], where input features are encoded using sine and cosine functions. More precisely, a feature  $p$  is encoded as a stack of trigonometric time-series

$$[\sin(2^0\pi p \odot t), \cos(2^0\pi p \odot t), \dots, \sin(2^{K-1}\pi p \odot t), \cos(2^{K-1}\pi p \odot t)] \quad (2)$$

for some integer  $K > 0$ , where the frequencies are modulated by the feature  $p$ . Here  $t$  denotes the discretized time coordinates and  $\odot$  indicates element-wise multiplication with appropriate broadcasting.

The use of sine and cosine functions for encoding features is particularly synergistic with FNOs, which operate in the frequency domain to learn mappings between functions. FNOs leverage the Fourier transform to represent and manipulate data as sums of sine and cosine functions, effectively learning complex patterns by capturing interactions among frequency components. NeRF-style encodings lead to a representation of the input features that aligns naturally with the spectral approach of FNOs, enhancing their ability to learn high-frequency dynamics. In this context, the sinusoidal embeddings can be thought of as a form of spectral lifting, translating low-dimensional inputs into a richer representation in the frequency domain that FNOs can more efficiently process.

In our PVALB neuron example, we encode separately the model features  $I_{thr}$  and  $s_{thr}$ , and the current injection, using this strategy and stack the resulting embeddings along the channel dimension of the input to the FNO. To ensure periodicity and promote distant representations of distinct HoF models in the feature space, we normalize both  $I_{thr}$  and  $s_{thr}$  to the range  $[0.5, 3.5]^2$  before embedding. Figure 3 displays the latent representations in normalized  $(I_{thr}, s_{thr})$ -space of the 60 HoF PVALB models used in our numerical experiments. This figure also shows that using only one of  $I_{thr}$  and  $s_{thr}$  instead of both would not allow to distinguish very well between different HoF models.

### 3.4 NOBLE: Neural Operator with Biologically-informed Latent Embeddings

We introduce the Neural Operator with Biologically-informed Latent Embeddings (**NOBLE**), for modeling neuronal voltage dynamics in response to current injections. **NOBLE** offers a scalable alternative to computationally intensive numerical solvers for biophysically detailed, PDE-based neuron models. It learns a direct mapping from input currents and an interpretable, continuous latent space of neuron features to voltage traces (Figure 1A). A key feature of **NOBLE** is its use of biologically-informed embeddings to generalize across variability in biological neuron models.

At its core, **NOBLE** is based on a temporal neural operator. The discretization invariance of neural operators allows for **NOBLE** to learn on low-resolution data and infer somatic voltage dynamics at higher resolutions. By combining a neural operator with the proposed continuous interpretable embedding, **NOBLE** learns a continuous operator over the space of bio-realistic neuron models.

The proposed **NOBLE** framework offers several key advantages that set it apart from existing approaches:

- While prior works model neuronal variability using many models, each generating a single prediction, **NOBLE** proposes a unifying framework combining all such models into one. Instead of training separate models for each experimentally observed neuronal variation, **NOBLE** learns directly the ensemble dynamics with the neuronal variability. Conditioned on a particular electrophysiological feature, **NOBLE** generates one realization of the variability in a biological neuron model.
- **NOBLE** can interpolate between model features in its embedding space to generate new biorealistic neuron models. This is a striking property that no prior work offers, (e.g. interpolation between parameters of PDE models offers no guarantee of validity for the resulting model).
- With the proposed continuous embedding space, **NOBLE** can be used to generate infinitely many different instances of neuron dynamics (both spiking and subthreshold), since it can interpolate between model features to generate new neuron models.
- The biologically-informed latent space of **NOBLE**, combined with its ability to generate arbitrarily many biophysically realistic neuron models, enables fine-grained exploration of neural dynamics. By sampling models across this space, **NOBLE** can produce somatic responses to diverse inputs, allowing for visualization techniques, such as heat maps and surface plots, to reveal how latent features influence electrophysiological behavior.

We demonstrate these novel characteristics of **NOBLE** with extensive numerical experiments, next.

## 4 Results

### 4.1 Experimental Setup

We consider the PVALB neuron modeling example introduced in Section 2.1. Given a PVALB model  $\text{HoF}_\ell$  and a current injection  $I$ , the neural operator in **NOBLE** takes the stacked embeddings of the (normalized) model features  $I_{thr}$  and  $s_{thr}$  associated to  $\text{HoF}_\ell$  (with  $K = 1$  frequency) and  $I$  (with  $K = 9$  different frequencies), and outputs the somatic voltage response. We have access to 60 HoF models  $\{\text{HoF}_i\}_{i=1}^{60}$  obtained using a multi-objective evolutionary optimization strategy. We use 50 of them during training,  $\{\text{HoF}_i^{train}\}_{i=1}^{50}$ , and keep the remaining 10 HoF models  $\{\text{HoF}_i^{test}\}_{i=1}^{10}$  for testing. Figure 3 displays the latent representations in normalized  $(I_{thr}, s_{thr})$ -space of these HoF models, and in particular distinguishes between those included and excluded during training. The training dataset is composed of 37,800 samples, where the current injections are sampled as described in Section 3.2, each of which is associated randomly to one of  $\{\text{HoF}_i^{train}\}$ . The samples are generated using a numerical solver [41] built on the NEURON simulation environment [42].

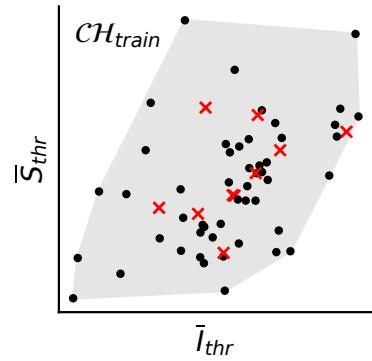


Figure 3: Latent representations in normalized  $(I_{thr}, s_{thr})$ -space of  $\{\text{HoF}_i^{train}\}$  (black dots) and  $\{\text{HoF}_i^{test}\}$  (red crosses) models. The latter lie in the convex hull  $\mathcal{CH}_{train}$  of  $\{\text{HoF}_i^{train}\}$ .

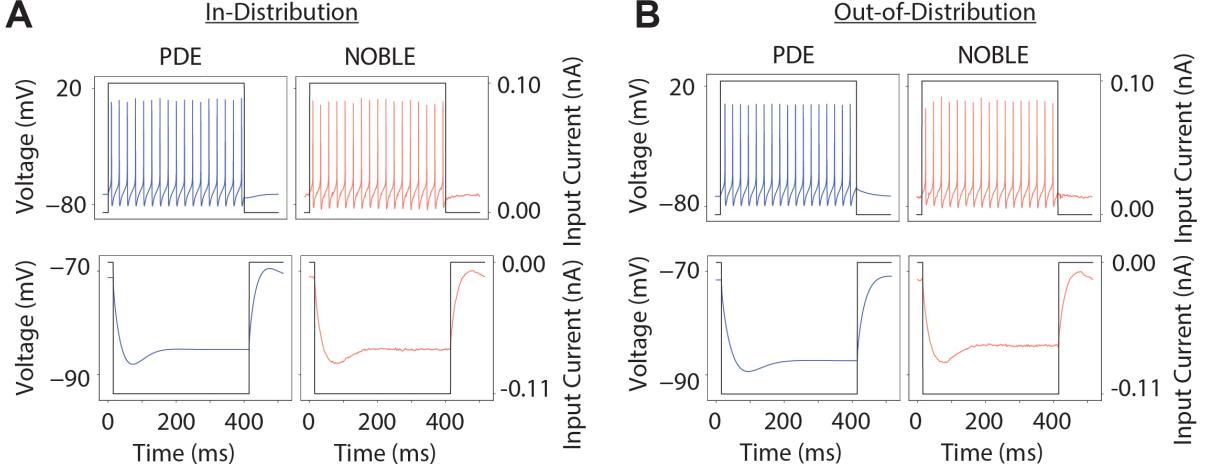


Figure 4: Successful predictions of **NOBLE** for model features experienced during training (In-Distribution) and model features excluded from training (Out-of-Distribution).

We choose the neural operator in **NOBLE** to be a 1D FNO with 12 layers, each with 24 hidden channels and 256 Fourier modes, and use the FNO implementation of the NeuralOperator library [46]. The resulting **NOBLE** with 1.8M trainable parameters is trained in PyTorch for 350 epochs to minimize the relative L4 error to the reference HoF simulations, using the Adam optimizer with learning rate 0.008, and the ReduceLROnPlateau scheduler with factor 0.4 and patience 4.

The trained **NOBLE** generate predictions significantly faster than the reference numerical solver [41]. We record the time necessary to generate 10,000 predictions on a workstation equipped with a single NVIDIA RTX 4090 GPU (24GB VRAM), an AMD Ryzen 9 7900X CPU, and 64GB of system RAM. The numerical solver only generates a single prediction at a time, and takes roughly 36,200 seconds to generate 10,000 predictions. When doing one inference at a time with **NOBLE** (i.e. batch size of 1), we generate 10,000 predictions in 157 seconds, i.e. a speedup of approximately 230 $\times$  compared to the solver. In addition, **NOBLE** can easily be accelerated on a single GPU by generating multiple predictions at the same time. In particular, with a batch size of 1000, **NOBLE** generates 10,000 predictions in 8.59 seconds, i.e. a speedup of approximately 4200 $\times$  compared to the solver.

## 4.2 Testing on HoF Models Experienced during Training

We begin by validating that the trained **NOBLE** can accurately replicate the somatic voltage responses of the HoF models  $\{\text{HoF}_i^{\text{train}}\}$  used for training, when tested on current injections that have not been experienced during training. As shown in Figure 4A, the voltage traces exhibit minimal differences, confirming that **NOBLE** works and generalizes well to unseen current injections. This is also confirmed by a L2 relative error of 2.18% on the test set with the  $\{\text{HoF}_i^{\text{train}}\}$  models. We further generate the F-I curves using the trained **NOBLE** for  $\{\text{HoF}_i^{\text{train}}\}$ , and compare them to the reference F-I curves produced by the numerical solver using the same HoF models. As shown in Figure 5B (top), the curves from both methods remain close, although for 3 out of the 50 HoF models, the **NOBLE** predictions exhibit higher variations in firing rate for high current amplitude.

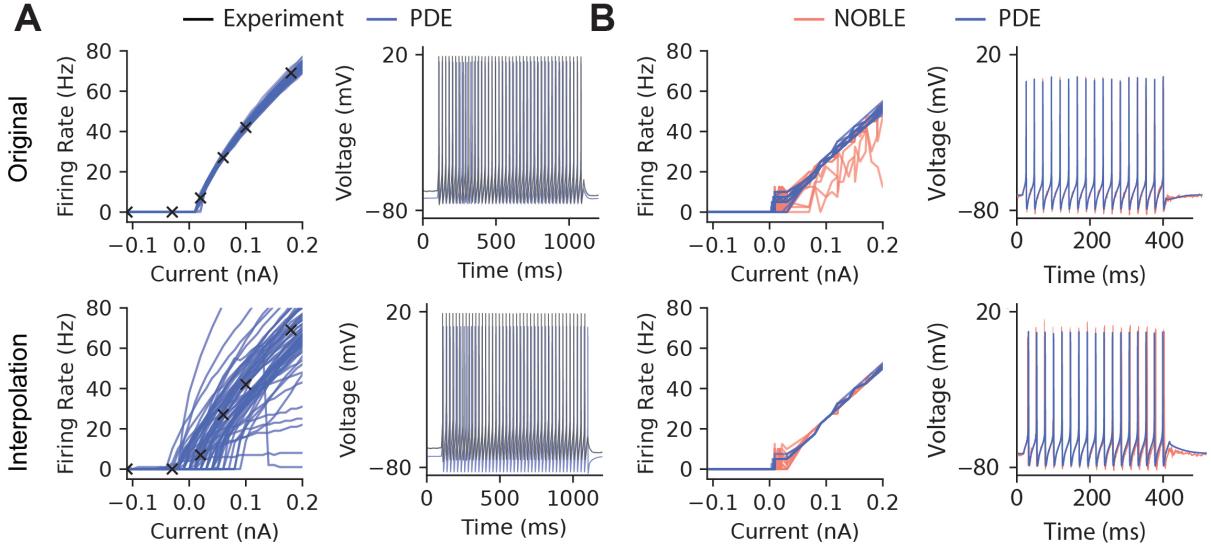


Figure 5: F-I curves and distributions of voltage traces for **A**) experimental versus HoF simulations (PDE) and **B**) NOBLE predictions versus HoF simulations (PDE). The top row shows the results for  $\{\text{HoF}_i^{\text{train}}\}$ , while the bottom row shows the results for interpolation between these models.

### 4.3 Interpolating Between Models

We now test NOBLE with the neuron model features associated with the HoF models  $\{\text{HoF}_i^{\text{test}}\}$  that were not used for training. Note that these models are located within the convex hull  $\mathcal{CH}_{\text{train}}$  of the HoF models  $\{\text{HoF}_i^{\text{train}}\}$  used for training, in the embedding space of the PVALB neuron models. Given a model  $\text{HoF}_k^{\text{test}}$  excluded during training, we sample randomly 50 points in the model latent space close to the defining  $(I_{\text{thr}}, s_{\text{thr}})$  features of  $\text{HoF}_k^{\text{test}}$ . This is illustrated in Figure 12 where we show the small neighborhood around a given model in which the 50 samples are sampled. In particular, none of these 50 new models have been experienced during training. Given a current injection, we then run the 50 corresponding NOBLE inferences in parallel to produce a distribution of predicted somatic responses, and compare to the somatic response obtained using the numerical solver with  $\text{HoF}_k^{\text{test}}$ . Figures 4B and 5B (bottom) show that the voltage traces are very similar, confirming that NOBLE interpolates correctly between models experienced during training, and can be used to obtain new biophysically-realistic models. By contrast, the underlying parameterizations of the HoF models obtained using the multi-objective evolutionary algorithms lack a consistent structure that would allow for meaningful interpolation to discover new biophysically-realistic models. This is exemplified in Figure 5A (bottom) where we see that both voltage responses and F-I curves are completely off when interpolating directly in-between the PDE parameters of the HoF models.

### 4.4 Ensemble Predictions

We now exhibit how the single trained NOBLE can be used for ensemble predictions. Given a current injection, we can first run 50 inferences in parallel of NOBLE for the PVALB HoF models  $\{\text{HoF}_i^{\text{train}}\}_{i=1}^{50}$  to produce 50 different predicted somatic responses. In Figure 6, we compare these 50 predictions with the corresponding 50 predictions obtained using the numerical solver on  $\{\text{HoF}_i^{\text{train}}\}_{i=1}^{50}$ . We see that the distribution of curves are very similar, as expected.

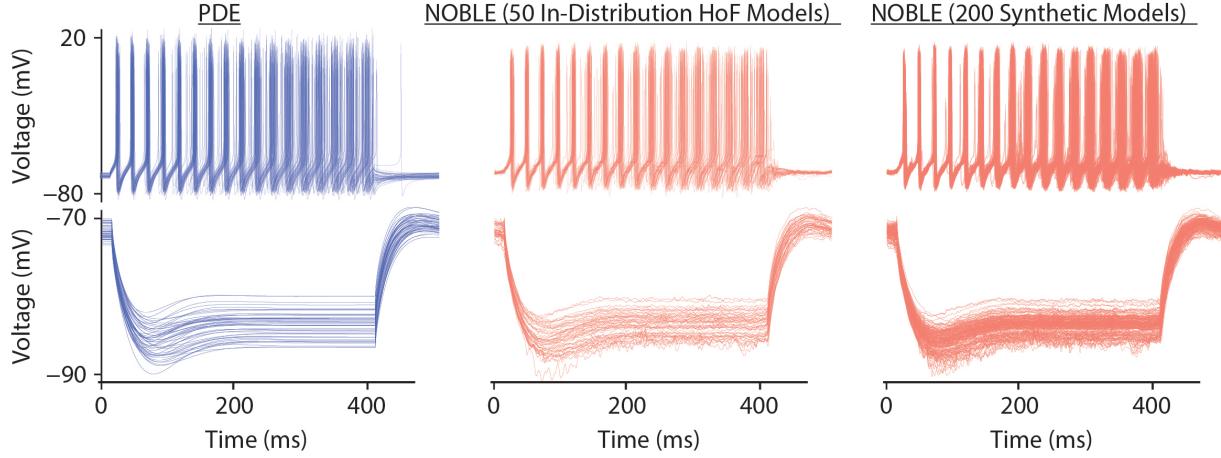


Figure 6: Distributions of voltage traces of  $\{\text{HoF}_i^{\text{train}}\}$  obtained using the numerical solver (left), NOBLE for  $\{\text{HoF}_i^{\text{train}}\}$  (middle), and NOBLE for 200 random samples from the model latent space.

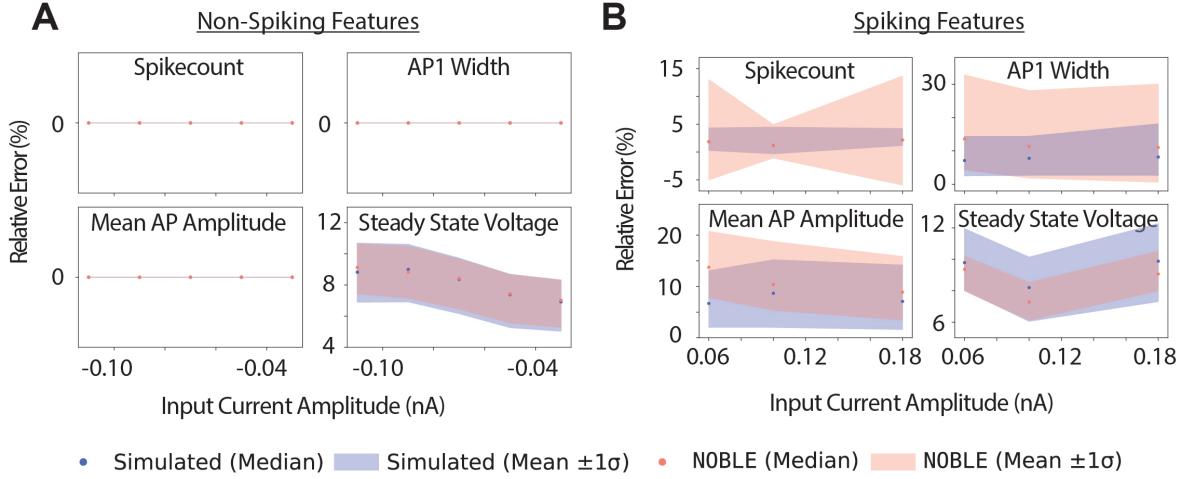


Figure 7: Distributions of key electrophysiological features for ensemble predictions using PDE numerical simulations and synthetic NOBLE predictions when compared to experimental results.

Since NOBLE enables interpolation between the HoF models used for training, we can generate novel, biophysically realistic neuron models and are no longer limited to the original 50 HoF models: the trained NOBLE can produce voltage responses for any neuron model whose representation lies within the convex hull  $\mathcal{CH}_{\text{train}}$  of the training set  $\text{HoF}_i^{\text{train}}$ . We demonstrate this by querying 200 novel models whose features are sampled randomly within  $\mathcal{CH}_{\text{train}}$ . As shown in Figure 6, the distribution of curves remains very similar, but the additional samples provide a denser coverage of the response space while maintaining biophysical realism, with no apparent artifacts or implausible predictions. Figure 7 also shows that the larger distribution of synthetic predictions retains a similar accuracy as the HoF models in terms of the 4 most important electrophysiological features when compared to those of experimental results. These results highlight the capacity of NOBLE to faithfully represent a diverse collection of biophysically realistic neuron models, while allowing for dense interpolation across the model space. As a result, NOBLE could pave the way for improved statistical analysis, and more reliable uncertainty quantification and robust predictive modeling of neuron dynamics.

## 5 Conclusion

We introduced **NOBLE** for learning the nonlinear somatic dynamics across a population of HoF models for a single neuron. Rather than training separate surrogates for each bio-realistic model, **NOBLE** learns a single neural operator that captures the inherent variability observed in experimental neuron recordings by mapping biologically interpretable embeddings to voltage responses from current injections. Demonstrated on HoF models of a PVALB neuron, **NOBLE** correctly captured the diverse neuron dynamics with a  $4200 \times$  speedup over traditional solvers, while maintaining accuracy across key electrophysiological features. **NOBLE** also allows for generating novel, bio-realistic neuron models through interpolation in the latent space, which was infeasible directly with HoF models. This allows for accurate predictions beyond the original set of HoF models, enabling faster model exploration and ensemble prediction. **NOBLE**'s interpretable latent space also offers new insights on how neuron characteristics affect neuron dynamics. While **NOBLE** successfully captures the dynamics across HoF models, it is currently limited to modeling single neurons. A natural extension of this work is to generalize **NOBLE** to multi-neuron settings, enabling applications such as neuron classification and predicting multi-neuron dynamics. Also, the embedding space used in this our experiments is low-dimensional, constructed from two interpretable features derived from a model's F-I curve. Although **NOBLE** demonstrates strong performance in this configuration, extending it to a learnable, higher-dimensional continuous embedding space remains an exciting direction for future work.

## Acknowledgements

L.G. was responsible for the complete technical implementation of this work. C.A.A. and A.A. conceptualized this work. L.G., V.D., and B.T. jointly developed the methodology of the novel **NOBLE** framework. P.H.W. and C.A.A. provided neuroscience-specific domain expertise, contextualizing the relevance and impact of this work within the broader field of neuroscience. P.H.W. supplied the biophysical PDE models and neuron morphologies. L.G. and P.H.W. produced the figures. L.G., V.D., B.T., P.H.W., and C.A.A. co-wrote the manuscript. C.A.A. and A.A. provided supervision and insightful editorial comments.

## Funding

A.A. is supported by the Bren Endowed Chair, ONR (MURI grant N00014-23-1-2654), and the AI2050 Senior Fellow program at Schmidt Sciences. C.A.A. is being supported by the National Institutes of Health R01 - NS120300 and R01 - NS130126. P.H.W. is being supported by the National Institutes of Health R01 - NS130126.

## References

- [1] Rodney J Douglas, Christof Koch, Misha Mahowald, Kevan AC Martin, and Humbert H Suarez. Recurrent excitation in neocortical circuits. *Science*, 269(5226):981–985, 1995.
- [2] György Buzsáki. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362–385, 2010.
- [3] Liqun Luo. Architectures of neuronal circuits. *Science*, 373(6559):eabg7285, 2021.
- [4] Kimberly Siletti, Rebecca Hodge, Alejandro Mossi Albiach, Ka Wai Lee, Song-Lin Ding, Lijuan Hu, Peter Lönnerberg, Trygve Bakken, Tamara Casper, Michael Clark, et al. Transcriptomic diversity of cell types across the adult human brain. *Science*, 382(6667):eadd7046, 2023.
- [5] Xi-Han Zhang, Kevin M Anderson, Hao-Ming Dong, Sidhant Chopra, Elvisha Dhamala, Prashant S Emani, Mark B Gerstein, Daniel S Margulies, and Avram J Holmes. The cell-type underpinnings of the human functional cortical connectome. *Nature Neuroscience*, 28(1):150–160, 2025.
- [6] Jim Berg, Staci A Sorensen, Jonathan T Ting, Jeremy A Miller, Thomas Chartrand, Anatoly Buchin, Trygve E Bakken, Agata Budzillo, Nick Dee, Song-Lin Ding, et al. Human neocortical expansion involves glutamatergic neuron diversification. *Nature*, 598(7879):151–158, 2021.
- [7] Thomas Chartrand, Rachel Dalley, Jennie Close, Natalia A Goriounova, Brian R Lee, Rusty Mann, Jeremy A Miller, Gabor Molnar, Alice Mukora, Lauren Alfiler, et al. Morphoelectric and transcriptomic divergence of the layer 1 interneuron repertoire in human versus mouse neocortex. *Science*, 382(6667):eadf0805, 2023.
- [8] Nathan W Gouwens, Staci A Sorensen, Fahimeh Baftizadeh, Agata Budzillo, Brian R Lee, Tim Jarsky, Lauren Alfiler, Katherine Baker, Eliza Barkan, Kyla Berry, et al. Integrated morphoelectric and transcriptomic classification of cortical gabaergic cells. *Cell*, 183(4):935–953, 2020.
- [9] Brian R Lee, Rachel Dalley, Jeremy A Miller, Thomas Chartrand, Jennie Close, Rusty Mann, Alice Mukora, Lindsay Ng, Lauren Alfiler, Katherine Baker, et al. Signature morphoelectric properties of diverse gabaergic interneurons in the human neocortex. *Science*, 382(6667):eadf6484, 2023.
- [10] Susan M Sunkin, Lydia Ng, Chris Lau, Tim Dolbeare, Terri L Gilbert, Carol L Thompson, Michael Hawrylycz, and Chinh Dang. Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic acids research*, 41(D1):D996–D1008, 2012.
- [11] Michael W Reimann, Costas A Anastassiou, Rodrigo Perin, Sean L Hill, Henry Markram, and Christof Koch. A biophysically detailed model of neocortical local field potentials predicts the critical role of active membrane currents. *Neuron*, 79(2):375–390, 2013.
- [12] Anatoly Buchin, Rebecca de Frates, Anirban Nandi, Rusty Mann, Peter Chong, Lindsay Ng, Jeremy Miller, Rebecca Hodge, Brian Kalmbach, Soumita Bose, et al. Multi-modal characterization and simulation of human epileptic circuitry. *Cell reports*, 41(13), 2022.
- [13] Anirban Nandi, Thomas Chartrand, Werner Van Geit, Anatoly Buchin, Zizhen Yao, Soo Yeun Lee, Yina Wei, Brian Kalmbach, Brian Lee, Ed Lein, et al. Single-neuron models linking

- electrophysiology, morphology, and transcriptomics across cortical cell types. *Cell reports*, 40(6), 2022.
- [14] Kaisa Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer Science & Business Media, 1999.
  - [15] Shaul Druckmann, Yoav Banitt, Albert A Gidon, Felix Schürmann, Henry Markram, and Idan Segev. A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Frontiers in neuroscience*, 1:56, 2007.
  - [16] Werner Van Geit, Michael Gevaert, Giuseppe Chindemi, Christian Rössert, Jean-Denis Courcol, Eilif B Muller, Felix Schürmann, Idan Segev, and Henry Markram. Bluepyopt: leveraging open source software and cloud infrastructure to optimise model parameters in neuroscience. *Frontiers in neuroinformatics*, 10:17, 2016.
  - [17] Clayton P Mosher, Yina Wei, Jan Kamiński, Anirban Nandi, Adam N Mamelak, Costas A Anastassiou, and Ueli Rutishauser. Cellular classes in the human brain revealed in vivo by heartbeat-related modulation of the extracellular action potential waveform. *Cell reports*, 30(10):3536–3551, 2020.
  - [18] Pablo Achard and Erik De Schutter. Complex parameter landscape for a complex neuron model. *PLoS computational biology*, 2(7):e94, 2006.
  - [19] Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, pages 1–9, 2024.
  - [20] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.
  - [21] Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for Fourier neural operators. *J. Mach. Learn. Res.*, 22(1), 2021. ISSN 1532-4435.
  - [22] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.
  - [23] Vignesh Gopakumar, Stanislas Pamela, Lorenzo Zanisi, Zongyi Li, Anima Anandkumar, and MAST Team. Fourier neural operator for plasma modelling. *arXiv preprint arXiv:2302.06542*, 2023.
  - [24] Thorsten Kurth, Shashank Subramanian, Peter Harrington, Jaideep Pathak, Morteza Mardani, David Hall, Andrea Miele, Karthik Kashinath, and Animashree Anandkumar. FourCastNet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators. 2022. doi: 10.48550/arXiv.2208.05419.
  - [25] Gege Wen, Zongyi Li, Qirui Long, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. Real-time high-resolution CO<sub>2</sub> geological storage prediction using nested Fourier neural operators. *Energy Environ. Sci.*, 16:1732–1741, 2023. doi: 10.1039/D2EE04204E.

- [26] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/JMS Journal of Data Science*, 1(3):1–27, 2024.
- [27] Ryan Y. Lin, Julius Berner, Valentin Duruisseaux, David Pitt, Daniel Leibovici, Jean Kossaifi, Kamyar Azizzadenesheli, and Anima Anandkumar. Enabling automatic differentiation with mollified graph neural operators, 2025.
- [28] Richard FitzHugh. Mathematical models of threshold phenomena in the nerve membrane. *The bulletin of mathematical biophysics*, 17:257–278, 1955.
- [29] Jinichi Nagumo, Suguru Arimoto, and Shuji Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50(10):2061–2070, 1962.
- [30] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.
- [31] Johann Rudi, Julie Bessac, and Amanda Lenzi. Parameter estimation with dense and convolutional neural networks applied to the fitzhugh–nagumo ode. In *Mathematical and Scientific Machine Learning*, pages 781–808. PMLR, 2022.
- [32] Pavel V Kuptsov, Nataliya V Stankevich, and Elmira R Bagautdinova. Discovering dynamical features of hodgkin–huxley-type model of physiological neuron using artificial neural network. *Chaos, Solitons & Fractals*, 167:113027, 2023.
- [33] Wei-Hung Su, Ching-Shan Chou, and Dongbin Xiu. Deep learning of biological models from data: applications to ode models. *Bulletin of mathematical biology*, 83:1–19, 2021.
- [34] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.10.045.
- [35] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *ArXiv*, abs/1711.10561, 2017.
- [36] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics informed deep learning (part ii): Data-driven discovery of nonlinear partial differential equations. *ArXiv*, abs/1711.10566, 2017.
- [37] Yan Barbosa Werneck, Rodrigo Weber dos Santos, Bernardo Martins Rocha, and Rafael Sachetto Oliveira. Replacing the fitzhugh-nagumo electrophysiology model by physics-informed neural networks. In *International Conference on Computational Science*, pages 699–713. Springer, 2023.
- [38] Matteo Ferrante, Andera Duggento, and Nicola Toschi. Physically constrained neural networks to solve the inverse problem for neuron models. *arXiv preprint arXiv:2209.11998*, 2022.
- [39] Himanshu Pandey, Anshima Singh, and Ratikanta Behera. An efficient wavelet-based physics-informed neural networks for singularly perturbed problems. *arXiv preprint arXiv:2409.11847*, 2024.

- [40] Edoardo Centofanti, Massimiliano Ghiotto, and Luca F Pavarino. Learning the hodgkin–huxley model with operator learning techniques. *Computer Methods in Applied Mechanics and Engineering*, 432:117381, 2024.
- [41] Sergey L Gratiy, Yazan N Billeh, Kael Dai, Catalin Mitelut, David Feng, Nathan W Gouwens, Nicholas Cain, Christof Koch, Costas A Anastassiou, and Anton Arkhipov. Bionet: A python interface to neuron for modeling large-scale networks. *PloS one*, 13(8):e0201630, 2018.
- [42] Michael L Hines and Nicholas T Carnevale. Neuron: a tool for neuroscientists. *The neuroscientist*, 7(2):123–135, 2001.
- [43] Tae Kim, Stephen Thankachan, James T McKenna, James M McNally, Chun Yang, Jee Hyun Choi, Lichao Chen, Bernat Kocsis, Karl Deisseroth, Robert E Strecker, et al. Cortically projecting basal forebrain parvalbumin neurons regulate cortical gamma band oscillations. *Proceedings of the National Academy of Sciences*, 112(11):3535–3540, 2015.
- [44] Guillermo Gonzalez-Burgos, Raymond Y Cho, and David A Lewis. Alterations in cortical network oscillations and parvalbumin neurons in schizophrenia. *Biological psychiatry*, 77(12):1031–1040, 2015.
- [45] Mariano I Gabitto, Kyle J Travaglini, Victoria M Rachleff, Eitan S Kaplan, Brian Long, Jeanelle Ariza, Yi Ding, Joseph T Mahoney, Nick Dee, Jeff Goldy, et al. Integrated multimodal cell atlas of alzheimer’s disease. *Nature Neuroscience*, 27(12):2366–2383, 2024.
- [46] Jean Kossaifi, Nikola Kovachki, Zongyi Li, David Pitt, Miguel Liu-Schiavini, Valentin Duruisseaux, Robert Joseph George, Boris Bonev, Kamyar Azizzadenesheli, Julius Berner, and Anima Anandkumar. A library for learning neural operators, 2024.
- [47] Alan L Hodgkin. The local electric changes associated with repetitive action in a non-medullated axon. *The Journal of physiology*, 107(2):165, 1948.
- [48] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

## A The Fourier Neural Operator Architecture

Neural operators compose linear integral operators  $\mathcal{K}$  with pointwise non-linear activation functions  $\sigma$  to approximate highly non-linear operators. More precisely, we define the neural operator

$$\mathcal{G}_\theta = \mathcal{Q} \circ \sigma(W_L + \mathcal{K}_L + b_L) \circ \dots \circ \sigma(W_1 + \mathcal{K}_1 + b_1) \circ \mathcal{P} \quad (3)$$

where  $\mathcal{P}, \mathcal{Q}$  are the pointwise neural networks that encode the lower dimension function into a higher dimensional space and vice versa. The model stacks  $L$  layers of  $\sigma(W_l + \mathcal{K}_l + b_l)$  where  $W_l$  are pointwise linear operators (matrices),  $\mathcal{K}_l$  are integral kernel operators,  $b_l$  are bias terms, and  $\sigma$  are fixed activation functions. The parameters  $\theta$  consists of all the parameters in  $\mathcal{P}, \mathcal{Q}, W_l, \mathcal{K}_l, b_l$ . Kossaifi et al. [46] maintain a comprehensive open-source PyTorch library for learning neural operators, which serves as the foundation for our implementation. A Fourier neural operator (FNO) [22] is a neural operator using Fourier integral operator layers, which are defined via

$$(\mathcal{K}(\phi)v_t)(x) = \mathcal{F}^{-1}\left(R_\phi \cdot (\mathcal{F}v_t)\right)(x) \quad (4)$$

where  $R_\phi$  is the Fourier transform of a periodic function  $\kappa$  parameterized by  $\phi$ . On a uniform mesh, the Fourier transform  $\mathcal{F}$  can be implemented using the fast Fourier transform (FFT). Here is a depiction of the Fourier Neural Operator (FNO) [22]:

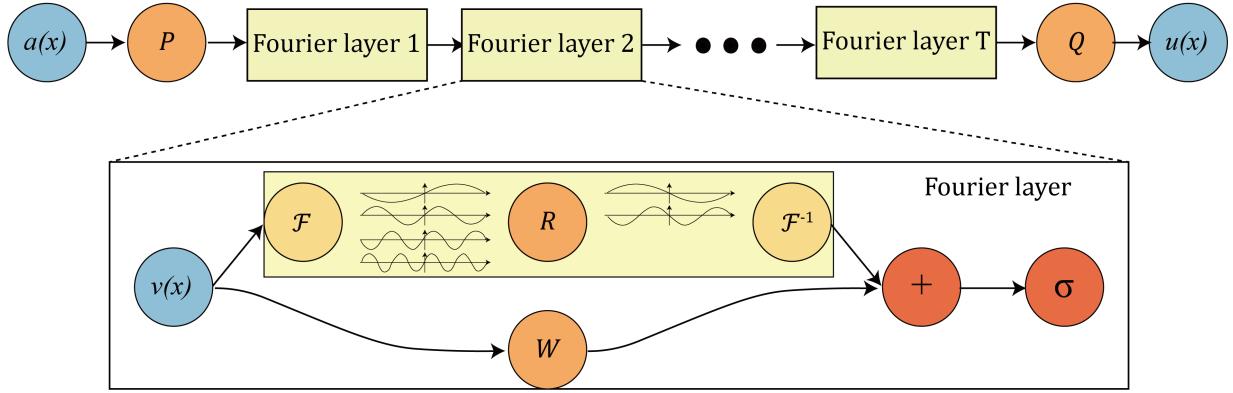


Figure 8: The Fourier Neural Operator (FNO) architecture (extracted from [22]).

## B Impact of Subsampling on Neuron Features

We present the results of the analysis conducted to determine the maximum subsampling factor that preserves the fidelity of extracted neuron features, mentioned in Section 3.1. The results are displayed in Figures 9 and 10 for the low-pass filtering followed by decimation in time subsampling strategy.

We first computed the relative error between the raw, non-subsampled HoF model voltage responses and the experimental data across all amplitudes. For each amplitude, we identified the minimum relative error across all HoF models, and then aggregated these minima to compute the mean and standard deviation. These serve as a reference for the inherent *worst-case* discrepancy between

simulations and experimental recordings in the absence of any subsampling. We visualize the mean as a solid black line and the standard deviation as dotted black lines.

Next, we repeated a similar analysis to quantify the additional relative errors introduced by subsampling. For each subsampling factor, we calculated the relative error between the subsampled and original HoF responses for each amplitude. These errors were then averaged across all HoF models, and the distribution of these averages is summarized using the mean (solid line), standard deviation (shaded region), and min/max (error bars).

This study shows that for most features, subsampling introduces negligible additional error. The most sensitive features were **AP1\_Width** and **AP1\_Peak**, which exhibited noticeable deviations at higher subsampling rates. To ensure we remain within the bounds of the intrinsic simulation-experiment discrepancy, we adopt a conservative downsampling factor of  $3\times$ , which maintains fidelity while reducing computational load.

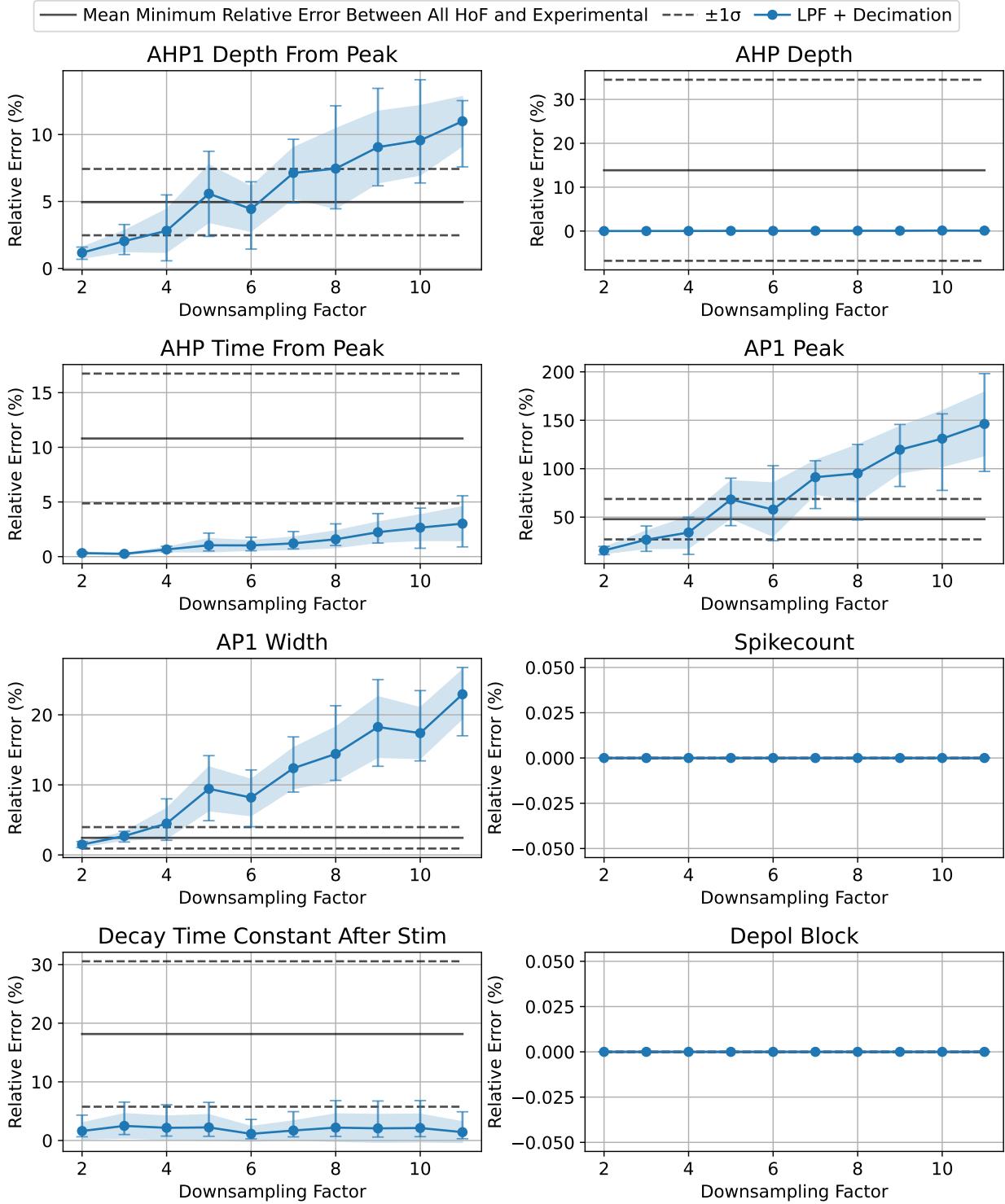


Figure 9: Analysis of the relative errors introduced in neuron feature computation as a function of subsampling factor, using low-pass filtering followed by decimation in time. The solid and dotted black lines indicate the mean and standard deviation, respectively, of the minimum relative error between non-subsampled HoF and experimental responses. The solid blue line, shaded region, and error bars represent the mean, standard deviation, and minimum–maximum statistics of the relative error between non-subsampled and subsampled HoF responses.

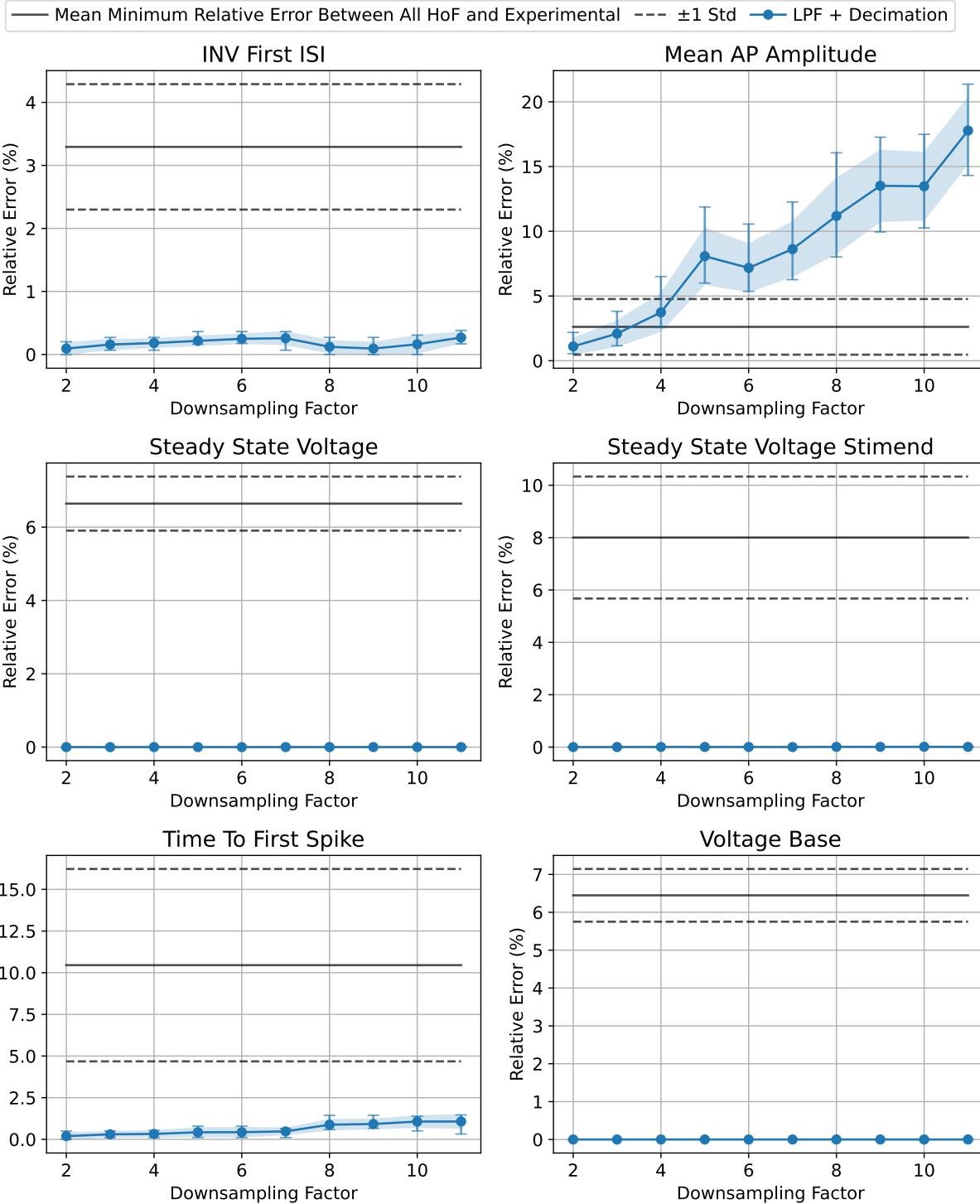


Figure 10: Analysis of the relative errors introduced in neuron feature computation as a function of subsampling factor, using low-pass filtering followed by decimation in time. The solid and dotted black lines indicate the mean and standard deviation, respectively, of the minimum relative error between non-subsampled HoF and experimental responses. The solid blue line, shaded region, and error bars represent the mean, standard deviation, and minimum–maximum statistics of the relative error between non-subsampled and subsampled HoF responses.

## C Input Current Amplitude Distribution

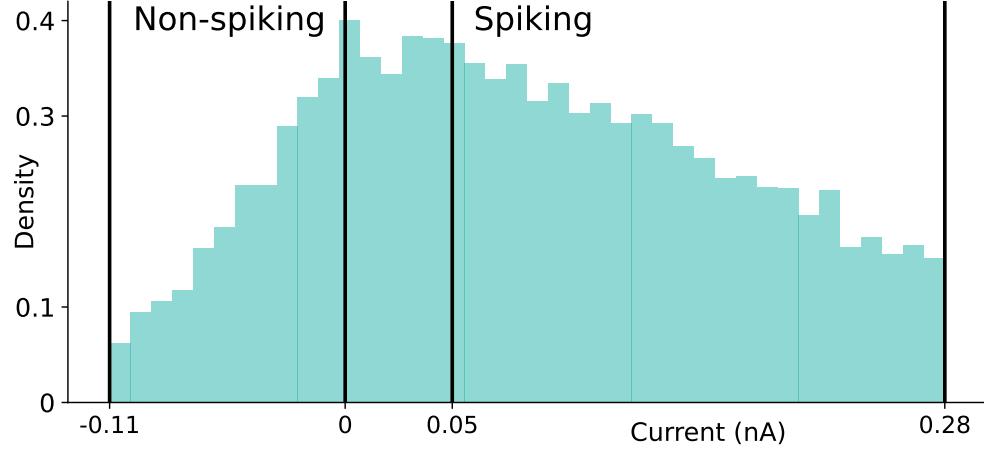


Figure 11: Distribution of square-pulse amplitudes in  $[-0.11, 0.28]$ nA considered. There is a spiking threshold (between 0 to 0.05nA) where neuron responses transition from being non-spiking to spiking.

## D Neighborhood Considered in Interpolation Experiments

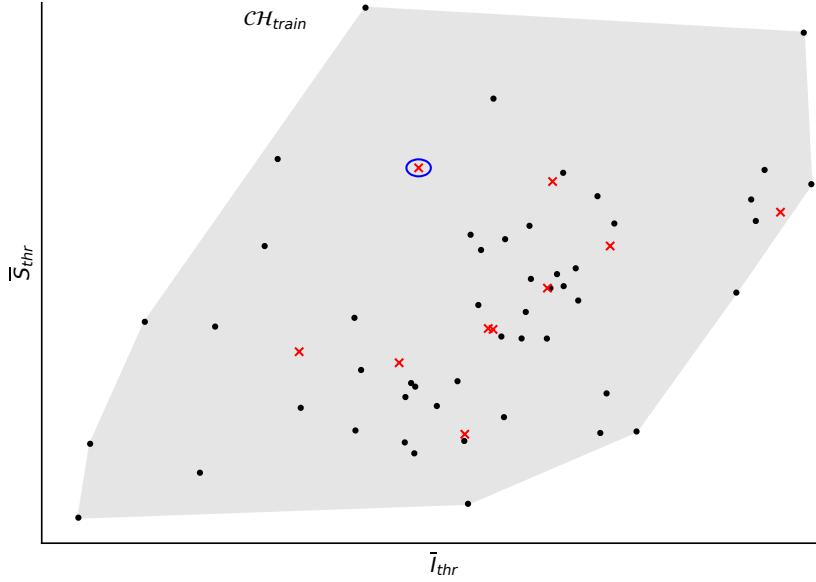


Figure 12: Latent representations in normalized  $(I_{thr}, s_{thr})$ -space of  $\{\text{HoF}_i^{train}\}$  (black dots) and  $\{\text{HoF}_i^{test}\}$  (red crosses) models. The latter lie in the convex hull  $\mathcal{CH}_{train}$  of the  $\{\text{HoF}_i^{train}\}$  models. In the interpolation experiment of Section 4.3, we sample 50 models in a small neighborhood around a given model  $\text{HoF}_k^{test}$ . We plot the boundary of this neighborhood (the blue ellipse) for an example of model  $\text{HoF}_k^{test}$ . Note that this neighborhood does not contain any model used for training, and in particular none of the 50 new models sampled in that neighborhood have been used for training.

## E Electrophysiological Features

For electrophysiological feature extraction and metrics, we use code from the Electophys Feature Extraction Library (eFEL) available at

<https://github.com/BlueBrain/eFEL>

The formulas, codes, and more details about each electrophysiological feature can be found at

<https://efel.readthedocs.io/en/latest/eFeatures.html>

We list below 15 important electrophysiological features and metrics of interest when constructing neuron models (where AP denotes action potential and AHP denotes after-hyperpolarization):

- **AHP\_depth**: Relative voltage values at the first AHP
- **AHP\_time\_from\_peak**: Time between AP peaks and first AHP depth
- **AHP1\_depth\_from\_peak**: Voltage difference between the first AP peak and first AHP depth
- **AP1\_peak**: The peak voltage of the first AP
- **AP1\_width**: Width of first spike at half spike amplitude, with the spike amplitude taken as the difference between the minimum between two peaks and the next peak
- **decay\_time\_constant\_after\_stim**: The decay time constant of the voltage right after the stimulus
- **depol\_block**: Check for a depolarization block. Returns 1 if there is a depolarization block or a hyperpolarization block, and returns 0 otherwise.
- **inv\_first\_ISI**: 1.0 over first interspike interval; returns 0 when no interspike interval
- **mean\_AP\_amplitude**: The mean of all of the AP amplitudes
- **steady\_state\_voltage**: The average voltage after the stimulus
- **steady\_state\_voltage\_stimend**: The average voltage during the last 10% of the stimulus duration.
- **time\_to\_first\_spike**: Time from the start of the stimulus to the maximum of the first peak
- **voltage\_base**: The average voltage during the last 10% of time before the stimulus
- **sag\_amplitude**: The difference between the minimal voltage and the steady state at the end of the stimulus
- **spikecount**: Number of spikes in the trace, including outside of stimulus interval