



OPEN ACCESS

EDITED BY

Alain Nogaret,
University of Bath, United Kingdom

REVIEWED BY

Andrey L. Shilnikov,
Georgia State University, United States
Ulrich Parlitz,
Max Planck Institute for Dynamics and
Self-Organization, Germany

*CORRESPONDENCE

Thiago B. Burghi
✉ tbb29@cam.ac.uk

RECEIVED 22 October 2024

ACCEPTED 18 July 2025

PUBLISHED 01 September 2025

CITATION

Burghi TB, Schapiro K, Ivanova M, Wang H,
Marder E and O'Leary T (2025) Quantitative
prediction of intracellular dynamics and
synaptic currents in a small neural circuit.
Front. Comput. Neurosci. 19:1515194.
doi: 10.3389/fncom.2025.1515194

COPYRIGHT

© 2025 Burghi, Schapiro, Ivanova, Wang,
Marder and O'Leary. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Quantitative prediction of intracellular dynamics and synaptic currents in a small neural circuit

Thiago B. Burghi^{1*}, Kyra Schapiro², Maria Ivanova²,
Huaxinyu Wang¹, Eve Marder² and Timothy O'Leary¹

¹Department of Engineering, University of Cambridge, Cambridge, United Kingdom, ²Volen National Center for Complex Systems, Brandeis University, Waltham, MA, United States

Fitting models to experimental intracellular data is challenging. While detailed conductance-based models are difficult to train, phenomenological statistical models often fail to capture the rich intrinsic dynamics of circuits such as central pattern generators (CPGs). A recent trend has been to employ tools from deep learning to obtain data-driven models that can quantitatively learn intracellular dynamics from experimental data. This paper addresses the general questions of modeling, training, and interpreting a large class of such models in the context of estimating the dynamics of a neural circuit. In particular, we use recently introduced Recurrent Mechanistic Models to predict the dynamics of a Half-Center Oscillator (HCO), a type of CPG. We construct the HCO by interconnecting two neurons in the Stomatogastric Ganglion using the dynamic clamp experimental protocol. This allows us to gather ground truth synaptic currents, which the model is able to predict—even though these currents are not used during training. We empirically assess the speed and performance of the training methods of teacher forcing, multiple shooting, and generalized teacher forcing, which we present in a unified fashion tailored to data-driven models with explicit membrane voltage variables. From a theoretical perspective, we show that a key contraction condition in data-driven dynamics guarantees the applicability of these training methods. We also show that this condition enables the derivation of data-driven frequency-dependent conductances, making it possible to infer the excitability profile of a real neuronal circuit using a trained model.

KEYWORDS

central pattern generator, artificial neural networks, machine learning, neural circuits, electrophysiology, system identification, dynamic clamp

1 Introduction

Constraining data-driven predictive models of biological neuronal circuits presents significant challenges. Neuronal circuits exhibit substantial variability in connectivity strength and intrinsic physiological characteristics across specimens, and they operate under time-varying inputs and uncertainties (Schulz et al., 2006, 2007; Marder, 2012; Anwar et al., 2022). These challenges become particularly visible in circuits with complex intrinsic dynamics, such as Central Pattern Generators (CPGs), where the interplay of membrane currents across multiple timescales defines circuit-level properties (Goaillard et al., 2009; Grashow et al., 2009; Franci et al., 2017; Drion et al., 2018), but the same issues are also at play in circuits whose outputs are less well-defined.

Until recently, two broad classes of modeling approaches were used to predict the activity of neural circuits. The first uses phenomenological stochastic neural dynamics to predict spike timing, sometimes with remarkable success (Paninski et al., 2005; Gerstner and Naud, 2009; Gerstner et al., 2014; Pillow and Park, 2016). These so-called statistical models can potentially fit data very well, but their general-purpose integrate-and-fire structure limits mechanistic insight; in particular this approach cannot capture the rich intrinsic dynamics present in circuits such as CPGs. Additionally, with few exceptions (e.g., Latimer et al., 2019), statistical models fail to account for the mechanistic effects of membrane conductances, which limits their interpretability and prevents their use in closed-loop experiments involving dynamic clamp or neuromodulation (Sharp et al., 1993; Prinz et al., 2004a). The second class of modeling approaches attempts to constrain biophysically detailed models to experimental data, and a number of methods have been developed for that purpose (Huys et al., 2006; Abarbanel et al., 2009; Buhry et al., 2011; Meliza et al., 2014; Lueckmann et al., 2017; Abu-Hassan et al., 2019; Alonso and Marder, 2019; Taylor et al., 2020; Gonçalves et al., 2020; Wells et al., 2024). Detailed models are interpretable, but suffer from excessive complexity that makes it difficult to fit them to data efficiently (Geit et al., 2008). In addition, no consensus exists on how much detail can be safely ignored while still capturing essential dynamical properties and accounting for experimentally observed constraints (Almog and Korngreen, 2016). Existing methodologies require precise measurements and high computational resources, making it impractical to use within the time constraints of an experiment. Finally, these models may capture plausible parameter subspaces of misspecified models rather than the “true” parameters (Prinz et al., 2004b).

To address these challenges, a third class of modeling approaches has recently emerged that leverages the power of artificial neural networks (ANNs) and deep learning to efficiently capture the complexity of intracellular dynamics in a quantitative fashion (Brenner et al., 2022; Durstewitz et al., 2023; Aguiar et al., 2024; Burghi et al., 2025a). In such approaches, the vector field governing the membrane dynamics is parametrized using specific types of ANNs. The resulting models can therefore be interpreted as particular classes of Recurrent Neural Networks (RNNs) or Universal Differential Equations (Chen et al., 2018; Rackauckas et al., 2021), depending on whether time is treated as discrete or continuous. The goal of the present work is to understand how to efficiently train discrete-time models from this broad data-driven class, and how to interpret them in electrophysiological terms. To answer these questions, we combine theoretical analysis with empirical investigation. To pursue the latter, we work with the recently introduced class of Recurrent Mechanistic Models or RMMs (Burghi et al., 2025a), a data-driven architecture based on structured state space models (Gu et al., 2022) and ANNs which is geared for rapid estimation of intracellular neuronal dynamics.

Our empirical results show that RMMs can be trained to quantitatively predict the membrane voltage and synaptic currents in a non-trivial small neuronal circuit known as a Half-Center

Oscillator (HCO). This circuit, composed of two reciprocally inhibitory neurons, autonomously generates anti-phasic bursting activity, and is the basis for oscillatory patterns in biology, such as the leech heartbeat (Calabrese et al., 1995; Wenning et al., 2004; Calabrese and Marder, 2024). Due to the complex intrinsic dynamics and extensive morphology of the cells involved in HCOs, fitting quantitatively predictive models to those circuits has historically been extremely challenging, even with access to high-performance computing and state-of-the-art numerical methods. While HCOs exist in nature, in this work we obtain one using the dynamic clamp technique (Sharp et al., 1993). This is done by creating artificial synapses between two Gastric Mill (GM) neurons in the Stomatogastric Ganglion of the crab *Cancer borealis*, similarly to previous work (Sharp et al., 1996; Grashow et al., 2009; Morozova et al., 2022). Defining the synaptic currents of the circuit ourselves provides ground truth connectivity data, which is used not for training but rather to assess the ability of RMMs to predict internal circuit connectivity. We show that RMMs can quantitatively predict synaptic currents from voltage measurements alone, that prediction accuracy depends on training algorithms, and that the accuracy improves when biophysical-like priors are introduced in the model.

A key contribution of this work is an empirical assessment of the speed and predictive performance of the training methods of teacher forcing (TF; Doya, 1992), multiple shooting (MS; Ribeiro et al., 2020b), and generalized teacher forcing (GTF; Doya, 1992; Abarbanel et al., 2009; Hess et al., 2023), which we present in a unified fashion tailored to models with an explicit membrane voltage variable (such as RMMs). From the theoretical perspective, we show how a key contraction property (Lohmiller and Slotine, 1998) of the internal neuronal dynamics guarantees the well-posedness of all three aforementioned training methods. The contraction property, which can be verified in a data-driven model by means of linear matrix inequalities, has been exploited in control theory for system identification (Manchester et al., 2021), for robustly learning data-driven models (Revay et al., 2024), and for estimating conductance-based models in real-time (Burghi and Sepulchre, 2024).

We also show that the contraction property enables a rigorous analysis of RNN-based data-driven models in terms of *frequency-dependent conductances*, which generalize the familiar neuronal input conductance, and are related to the sensitivity functions of control theory (Franci et al., 2019). Using RMMs, we obtain and interpret data-driven frequency-dependent conductances of the experimental HCO circuit.

In studying RNN-based methods using RMMs as a representative example, our results also extend the results of Burghi et al. (2025a), which focused on single-compartment dynamics. We find that our flexible, lightweight RMM approach enables us to build a predictive model of a complex neuronal circuit within the timeframe of an experiment. This paves the way for closed-loop, model-based manipulations of neural activity. Crucially, the model structure provides interpretable measurements of the circuit in quasi-real-time, facilitating novel types of diagnostic and scientific experiments on living neural circuits.

2 Results

Our results address three key questions related to data-driven neuronal circuit models where the membrane dynamics of individual neurons is parametrized with artificial neural networks (ANNs). First, *how to best learn such models efficiently*, that is, how to formulate an optimization algorithm capable of rapidly delivering good predictive models of the real neuronal circuit activity. Second, *whether such models can accurately predict unmeasured synaptic currents* inside a neuronal circuit. Third, *how to interpret such a model in terms of biophysically meaningful quantities*.

Our results are relevant to a wide class of data-driven models with ANN-based membrane dynamics. To facilitate working with sampled data, models are formulated in discrete time. The *input-output data* used to train a model is given by a sequence $\{v_t, u_t\}$ of vector-valued measured intracellular membrane voltages v_t and vector-valued measured injected electrical currents u_t , observed at discrete time points $t \in \mathbb{N}$. The dynamics of a circuit model with $n \geq 1$ neurons is described by the discrete-time state-space model

$$C \frac{\hat{v}_{t+1} - \hat{v}_t}{\delta} = -h_\theta(\hat{v}_t, x_t) + u_t \quad (1a)$$

$$x_{t+1} = f_\eta(\hat{v}_t, x_t) \quad (1b)$$

where $\hat{v}_t = (\hat{v}_{1,t}, \dots, \hat{v}_{n,t})$ is a state vector gathering all predicted membrane voltages, $u_t = (u_{1,t}, \dots, u_{n,t})$ is an input vector gathering all (measured) injected currents, and x_t is a state vector gathering internal (also known as hidden or latent) state variables. The system is parametrized by a fixed sampling period $\delta > 0$, a learnable diagonal membrane capacitance matrix C , and learnable piecewise differentiable mappings f_η and h_θ , whose parameters are the vectors θ and η , respectively. Notice that C can in principle be estimated separately through voltage-clamp (Taylor, 2012); this is not pursued in our paper.

The motivation behind Equation 1 comes from the simplest of all mechanistic constraints: the membrane dynamics is assumed to be reasonably modeled by Kirchoff's laws. Equation 1 thus describes the balance between a capacitive current

$$C \dot{\hat{v}} \approx C \delta^{-1} (\hat{v}_{t+1} - \hat{v}_t),$$

an injected current u , and a number of ionic and synaptic currents whose sum total is given by the output of h_θ . We are interested in models where the functions h_θ and f_η are constructed with ANNs. Models that follow this idea are Piecewise Linear Recurrent Neural Networks (Durstewitz et al., 2023), Recurrent Mechanistic Models (Burghi et al., 2025a), and traditional Recurrent Neural Networks (RNNs) where some of the states are explicitly identified with membrane voltages. All such models can be written in the form Equation 1 by re-labeling their states and re-parametrizing learnable variables. Notice that such models can be formulated in continuous-time following the frameworks of Neural (Chen et al., 2018) and Universal (Rackauckas et al., 2021) ODEs. Discretized conductance-based models in the Hodgkin-Huxley style (Izhikevich, 2007; Hodgkin and Huxley, 1952) fit the structure of Equation 1, but do not traditionally incorporate ANNs. The dynamics of HH-type ionic currents are instead based on detailed

modeling assumptions informed by voltage-clamp experiments. Abandoning the HH formalism raises the question of how to gain relevant insights about the physiology of a neuronal circuit. This is addressed at the end of Results.

Over the next sections we describe different training methods used for inferring the dynamics of an arbitrary neural circuit using models of the form Equation 1: teacher forcing (TF), multiple shooting (MS), and generalized teacher forcing (GTF). Our experimental results assess the efficiency of TF, MS, and GTF by working with a Half-Center Oscillator circuit, which introduced next.

2.1 Model of a half-center oscillator

Concretely, we consider the neuronal circuit given by the experimental preparation illustrated in Figure 1A. The two-neuron circuit, which is commonly known as a Half-Center Oscillator (HCO), consists of two cells with measured membrane voltages $v_{1,t}$ and $v_{2,t}$, interconnected via inhibitory synapses. In naturally occurring HCOs, synaptic currents are typically not measured due to the difficulty of obtaining them while the circuit is intact. For this reason, and to provide ground truth data against which synaptic current predictions of data-driven models can be validated, the HCO preparation used in this paper is created artificially with dynamic clamp (Sharp et al., 1993). Dynamic clamp is used to inject two electrical current components into the membrane of each neuron: a virtual inhibitory synaptic current $I_{syn,t}^{ij}$ and a virtual hyperpolarization-induced current $I_{h,t}^i$. In addition to dynamic clamp currents, each neuron is also excited with a noisy injected current $u_{i,t}$, which is used to explore the dynamics of the circuit and ensure that the data is informative enough for the purposes of data-driven modeling. See *applied current* in Methods for details. The particular HCO used in this paper is created using two cells in the Stomatogasteric Ganglion of the crab, and the procedure for forming this circuit with dynamic clamp is can be found in Morozova et al. (2022). A diagram of the resulting HCO is shown in Figure 1B, where artificial circuit elements introduced by dynamic clamp are depicted in color, while biological elements are depicted in gray.

To estimate the HCO dynamics, we use the Recurrent Mechanistic Model (RMM) framework of Burghi et al. (2025a). The RMM used throughout this paper is depicted in Figure 1C. It is described in succinct form by

$$C \frac{\hat{v}_{t+1} - \hat{v}_t}{\delta} = -\text{mlp}_\theta(\hat{v}_t, x_t) + u_t \quad (2)$$

$$x_{t+1} = Ax_t + B\hat{v}_t$$

where A and B are matrices of the appropriate dimensions, and $\text{mlp}_\theta(\hat{v}_t, x_t)$ is a multi-layer perceptron (MLP) whose inputs are given by $\hat{v}_t = (\hat{v}_{1,t}, \hat{v}_{2,t})$ and x_t . For a description of the model in terms of individual circuit neurons, see *modeling details* in Methods. It can be seen that the RMM in Equation 2 greatly constrains the more general formulation (Equation 1) by imposing a linear state-space dynamics of the internal states x_t , and by imposing a feed-forward artificial neural network structure on h_θ .

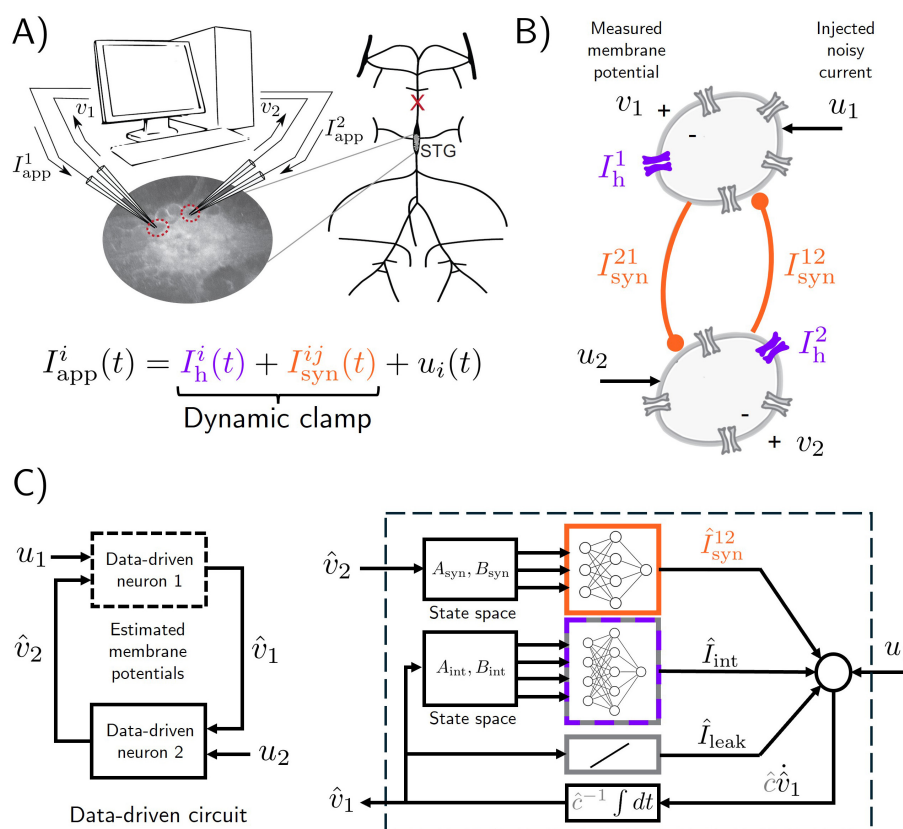


FIGURE 1

Predicting the intracellular membrane voltages and synaptic currents of a real neuronal circuit. (A) we estimate the dynamics of a Half-Center Oscillator (HCO) constructed experimentally with dynamic clamp. Two neurons from the Stomatogastric Ganglion of *Cancer borealis* are given virtual hyperpolarization-induced intrinsic currents I_h^i and virtual inhibitory synaptic currents I_{syn}^j , which are implemented via dynamic clamp Prinz et al. (2004a) (see Methods). In addition to the dynamic clamp currents, a noisy current u_i is also injected in each neuron. Illustration adapted from Morozova et al. (2022). (B) A diagram of the resulting HCO. Dynamic clamp currents are pictured in color; naturally occurring ionic currents are pictured in gray. (C) Conceptual diagram of a data-driven model used to learn the HCO dynamics, with a detailed diagram of neuron 1 on the right. The membrane dynamics of each of the neurons in the HCO is modeled using a combination of linear state-space models and multi-layer perceptrons, following the *Recurrent Mechanistic Model* paradigm of Burghi et al. (2025a) (see Methods for the specific model used in this figure). The data used to learn the model are the measured membrane voltages v_1 and v_2 and the noisy injected currents u_1 and u_2 .

Training the RMM means finding good values for C and θ so that the model is able to predict the response of the HCO to the injected noisy currents $u_{1,t}$ and $u_{2,t}$. For our purposes, the HCO response consists of the measured voltages $v_{1,t}$ and $v_{2,t}$ (which are used during training), and the synaptic currents $I_{syn,t}^{12}$ and $I_{syn,t}^{21}$ (which are not). Training the RMM allows us to assess various training algorithms that can also be used to train other types of data-driven models. RMMs are convenient for that purpose since they can be rapidly trained. In the results below we have intentionally limited the size of the RMM, as well as the amount of data used in training, so that a meaningful comparison between learning methods could be pursued.

2.2 Training data-driven models: a unified perspective

Training data-driven models of the form Equation 1 to learn complex dynamics such as spiking and bursting is in general not

trivial. This is because the *forward dynamics* of neuronal models, i.e. the mapping from applied current (u_t) to predicted membrane voltage (\hat{v}_t), is highly sensitive to changes in the model's parameters. The sensitivity leads to the problem of *exploding gradients* (Pascanu et al., 2013) when one approaches the learning problem naively. To avoid exploding gradients, the standard approach for training neuronal models is to exploit the structure given by Equation 1 and employ some variation or generalization of the method of *teacher forcing* (Doya, 1992). We approach the problem of training neuronal data-driven models from a unified perspective, relating the methods of *teacher forcing*, *multiple shooting*, and *generalized teacher forcing* to each other. Throughout this section, we consider an exponential contraction condition (Lohmiller and Slotine, 1998) on the internal dynamics of the generic model (Equation 1), which we show to be sufficient for the applicability of these training methods. To make this paper self-contained, this condition is recapped in the Methods section. In what follows, v_t and u_t denote the experimentally recorded voltage and applied current, respectively.

2.2.1 Recurrent and feed-forward teacher forcing (TF)

The idea in the teacher forcing method is to train a model of the form [Equation 1](#) by solving the optimization problem

$$\min_{\theta, \eta, \xi, C} \frac{1}{N} \sum_{t=0}^{N-1} \|v_{t+1} - \tilde{v}_{t+1}\|^2 + r_\rho(\theta, \eta) \quad (3a)$$

$$\text{subject to } \tilde{v}_{t+1} = v_t + \delta C^{-1}(-h_\theta(v_t, \tilde{x}_t) + u_t) \quad (3b)$$

$$\tilde{x}_{t+1} = f_\eta(v_t, \tilde{x}_t), \quad \tilde{x}_0 = \xi \quad (3c)$$

where $r_\rho(\theta, \eta)$ is a regularization term, and ρ is a positive hyperparameter. We shall use $r_\rho(\theta, \eta) = \rho \|\theta\|^2 + \rho \|\eta\|^2$ throughout the paper. [Equation 3](#) corresponds to minimizing the squared one-step-ahead voltage prediction error, with the one-step-ahead prediction \tilde{v}_{t+1} obtained by *forcing* the vector field of the model ([Equation 1](#)) with the “teacher” signal v_t . Notice that it is crucial to distinguish the predictions used for training (\tilde{v}_t), computed with [Equations 3b–3c](#), from the predictions of the simulated model (\hat{v}_t), computed with [Equation 1](#). Teacher forcing uses the former as a means to improve the latter, and it should be emphasized that good \tilde{v}_t predictions do not necessarily imply good \hat{v}_t predictions. Letting $\mathcal{L}(\theta, \eta, \xi, C)$ denote the overall loss function of [Equation 3](#), obtained by simulating [Equation 3b–3c](#) and substituting the result in [Equation 3a](#), the problem is usually solved using some variant of the basic gradient descent (GD) method, which in our context can be written as

$$\theta_{k+1} = \theta_k - \beta \nabla_\theta \mathcal{L}(\theta_k, \eta_k, \xi_k) \quad (4)$$

for θ , and analogously for η , ξ , and C (here, β is the step size and $k = 0, 1, 2, \dots$ are the training epochs). In non-convex problems, two conditions are necessary for finding good data-driven model parameters with gradient descent or any of its enhanced versions, such as ADAM ([Kingma and Ba, 2017](#)). First, one must choose a reasonable set of initial training parameters ξ_0 , θ_0 , η_0 , and C_0 . Second, one must ensure that the gradients in $\nabla \mathcal{L}(\theta_k, \eta_k, \xi_k)$ remain well-behaved throughout training.

TF as stated in [Equation 3](#) involves temporal *recurrence* in the dynamic constraints ([Equation 3b–3c](#)). Consequently, computing the gradients in [Equation 4](#) requires *backpropagating through time*, which might be computationally costly depending on the length of the dataset. Recurrence can be eliminated from the problem by foregoing the learning of the initial state ξ and the internal dynamics parameters η . If both ξ and η remain constant during training, TF becomes a *feed-forward* learning problem. One can then divide training into two steps: first, simulate

$$\tilde{x}_{t+1}^0 = f_{\eta_0}(v_t, \tilde{x}_t^0), \quad \tilde{x}_0^0 = \xi_0 \quad (5)$$

for fixed η_0 and ξ_0 , then, solve the (unconstrained) problem

$$\min_{\theta, C} \frac{1}{N} \sum_{t=0}^{N-1} \|v_{t+1} - v_t - \delta C^{-1}(-h_\theta(v_t, \tilde{x}_t^0) + u_t)\|^2 + r_\rho(\theta) \quad (6)$$

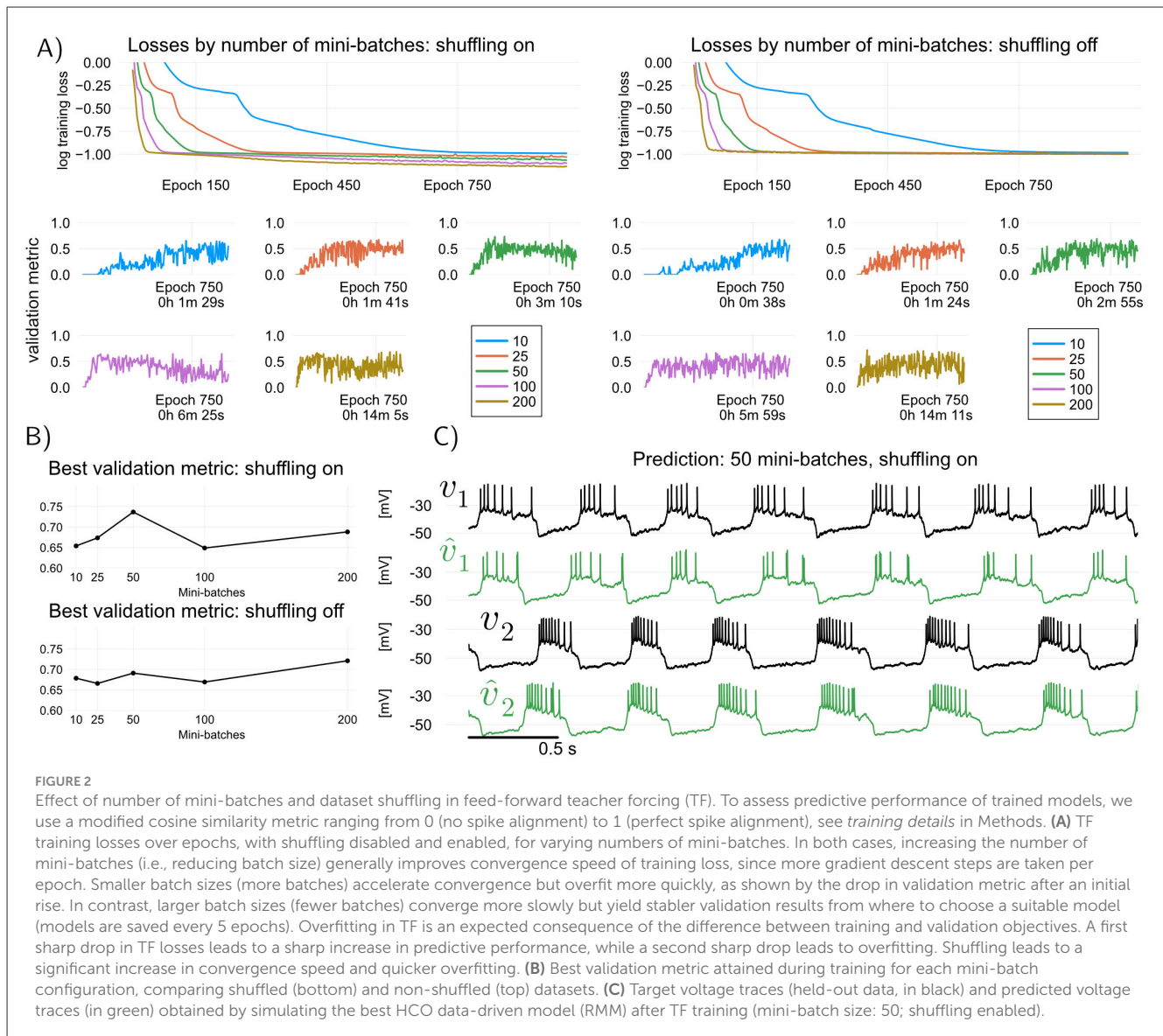
using the fixed time series of internal states \tilde{x}_t^0 . We say [Equation 6](#) is a feed-forward problem because it reduces to training the

mapping $C^{-1}(h_\theta(v_t, \tilde{x}_t^0) + u_t)$, which in data-driven models such as [Equation 2](#) will usually contain feed-forward ANNs. Solving this problem is significantly faster than solving its recurrent counterpart ([Equation 3](#)), since backpropagating through time is no longer required—backpropagation is limited to the layers of h_θ . This idea, which we call *feed-forward TF*, has previously been used to train conductance-based models by taking existing ion channel models, and training the voltage dynamics with fixed ion channel kinetics ([Huys et al., 2006](#)).

An exponential contraction condition of the internal states (see *contraction theory* in Methods) guarantees that both recurrent and feed-forward versions of TF are well-posed. In the former, it precludes exploding gradients, since one can show that if the internal dynamics is contracting, uniformly in v and η , then the loss gradients $\nabla \mathcal{L}$ remain bounded. In addition, in both versions of TF above, contraction helps to improve learning performance by using *warmed up initial conditions* ξ_0 (see *training details* in Methods).

In feed-forward TF, two tricks can be employed to speed up training and yield models that generalize better on unseen data. First, one can use *mini-batching* ([Li et al., 2014](#)) to perform GD parameter updates that only use part of the dataset. This is done by taking gradients of partial sums of the total loss function ([Equation 6](#)), resulting in better exploration of the loss landscape. Second, one can *shuffle* ([Meng et al., 2019](#)) the dataset (that is, randomly permute the data points) *before* partitioning the summation in the cost ([Equation 6](#)) into mini-batches; this can sometimes lead to better generalization. In [Figure 2](#) we apply feed-forward TF to the RMM data-driven model of the HCO, and compare the effect of mini-batch size and shuffling on the predictive power of trained models. Models are obtained with $\rho = 5 \times 10^{-8}$, which was selected after a simple grid search over regularization constants. Validation traces are shown in terms of a modified cosine similarity metric which is applied after filtering and smoothing out the spikes in the dataset, see Methods. It is seen that in general, increasing the number of batches in a dataset (decreasing the mini-batch size) is beneficial up to a point. It can also be seen that shuffling the dataset during training results in faster improvements on the validation metric, but quicker overfitting due to the significant increase in convergence speed.

The most distinctive aspect of TF training is its susceptibility to overfitting, which can be observed in [Figure 2A](#) and is more prominent when shuffling is enabled. This is because the TF problem in [Equation 3a](#) or 6 minimizes discrepancies in the map (the right-hand side) of the voltage dynamics in [Equation 1](#), whereas the validation metric assesses discrepancies in the simulated voltage trace. By saving “snapshots” of trained models during training with feed-forward TF, overfitting can be avoided, and reasonably good models obtained in a short amount of time. This is shown in [Figure 2B](#), where the performance of the best model over the course of training for different number of mini-batches is plotted. As [Figure 2C](#) shows, a validation metric of around 0.73 results in visually adequate good predictions, with a few intra-burst spikes of one of the HCO neurons being missed over a few bursts. Increasing model size, choosing a better exploratory injected current u_t , and conducting an extensive hyperparameter search are all strategies that can be pursued to improve models obtained with TF.



2.2.2 Multiple shooting trades off speed for performance

Teacher forcing is significantly limited by the quality of the data. This becomes evident when one interprets the loss (Equation 6) in terms of a target voltage difference $v_{t+1} - v_t$. This difference is a high-pass filtering operation that amplifies the variance in measurement noise, which is ubiquitously present in neurophysiological applications. To deal with this problem, one must take into account the recurrence between voltage and internal dynamics, which is accomplished by generalizing teacher forcing. One way to do so is to employ *multiple shooting*, which is based on a classical method for finding numerical solutions of differential equations (Ribeiro et al., 2020b).

The idea in multiple shooting is to divide a dataset of length N into $N_s = \lfloor N/s \rfloor$ intervals or “shots” of size $s > 1$, and then train the model by simulating it in parallel over each of the shots separately (we use $\lfloor \cdot \rfloor$ for the round-down operator). Figure 3A illustrates this idea. Mathematically, when applied to models of the

form Equation 1, multiple shooting takes the form

$$\min_{\theta, \eta, \xi^{(n)}, v^{(n)}, C} \frac{1}{sN_s} \sum_{n=0}^{N_s-1} \sum_{t=0}^{s-1} \|v_{t+ns} - \check{v}_t^{(n)}\|^2 + r_\rho(\theta, \eta) + \rho_v \|\check{v}_s^{(n)} - v^{(n+1)}\|^2 + \rho_x \|\check{x}_s^{(n)} - \xi^{(n+1)}\|_D^2 \quad (7a)$$

$$\text{subject to } \check{v}_{t+1}^{(n)} = \check{v}_t^{(n)} + \delta C^{-1}(-h_\theta(\check{v}_t^{(n)}, \check{x}_t^{(n)}) + u_{t+ns}), \quad \check{v}_0^{(n)} = v^{(n)} \quad (7b)$$

$$\check{x}_{t+1}^{(n)} = f_\eta(\check{v}_t^{(n)}, \check{x}_t^{(n)}), \quad \check{x}_0^{(n)} = \xi^{(n)} \quad (7c)$$

where Equations 7b–7c describe the N_s simulations involved in the learning algorithm, run for $t = 0, 1, \dots, s-1$. In Equation 7a, the initial conditions of the voltage and internal states in each shot, $v^{(n)}$ and $\xi^{(n)}$ respectively, must be learned. Assuming that the contraction condition holds, initial GD conditions $\xi_0^{(n)}$ can be chosen effectively by warming up the model (see Methods); for voltage, one can simply take $v_0^{(n)} = v_{ns}$. When $s = N$, multiple

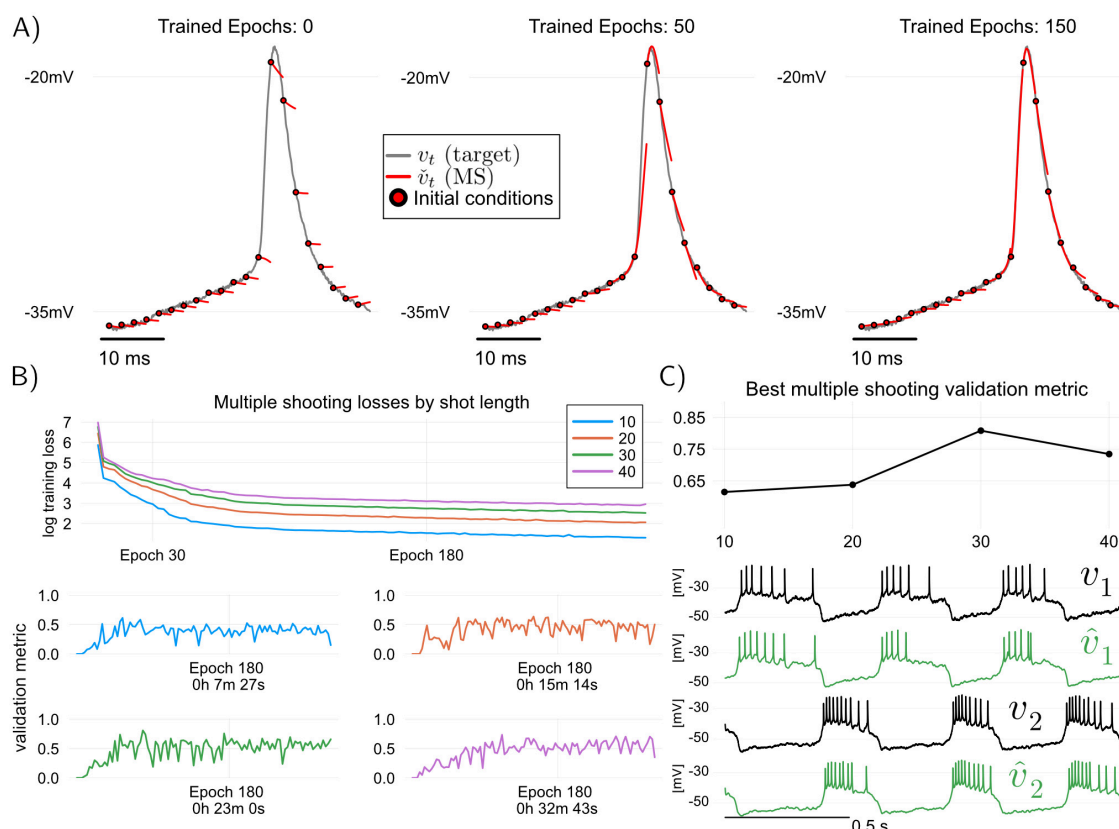


FIGURE 3

Training and validating the RMM (Equation 2) using multiple shooting, which trades off training speed for predictive performance. Here, modified angular separation in the interval $[0, 1]$ is used as validation metric, see *training details* in Methods. (A) Illustration of predictive “shots” during multiple shooting training. As training progresses, initial shot conditions as well as model parameters are learned. (B) Multiple shooting training loss (top) and modified angular separation (validation metric) of trained models, plotted for four different shot sizes. Increasing the shot length, training losses and training time increase, but predictive power of trained models (as measured by angular separation) may also increase. One advantage of Multiple shooting over Teacher Forcing is the relatively stable and monotonic increase in validation performance as a function of training epoch. (C) Maximum validation metric attained for each of the shot sizes plotted, and voltage traces of validation target (held-off data) and predictions of the best model (shot size of 30 samples). All models were trained starting from random MLP weights (Methods).

shooting becomes equivalent to minimizing the discrepancy between the measured v and the simulated \hat{v} from Equation 1 over the entire dataset, which requires backpropagating through many bursts and spikes, and likely leads to exploding gradients. For $s = 1$, and fixed $v^{(n)} = v_n$, multiple shooting becomes equivalent to the dynamic TF problem (Equation 3a). Hence multiple shooting provides a means to trade off the speed of feed-forward TF for the enhanced predictive power of models trained with backpropagation through time, while avoiding exploding gradients due to the limited simulation horizon of each shot (Ribeiro et al., 2020a). It is worth noticing the term involving $\rho_x > 0$ in the cost function (Equation 7a): it is a regularizing term that attempts to drive the final internal states in each shot toward the initial states of the next shot.

When using multiple shooting in practice, it is not obvious how to choose values for the hyperparameters s (shot length), ρ (parameter regularization) and ρ_v , ρ_x (final state regularization). Such questions are nearly impossible to answer theoretically, and one must resort to using hyperparameters that are found empirically to be the best for fitting a given neuronal circuit. Fixing the regularization constants $\rho = 5 \times 10^{-9}$, and $\rho_v = \rho_x = 500$,

Figure 3 illustrates the effect of increasing the shot length s on training and validation with multiple shooting. In Figure 3B, It can be seen that by increasing the shot length s , predictive power of the models may increase, but this comes at the cost of increased training times. In particular, for the hyperparameters tested in this paper, MS results in a better model than TF (modified cosine similarity of 0.8 for MS versus 0.73 for TF, cf Figures 2, 3). One advantage of multiple shooting over TF is the fact that increases in the validation metric are relatively monotonic as a function of epochs. This means that, to obtain a good model, one can save “model snapshots” at a lower rate during training.

2.2.3 Generalized teacher forcing yields a filter for the membrane dynamics

Generalized teacher forcing (GTF), first suggested by Doya (1992) and recently revisited in Hess et al. (2023), extends TF by forcing the model dynamics with a convex combination of the data and the model’s outputs. GTF was formulated with the same goal as multiple shooting: to avoid exploding gradients when

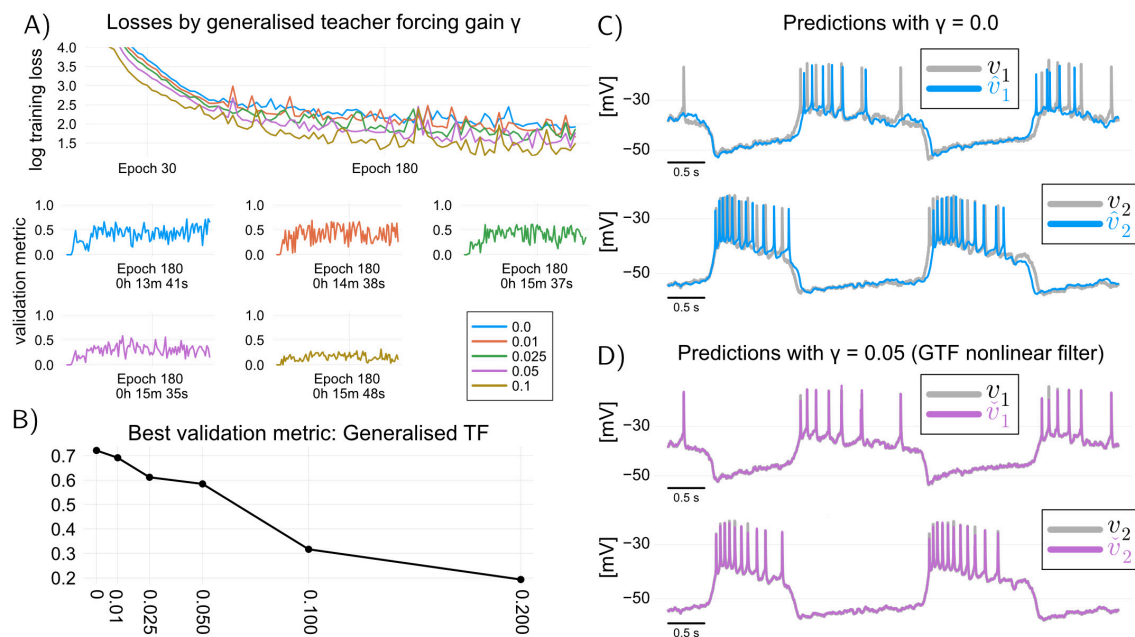


FIGURE 4

Generalized teacher forcing yields a data-driven filter or observer for the voltage dynamics. (A) learning curves when training the RMM (Equation 2) using GTF in combination with multiple shooting (shot length $s = 20$ samples). (B) There is no obvious benefit to using GTF if the purpose is to use data-driven models solely for simulation: trained model performance deteriorates for positive values of γ . (C) Simulation of trained model with $\gamma = 0$ (equivalent to MS). (D) Simulation of the filter equations for the model trained with GTF and $\gamma = 0.05$. GTF provides a nonlinear filter of the voltage dynamics which can be used to improve the quality of recordings, as well as to obtain better estimates of the internal states of the model.

learning recurrent models. When specialized to neuronal data-driven models, GTF can be formulated by replacing the voltage data v_t in the TF problem (Equation 3) by a convex combination of v_t and voltage predictions \tilde{v}_t . GTF has also been proposed, although not under that name, to learn continuous-time conductance-based models in Abarbanel et al. (2009) (see also Brookings et al., 2014). To learn the dynamics of discrete-time neuronal models, the GTF formulations of both Doya (1992) and Abarbanel et al. (2009) can be reconciled through the following problem:

$$\min_{\theta, \eta, \xi} \frac{1}{N} \sum_{t=0}^{N-1} \|v_t - \tilde{v}_t\|^2 + r_\rho(\theta, \eta) \quad (8a)$$

$$\text{subject to } \tilde{v}_{t+1}^- = \tilde{v}_t + \delta C^{-1} (-h_\theta(\tilde{v}_t, \tilde{x}_t) + u_t) \quad (8b)$$

$$\tilde{x}_{t+1} = f_\eta(\tilde{v}_t, \tilde{x}_t) \quad (8c)$$

$$\tilde{v}_t = \tilde{v}_t^- + \gamma(v_t - \tilde{v}_t^-) \quad (8d)$$

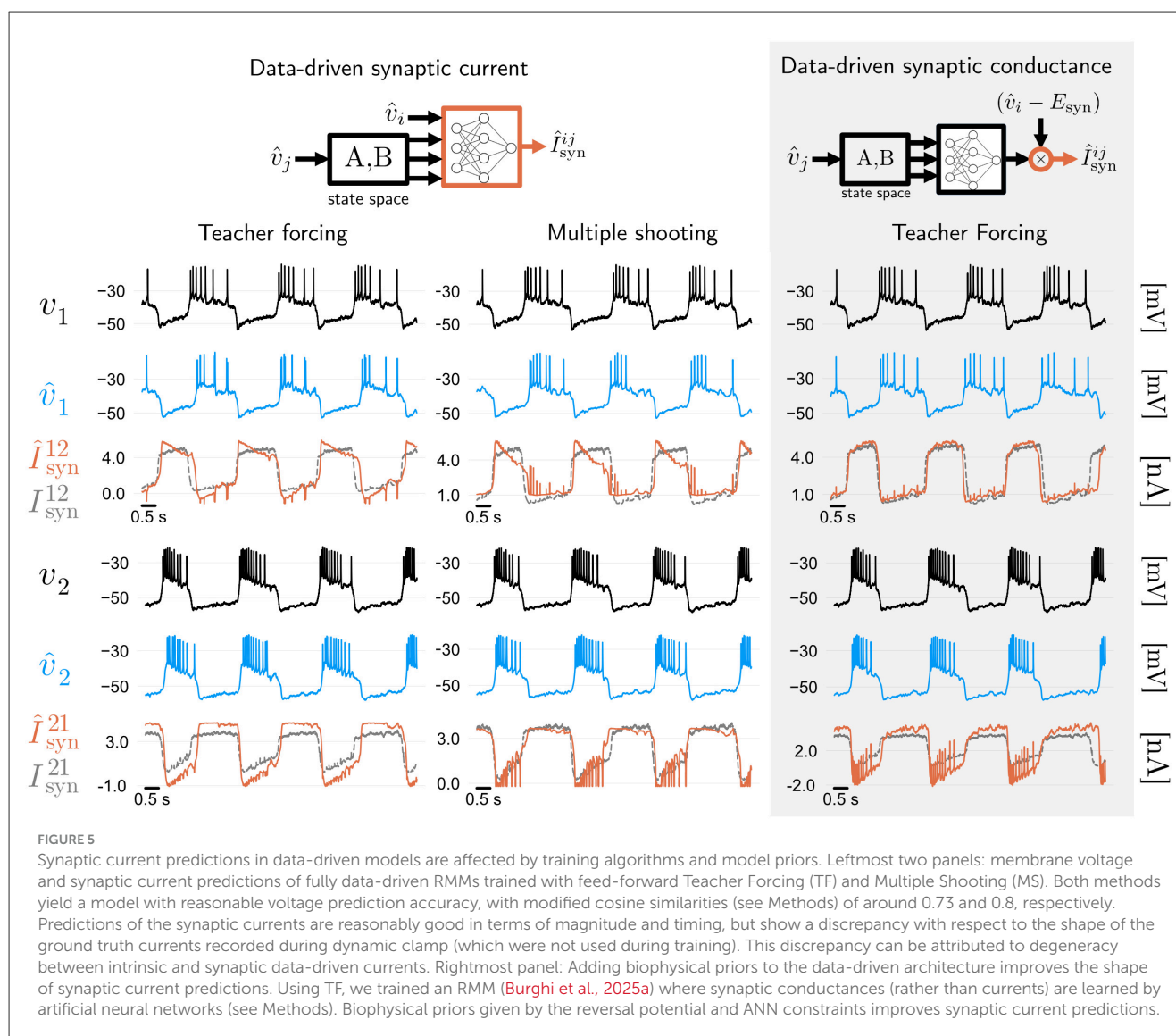
$$\tilde{v}_0^- = v_0, \quad \tilde{x}_0 = \xi$$

where $\gamma \in (0, 1)$ is a hyperparameter of the problem. Notice that setting $\gamma = 0$ turns (Equation 8) into a standard simulation-error system identification problem (Ljung, 1999), usually solved with GD and backpropagation through time; setting $\gamma = 1$ recovers the TF problem (Equation 3). Notice also that Equation 8d, which implements a convex combination of measurements and predictions, can also be interpreted as an *update* of the prediction \tilde{v}_{t+1}^- using the data v_{t+1} . The latter allows interpreting GTF for neuronal systems a method to learn a *filter* or *observer* (signal processing and control engineering terminologies, respectively) for

the neuronal membrane voltage. This is how Abarbanel et al. (2009) interpreted their continuous-time version of the problem, which was applied to conductance-based models. The control theoretic interpretation allows one to explore further generalizations of teacher forcing involving, for instance, the Extended Kalman Filter, in which γ is made adaptive, see Burghi and Sepulchre (2024).

Because GTF in neuronal models exploits measured voltages but not measured internal states (the latter are not available), exploding gradients can in general still be observed during training. However, here we find another benefit of the contraction assumption (see *contraction theory* in Methods). If a neuronal data-driven model possesses a uniformly contracting internal dynamics, then there is a value of γ sufficiently close to 1 such that exploding gradients are precluded during gradient descent. But even if γ is chosen to avoid exploding gradients, one may still find that solving the problem (Equation 8) takes too much time, especially when the dataset length N is large. For that reason, one can *combine GTF with multiple shooting*. Doing so is trivial: the multiple shooting cost (Equation 7a) can be obtained by partitioning the GTF cost (Equation 8a) according to a desired shot length, and using the GTF constraints with $\gamma = 0$ within each shot. The combined GTF-multiple-shooting problem for $\gamma > 0$ is obtained in the same way by keeping the GTF constraints (Equations 8b–8d) with $\gamma > 0$.

An important question is whether combining GTF with multiple shooting yields better models than multiple shooting alone. Figure 4A illustrates typical training curves obtained with the combined GTF-multiple-shooting method, for different values of γ . It can be seen that increasing the value of $\gamma > 0$ in general leads



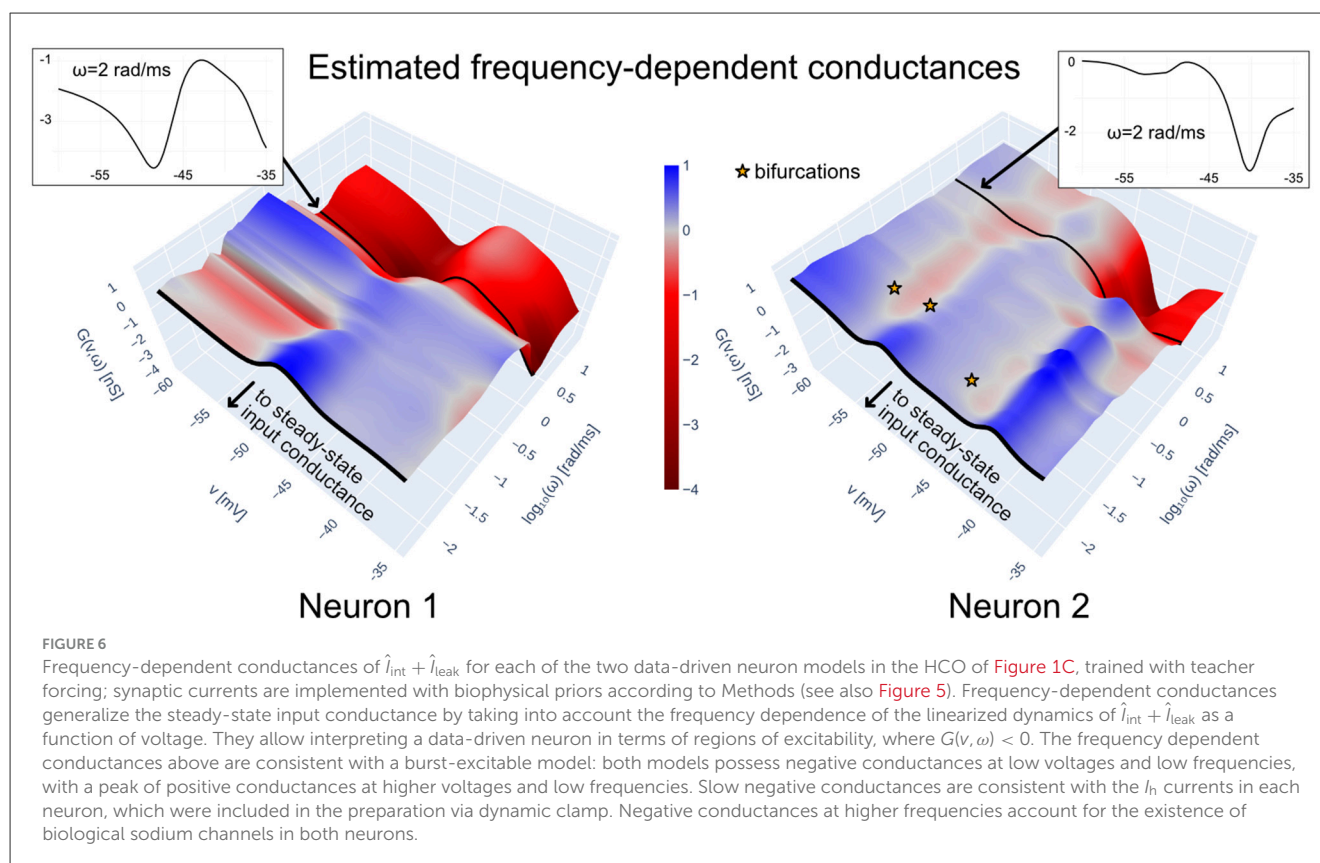
to poorer validation performance. This performance deterioration can be explained by the way we validate the models: in validation, models are not allowed to use the recorded voltages, and hence to obtain the validation curves of Figure 4A we have used predictions based on simulation of Equation 2, which does not contain the update step (Equation 8d) used to train a model with GTF. In other words, we have trained such models with $\gamma > 0$, but validated them with $\gamma = 0$; a loss in performance is hence expected. This empirical observation seemingly questions the usefulness of GTF, at least in the context where it is used in conjunction with multiple shooting for speeding up training times.

In fact, GTF should not be contrasted with MS, but rather be viewed as method for training a discrete-time *nonlinear filter* of the membrane dynamics, which can be used to obtain improved measurements of the membrane potential, as well as improved estimates of its internal states. The filter is simply the Luenberger observer given by Equations 8b–8d. Figures 4C, D illustrate this point: we can see that the predictions \hat{v}_t of a simulation model

($\gamma = 0.0$) obtained with GTF can be drastically improved by generating predictions \check{v}_t from the trained filter ($\gamma = 0.05$). The fact that very small values of γ result in good filter predictions suggests that predictive discrepancies in models trained with TF or multiple shooting (where $\gamma = 0$) can be at least in part attributed to *unmeasured input disturbances* that have not been taken into account when training those models; in fact, one interpretation of the correction term in an observer is that it provides an estimate of such disturbances (Sontag, 1998). Combining GTF with rapidly trainable data-driven models hence provides a means to filter out such disturbances, in real-time, if necessary.

2.2.4 Training methods and model priors affect synaptic current predictions

If intracellular voltage predictions of a neural circuit are all one cares about, then the takeaway from Figures 2–4 is that MS improves on TF, and GTF yields a nonlinear filter of the neuronal



membrane, which can be used in real-time. However, how good are the predictions of the synaptic currents, which are not used during training? In the leftmost two panels of Figure 5, we show both voltage and synaptic current predictions of the best model trained with TF from Figure 2 (modified angular separation 0.73), and the best model trained with MS from Figure 3 (modified angular separation 0.8). While the model trained with MS predicts voltage better, it can be seen that synaptic current predictions of the TF and MS models suffer from the same problem: while those predictions are accurate in terms of magnitude and timing, the shape of the first synaptic current in both cases is inaccurate. This shows that the fully data-driven RMM from Equation 3.1 in Methods has a degenerate dynamics: some share of the ground truth synaptic current of the HCO is, in the RMM, supplied by the intrinsic currents. Given that the synaptic current is not used during training, this is a consequence of the fact that the RMM has very few priors—in fact, both its intrinsic and synaptic currents have the very same model structure.

We investigate whether synaptic current predictions can be improved by introducing biophysical priors in the data-driven model. This can be accomplished by employing *data-driven conductances* of RMMs (Burghi et al., 2025a). We replace the fully data-driven synaptic current (Equation 3.1 in Methods) by a biophysically informed version containing data-driven conductances (Equation 19 in Methods). After training the resulting RMM with teacher forcing, we obtain the predictions shown on the rightmost panel of Figure 5 (modified angular

separation of around 0.75). It can be seen that including a biophysical prior on the data-driven model's synapse is enough to recover the shape of the first synapse to a high degree of accuracy, while the second synapse is predicted with a small error. This result shows that when it comes to predicting the internal currents of a neural circuit, one cannot be guided by the predictive accuracy of external (measured) signals alone.

2.3 Interpreting data-driven models of intracellular dynamics

A downside of using fully data-driven models of the form (Equation 1) is the fact that it is not trivial to interpret the model's dynamics in biophysical terms. We show that data-driven models can still be interpreted in quasi-biophysical terms by means of a *frequency-dependent conductance*. This is a generalization of the familiar *input conductance* used by electrophysiologists to study neuronal excitability (Mauro et al., 1970; Franci et al., 2019). Frequency-dependent conductances uncover excitable properties of the membrane at different timescales, serving a purpose similar to that of the early and late IV curves originally introduced by Hodgkin and Huxley (1952). This section discusses such conductances by assuming that Equation 1 is the model of a single neuron ($n = 1$).

2.3.1 Frequency-dependent conductances

We derive frequency-dependent conductances of a data-driven model satisfying the contraction assumption (Methods). This assumption ensures that the system (Equation 1) has a (steady-state) IV curve given by

$$u_{\infty}(v) = h(v, x_{\infty}(v)) \quad (9a)$$

where $x_{\infty}(v)$ is the unique equilibrium state of Equation 1b as a function of v , that is, it satisfies

$$x_{\infty}(v) = f(v, x_{\infty}(v)) \quad (9b)$$

for any v . The fact that this unique equilibrium exists is a consequence of the contraction of the internal dynamics (Equation 1b), which for constant v becomes an autonomous system (Lohmiller and Slotine, 1998, Section 3). The derivative of the IV curve (Equation 9a) is the (steady-state) *conductance-voltage* (GV) curve. It is given by

$$G_{\infty}(v) = \left[\frac{\partial h}{\partial v} + \frac{\partial h}{\partial x} \frac{\partial x_{\infty}}{\partial v} \right]_{v=v, x=x_{\infty}(v)} \quad (10)$$

which follows from the chain rule. The *admittance-voltage* (YV) curve is an extension of the GV curve into the frequency domain. For discrete-time circuits, it is given by

$$Y(v, \omega) = \left[\frac{\partial h}{\partial v} + \frac{\partial h}{\partial x} \left(e^{j\omega\delta} I - \frac{\partial f}{\partial x} \right)^{-1} \frac{\partial f}{\partial v} \right]_{v=v, x=x_{\infty}(v)} \quad (11)$$

where I is the identity matrix. For an explanation of Equation 11, see *admittance-voltage curve* in Methods. The YV curve is complex-valued, and hence can be decomposed into

$$G(v, \omega) = \text{Re}[Y(v, \omega)], \quad B(v, \omega) = \text{Im}[Y(v, \omega)] \quad (12)$$

which are the *frequency-dependent conductance* and *susceptance*, respectively. In Methods, we show that

$$G(v, 0) = G_{\infty}(v) \quad (13)$$

which formally demonstrates that $G(v, \omega)$ extends the GV curve into the frequency domain.

2.3.2 Voltage-frequency regions of excitability

Electrophysiologists have long inferred neuronal excitability from experimentally recorded steady-state IV curves exhibiting regions of negative slope—negative conductances. These are regions where, locally, ionic currents tend to collectively destabilize the membrane dynamics. In a data-driven model, this occurs at voltages where $G_{\infty}(v) < 0$. It is however well-known that negative steady-state conductances are by no means necessary for excitable behavior. Frequency-dependent conductances generalize the previous statements by taking membrane dynamics into account. Formally, we can infer that a model is (locally) excitable on a voltage range V if we can find a range of frequencies where

$$G(v, \omega) < 0 \quad (14)$$

for all $(v, \omega) \in V \times \mathbb{R}$. If $0 \in \mathbb{R}$, then Equation 14 includes the steady-state case $G_{\infty}(v) < 0$. If V contains positive frequencies, then we can say that the model presents *frequency preference*: the neuron responds excitably to frequencies where Equation 14 holds (see Izhikevich, 2007, p. 232 for a dynamical systems theory point of view). Control theory provides an interesting interpretation of the local excitability condition (Equation 14). A stable linear system is said to be *passive* if their complex-valued frequency response is confined to the closed right half of the complex plane (Goodwin and Sin, 1984). Since the admittance $Y(v, \omega)$ represents the frequency response of the (linearized) total membrane current, we can interpret Equation 14 as a passivity-breaking condition on the total membrane current dynamics. In other words, Equation 14 indicates the presence of *active* currents acting at timescales given by $2\pi/\omega$.

To illustrate the above points, Figure 6 shows the frequency-dependent conductance of the total intrinsic current of each of the neurons in the HCO model constructed with data-driven conductance-based synapses (see Equations 3.1 and 19 in Methods; the trained model corresponds to the rightmost pane of Figure 5). Intrinsic admittances of each neuron are computed with Equations 11–12 using $h = \hat{I}_{\text{int}}^i + \hat{I}_{\text{leak}}^i$ for $i = 1, 2$. The frequency-dependent conductances in Figure 6 are in accordance to the excitability features of the biological neurons in the HCO. Both neuron models are locally excitable (red regions) at low voltages and low frequencies, a feature we can attribute to the I_h (hyperpolarization-induced) present in the preparation. Furthermore, the conductance of Neuron 2 near -40 mV is considerably more negative than that of Neuron 1; this agrees with Neuron 1's sharp decrease in intra-burst excitability around that voltage (cf. Figure 4).

2.3.3 Ultrasensitivity and bifurcations

In neurons with non-monotonic IV curves, containing negative steady-state conductances, one finds that the onset of spiking or bursting occurs at the boundary of the voltage interval where $G_{\infty}(v) < 0$. Frequency-dependent conductances enable a generalization of this statement. Formally, any voltage \bar{v} where

$$G_{\infty}(\bar{v}) = 0 \text{ and } G_{\infty}'(\bar{v}) = 0 \quad (15)$$

lies at a *saddle-node* (also called *fold*) bifurcation point (Izhikevich, 2007). In such cases, increasing the applied current of an initially quiescent model past $u_{\infty}(\bar{v})$ pushes the voltage past the threshold \bar{v} toward spiking or bursting. For excitable models with a monotonic steady-state IV curve, Equation 15 cannot be satisfied, and a more general condition is required. In Methods, we show that the condition

$$C(\cos(\bar{\omega}\delta) - 1) + \delta G(\bar{v}, \bar{\omega}) = 0 \quad (16)$$

is a necessary condition for a saddle-node or Neimark-Sacker (the discrete-time version of the Hopf) bifurcation when $\bar{\omega} = 0$ or $\bar{\omega} > 0$, respectively, generalizing the condition (Equation 15) by means of the relation (Equation 13). The candidate bifurcation point $(\bar{v}, \bar{\omega})$ can be found by solving Equation 30 in Methods. Critically, for small sampling time δ , Equation 16 occurs near the boundary

of the regions of negative conductance given by Equation 14.¹ Borrowing terminology from Franci et al. (2019), in this region of the frequency-voltage space, the model can be *ultra-sensitive*.

Returning to Figure 6, we have plotted the candidate bifurcation points of Neuron 2 in the HCO model; recalling that the admittance is computed with $h = \hat{I}_{\text{int}}^2 + \hat{I}_{\text{leak}}^2$, this analysis treats the synaptic current from Neuron 1 as an external input. It can be seen that those points occur close to where the conductance flips sign. Furthermore, we can see that a bifurcation is predicted near -54 mV, which agrees well with the onset of the Neuron 2 bursts seen in Figure 5.

3 Methods

3.1 Modeling details

The RMM (Equation 2) used to model the HCO circuit of Figure 1 is given by

$$c_i \frac{\hat{v}_{i,t+1} - \hat{v}_{i,t}}{\delta} = - \underbrace{\text{mlp}(\hat{v}_{i,t}, w_{i,t}; \theta_i^{\text{int}})}_{\hat{I}_{\text{int},t}^{ij}} - \underbrace{\text{mlp}(\hat{v}_{i,t}, z_{i,t}; \theta_i^{\text{syn}})}_{\hat{I}_{\text{syn},t}^{ij}} - \underbrace{\theta_i^{\text{leak},1} \hat{v}_{i,t} + \theta_i^{\text{leak},2}}_{\hat{I}_{\text{leak},t}^{ij}} + u_{i,t}$$

$$w_{i,t+1} = A_i^{\text{int}} w_{i,t} + B_i^{\text{int}} \hat{v}_{i,t}$$

$$z_{i,t+1} = A_i^{\text{syn}} z_{i,t} + B_i^{\text{syn}} \hat{v}_{j,t}$$

with $\delta = 0.1$ the sampling period of the data. Here, $\text{mlp}(\hat{v}_{i,t}, w_{i,t}; \theta_i^{\text{int}})$ is a multi-layer perceptron (MLP) representing the sum total of intrinsic currents flowing through neuron i , while $\text{mlp}(\hat{v}_{i,t}, z_{i,t}; \theta_i^{\text{syn}})$ represents the total synaptic current flowing from neuron j to neuron i . The MLPs are given by the chained composition of matrix multiplications and bias additions with sigmoidal activation functions, which are in turn composed with an input normalization layer. Mathematically,

$$\text{mlp}(v, w; \theta) := W_{L+1} \circ \left(\bigcirc_{\ell=1}^L \sigma_{\ell}(W_{\ell} \cdot + b_{\ell}) \right) \circ q(v, w) \quad (17)$$

where: the circle operator denotes function composition, W_{ℓ} and b_{ℓ} are learnable layer weights and biases, constituting the parameter vector θ ; $\sigma_{\ell} = \tanh$ for all ℓ are activation functions applied elementwise to their inputs; and $q(v, w)$ is an affine normalization function mapping the minimum and the maximum values of v and w_i of the training dataset to -1 and $+1$, respectively (the w states of the training dataset are obtained by warming up; see *training details*). The use of an MLP (as opposed to a single-layer perceptron) can be justified on the basis that, since the internal dynamics in Equation 3.1 is linear, more complexity in the voltage dynamics may be desirable (empirically, few layers are required for satisfactory results). In this paper, the intrinsic current MLP has $L^{\text{int}} = 3$ layers with 20 activation functions in each layer, and the synaptic current MLP has $L^{\text{syn}} = 3$ layers with 10 activation

functions in each layer. The parameters $\theta^{\text{leak},1} = (\tilde{\theta}^{\text{leak},1})_+$ and $\theta^{\text{leak},2}$ are used to represent the leak current, with the former being constrained to be positive by the softplus function $(\cdot)_+$ so that it implements a positive leak conductance. MLP weights were initialized with the Nguyen-Widrow initialization heuristic (Nguyen and Widrow, 1990), with biases initialized spread over the interval -1 to 1 .

Following Burghi et al. (2025a), the linear systems in the internal dynamics of Equation 3.1 are constructed so as to have orthogonal impulse responses (Heuberger et al., 2005), which is done to produce a rich set of state trajectories following a voltage spike. To construct the discrete-time system matrices A and B , we consider prior knowledge about the continuous time constants relevant to neural dynamics. Starting from a set of continuous time constants $\{\tau_k\}_{k=1,\dots,n}$, we first discretize the τ_k with the stability-preserving transformation

$$\lambda_k = \exp(-\delta/\tau_k)$$

(recall δ is the sampling period). We then set $A := A_n$ and $B := B_n$, with A_n and B_n being computed according to

$$A_1 = \lambda_1, \quad B_1 = \sqrt{1 - \lambda_1^2}, \quad \Gamma_1 = B_1, \quad \Delta_1 = -\lambda_1$$

$$A_i = \begin{bmatrix} A_{i-1} & 0 \\ \sqrt{1 - \lambda_i^2} \Gamma_{i-1} & \lambda_i \end{bmatrix}, \quad B_i = \begin{bmatrix} B_{i-1} \\ \sqrt{1 - \lambda_i^2} \Delta_{i-1} \end{bmatrix} \quad (18)$$

$$\Gamma_i = \begin{bmatrix} -\lambda_i \Gamma_{i-1} & \sqrt{1 - \lambda_i^2} \end{bmatrix}, \quad \Delta_i = -\lambda_i \Delta_{i-1}$$

for $i = 2, \dots, n$. The resulting state space system responds to an impulse function with orthogonal state trajectories, see (Heuberger et al., 2005, Chapter 2) for details and the SI appendix of Burghi et al. (2025a) for an illustration. To construct A^{int} and A^{syn} , we define the two time constant subsets

$$\tau^{\text{fast+slow}} = \{0.2k\}_{k=1,\dots,8} \cup \{2.0k\}_{k=1,\dots,8} \cup \{10 + 10k\}_{k=1,\dots,8}$$

$$\tau^{\text{ultraslow}} = \{200k - 100\}_{k=1,\dots,8}$$

which were chosen by sampling the range of the time constant functions found in the gating kinetics of conductance-based models of neurons from the STG (Liu et al., 1998) [see the SI appendix of Burghi et al. (2025a) for plots of such functions]. The matrix A^{int} is constructed as in Equation 18 using $\tau^{\text{fast+slow}} \cup \tau^{\text{ultraslow}}$, while A^{syn} is constructed with $\tau^{\text{fast+slow}}$ only. This results in a total of 96 internal states in the full data-driven HCO model. In this paper the A and B matrices are kept fixed to illustrate how feed-forward teacher forcing (which precludes learning those matrices) can be related to MS and GTF. While it is in principle possible to learn the components of A and B during MS and GTF, we do not pursue this in this paper.

3.1.1 Synaptic current with data-driven conductances

To obtain an HCO model with biophysically informed data-driven synapses (Burghi et al., 2025a), we replace the synaptic current model in Equation 3.1 by

$$\hat{I}_{\text{syn},t}^{ij} = \tilde{g}^{ij} \sigma_+(h_+(z_{i,t}); \theta_i^{\text{syn}})(\hat{v}_{i,t} - E^{\text{syn}})$$

$$z_{i,t+1} = \text{diag}(\lambda[A_i^{\text{syn}}])z_{i,t} + (1 - \lambda[A_i^{\text{syn}}])\hat{v}_{j,t} \quad (19)$$

¹ Had we opted for a *passivity-preserving* discretization of the capacitive current (instead of a forward-Euler one), condition (Equation 16) would occur precisely at the boundary of Equation 14.

where we have used $\lambda[\cdot]$ to represent the vector of eigenvalues of a matrix, and $\mathbf{1}$ to represent the all-ones vector. In Equation 19, $\sigma_+(\cdot; \theta_i^{\text{syn}})$ is a single-layer perceptron whose weight matrices are constrained to be elementwise positive, and $h_+(\cdot)$ is a fixed min-max-type normalization layer constructed so that, at random parameter initialization, the output of σ_+ is approximately 0 at -60.0 mV, and approximately 1 at -40.0 ; this is done to mimic the activation range of the synapse used in the dynamic clamp experiment [see *constructing data-driven conductances* in the Methods section of Burghi et al. (2025a) for details]. The activation ranges above are not fine-tuned, so that the model can be said to be parsimonious one. The scalar \bar{g} is a learnable maximal conductance parameter, and E^{syn} is a fixed Nernst potential with value set to that of the synapse implemented in the HCO with dynamic clamp.

3.2 Contraction theory

We identify a tractable mathematical condition on the generic model (Equation 1) that guarantees the well-posedness of training and interpretability methods:

Exponential contraction condition: In the generic data-driven model of Equation 1, the internal dynamics states x_t in Equation 1b are *exponentially contracting* (Lohmiller and Slotine, 1998), uniformly in the voltage variable v_t and the parameters η .

Exponential contraction implies that for any given trajectory of v_t , the distance between two trajectories of x_t starting at different points of the state space decreases exponentially fast as $t \rightarrow \infty$ (Theorem 2 of Lohmiller and Slotine, 1998). A key technical feature of exponential contraction is that it can be guaranteed to hold via a sufficient condition involving a Lyapunov-type linear matrix inequality, which can be verified efficiently. Mathematically, contraction of the internal states is guaranteed if one can find a constant positive definite matrix² P and a positive constant $0 < \alpha < 1$ such that the Jacobian of the internal dynamics vector field satisfies the linear matrix inequality (LMI) given by

$$\frac{\partial f_\eta}{\partial x}(v, x)^\top P \frac{\partial f_\eta}{\partial x}(v, x) \preceq \alpha P \quad (20)$$

for all x and v (Lohmiller and Slotine, 1998). Intuitively, contraction is a strong stability property which makes a nonlinear system behave in many ways similarly to a stable linear one; for a tutorial, see Jouffroy and Fossen (2010). For data-driven models, one can guarantee that the internal states x_t contract throughout the learning procedure and by constraining the internal dynamics (Equation 1b) so that the matrix inequality holds uniformly in the parameters η .

When the internal dynamics of a data-driven model is linear, as is the case of Equation 2, then it is well known (Goodwin and Sin, 1984) that the LMI condition (Equation 20) becomes equivalent to the simpler condition below:

Exponential contraction condition (simple case): In the Recurrent Mechanistic Model of (Equation 2), we have

$$|\lambda_i[A]| < 1, \quad i = 1, \dots, \dim(x_t) \quad (21)$$

where $\lambda_i[A]$ denotes the i^{th} eigenvalue of A . In other words, the dynamics of x_t with v_t seen as an input is asymptotically stable.

The stability condition above is simple to enforce in linear systems, and is guaranteed by the constructive procedure used to obtain A described in the previous section.

3.2.1 Contraction and training methods

While weaker conditions with Lyapunov exponents can also achieve the effect of trajectory convergence seen in contracting systems, what makes dealing with contraction matrix inequalities such as Equation 20 attractive is the fact that it can be used to show that training with TF, MS, and GTF is tractable.

When the contracting internal dynamics assumption above is satisfied, then it is possible to guarantee that exploding gradients will not be observed in teacher forcing (TF) nor in generalized teacher forcing (GTF), regardless of the time horizon of the dataset. In TF, this can be shown by applying the chain rule to the loss gradient $\nabla \mathcal{L}$, and then using the LMI (Equation 20) to ensure that the gradient does not grow unbounded as the number of datapoints grows as $N \rightarrow \infty$. For instance, when computing $\nabla_\eta \mathcal{L}$, one obtains a term $\nabla_\eta \hat{x}_{N-1} \nabla_x h_\theta(\bar{v}_{N-1}, \hat{x}_{N-1})(\bar{v}_N - \hat{v}_N)$ that could potentially grow unbounded. This is because $\nabla_\eta \hat{x}_{N-1}$ is given by the recurrence relation

$$\nabla_\eta \hat{x}_t = \nabla_\eta \hat{x}_{t-1} \nabla_x f_\eta(\bar{v}_{t-1}, \hat{x}_{t-1}) + \nabla_\eta f_\eta(\bar{v}_{t-1}, \hat{x}_{t-1}) \quad (22)$$

which is a (potentially unstable) time-varying linear system. But the contraction LMI (Equation 20) ensures that this dynamical system has bounded states; this can be shown by taking $(\nabla_\eta \hat{x}_t^\top)P(\nabla_\eta \hat{x}_t)$ as a Lyapunov function for Equation 22 (see Khalil, 2002, Chapter 4 for an introduction to bounded-input-bounded-state stability analysis). In GTF, the situation becomes complicated by the interaction between \bar{v}_t and \hat{x}_t . But using contraction analysis, it can be claimed that under the contracting internal states assumption, taking γ close enough to 1 ensures that the dynamical system given by Equation 8b–8d is also contracting, and hence gradients cannot explode, similarly to the TF case. The proof of this claim is a simple discrete-time reformulation of Proposition 2 in Burghi et al. (2021).

3.3 Training details

All models in this paper were programmed in Julia using the Flux.jl package (Innes, 2018) for implementing MLPs as well as the Adam (Kingma and Ba, 2017) gradient descent routine. The code used to implement this model can be found on github.com/thiagoburghi/RMM/tree/HCO. The Adam moment parameters used in this paper are given by $\beta_1 = 0.9$ and $\beta_2 = 0.999$ in all cases. For models trained with TF in Figures 2, 5, and 6, we used a step size of $\eta = 0.001$. For models trained with MS in Figures 3 and 5, we used a scheduler for the step size, starting at

² A positive definite (semi-definite) matrix is a symmetric matrix whose eigenvalues are all positive (non-negative). Negative definite and semi-definite matrices are defined analogously.

$\eta = 0.01$ and changing to $\eta = 0.005$ at 50 epochs and $\eta = 0.0025$ at 100 epochs (this was done to rapidly decrease the large initial error in shot predictions). For models trained with GTF in [Figure 4](#), we used a fixed step size of $\eta = 0.01$. The D -norm in [Equation 7a](#) is a weighted 2-norm; we have used a diagonal weight matrix D whose nonzero entries are given by the inverse of the DC gain $(I - A)^{-1}B$. MS and GTF were implemented by partitioning the dataset into 48 mini-batches, and applying the methods as described in [Equations 7 and 8](#) within each mini-batch before taking a gradient step. Each epoch in [Figures 2–4](#) contains one gradient update for each mini-batch; mini-batches in MS and GTF were shuffled between each epoch (notice this is different from shuffling in TF, which also shuffles samples within mini-batches). All models were trained from randomly initialized MLP parameters.

3.3.1 Warming up initial conditions

Warming up the internal states' initial conditions is one of the most important steps in training data-driven models using the methods described in this paper. Consider a generic model in the form ([Equation 1](#)) satisfying the exponential contraction condition ([Equation 20](#)). Given the voltage data v_t and an initial parameter vector η_0 , let

$$\bar{x}_{t+1}^0 = f_{\eta^0}(v_t, \bar{x}_t^0), \quad t = -M, -(M-1), \dots, -1, 0, 1, \dots, N-1$$

where $v_{-M}, v_{-(M-1)}, \dots, v_{-1}$ is an initial portion of the data that is discarded during training, and v_0 and v_{N-1} are the first and last data points used during training, respectively. Under the assumption of contraction, the value of \bar{x}^{-M} does not matter, as it is forgotten along with the internal dynamics' transients. The warmed up states can be used as follows: in the teacher forcing problems ([Equations 3a and 5](#)), one simply chooses $\xi^0 = \bar{x}_0^0$, and in the multiple shooting problem ([Equation 7a](#)) (and when combining it with GTF), one takes $\xi^{(n)} = \bar{x}_{n_s}^0$ for $n = 0, 1, \dots, N_s$.

3.4 Applied current

To create the HCO, the two GM cells were virtually coupled together with dynamic clamp ([Sharp et al., 1993](#)). This was achieved using a custom RTXI module ([Patel et al., 2017](#)) to inject voltage-dependent currents into each cell. The total applied current injected in the i^{th} HCO neuron is given by

$$I_{\text{app},t}^i = \underbrace{I_{h,t}^i + I_{\text{syn},t}^i}_{\text{dynamic clamp}} + u_{i,t} \quad (23)$$

where $(i, j) \in \{(1, 2), (2, 1)\}$. In [Equation 23](#), the current components $I_{h,t}^i$ and $I_{\text{syn},t}^i$ are a virtual hyperpolarization-induced current and a synaptic current defined with conductance-based models, while the current $u_{i,t}$ is a stochastic current used to perturb the HCO dynamics and train the data-driven model. Each individual current component in [Equation 23](#) was saved separately by RTXI while measuring the responses of the HCO.

TABLE 1 Parameters used to create the HCO with dynamic clamp.

i	\bar{g}_{syn}^i	E_{syn}^i	v_{syn}^i	\bar{g}_h^i	E_h^i	v_h^i	$\bar{\tau}_i$
1	105	-80	-50 mV	225	-10	-45 mV	1,000 ms
2	150	-80	-44 mV	120	-10	-45 mV	1,500 ms

3.4.1 Noisy injected current

The current vector u_t in [Equations 1, 2, 3, 7 and 8](#) represents a measured injected current used to excite the dynamics of a neural circuit and facilitate the training of a data-driven model. In the context of our HCO experiment, $u_t = (u_{1,t}, u_{2,t})$ with $u_{i,t}$ being defined as a stochastic component of the applied current in [Equation 23](#). Each $u_{i,t}$ is given by a discretized Ornstein-Uhlenbeck (OU) process with nonzero mean,

$$u_{i,t+1} = u_{i,t} + \delta a(\mu_i - u_{i,t}) + b\sqrt{\delta}\varepsilon_{i,t} \quad (24)$$

where $\delta = 0.1$ ms is the sampling period, μ_i is the process mean, a and b are process parameters determining the process rate and variance, respectively, and $\varepsilon_{i,t}$ are i.i.d. normally distributed random variables with zero mean and unity standard deviation. All the models in this paper were trained and validated with the same dataset, which consisted in two trials of around two minutes each, obtained by changing the parameters a and b . We fixed the means μ_i to values where a rhythm could be observed (see also [dynamic clamp](#) below), and then recorded two minutes of data with the following combinations of OU process parameters: $(a, b) = (0.04, 0.1), (0.02, 0.025)$. Training was performed using the first 75% percent of the dataset (around 165 seconds), while validation was performed using the last 25% of one of the trials (around 27 seconds).

3.4.2 Dynamic clamp currents

Similarly to [Morozova et al. \(2022\)](#), we used the synapse model

$$I_{\text{syn},t}^{ij} = -\bar{g}_{\text{syn}}^i z_{i,t} (v_{i,t} - E_{\text{syn}}^i) \quad (25)$$

$$z_{i,t+1} = z_{i,t} + \frac{\delta}{50(1.1 - \sigma_{\text{syn}}^i(v_{j,t}))} (-z_{i,t} + \sigma_{\text{syn}}^i(v_{j,t}))$$

where δ is the sampling period, and $\sigma_{\text{syn}}^i(v_{j,t}) = (1 + \exp(-(v_{j,t} - v_{\text{syn}}^i)/2))^{-1}$ is the logistic synaptic activation function. Parameters of each of the two synapses can be found in [Table 1](#). The hyperpolarization-induced current was given by

$$I_{h,t}^i = -\bar{g}_h^i w_{i,t} (v_{i,t} - E_h^i) \quad (26)$$

$$w_{i,t+1} = w_{i,t} + \frac{\delta}{\bar{\tau}_i \tau_h(v_{i,t}) + 0.1} (-w_{i,t} + \sigma_h^i(v_{i,t}))$$

with $\tau_h(v_{i,t}) = (1 + \exp(-(v_{i,t} + 110)/13))^{-1}$ and $\sigma_h^i(v_{i,t}) = (1 + \exp(-(v_{i,t} - v_h^i)/7))^{-1}$ both given by logistic nonlinearities. The specific parameters of each hyperpolarization-induced current are found in [Table 1](#). Synaptic and H-current parameters were chosen starting from the values indicated in [Morozova et al. \(2022\)](#), and then varying the parameters shown in [Table 1](#) until a HCO rhythm was observed under dynamic clamp.

3.4.3 Validation

In Figures 2–4, the *training loss* corresponds to the loss function of the method being used to train the model, that is, Equations 6, 7a or 8a. The *validation metric* is a modified angular separation metric used to assess the predictive power of the models when they are used for predicting the voltage time series, which is done by simulating (Equation 2). The angular separation between two smoothed spike trains is a traditional metric for comparing the similarity between two spike trains (Gerstner et al., 2014), and is given by $\langle y, \hat{y} \rangle / (\|y\| \|\hat{y}\|)$ where y and \hat{y} are vector-valued spike train sequences, $\langle \cdot, \cdot \rangle$ is the inner product, and $\|\cdot\|$ is the norm of a vector. In this work we modified this metric according to

$$\text{modified angular separation} = \frac{\langle y, \hat{y} \rangle}{\max\{\|y\|^2, \|\hat{y}\|^2\}}$$

so as to take magnitude (as well as shape) of the smoothed spike trains into account. The modified metric was empirically determined to work better for bursting signals. Here, $y = \{y_0, \dots, y_N\}$ and $\hat{y} = \{\hat{y}_0, \dots, \hat{y}_N\}$ are smoothed spike train time series obtained as follows: (i) both voltage time series v and \hat{v} are bandpass-filtered forwards and backwards in time by a 3rd order Butterworth filter with cutoff frequencies at 1/50 and 1/2 kHz to remove the slow bursting wave (forwards-and-backwards filtering is used to avoid phase distortion) (ii) the result is thresholded using a relu nonlinearity with threshold set at 5 mV to extract the spikes from the resulting high-frequency spiking signal (iii) the maxima of thresholded spikes is used to define a spike train of impulses (maxima within 5 samples counted as a single spike) (iv) the resulting spike train is convolved with a Laplace kernel given by $\exp(-|t|/500) / \sum_t |\exp(-|t|/500)|$. Notice that the modified angular separation always lies between 0 and 1, with 0 corresponding to completely disjoint smoothed spike trains, and 1 corresponding to a perfect prediction.

3.5 Admittance-voltage curve

The YV curve $Y(v, \omega)$ in Equation 11 describes how a small sinusoidal signal of frequency ω around a constant set-point v is amplified by the total intrinsic current of the neuron (the sum of the leak and ionic currents, and, in a neuronal circuit, of the synaptic currents). Mathematically, this signal can be represented as

$$\tilde{v}_t = \varepsilon \sin(\omega \delta t) + v, \quad t = 0, 1, 2, \dots$$

with constant v and small $\varepsilon > 0$ (recall δ is the sampling period of the measurements). The total intrinsic current of a neuron (or circuit) is the dynamical system

$$\begin{aligned} x_{t+1} &= f_\eta(v_t, x_t) \\ y_t &= h_\theta(v_t, x_t) \end{aligned} \quad (27)$$

with input given by v_t and output (the current) given by y_t . The YV curve (Equation 11) is the *transfer function* of the *linearized* system above (Åström and Murray, 2008). It can be used to determine that the approximate response of the system to the sinusoidal signal, according to

$$y_t = \varepsilon |Y(v, \omega)| \sin(\omega \delta t + \angle Y(v, \omega)) + u_\infty(v), \quad t = 0, 1, 2, \dots \quad (28)$$

where $|Y(v, \omega)|$ is the *gain* and $\angle Y(v, \omega)$ the *phase* of the system's transfer function.

To derive the extension of conductances into the frequency domain, we prove the following statement:

Proposition: Under the contraction assumption, we have

$$Y(v, 0) = G_\infty(v) \quad (29)$$

for all v . Furthermore, if there exist $\bar{\omega} \geq 0$ and $\bar{v} \in \mathbb{R}$ such that

$$C(e^{j\bar{\omega}\delta} - 1) + \delta Y(\bar{v}, \bar{\omega}) = 0 \quad (30)$$

then \bar{v} is a candidate fold bifurcation point ($\bar{\omega} = 0$) or a candidate Neimark-Sacker bifurcation point ($\bar{\omega} > 0$).

Proof of the proposition: The contraction assumption ensures that, for constant v , the nonlinear system (Equation 27) has a unique equilibrium point $x_\infty(v)$ which is asymptotically stable (Theorem 2 of Lohmiller and Slotine, 1998). This equilibrium point is given implicitly by Equation 9b. The Converse Lyapunov Theorem (Khalil, 2002) then guarantees that the eigenvalues of the Jacobian matrix $\partial f_\eta / \partial x|_{x=x_\infty(v)}$ are strictly inside the unit circle for all v . It follows that the matrix

$$\frac{\partial}{\partial x} (x - f_\eta(v, x)) = I - \frac{\partial}{\partial x} f_\eta(v, x)$$

is invertible at $(v, x_\infty(v))$ for all v . We can thus use the implicit function theorem (Rudin, 1976) to find that for each v , the solution $x_\infty(v)$ of the equilibrium equation (Equation 9b) satisfies

$$\frac{\partial}{\partial v} x_\infty(v) = \left(I - \frac{\partial}{\partial x} f_\eta \right)^{-1} \frac{\partial}{\partial v} f_\eta$$

with partial derivatives evaluated at $(v, x_\infty(v))$. The relation (Equation 29) follows from substituting this expression in Equation 10 and comparing the result to Equation 11.

The fact that Equation 30 is a necessary condition for a bifurcation can be established by applying Schur's complement to express the determinant of the matrix arising from the condition that an eigenvalue of the Jacobian of the linearized model (Equation 1) lies on the unit circle. Finally, Equation 13 and the bifurcation condition (Equation 16) follow directly from taking the real parts of Equations 29 and 30.

4 Materials and equipment

4.1 Animals

Cancer borealis were purchased from Commercial Lobster (Boston, MA) between May 2024 and March 2025 and maintained in tanks containing artificial seawater (Instant Ocean) at 11°C–13°C. Animals were anesthetized on ice for 30 minutes before the stomatogastric nervous system was dissected and pinned in a petri dish coated with Sylgard (Dow Corning) as described previously (Gutierrez and Grashow, 2009). The stomatogastric ganglion (STG) was desheathed to allow intracellular recording. All preparations included the STG, the esophageal ganglia, and two commissural ganglia. The nervous system was kept in physiological saline

composed of 440 mM NaCl, 11 mM KCl, 13 mM CaCl_2 , 26 mM MgCl_2 , 11 mM Trizma base, and 5 mM Maleic acid, pH 7.4–7.5 at 23°C (7.7–7.8 pH at 11°C). All reagents were purchased from Sigma-Aldrich.

4.2 Electrophysiology

Extracellular recordings were made by placing 90% Vaseline 10% mineral oil solution wells around the upper lateral ventricular nerve (*lvn*) and stomatogastric nerve (*stn*). Stainless-steel pin electrodes were placed within the *lvn* well to monitor spiking activity and amplified using Model 1700 Differential AC Amplifiers (A-M Systems). Intracellular recordings from the soma of two Gastric Mill (GM) cells were made using two-electrode current clamp with 5–20 M sharp glass microelectrodes filled with 0.6 M K_2SO_4 and 20 mM KCl solution. Intracellular signals were amplified with an Axoclamp 900 A amplifier (Molecular Devices, San Jose). Cells were identified by matching intracellular activity with activity on the *lvn*. All amplified signals were digitized using a Digidata 1440 digitizer (Molecular Devices, San Jose) and recorded using pClamp data acquisition software (Molecular Devices, San Jose, version 10.5), sampling frequency of 10 kHz.

After identification of two GM cells, the preparation was disconnected from the descending modulation by adding a solution of 750 mM sucrose and 10^{-7}M tetrodotoxin (Sigma) in the *stn* well. Glutamatergic synapses were blocked by adding 10^{-5}M picrotoxin (Sigma) to the perfusion solution. For the purposes of real-time data processing the intracellular signals were acquired with NI PCIe-6259 M Series Multifunction DAQ (National Instruments) at 10 kHz sampling frequency, and used in Real-Time eXperiment Interface (RTXI) software version 3.0.

4.3 Hardware

All computational work was done in a desktop computer set up with an AMD Ryzen 7 7800X3D 8-Core Processor, 64 Gigabytes of RAM, and a NVIDIA GeForce RTX 4090 graphics card with 24 GB of VRAM. Models were trained in the GPU using the Julia library CUDA.jl in conjunction with the module Flux.jl.

5 Discussion

What criteria should one use to determine the predictive quality of a data-driven model? Invariably, machine learning and AI approaches use a loss of some kind, and ask how well the model predictions generalize with respect to this loss on unseen data. The choice of loss can be crucial: evaluating the point-wise difference of values on a timeseries can emphasize quantitative variation in a dataset that might be an idiosyncratic feature of *that* dataset. This danger is well known, and places the burden on the experimentalist to measure as broad and ‘representative’ a dataset as is feasible.

5.1 Deep-learning models for intracellular dynamics

In this paper, we have considered ANN-based models aimed at *quantitative prediction* of intracellular dynamics. Our understanding of quantitative prediction is specific, and helps to narrow the class of data-driven models considered in our work. Throughout this paper, we have implicitly equated the *membrane dynamics* of a neuron with its *system behavior*, a concept borrowed from control theory (Willemss, 2007). Following Sepulchre et al. (2018), the system behavior of the membrane is the set of all applied current–membrane voltage trajectory pairs that are consistent with the membrane’s biophysics. Concretely, this implies that quantitatively accurate predictions of the membrane dynamics must be consistent with current-clamp and voltage-clamp recordings alike. This view is useful because it unambiguously excludes from consideration many types of data-driven models which can otherwise be said to be predictive (for instance, most integrate-and-fire models, whose suprathreshold behavior is not well-defined).

Following this view, quantitatively predictive models have until recently been almost exclusively “detailed”, “biophysical”, or “conductance-based”, that is, consistent with the mechanistic modeling paradigm established by Hodgkin and Huxley in their seminal work (Hodgkin and Huxley, 1952). However, because conductance-based models are derived from a highly simplified low-dimensional description of neuronal dynamics, their quantitative predictive power is limited. Data-driven models forfeiting quantitative predictions of the membrane dynamics, such as Spike Response Models (Gerstner et al., 2014) and Generalized Linear Models (Pillow and Park, 2016), have repeatedly been shown to beat conductance-based models at predicting particular features of neuronal excitability, most notably spike timing. The ANN-based models considered in this paper can improve on the quantitative prediction of conductance-based models, while keeping the data-driven flexibility of “black-box” statistical models. As we have seen, ANN-based frameworks such as RMMs (Burghi et al., 2025a) further allow biophysical knowledge to inform model architecture. The HCO circuit used in this paper had a known connectivity, allowing us to exploit knowledge of synaptic dynamics to improve on synaptic current predictions. In circuits with unknown connectivity, we have shown that models with minimal biophysical priors can still be trusted to provide reasonable (if not perfectly accurate) predictions of internal circuit currents.

The modeling paradigm studied in this paper prioritizes quantitative analytical predictions over fine biophysical detail, while avoiding the pitfalls of purely statistical models. Recent papers following this paradigm have focused on different aspects of the learning and interpretation problem, but are similar in spirit (Brenner et al., 2022; Durstewitz et al., 2023; Aguiar et al., 2024; Burghi et al., 2025a). Such approaches can in principle be pursued in continuous-time, under the framework of Neural (Chen et al., 2018) and Universal (Rackauckas et al., 2021) ODEs. Interesting questions concerning the potential efficiency gains of continuous-time implementations are left for future work.

5.2 Data-driven models can be interpreted using electrophysiology tools

There is a long tradition in electrophysiology of using voltage clamp to study neural excitability via the spectral properties of the neuronal membrane and its currents; see for instance (Mauro et al., 1970; Juusola, 1994; Magnani and Moore, 2011). We have seen that frequency-dependent conductances used to study neural excitability can also be employed to *interpret* a data-driven model which, at first, resembles a black box. This type of analysis was enabled by two model properties: the separation of model states into voltage and internal states, and the contraction of the internal dynamics.

Our results enable relevant extensions of this frequency-domain methodology. First, one could seek to interpret data-driven models with timescale-separated admittances, similarly what is done in Franci et al. (2019) for conductance-based models. The linearity of the internal dynamics of RMMs can readily be exploited for that purpose. Second, instead of analysing admittances neuron by neuron, one could study a data-driven neural circuit by treating its ionic currents as a *multiple-input-multiple-output* system, in which case admittances become matrix-valued (Goodwin and Sin, 1984; Khalil, 2002). Developing these extensions is left for future work.

It is important to distinguish the frequency-dependent conductances obtained from our model from those that could be in principle obtained experimentally with voltage-clamp. For many types of neurons with complex morphology, including the STG neurons used in this paper, voltage-clamp recordings made at a single site (e.g., the soma) leads to distorted recorded currents; see, for instance, Bar-Yehuda and Korngreen (2008); Poleg-Polsky and Diamond (2011); Taylor (2012). Hence, frequency-dependent conductances should be regarded primarily as a tool for model interpretation, rather than for predicting precise experimental I-V relations.

5.3 Training methods

While we have emphasized that TF, multiple-shooting and GTF can be understood from a unified perspective, other approaches in the literature can also be associated to these methods. In particular, feed-forward TF as discussed here can be related to the idea of *reservoir computing* (Lukoševičius and Jaeger, 2009). The conceptual difference to the way reservoir computing is usually introduced, is that in data-driven models for intracellular dynamics, the reservoir is an *internal* dynamics, given by Equation 1b, which can be fixed to provide a reservoir of internal states. In neuronal models, learning the “readout” function h_θ is not the goal but rather a means to learn the forward dynamics (Equation 1). In other words, the task of the reservoir network is to learn the *inverse* dynamics of the neuron or circuit: the dynamics of intrinsic and synaptic ionic currents.

Our paper connects the areas of data-driven models for neural circuit dynamics to contraction theory (Lohmiller and Slotine, 1998) from control engineering. There is a growing

literature in systems theory dealing with learning contracting data-driven systems such as recurrent neural networks and equilibrium networks; see, for instance, (Manchester, 2018; Revay and Manchester, 2024; Pauli et al., 2024).

Data availability statement

The raw data supporting the conclusions of this article can be found in Burghi et al. (2025b), <https://doi.org/10.17863/CAM.120457>.

Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because ethical approval is not required for research on Cancer Borealis in the United States of America.

Author contributions

TB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. KS: Methodology, Writing – original draft. MI: Methodology, Writing – original draft. HW: Investigation, Methodology, Validation, Writing – review & editing. EM: Project administration, Resources, Supervision, Writing – review & editing. TO: Conceptualization, Project administration, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition, Investigation.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Kavli Foundation (Thiago Burghi, Harry Wang, Timothy O’Leary) and the National Institutes of Health (Kyra Schapiro, Maria Ivanova, Eve Marder) (NIH grant numbers R35 NS097343 and 5T32NS007292-31).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abarbanel, H., Creveling, D., Farsian, R., and Kostuk, M. (2009). Dynamical State and Parameter Estimation. *SIAM J. Appl. Dyn. Syst.* 8, 1341–1381. doi: 10.1137/090749761
- Abu-Hassan, K., Taylor, J. D., Morris, P. G., Donati, E., Bortolotto, Z. A., Indiveri, G., et al. (2019). Optimal solid state neurons. *Nat. Commun.* 10:5309. doi: 10.1038/s41467-019-13177-3
- Aguiar, M., Das, A., and Johansson, K. H. (2024). "Learning flow functions of spiking systems," in *Proceedings of the 6th Annual Learning for Dynamics & Control Conference* (New York: PMLR), 591–602.
- Almog, M., and Korngreen, A. (2016). Is realistic neuronal modeling realistic? *J. Neurophysiol.* 116, 2180–2209. doi: 10.1152/jn.00360.2016
- Alonso, L. M., and Marder, E. (2019). Visualization of currents in neural models with similar behavior and different conductance densities. *Elife* 8:e42722. doi: 10.7554/eLife.42722
- Anwar, H., Martinez, D., Bucher, D., and Nadim, F. (2022). Inter-animal variability in activity phase is constrained by synaptic dynamics in an oscillatory network. *eNeuro* 9:ENEURO.0027-22.2022. doi: 10.1523/ENEURO.0027-22.2022
- Aström, K. J., and Murray, R. M. (2008). *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton: Princeton University Press.
- Bar-Yehuda, D., and Korngreen, A. (2008). Space-clamp problems when voltage clamping neurons expressing voltage-gated conductances. *J. Neurophysiol.* 99, 1127–1136. doi: 10.1152/jn.01232.2007
- Brenner, M., Hess, F., Mikhael, J. M., Bereska, L. F., Monfared, Z., Kuo, P.-C., et al. (2022). "Tractable dendritic RNNs for reconstructing nonlinear dynamical systems," in *Proceedings of the 39th International Conference on Machine Learning* (New York: PMLR), 2292–2320.
- Brookings, T., Goeritz, M. L., and Marder, E. (2014). Automatic parameter estimation of multicompartmental neuron models via minimization of trace error with control adjustment. *J. Neurophysiol.* 112, 2332–2348. doi: 10.1152/jn.00007.2014
- Buhry, L., Grassia, F., Giremus, A., Grivel, E., Renaud, S., and Saighi, S. (2011). Automated parameter estimation of the hodgkin-huxley model using the differential evolution algorithm: application to neuromimetic analog integrated circuits. *Neural Comput.* 23, 2599–2625. doi: 10.1162/NECO_a_00170
- Burghi, T. B., Ivanova, M., Morozova, E. O., Wang, H., Marder, E., and O'Leary, T. (2025a). Rapid, interpretable data-driven models of neural dynamics using recurrent mechanistic models. *Proc. Natl. Acad. Sci. U.S.A.* 122:e2426916122. doi: 10.1073/pnas.2426916122
- Burghi, T. B., Schapiro, K., Ivanova, M., Wang, H., Marder, E., and O'Leary, T. (2025b). *Research Data Supporting 'Quantitative Prediction of Intracellular Dynamics and Synaptic Currents in a Small Neural Circuit'*. Apollo - University of Cambridge Repository. doi: 10.17863/CAM.120457
- Burghi, T. B., Schoukens, M., and Sepulchre, R. (2021). Feedback identification of conductance-based models. *Automatica* 123:109297. doi: 10.1016/j.automatica.2020.109297
- Burghi, T. B., and Sepulchre, R. (2024). Adaptive observers for biophysical neuronal circuits. *IEEE Trans. Automat. Contr.* 69, 5020–5033. doi: 10.1109/TAC.2023.3344723
- Calabrese, R. L., and Marder, E. (2024). Degenerate neuronal and circuit mechanisms important for generating rhythmic motor patterns. *Physiol. Rev.* 105, 95–135. doi: 10.1152/physrev.00003.2024
- Calabrese, R. L., Nadim, F., and Olsen, Ø. H. (1995). Heartbeat control in the medicinal leech: a model system for understanding the origin, coordination, and modulation of rhythmic motor patterns. *J. Neurobiol.* 27, 390–402. doi: 10.1002/neu.480270311
- Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). "Neural ordinary differential equations," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, QC: Curran Associates, Inc.), 6572–6583.
- Doya, K. (1992). "Bifurcations in the learning of recurrent neural networks," in *1992 IEEE International Symposium on Circuits and Systems* (San Diego, CA: IEEE), 2777–2780.
- Drion, G., Dethier, J., Franci, A., and Sepulchre, R. (2018). Switchable slow cellular conductances determine robustness and tunability of network states. *PLoS Comput. Biol.* 14:e1006125. doi: 10.1371/journal.pcbi.1006125
- Durstewitz, D., Koppe, G., and Thurm, M. I. (2023). Reconstructing computational system dynamics from neural data with recurrent neural networks. *Nat. Rev. Neurosci.* 24, 693–710. doi: 10.1038/s41583-023-00740-7
- Franci, A., Drion, G., and Sepulchre, R. (2017). Robust and tunable bursting requires slow positive feedback. *J. Neurophysiol.* 119, 1222–1234. doi: 10.1152/jn.00804.2017
- Franci, A., Drion, G., and Sepulchre, R. (2019). "The sensitivity function of excitable feedback systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)* (Nice: IEEE), 4723–4728.
- Geit, W. V., Schutter, E. D., and Achard, P. (2008). Automated neuron model optimization techniques: a review. *Biol. Cybern.* 99:241–251. doi: 10.1007/s00422-008-0257-6
- Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge: Cambridge University Press.
- Gerstner, W., and Naud, R. (2009). How good are neuron models? *Science* 326, 379–380. doi: 10.1126/science.1181936
- Goaillard, J.-M., Taylor, A. L., Schulz, D. J., and Marder, E. (2009). Functional consequences of animal-to-animal variation in circuit parameters. *Nat. Neurosci.* 12, 1424–1430. doi: 10.1038/nn.2404
- Gonçalves, P. J., Lueckmann, J.-M., Deistler, M., Nonnenmacher, M., Öcal, K., Bassetto, G., et al. (2020). Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife* 9:e56261. doi: 10.7554/eLife.56261
- Goodwin, G. C., and Sin, K. S. (1984). *Adaptive Filtering Prediction and Control*. Mineola, NY: Dover Publications.
- Grashow, R., Brookings, T., and Marder, E. (2009). Reliable neuromodulation from circuits with variable underlying structure. *Proc. Nat. Acad. Sci.* 106, 11742–11746. doi: 10.1073/pnas.0905614106
- Gu, A., Goel, K., and Ré, C. (2022). Efficiently modeling long sequences with structured state spaces. *arXiv [Preprint]*. doi: 10.48550/arXiv.2111.00396
- Gutierrez, G. J., and Grashow, R. G. (2009). Cancer borealis stomatogastric nervous system dissection. *J. Visual. Exp.* 25:1207. doi: 10.3791/1207
- Hess, F., Monfared, Z., Brenner, M., and Durstewitz, D. (2023). "Generalized teacher forcing for learning chaotic dynamics," in *Proceedings of the 40th International Conference on Machine Learning* (Honolulu, Hawaii: JMLR.org), 13017–13049.
- Heuberger, P. S. C., van den Hof, P. M. J., and Wahlberg, B. (2005). *Modelling and Identification with Rational Orthogonal Basis Functions*. London: Springer.
- Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764
- Huys, Q. J. M., Ahrens, M. B., and Paninski, L. (2006). Efficient estimation of detailed single-neuron models. *J. Neurophysiol.* 96, 872–890. doi: 10.1152/jn.00079.2006
- Innes, M. (2018). Flux: Elegant machine learning with Julia. *J. Open Source Softw.* 3:602. doi: 10.21105/joss.00602
- Izhikevich, E. M. (2007). *Dynamical Systems in Neuroscience*. Cambridge, MA: MIT Press.
- Jouffroy, J., and Fossen, T. I. (2010). A tutorial on incremental stability analysis using contraction theory. *Model. Identif. Cont.* 31, 93–106. doi: 10.4173/mic.2010.3.2
- Juusola, M. (1994). Measuring complex admittance and receptor current by single electrode voltage-clamp. *J. Neurosci. Methods* 53, 1–6. doi: 10.1016/0165-0270(94)90137-6
- Khalil, H. K. (2002). *Nonlinear Systems*. Upper Saddle River, NJ: Prentice Hall.
- Kingma, D. P., and Ba, J. (2017). Adam: a method for stochastic optimization. *arXiv [Preprint]*. doi: 10.48550/arXiv.1412.6980

- Latimer, K. W., Rieke, F., and Pillow, J. W. (2019). Inferring synaptic inputs from spikes with a conductance-based neural encoding model. *Elife* 8:e47012. doi: 10.7554/eLife.47012
- Li, M., Zhang, T., Chen, Y., and Smola, A. J. (2014). "Efficient mini-batch training for stochastic optimization," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14* (New York, NY: Association for Computing Machinery), 661–670.
- Liu, Z., Golowasch, J., Marder, E., and Abbott, L. F. (1998). A model neuron with activity-dependent conductances regulated by multiple calcium sensors. *J. Neurosci.* 18, 2309–2320. doi: 10.1523/JNEUROSCI.18-07-02309.1998
- Ljung, L. (1999). *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice Hall PTR.
- Lohmiller, W., and Slotine, J.-J. E. (1998). On contraction analysis for non-linear systems. *Automatica* 34, 683–696. doi: 10.1016/S0005-1098(98)00019-3
- Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., and Macke, J. H. (2017). "Flexible statistical inference for mechanistic models of neural dynamics," in *31st Conference on Neural Information Processing Systems (NIPS 2017)* (Long Beach, CA: NIPS).
- Lukoševičius, M., and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Comp. Sci. Rev.* 3, 127–149. doi: 10.1016/j.cosrev.2009.03.005
- Magnani, C., and Moore, L. E. (2011). Quadratic sinusoidal analysis of voltage clamped neurons. *J. Comput. Neurosci.* 31, 595–607. doi: 10.1007/s10827-011-0325-0
- Manchester, I. R. (2018). "Contracting nonlinear observers: convex optimization and learning from data," in *2018 American Control Conference (ACC)* (Milwaukee, WI: IEEE), 1873–1880.
- Manchester, I. R., Revay, M., and Wang, R. (2021). "Contraction-based methods for stable identification and robust machine learning: a tutorial," in *2021 60th IEEE Conference on Decision and Control (CDC)* (Austin, TX: IEEE), 2955–2962.
- Marder, E. (2012). Neuromodulation of neuronal circuits: back to the future. *Neuron* 76, 1–11. doi: 10.1016/j.neuron.2012.09.010
- Mauro, A., Conti, F., Dodge, F., and Schor, R. (1970). Subthreshold behavior and phenomenological impedance of the squid giant axon. *J. General Physiol.* 55, 497–523. doi: 10.1085/jgp.55.4.497
- Meliza, C. D., Kostuk, M., Huang, H., Nogaret, A., Margoliash, D., and Abarbanel, H. D. I. (2014). Estimating parameters and predicting membrane voltages with conductance-based neuron models. *Biol. Cybern.* 108, 495–516. doi: 10.1007/s00422-014-0615-5
- Meng, Q., Chen, W., Wang, Y., Ma, Z.-M., and Liu, T.-Y. (2019). Convergence analysis of distributed stochastic gradient descent with shuffling. *Neurocomputing* 337, 46–57. doi: 10.1016/j.neucom.2019.01.037
- Morozova, E., Newstein, P., and Marder, E. (2022). Reciprocally inhibitory circuits operating with distinct mechanisms are differently robust to perturbation and modulation. *Elife* 11:e74363. doi: 10.7554/eLife.74363
- Nguyen, D., and Widrow, B. (1990). "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *1990 IJCNN International Joint Conference on Neural Networks*, 21–26.
- Paninski, L., Pillow, J., and Simoncelli, E. (2005). Comparing integrate-and-fire models estimated using intracellular and extracellular data. *Neurocomputing* 65, 379–385. doi: 10.1016/j.neucom.2004.10.032
- Pascanu, R., Mikolov, T., and Bengio, Y. (2013). "On the difficulty of training recurrent neural networks," in *Proceedings of the 30th International Conference on Machine Learning* (Atlanta, GA: ICML'13), 1310–1318.
- Patel, Y., George, A., White, J., Dorval, A., Christini, D., and Butera, R. (2017). Hard real-time closed-loop electrophysiology with the Real-Time eXperiment Interface (RTXI). *PLoS Comput. Biol.* 13:e1005430. doi: 10.1371/journal.pcbi.1005430
- Pauli, P., Wang, R., Manchester, I. R., and Allgöwer, F. (2024). "Lipschitz-bounded 1D convolutional neural networks using the Cayley transform and the controllability Gramian," in *2023 62nd IEEE Conference on Decision and Control (CDC)* (Singapore: IEEE).
- Pillow, J., and Park, M. (2016). "Adaptive Bayesian methods for closed-loop neurophysiology," in *Closed Loop Neuroscience* (Cambridge, MA: Academic Press), 3–18.
- Poleg-Polsky, A., and Diamond, J. S. (2011). Imperfect space clamp permits electrotonic interactions between inhibitory and excitatory synaptic conductances, distorting voltage clamp recordings. *PLoS ONE* 6:e19463. doi: 10.1371/journal.pone.0019463
- Prinz, A. A., Abbott, L. F., and Marder, E. (2004a). The dynamic clamp comes of age. *Trends Neurosci.* 27, 218–224. doi: 10.1016/j.tins.2004.02.004
- Prinz, A. A., Bucher, D., and Marder, E. (2004b). Similar network activity from disparate circuit parameters. *Nat. Neurosci.* 7, 1345–1352. doi: 10.1038/nn1352
- Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., et al. (2021). *Universal Differential Equations for Scientific Machine Learning*.
- Revay, M., and Manchester, I. (2020). "Contracting implicit recurrent neural networks: stable models with improved trainability," in *Proceedings of Machine Learning Research* (Arlington, VA: The cloud).
- Revay, M., Wang, R., and Manchester, I. R. (2024). Recurrent equilibrium networks: flexible dynamic models with guaranteed stability and robustness. *IEEE Trans. Automat. Contr.* 69, 2855–2870. doi: 10.1109/TAC.2023.3294101
- Ribeiro, A. H., Tiels, K., Aguirre, L. A., and Schön, T. (2020a). "Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (New York: PMLR), 2370–2380.
- Ribeiro, A. H., Tiels, K., Umenberger, J., Schön, T. B., and Aguirre, L. A. (2020b). On the smoothness of nonlinear system identification. *Automatica* 121:109158. doi: 10.1016/j.automatica.2020.109158
- Rudin, W. (1976). *Principles of Mathematical Analysis*. New York City: McGraw-Hill.
- Schulz, D. J., Goaillard, J.-M., and Marder, E. (2006). Variable channel expression in identified single and electrically coupled neurons in different animals. *Nat. Neurosci.* 9, 356–362. doi: 10.1038/nn1639
- Schulz, D. J., Goaillard, J.-M., and Marder, E. E. (2007). Quantitative expression profiling of identified neurons reveals cell-specific constraints on highly variable levels of gene expression. *Proc. Nat. Acad. Sci.* 104, 13187–13191. doi: 10.1073/pnas.0705827104
- Sepulchre, R., Drion, G., and Franci, A. (2018). "Excitable behaviors," in *Emerging Applications of Control and Systems Theory* (Cham: Springer International Publishing), 269–280.
- Sharp, A. A., O'Neil, M. B., Abbott, L. F., and Marder, E. (1993). Dynamic clamp: Computer-generated conductances in real neurons. *J. Neurophysiol.* 69, 992–995. doi: 10.1152/jn.1993.69.3.992
- Sharp, A. A., Skinner, F. K., and Marder, E. (1996). Mechanisms of oscillation in dynamic clamp constructed two-cell half-center circuits. *J. Neurophysiol.* 76, 867–883. doi: 10.1152/jn.1996.76.2.867
- Sontag, E. D. (1998). *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. (New York, NY: Springer).
- Taylor, A. L. (2012). What we talk about when we talk about capacitance measured with the voltage-clamp step method. *J. Comput. Neurosci.* 32, 167–175. doi: 10.1007/s10827-011-0346-8
- Taylor, J. D., Winnall, S., and Nogaret, A. (2020). Estimation of neuron parameters from imperfect observations. *PLoS Comput. Biol.* 16:e1008053. doi: 10.1371/journal.pcbi.1008053
- Wells, S. A., Taylor, J. D., Morris, P. G., and Nogaret, A. (2024). Inferring the dynamics of ionic currents from recursive piecewise data assimilation of approximate neuron models. *PRX Life* 2:023007. doi: 10.1103/PRXLife.2.023007
- Wenning, A., Cymbalyuk, G. S., and Calabrese, R. L. (2004). Heartbeat control in leeches. I. constriction pattern and neural modulation of blood pressure in intact animals. *J. Neurophysiol.* 91, 382–396. doi: 10.1152/jn.00526.2003
- Willems, J. C. (2007). The behavioral approach to open and interconnected systems. *IEEE Cont. Syst. Magaz.* 27, 46–99. doi: 10.1109/MCS.2007.906923