Universal Differential Equations as a Common Modeling Language for Neuroscience

Ahmed El-Gazzar (ahmed.elgazzar@donders.ru.nl)
Department of Machine Learning and Neural Computing,
Donders Institute for Brain, Cognition and Behaviour, Radboud University,
Nijmegen, the Netherlands

Marcel van Gerven (marcel.vangerven@donders.ru.nl)
Department of Machine Learning and Neural Computing,
Donders Institute for Brain, Cognition and Behaviour, Radboud University,
Nijmegen, the Netherlands

Abstract

The rise of large-scale neuroscience datasets has driven widespread adoption of deep neural networks (DNNs) as models of biological neural systems. While DNNs can approximate functions directly from data circumventing the need for mechanistic modeling, they risk producing implausible and difficult-to-interpret models. In this paper, we argue for universal differential equations (UDEs) as a unifying approach for model development and validation in neuroscience. UDEs view differential equations as parameterizable, differentiable mathematical objects that can be augmented and trained with scalable deep learning techniques. This synergy facilitates the integration of classical mathematical modeling with emerging advancements in AI into a potent framework. We provide a primer on this burgeoning topic in scientific machine learning and describe a generative modeling recipe for fitting UDEs on neural and behavioral data. Our goal is to show how UDEs can fill in a critical gap between mechanistic, phenomenological, and data-driven models in neuroscience and highlight their potential to address inherent challenges across diverse applications such as understanding neural computation, controlling neural systems, neural decoding, and normative modeling.

Keywords: Differentiable modeling; Universal differential equations; Neural computation; Dynamical systems

Introduction

As holds for all the natural sciences, modern neuroscience is a scientific discipline whose advancement is fueled by both theoretical and experimental research. From a theoretical standpoint, we have witnessed important developments, ranging from detailed mechanistic models of specific neural circuits (S. S. Kim, Rouault, Druckmann, & Jayaraman, 2017; Izhikevich & Edelman, 2008; Felleman & Van Essen, 1991; Bliss & Collingridge, 1993) to grand unified theories of brain function (Van Gelder, 1998; Friston, 2009; Hawkins, 2021; Miller & Cohen, 2001). At the same time, from an experimental standpoint, advances in neurotechnolgy are allowing us to measure (Steinmetz et al., 2021; Urai, Doiron, Leifer, & Churchland, 2022; Machado, Kauvar, & Deisseroth, 2022) and manipulate (Deisseroth, 2015; Lozano et al., 2019) the activity of multiple neurons at an unprecedented scale.

A critical question is how to effectively integrate theoretical and empirical insights to expand our grasp of neural mechanisms and advance practical applications. In this perspective paper, we argue for universal differential equations (UDEs) as a unifying approach for neuroscience (Rackauckas et al., 2020). UDEs embrace the dynamical systems perspective on neuroscience, where neural systems are viewed as dynamical systems whose flow (dynamics) can be described in terms of systems of differential equations (DEs) (Izhikevich, 2007; Favela, 2021; Durstewitz, Koppe, & Thurm, 2023). Unlike conventional DEs, UDEs can be (partly or fully) estimated from data by marrying dynamical systems theory with machine learning. This formulation allows the integration of a-priori knowledge about the system along with high-capacity function approximators to model complex systems in the absence of large-scale datasets. Consequently, UDEs are rapidly garnering attention across scientific domains where the datasets are still relatively scarce, and mechanistic, theory-driven models are prevalent, yet fall short in accounting for data variance (Rackauckas et al., 2020; AlQuraishi & Sorger, 2021; Lai, Mylonas, Nagarajaiah, & Chatzi, 2021; Karniadakis et al., 2021). Similarly in neuroscience, differential equations are ubiquitous, underpinning the majority of theoretical, biophysical, and phenomenological models (Izhikevich, 2007). And despite their advances, existing measurements tools only provide sparse and noisy representation of the underlying neural mechanisms, which require both appropriate numerical tools and expert knowledge to guide modeling. In our view, UDEs provide a unique opportunity to bridge different modeling techniques, spanning various biological and abstraction scales in a unified framework to propel both fundamental and applied neuroscience.

To motivate UDEs, we begin with a critique on the current landscape of data-driven dynamical systems in neuroscience, highlighting key applications, and challenges, culminating in the motivation for hybrid approaches. Next, we delve into the taxonomy of UDEs in the context of stochastic dynamical systems and show how these mathematical objects provide a spectrum of modeling techniques familiar to the neuroscientist spanning from traditional mechanistic white-box models to sophisticated black-box deep learning models. We provide a general recipe for domain-informed training of UDEs for neural system identification and examine the benefits of UDE-based models in emerging applications within the field. We conclude by discussing current challenges and potential future directions. Through this discourse, we argue that UDEs, when augmented with modern machine learning techniques, can serve as the foundational building block for multi-scale modeling in neuroscience, establishing a common language for theory formation and model development.

Dynamical Systems in Neuroscience

From Mechanistic to Data-Driven Models

A prevalent perspective in neuroscience is viewing the brain as a dynamical system, availing the comprehensive toolbox of dynamical systems theory (DST) to the field (Van Gelder, 1998; Izhikevich, 2007; Deco, Jirsa, Robinson, Breakspear, & Friston, 2008; Breakspear, 2017; Favela, 2021). DST enables the formalization of mechanistic models as systems of differential equations (Hodgkin & Huxley, 1952; FitzHugh, 1961; Izhikevich, 2003) and provides intuitive geometrical and topological representations of neural systems (Deco & Jirsa, 2012; Khona & Fiete, 2022). This framework also facilitates the adoption of phenomenological models from statistical physics (Wilson & Cowan, 1972; Kuramoto, 1975; Buzsaki & Draguhn, 2004). However, both approaches have limitations: mechanistic models are laborious to develop and often lack detail, while phenomenological models provide only abstract descriptions of neural processes (Ramezanian-Panahi et al., 2022).

The availability of large-scale datasets has spurred the exploration of data-driven dynamical systems (Landhuis, 2017). These methods minimize a-priori assumptions, instead leveraging rich data to guide model identification (S. L. Brunton & Kutz, 2019). This approach is particularly valuable in neuroscience (B. W. Brunton & Beyeler, 2019), where systems are complex, theoretical frameworks are nascent, and measurement tools provide incomplete mechanistic insights. Consequently, data-driven systems, particularly deep recurrent neural networks (RNNs), are increasingly adopted across neuroscience: from probing cognitive functions (Durstewitz et al., 2023; Vyas, Golub, Sussillo, & Shenoy, 2020), to developing neurostimulation profiles (Tang & Bassett, 2018; Acharya, Ruf, & Nozari, 2022), neural decoding (Willett et al., 2023; Metzger et al., 2023), and clinical applications (Bystritsky, Nierenberg, Feusner, & Rabinovich, 2012; Roberts, Friston, & Breakspear, 2017)

Challenges of Data-Driven Methods

The shift towards data-driven methodologies in neuroscience introduces significant technical challenges. These range from data-centric challenges such as high dimensionality, partial observability, non-linearity, process and measurement noise, non-stationarity, and data scarcity, to modeling hurdles such as uncertainty quantification, non-identifiability, and interpretability issues (Durstewitz et al., 2023). This landscape has resulted in a plethora of specialized technical advancements driven by distinct theoretical and practical frameworks (B. W. Brunton & Beyeler, 2019; Hurwitz, Kudryashova, Onken, & Hennig, 2021; Ramezanian-Panahi et al., 2022). A symptom of this status quo is the prevalent dichotomy between model expressivity and interpretability. As researchers opt for more expressive models to capture the intricacies of neural dynamics, they encounter interpretability challenges. This is further exacerbated by optimization challenges that arise either due to the models (e.g. exploding/vanishing gradients in RNNs) or the behavior of the system (e.g. chaos and non-stationarity), entailing highly technical solutions that further fragments neuroscientific practice.

Additionally, while the allure of utilizing unbiased expressive models is initially appealing, in the absence of large-scale curated datasets, eschewing prior knowledge often results in ill-posed problems and implausible solutions as highlighted in recent studies (Kao, 2019; Alber et al., 2019; Genkin, Hughes, & Engel, 2021; Genkin, Shenoy, Chandrasekaran, & Engel, 2023). In practical terms, this means that the models become prone to overfitting on spurious correlations and exhibit high sensitivity to design choices that are peripheral to the main task at hand, ultimately leading to issues in generalization and replication across datasets, tasks, and subjects (Maheswaranathan, Williams, Golub, Ganguli, & Sussillo, 2019; Schaeffer, Khona, & Fiete, 2022; Hurwitz et al., 2021; Han, Poggio, & Cheung, 2023).

New Frontiers

Neural differential equations (NDEs) (Chen, Rubanova, Bettencourt, & Duvenaud, 2018; Kidger, 2022) have emerged as a powerful tool of choice to implement data-driven dynamical systems. NDEs represent an emerging family of continuous models that utilize neural networks to parameterize the vector fields of differential equations. This integration marries the expressive power of neural networks with the rigorous theoretical foundations established by decades of research in differential equations and dynamical systems theory. While originally popularized as deep neural network models with continuous depth (Chen et al., 2018), recent advancements have burgeoned into a rich spectrum of continuous-time architectures rooted in dynamical systems theory (Tzen & Raginsky, 2019a; Morrill, Salvi, Kidger, & Foster, 2021; Z. Li et al., 2020; Jia & Benson, 2019; Poli et al., 2019; Kidger, Morrill, Foster, & Lyons, 2020). Recently, NDEs are being increasingly adopted in computational and systems neuroscience (T. D. Kim, Luo, Pillow, & Brody, 2021; Sedler, Versteeg, & Pandarinath, 2022; Geenjaar et al., 2023; Versteeg, Sedler, McCart, & Pandarinath, 2023). While this is promising, their current application have only focused on black-box, explicitly discretized versions that do not capture the broader potential of NDEs as a pathway towards a unified scientific modeling language (Shen et al., 2023; AlQuraishi & Sorger, 2021; Wang et al., 2023). This untapped potential can be realized by conceptualizing differential equations as parameterizable, differentiable mathematical objects amenable to augmentation and training via scalable machine learning techniques. Traditional DEs and NDEs can thus be viewed as special cases at the extreme ends of a spectrum.

Universal Differential Equations

Mathematical Formulation

A UDE is a mathematical model that extends a traditional differential equation by incorporating free parameters whose values can be learnt from data. By including free parameters, a UDE can act as a universal approximator(Cybenko, 1989; Hornik, Stinchcombe, & White, 1989), meaning that it is able to approximate any dynamical system. In their most general form, UDEs are parameterized forced stochastic delay partial differential equations (Rackauckas et al., 2020). In this paper, we focus our attention on parameterized forced stochastic differential equations (SDEs). SDEs extend ordinary differential equations by incorporating stochastic processes, enabling the modeling of dynamical systems subject to uncertainty. The key to this extension is the inclusion of a stochastic term that represents random fluctuations arising from either intrinsic or extrinsic factors. A forced SDE makes explicit how the (multi-dimensional) state $x(t)$ of a system of interest changes as a function of control inputs $u(t)$ and (Brownian) process noise $W(t)$ with t the time index. This can be succinctly represented as

$$\mathrm{d}x(t) = \mu_\theta(x(t), u(t))\mathrm{d}t + \sigma_\theta(x(t), u(t))\mathrm{d}W(t)\,, \qquad (1)$$

where μ and σ are drift and diffusion functions, representing the deterministic and stochastic parts of the time evolution of the system. Both μ and σ are parameterized by θ, which are the (learnable) free parameters of the system. We often drop the time index from our notation and write (1) more compactly as $\mathrm{d}x = \mu_\theta(x,u)\mathrm{d}t + \sigma_\theta(x,u)\mathrm{d}W$.

SDEs offer considerable flexibility for modeling dynamical systems. This adaptability stems from the diffusion term's configuration and Brownian motion properties (Oksendal, 2013; Särkkä & Solin, 2019). For instance, in cases where σ is a constant matrix or a state-independent function, the noise becomes *additive*, rendering it suitable for modeling extrinsic uncertainties such as external, unobserved interactions. Conversely, when σ is a function of the system's state, the noise becomes *multiplicative*, which varies with the system's state, capturing intrinsic uncertainties, such as uncertainties in drift term parameters. These nuances provide a comprehensive framework for modeling complex dynamical systems with varying types of uncertainty. It is at the modeler's discretion to define the functional form of μ and σ. Ultimately, this functional form should accurately capture the (uncertain) evolution of the state of the system. This is evaluated by computing the solution to Eq. 1, which is a distribution over paths $x(t)$ within some range $t \in [0,T]$. When this functional form is unknown, a feed-forward neural network becomes a conventional choice.

Fitting a UDE to Data

The key idea behind efficient and scalable training of UDEs is the incorporation of a numerical solver within a differentiable computational graph (Fig. 1a). This setup enables gradient back-propagation through the solution of the differential equation, enabling fitting the UDE parameters to observed data given a suitable cost function. There are two primary strategies for this purpose: i) *discretize-then-optimize*, which involves storing and gradient-backpropgation through all intermediate steps of the solver, providing exact gradients and ii) *optimize-then-discretize*, utilizing the adjoint-method (Chen et al., 2018) to approximate gradients at fixed memory cost. Effectively, this setup enables the training of a UDE-based model using standard loss functions similar to those used in discrete deep learning models. Nonetheless, given the stochastic nature of a UDE of the form in Eq. 1, UDEs are typically trained as generative models with generative modeling objectives such as the evidence lower bound (X. Li, Wong, Chen, & Duvenaud, 2020; Course & Nair, 2023a), adversarial loss (Kidger, Foster, Li, & Lyons, 2021), or, very recently, matching objectives (Bartosh, Vetrov, & Naesseth, 2025).

A Continuum of Models

The UDE formulation naturally encompasses a spectrum of modeling approaches from traditional white-box mechanistic models to contemporary expressive black-box deep learning models (Fig. 1b). Several modeling scenarios can thus be phrased as a UDE training problem. Here we provide some examples of these scenarios, where we use a subscript θ to indicate free parameters.

Differential Equations with Known Unknowns When the structure of the system is known but some parameters are unknown, training a UDE amounts to estimating these parameters from data, balancing interpretability and adaptability while reducing search space (Linial, Ravid, Eytan, & Shalit, 2021; Djeumou, Neary, & Topcu, 2023; Abrevaya et al., 2023).

For instance, the Ornstein-Uhlenbeck (OU) process models a neuron's membrane potential (Laing & Lord, 2009): (Laing & Lord, 2009):

$$\mathrm{d}x = a(m-x)\mathrm{d}t + b\mathrm{d}W\,, \qquad (2)$$

where x denotes the membrane potential and $\theta = (a,m,b)$ are the free parameters. Here the OU process provides the structure of the model dynamics, while the values of the parameters θ are estimated by fitting the UDE on empirical observations.

Differential Equations with Learnable Uncertainty In this setup, the structure of the deterministic dynamics is known or assumed, with unknown parameters, and a function approximator is used to capture intrinsic and/or extrinsic uncertainty about the model.

For example, consider the modern interpretation of a Wilson-Cowan model (Wilson & Cowan, 1972), used to describe the the average firing rates of a group of neurons (Sussillo, 2014). This model can be phrased as a UDE

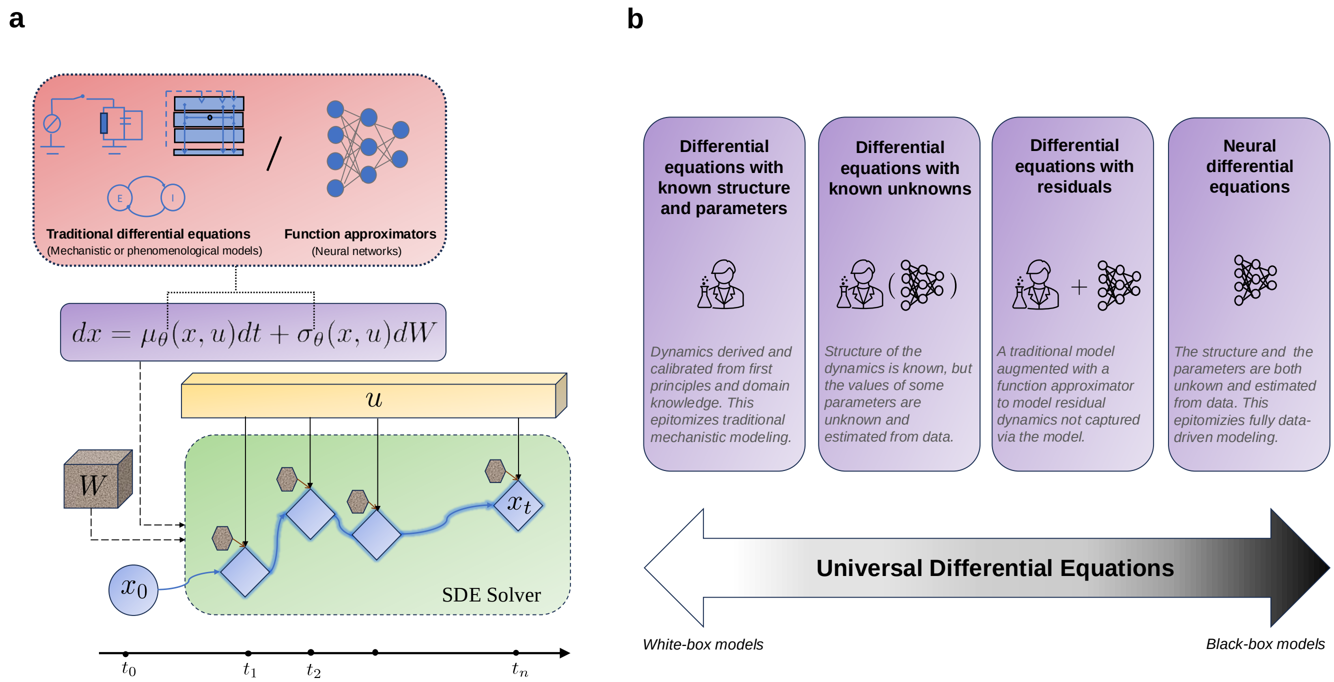

Figure 1: **Universal Differential Equations.** a) A schematic illustration of a universal differential equation. The vector field of the differential equation is defined via either an existing model from the literature, or a differentiable universal approximator (e.g. a neural network) or a combination of both. The numerical solver is an SDE-compatible solver, which takes in the initial condition x_0, a Wiener process generator W, the forcing signal u, the functions defining the vector fields of the SDE μ and σ, along with their parameters θ. The solver computes the solution at time t. The parameters of the differential equation can then be trained either via automatic differentiation or the adjoint method. This setup enables the use of a UDE either as universal function approximator on their own or as a part in a differentiable computational graph. b) The formulation of a UDE encompasses a spectrum of modeling techniques from white-box traditional models to data-driven black box models. This flexibility can foster interoperability between different methodological efforts, and offer a principled approach to balance between data adaptability and scientific rationale in model development.

to capture stochastic dynamics not captured by the original model as follows:

$$\mathrm{d}x = \frac{1}{\tau}(-x + Jr(x) + Bu)\mathrm{d}t + \sigma_\beta(x,u)\mathrm{d}W\,, \tag{3}$$

where x represents the neurons' synaptic currents and $\theta = (\tau, J, B, \beta)$ are the free parameters. The function σ is a differentiable function approximator (a neural network) that captures both how the dynamics respond to external unobserved inputs (extrinsic uncertainty) and how the dynamics evolve subject to uncertainty about the model parameters (intrinsic uncertainty). Hence, θ denotes the parameters of the traditional model and the function approximator. These parameters are jointly learned by fitting the UDE on observations. This setup allows leveraging interpretable mechanistic deterministic models while embracing the complex stochastic nature that arise empirically when modeling complex systems from partial or noisy observations.

Differential Equations with Residuals This approach augments a traditional model by learning unknown components via a function approximator. For instance, the Kuramoto model (Kuramoto, 1975), widely used to study neural synchronization, assumes homogeneous oscillators. A UDE formulation can account for heterogeneity via:

$$\mathrm{d}x = \left(\omega + \frac{K}{N}\sum_{j=1}^{N}\sin(x_j - x) + f_\alpha(x)\right)\mathrm{d}t + \Sigma\mathrm{d}W\,, \tag{4}$$

where $f_\alpha(x)$ is a trainable function that corrects model discrepancies.

Neural Differential Equations At the extreme end of the spectrum, both the structure and parameters of the system are unknown, and UDEs fully rely on neural networks to learn the governing equations (Kidger, 2022; Tzen & Raginsky, 2019b, 2019a; X. Li et al., 2020):

$$\mathrm{d}x = \mu_\alpha(x,u)\mathrm{d}t + \sigma_\beta(x,u)\mathrm{d}W\,, \tag{5}$$

where μ and σ are neural networks with parameters $\theta = (\alpha, \beta)$. This equation can be viewed as a stochastic, continuous-time generalization of discrete-time deep recurrent neural networks prevalent in contemporary machine learning research (Kidger, 2022; Tzen & Raginsky, 2019b, 2019a; X. Li et al., 2020).

Towards Informed Stochastic Models

The presented UDE configurations fill a spectrum between white-box and black-box models under a unified formulation. Intuitively, as one progresses from white-box models towards black-box models, the reliance on empirical data for model identification increases correspondingly, inversely proportional to the number of presupposed assumptions about the underlying dynamics (the more correct the model, the less data needed, and vice versa). In practice, it should be expected that a certain degree of knowledge or hypothesis about the studied system is available. This knowledge should not be constrained to the structure of the dynamics, but could cover all aspects of the computational model (e.g., dimensionality, information about the stimulus or observation modality, scale of noise, expected dynamics, etc.). UDEs simply serve as a universal tool for evaluating this knowledge, or augmenting them to develop scalable models that can be used in downstream applications.

Crucially, UDEs conceptualize neural processes as continuous-time stochastic processes. This perspective can bring computational models closer to the complex nature of neural processes. This is imperative when modeling neural dynamics, where stochasticity can be traced from the molecular level, with stochastic behaviors in ion channels and synaptic transmission (Hille, 1978; Sakmann, 2013), to the cellular scale where neurons demonstrate unpredictable firing patterns (Tuckwell, 1988). Importantly, stochasticity is not confined to the micro-scale as it escalates to the level of neural populations, where the effects of noise and randomness are not merely incidental but play a crucial role in the functioning and organization of neural systems (Rolls & Deco, 2010; Faisal, Selen, & Wolpert, 2008). The following section delves into leveraging UDEs to develop differentiable, informed, probabilistic models for neural system identification.

Neural System Identification

Consider a neural system whose state evolves as a continuous-time stochastic process $\{x(t)\colon 0 \leq t \leq T\}$ that is potentially modulated by exogenous input $u(t)$. In practice, we do not observe x or u directly. Instead, we record discrete-time stimuli $v_n = v(t_n)$ and neural or behavioral outputs $y_n = y(t_n)$ at timepoints $t_1, \ldots, t_N$ with $0 \leq t_n \leq T$. We use $v_{1:N}$ and $y_{1:N}$ to denote these observations across timepoints. Let $\tau = (v_1, y_1, \ldots, v_N, y_N)$ denote a trajectory of stimuli and responses and assume that we have access to a dataset $\mathcal{D} = \{\tau^1, \ldots, \tau^K\}$ consisting of k such trajectories. The overarching goal of neural system identification is to estimate the latent neural states $x(t)$ from the data $\mathcal{D}$.

We propose to model the underlying stochastic process x as the solution of a latent UDE, and frame the problem of system identification as a posterior inference problem of the distribution $p(x|y,v)$, which we tackle via variational inference. Accurate resolution of this problem yields multiple benefits. First, it allows inference of the latent states of the system. Second, it allows reconstructing and predicting the system's behavior under various conditions. Third, it provides an expressive probabilistic modeling framework that quantifies uncertainty and incorporates prior knowledge, facilitating robust hypothesis generation and testing. Recent advances in variational inference for stochastic differential equations (SDEs) (X. Li et al., 2020; Tzen & Raginsky, 2019a, 2019b; Ryder, Golightly, McGough, & Prangle, 2018; Course & Nair, 2023b) make this formulation tractable and appealing for complex neural data, particularly in naturalistic settings. Below, we outline a flexible architecture for building such models along with the training objective. A general overview of the architecture is illustrated in Figure 2.

Model Architecture

Stimulus Encoder When the stimulus v is high-dimensional (e.g., images, video, text), it is computationally expensive to integrate it directly into a continuous-time dynamics model. Instead, we map v to a (potentially lower-dimensional) continuous-time representation u. Formally,

$$u(t) = \pi \{\alpha_\theta (v_\tau)\}_{\tau \leq t} \,, \tag{6}$$

where $\alpha_\theta\colon \mathbb{R}^{d_v} \to \mathbb{R}^{d_u}$ is an encoding function (trainable or pre-trained) and $\pi\colon \mathbb{R}^{d_u} \times [0,T] \to \mathbb{R}^{d_u}$ is an interpolation scheme(e.g., piecewise-constant, spline-based) chosen to match the temporal resolution requirements. In simpler cases (low-dimensional v), α may be the identity.

Recognition Model The objective of this module is to approximate the posterior of the initial condition $p(x_0|y,c)$ of the system. To accomplish this, we define a mapping function that uses the observed data to infer x_0.

$$p(x_0|y,c) = \zeta_\phi (y_{c:0}, u_{c:0}) \,, \tag{7}$$

where ζ is a sequential model with parameters ϕ and $c \in [0,N]$ denotes the end of the observation interval used for estimating the initial condition. Note that notations $y_{c:0}, u_{c:0}$ indicate that the intervals are reversed in time. The choice of c should depend on the nature of the dynamics or context of the application. For example, in stationary settings, it might suffice to have $c \ll N$. It is also important to consider, which phenomena is under study. In most cognitive experiments, pre-task recordings exist and can be utilized for this purpose. In general it is important to ensure that ζ is not overly parameterized to avoid encoding future information about the dynamics as recommended by (Massaroli, Poli, Park, Yamashita, & Asama, 2020).

Process Model The goal of this module is to learn the distribution of the latent stochastic process $p(x|u)$. This is done by employing a UDE to model the temporal evolution of the initial

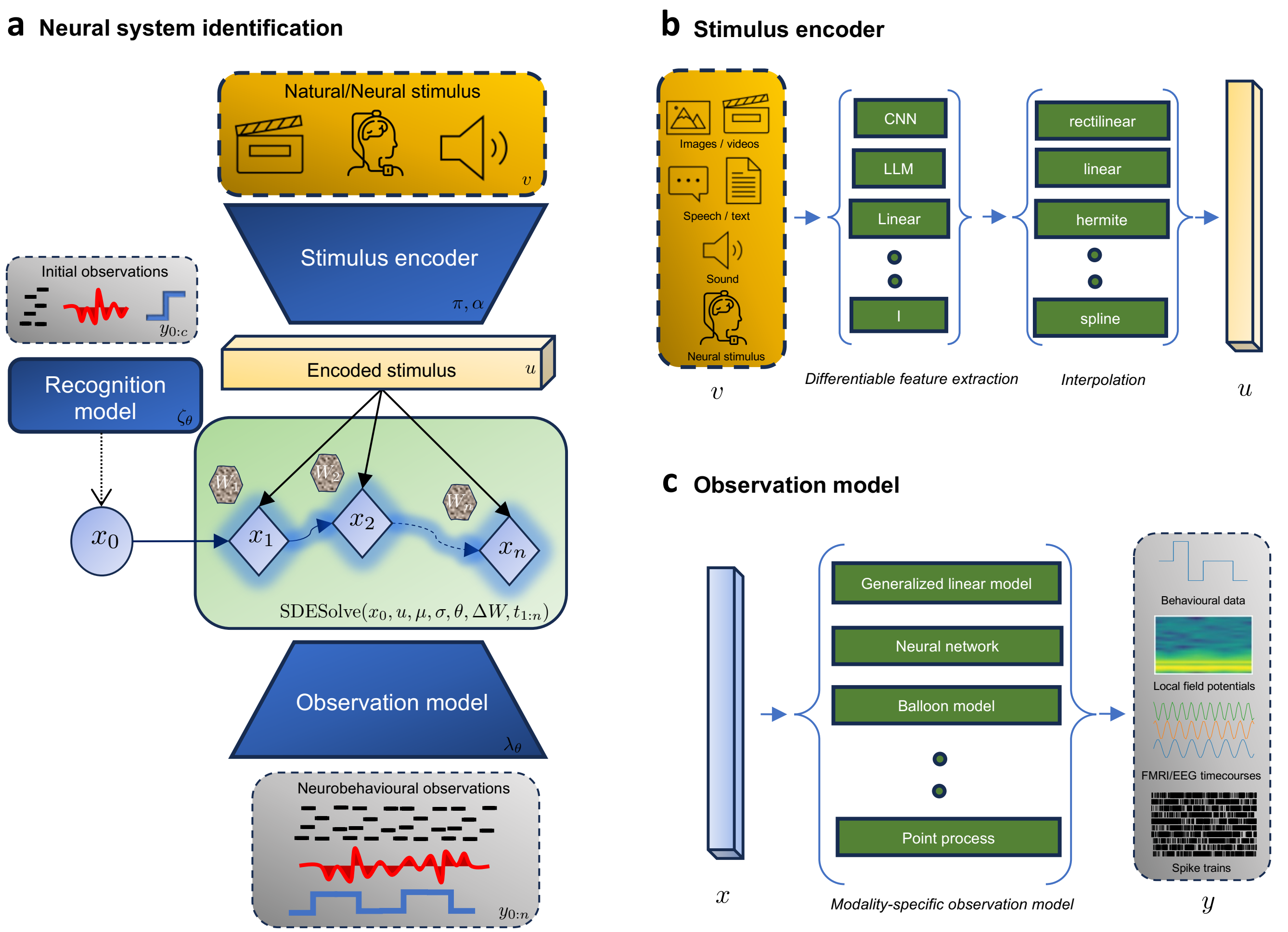

Figure 2: **Framework for neural system identification** a) Shows the forward pass (generative mode) during the encoding of a (high-dimensional) stimulus v into neurobehavioral observations y. This is done through a fully differentiable graph, which consists of i) a stimulus encoder to encode the stimulus into a lower dimensional continuous representation, ii) a recognition model to infer the hidden initial state x_0, iii) a latent dynamics model to model the temporal evolution of the dynamics, and iv) an observation model to map the latent states into observations. b) Illustrates the formulation of the informed stimulus encoder which is tasked with learning a lower dimensional continuous representation u from the discrete (high-dimensional) stimulus signal v. c) Illustrates examples of modality-specific observation models to map the latent process into neurobehavioral measurements.

state x_0, subject to external control u, and Brownian motion W. This is expressed as before as

$$dx = \mu_\theta(x,u)dt + \sigma_\theta(x,u)dW\,. \tag{8}$$

The design of the UDE should be dependent on domain knowledge about the system in question and the downstream application of the model.

Observation Model Observation models, also known as measurement or emission models, define the probabilistic relationship between the latent states of a system and the observed data. The observation model is formalized as follows:

$$p(y|x) = \lambda_\theta(x(t), \varepsilon(t))\,, \tag{9}$$

where $\lambda_\theta : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y}$ is the observation function and $\varepsilon(t)$ is observation noise.Biophysical constraints, measurement noise, and interpretability demands inform the choice of λ. For instance, Poisson or other point-process models may be used for spike trains (Heeger et al., 2000), nonlinear Gaussian models for local field potentials (Herreras, 2016), and specialized emission models for fMRI to account for hemodynamic responses (Friston, Mechelli, Turner, & Price, 2000).

Training Objective

To perform variational inference in the context of latent continuous stochastic processes, we need to define a reasonable and tractable family of path distributions for the approximate posterior. Following the approach of (X. Li et al., 2020), we may employ an SDE to represent this. Specifically, here we can define our process model that we wish to learn in Eq. 8 as our prior and the approximate posterior as another black-box UDE. Note that the term prior here refers to our main generative UDE, and the approximate posterior is an auxiliary UDE that is used only during training. This approximate posterior UDEs can be written as:

$$d\tilde{x} = \mu_\phi(\tilde{x}, y, u)dt + \sigma_\theta(\tilde{x}, u)dW\,, \qquad (10)$$

where ϕ are the variational parameters. Note that both the prior and the approximate posterior share the same diffusion σ_θ, a decision which guarantees that the Kullback-Leibler (KL) divergence between the two probability measures they induce is finite (under some mild conditions) (X. Li et al., 2020). This KL divergence can be defined using Girsanov's theorem (Girsanov, 1960) as

$$D_{\mathrm{KL}}(Q\,||\,P) = \mathbb{E}_{\tilde{x}}\left[\int_0^\tau \frac{1}{2}\,\|\Delta(\tilde{x}, y, u)\|^2\,dt\right] \qquad (11)$$

where $\Delta(\tilde{x}, y, u) = \sigma_\theta(\tilde{x}, u)^{-1}\left(\mu_\phi(\tilde{x}, y, u) - \mu_\theta(\tilde{x}, u)\right)$ with P and Q denoting the measures induced by the prior and approximate posterior, respectively. Using this we can define an evidence lower bound (ELBO) on the conditional marginal likelihood of the observations:

$$\text{ELBO}(\theta, \phi; v, y) = \mathbb{E}_{\tilde{x}}\left[\log p(y|\tilde{x})\right] - D_{KL}(Q||P) \qquad (12)$$

Maximizing this ELBO via stochastic gradient methods jointly updates θ (generative model) and ϕ (inference model). This approach provides a powerful and flexible framework for neural system identification: it marries domain-driven modeling choices with scalable data-driven training, enabling modeling latent, nonlinear, and stochastic neural processes.

Opportunities in Neuroscience

Universal Differential Equations (UDEs), trained for neural system identification, offer a powerful alternative to existing data-driven models in neuroscience. We highlight four key applications, emphasizing the advantages of UDE integration.

Explaining Cognition and Behavior

A core challenge in neuroscience is linking brain activity to cognitive and behavioral functions. Latent variable models (LVMs) have revealed that these functions arise from coordinated, low-dimensional neural population activity governed by latent states (Mante, Sussillo, Shenoy, & Newsome, 2013; Churchland et al., 2012; Elsayed & Cunningham, 2017). However, LVMs are often limited to simple, stereotyped behaviors and face challenges in defining input-output structures, interneuronal influences, and incorporating anatomical constraints (Urai et al., 2022; Hurwitz et al., 2021; Vyas et al., 2020). UDEs can address these limitations by:

- **Structured and Multi-Scale Dynamics:** UDEs can be used to structure the latent space, leveraging the rich history of differential equations in neuroscience to improve model expressiveness and interpretability. Additionally, UDEs allows building tractable multi-scale models, combining mechanistic models at one scale (e.g., single-neuron dynamics) with learned dynamics at another (e.g., inter-population interactions) (Fig. 3a).

- **Dynamical Systems Analysis:** Analyzing the trained UDE's vector field (fixed points, limit cycles, etc.) can provide insights into the computational mechanisms underlying cognition and behavior, similar to analyses of RNN models (Sussillo & Barak, 2013; Vyas et al., 2020).

- **Modeling Complex Noise:** UDEs can capture stochasticity arising from unobserved interactions and intrinsic properties by using a neural network in the diffusion term. This surpasses common simplified noise models (e.g., additive Gaussian noise) (Linderman et al., 2017; Laing & Lord, 2009; Pandarinath et al., 2018), which are often inadequate for higher-order brain functions where noise plays a crucial role in neural coding and behavior (Rolls & Deco, 2010; Faisal et al., 2008).

- **Model Comparison:** UDEs facilitate model comparison by allowing researchers to configure the prior UDE to reflect specific theoretical hypotheses about neural processes. The likelihood of observed data can then be used to evaluate the relative effectiveness of different dynamical models, providing a more scalable approach than traditional Bayesian model comparison (Grimmer, 2011).

Neural Control

The intersection of neuroscience and control theory is driven by applications like brain-computer interfaces (BCIs) and neurostimulation (Yang, Connolly, & Shanechi, 2018; Acharya et al., 2022). Closed-loop control is crucial for reliability, safety, and efficiency (Ramirez-Zamora et al., 2018; Särkkä & Solin, 2019), and model-based approaches enable in-silico validation and causal analysis (Rueckauer & van Gerven, 2023; Imbens & Rubin, 2015). The brain's complexity, however, presents unique challenges for model-based control, including high dimensionality, stochasticity, and limited data availability (Schiff, 2011). UDEs can offer solutions by:

- **Balancing Expressiveness and Data Needs:** UDEs integrate mechanistic and data-driven models, providing a compromise between fully mechanistic models (low data needs, potentially lower accuracy) and purely data-driven models (high data needs, potentially higher accuracy).

- **Uncertainty Quantification:** Latent UDEs can estimate and disentangle various uncertainties (epistemic and aleatoric), critical for safe and reliable control.

- **Real-time Adaptability:** UDEs combined with adaptive numerical solvers, provide a prediction accuracy/computation trade-off essential for online application. Their continuous-time nature allows for handling irregularly-sampled data

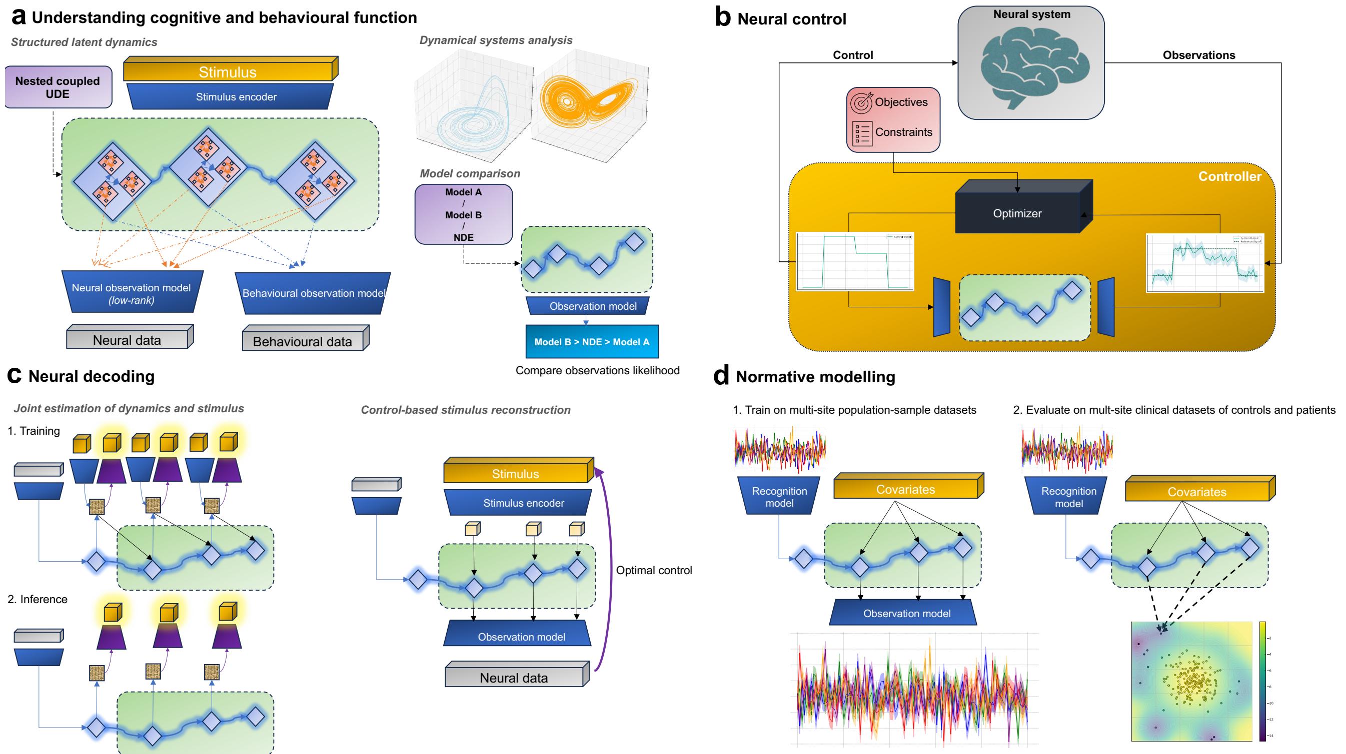

Figure 3: **UDE-based models across different applications in neuroscience a:** UDEs can be used to represent the underlying latent dynamical system to understand how neural dynamics give rise to computations and ultimately behaviour. **b:** UDEs can be used for model-based closed loop control of neural systems. **c:** UDE models trained for neural encoding can also be leveraged for neural decoding of the stimulus from neural observations. **d:** UDE models can be employed for capturing population-average neural dynamics utilized for patients stratification in a normative modeling framework.

and development of continuous-time control strategies (Lewis, Vrabie, & Syrmos, 2012). Figure 3b demonstrates a model predictive control approach of applying UDEs for neural control.

Neural Decoding

Neural decoding aims to predict external stimuli from recorded brain activity, with applications ranging from communication interfaces to understanding brain-stimulus interactions (Rieke, Warland, Van Steveninck, & Bialek, 1999; Horikawa, Tamaki, Miyawaki, & Kamitani, 2013; Anumanchipalli, Chartier, & Chang, 2019; Seeliger, Güçlü, Ambrogioni, Güçlütürk, & van Gerven, 2018; Metzger et al., 2023). Current approaches rely on regression or (approximate) Bayesian methods (Warland, Reinagel, & Meister, 1997; Horikawa et al., 2013; Anumanchipalli et al., 2019; Pillow, Ahmadian, & Paninski, 2011). UDEs trained via variational inference can be adapted for the latter via:

- **Input-Output Inversion:** A straightforward approach is to invert the input and output described during systems identification.
- **Unified Encoding-Decoding:** Another approach is to use the same model for both encoding and decoding (Paninski, Pillow, & Lewi, 2007; Kriegeskorte & Douglas, 2019). One route would be to extend the variational inference setup introduced to estimate the posterior distribution $p(v \mid y)$. Another relevant approach would be to frame stimulus inference as an optimal control problem similar to (Schimel, Kao, Jensen, & Hennequin, 2021). Figure 3c demonstrates both examples.

Normative Modeling

Normative modeling characterizes brain variation and assesses individual deviations, proving valuable in clinical and developmental neuroscience, particularly in psychiatry (Marquand, Wolfers, Mennes, Buitelaar, & Beckmann, 2016; Insel et al., 2010; Bethlehem et al., 2022). Current normative models often focus on static measures, while dynamic models of functional neuroimaging data remain challenging due to high dimensionality, inter-subject variability, and noise (Marquand et al., 2019; Rutherford et al., 2022). UDEs offer a solution by capturing individual variability through the diffusion term, while the drift term reflects the population average that can be personalized by including covariates as fixed arguments. An example is demonstrated in Figure 3d.

Discussion

There is a growing consensus that solutions to complex science and engineering problems require novel methodologies that are able to integrate traditional mechanistic modeling approaches and domain expertise with current machine learning and optimization techniques (Raissi, Perdikaris, & Karniadakis, 2019; Alber et al., 2019; Willard, Jia, Xu, Steinbach, & Kumar, 2022; Cuomo et al., 2022; AlQuraishi & Sorger, 2021). In this vein, we outlined the potential of universal differential equations as a framework to facilitate this integration in neuroscience. Our endeavor is centered around establishing a common modeling language across the field that can unify existing efforts in alignment with the current paradigm shift happening across scientific disciplines where neural networks and dynamical systems are viewed as two sides of the same coin. At first glance, these insights might not be entirely novel or exciting for a field such as computational neuroscience, which has been successfully applying RNNs and their variants as discretization of ODE models for over a decade. What is exciting however, is that this offers a fresh perspective and an opportunity to connect models from different scales of organization and levels of abstraction in neuroscience under one potent framework that aligns with ongoing developments in the fields of generative modeling and scientific computing.

Acknowledgements

This publication is part of the project Dutch Brain Interface Initiative (DBI2) with project number 024.005.022 of the research programme Gravitation, which is financed by the Dutch Ministry of Education, Culture and Science (OCW) via the Dutch Research Council (NWO).

References

Abrevaya, G., Ramezanian-Panahi, M., Gagnon-Audet, J.-C., Polosecki, P., Rish, I., Dawson, S. P., . . . Dumas, G. (2023). Effective Latent Differential Equation Models via Attention and Multiple Shooting. In *The symbiosis of deep learning and differential equations iii.*

Acharya, G., Ruf, S. F., & Nozari, E. (2022). Brain modeling for control: A review. *Frontiers in Control Engineering*, *3*, 1046764.

Alber, M., et al. (2019). Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digital Medicine*, *2*(1), 115.

AlQuraishi, M., & Sorger, P. K. (2021). Differentiable biology: using deep learning for biophysics-based and data-driven modeling of molecular mechanisms. *Nature methods*, *18*(10), 1169–1180.

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, *568*(7753), 493–498.

Bartosh, G., Vetrov, D., & Naesseth, C. A. (2025). Sde matching: Scalable and simulation-free training of latent stochastic differential equations. *arXiv preprint arXiv:2502.02472*.

Bethlehem, R. A., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., . . . others (2022). Brain charts for the human lifespan. *Nature*, *604*(7906), 525–533.

Bliss, T. V., & Collingridge, G. L. (1993). A Synaptic Model of Memory: Long-Term Potentiation in the Hippocampus. *Nature*, *361*(6407), 31–39.

Breakspear, M. (2017). Dynamic Models of Large-Scale Brain Activity. *Nature Neuroscience*, *20*(3), 340–352.

Breakspear, M., Heitmann, S., & Daffertshofer, A. (2010). Generative models of cortical oscillations: neurobiological implications of the Kuramoto model. *Frontiers in Human Neuroscience*, *4*, 190.

Brunton, B. W., & Beyeler, M. (2019). Data-Driven Models in Human Neuroscience and Neuroengineering. *Current Opinion in Neurobiology*, *58*, 21–29.

Brunton, S. L., & Kutz, J. N. (2019). *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press.

Buzsaki, G., & Draguhn, A. (2004). Neuronal Oscillations in Cortical Networks. *Science*, *304*(5679), 1926–1929.

Bystritsky, A., Nierenberg, A., Feusner, J., & Rabinovich, M. (2012). Computational non-linear dynamical psychiatry: A new methodological paradigm for diagnosis and course of illness. *Journal of Psychiatric Research*, *46*(4), 428–435.

Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, *31*.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, *487*(7405), 51–56.

Course, K., & Nair, P. (2023a). Amortized reparametrization: efficient and scalable variational inference for latent sdes. *Advances in Neural Information Processing Systems*, *36*, 78296–78318.

Course, K., & Nair, P. B. (2023b). State Estimation of a Physical System with Unknown Governing Equations. *Nature*, *622*(7982), 261–267.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, *92*(3), 88.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303–314.

Deco, G., & Jirsa, V. K. (2012). Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors. *Journal of Neuroscience*, *32*(10), 3366–3375.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. (2008). The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. *PLoS Computational Biology*, *4*(8), e1000092.

Deisseroth, K. (2015). Optogenetics: 10 Years of Microbial Opsins in Neuroscience. *Nature*

Neuroscience, *18*, 1213-1225. Retrieved from https://www.nature.com/articles/nn.4091 doi: 10.1038/nn.4091

Djeumou, F., Neary, C., & Topcu, U. (2023). How to learn and generalize from three minutes of data: Physics-constrained and uncertainty-aware neural stochastic differential equations. *arXiv preprint arXiv:2306.06335*.

Durstewitz, D., Koppe, G., & Thurm, M. I. (2023). Reconstructing computational system dynamics from neural data with recurrent neural networks. *Nature Reviews Neuroscience*, 1–18.

Elsayed, G. F., & Cunningham, J. P. (2017). Structure in neural population recordings: An expected byproduct of simpler phenomena? *Nature Neuroscience*, *20*(9), 1310–1318.

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature reviews neuroscience*, *9*(4), 292–303.

Favela, L. H. (2021). The Dynamical Renaissance in Neuroscience. *Synthese*, *199*(1-2), 2103–2127.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47.

FitzHugh, R. (1961). Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical Journal*, *1*(6), 445–466.

Friston, K. J. (2009). The Free-Energy Principle: A Rough Guide to the Brain? *Trends in Cognitive Sciences*, *13*, 293-301.

Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000). Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. *NeuroImage*, *12*(4), 466–477.

Geenjaar, E., Kim, D., Ohib, R., Duda, M., Kashyap, A., Plis, S., & Calhoun, V. (2023). Learning low-dimensional dynamics from whole-brain data improves task capture. *arXiv preprint arXiv:2305.14369*.

Genkin, M., Hughes, O., & Engel, T. A. (2021). Learning non-stationary langevin dynamics from stochastic observations of latent trajectories. *Nature communications*, *12*(1), 5986.

Genkin, M., Shenoy, K. V., Chandrasekaran, C., & Engel, T. A. (2023). The dynamics and geometry of choice in premotor cortex. *BioRxiv*.

Girsanov, I. V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, *5*(3), 285–301.

Grimmer, J. (2011). An introduction to bayesian inference via variational approximations. *Political Analysis*, *19*(1), 32–47.

Han, Y., Poggio, T. A., & Cheung, B. (2023). System Identification of Neural Systems: If We Got It Right, Would We Know? In *International conference on machine learning* (pp. 12430–12444).

Hawkins, J. (2021). *A Thousand Brains: A New Theory of Intelligence*. Basic Books.

Heeger, D., et al. (2000). Poisson model of spike generation. *Handout, University of Standford*, *5*(1-13), 76.

Herreras, O. (2016). Local field potentials: Myths and misunderstandings. *Frontiers in Neural Circuits*, *10*, 101.

Hille, B. (1978). Ionic channels in excitable membranes. current problems and biophysical approaches. *Biophysical journal*, *22*(2), 283–294.

Hodgkin, A. L., & Huxley, A. F. (1952). A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve. *The Journal of Physiology*, *117*(4), 500.

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, *340*(6132), 639–642.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Hurwitz, C., Kudryashova, N., Onken, A., & Hennig, M. H. (2021). Building population models for large-scale neural recordings: Opportunities and pitfalls. *Current Opinion in Neurobiology*, *70*, 64–73.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... Wang, P. (2010). *Research domain criteria (rdoc): toward a new classification framework for research on mental disorders* (Vol. 167) (No. 7). Am Psychiatric Assoc.

Izhikevich, E. M. (2003). Simple Model of Spiking Neurons. *IEEE Transactions on Neural Networks*, *14*(6), 1569–1572.

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. MIT press.

Izhikevich, E. M., & Edelman, G. M. (2008). Large-Scale Model of Mammalian Thalamocortical Systems. *Proceedings of the National Academy of Sciences*, *105*(9), 3593–3598.

Jia, J., & Benson, A. R. (2019). Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems*, *32*.

Kao, J. C. (2019). Considerations in using recurrent neural networks to probe neural dynamics. *Journal of Neurophysiology*, *122*(6), 2504–2521.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440.

Khona, M., & Fiete, I. R. (2022). Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, *23*(12), 744–766.

Kidger, P. (2022). On neural differential equations. *arXiv preprint arXiv:2202.02435*.

Kidger, P., Foster, J., Li, X., & Lyons, T. J. (2021). Neural sdes as infinite-dimensional gans. In *International conference on machine learning* (pp. 5453–5463).

Kidger, P., Morrill, J., Foster, J., & Lyons, T. (2020). Neural controlled differential equations for irregular time series.

Advances in Neural Information Processing Systems, *33*, 6696–6707.

Kim, S. S., Rouault, H., Druckmann, S., & Jayaraman, V. (2017). Ring Attractor Dynamics in the Drosophila Central Brain. *Science*, *853*, 849-853.

Kim, T. D., Luo, T. Z., Pillow, J. W., & Brody, C. D. (2021). Inferring latent dynamics underlying neural population activity via neural differential equations. In *International conference on machine learning* (pp. 5551–5561).

Kriegeskorte, N., & Douglas, P. K. (2019). Interpreting encoding and decoding models. *Current Opinion in Neurobiology*, *55*, 167–179.

Kuramoto, Y. (1975). Self-entrainment of a population of coupled non-linear oscillators. In *International symposium on mathematical problems in theoretical physics: January 23–29, 1975, kyoto university, kyoto/japan* (pp. 420–422).

Lai, Z., Mylonas, C., Nagarajaiah, S., & Chatzi, E. (2021). Structural identification with physics-informed neural ordinary differential equations. *Journal of Sound and Vibration*, *508*, 116196.

Laing, C., & Lord, G. J. (2009). *Stochastic methods in neuroscience*. OUP Oxford.

Landhuis, E. (2017). Big Brain, Big Data. *Nature*, *541*.

Lewis, F. L., Vrabie, D., & Syrmos, V. L. (2012). *Optimal control*. John Wiley & Sons.

Li, X., Wong, T.-K. L., Chen, R. T., & Duvenaud, D. (2020). Scalable gradients for stochastic differential equations. In *International conference on artificial intelligence and statistics* (pp. 3870–3882).

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.

Linderman, S., Johnson, M., Miller, A., Adams, R., Blei, D., & Paninski, L. (2017). Bayesian learning and inference in recurrent switching linear dynamical systems. In *Artificial intelligence and statistics* (pp. 914–922).

Linial, O., Ravid, N., Eytan, D., & Shalit, U. (2021). Generative ode modeling with known unknowns. In *Proceedings of the conference on health, inference, and learning* (pp. 79–94).

Lozano, A. M., Lipsman, N., Bergman, H., Brown, P., Chabardes, S., Chang, J. W., ... others (2019). Deep Brain Stimulation: Current Challenges and Future Directions. *Nature Reviews Neurology*, *15*(3), 148–160.

Machado, T. A., Kauvar, I. V., & Deisseroth, K. (2022). Multiregion Neuronal Activity: The Forest and the Trees. *Nature Reviews Neuroscience*, *23*(11), 683–704.

Maheswaranathan, N., Williams, A., Golub, M., Ganguli, S., & Sussillo, D. (2019). Universality and individuality in neural dynamics across large populations of recurrent networks. *Advances in Neural Information Processing Systems*, *32*.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84.

Marquand, A. F., Kia, S. M., Zabihi, M., Wolfers, T., Buitelaar, J. K., & Beckmann, C. F. (2019). Conceptualizing mental disorders as deviations from normative functioning. *Molecular psychiatry*, *24*(10), 1415–1424.

Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J., & Beckmann, C. F. (2016). Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. *Biological psychiatry: cognitive neuroscience and neuroimaging*, *1*(5), 433–447.

Massaroli, S., Poli, M., Park, J., Yamashita, A., & Asama, H. (2020). Dissecting neural odes. *Advances in Neural Information Processing Systems*, *33*, 3952–3963.

Matthews, P. C., Mirollo, R. E., & Strogatz, S. H. (1991). Dynamics of a large system of coupled nonlinear oscillators. *Physica D: Nonlinear Phenomena*, *52*(2-3), 293–331.

Metzger, S. L., Littlejohn, K. T., Silva, A. B., Moses, D. A., Seaton, M. P., Wang, R., ... others (2023). A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, 1–10.

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(1), 167–202.

Morrill, J., Salvi, C., Kidger, P., & Foster, J. (2021). Neural rough differential equations for long time series. In *International conference on machine learning* (pp. 7829–7838).

Oksendal, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., ... others (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, *15*(10), 805–815.

Paninski, L., Pillow, J., & Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Progress in Brain Research*, *165*, 493–507.

Pei, F., Ye, J., Zoltowski, D., Wu, A., Chowdhury, R. H., Sohn, H., ... others (2021). Neural latents benchmark'21: evaluating latent variable models of neural population activity. *arXiv preprint arXiv:2109.04463*.

Pillow, J. W., Ahmadian, Y., & Paninski, L. (2011). Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Computation*, *23*(1), 1–45.

Poli, M., Massaroli, S., Park, J., Yamashita, A., Asama, H., & Park, J. (2019). Graph neural ordinary differential equations. *arXiv preprint arXiv:1911.07532*.

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., ... Edelman, A. (2020). Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707.

Ramezanian-Panahi, M., Abrevaya, G., Gagnon-Audet, J.-C., Voleti, V., Rish, I., & Dumas, G. (2022). Generative models of brain dynamics. *Frontiers in Artificial Intelligence*, *5*, 807406.

Ramirez-Zamora, A., Giordano, J. J., Gunduz, A., Brown, P., Sanchez, J. C., Foote, K. D., . . . others (2018). Evolving applications, technological challenges and future opportunities in neuromodulation: Proceedings of the fifth annual deep brain stimulation think tank. *Frontiers in Neuroscience*, 734.

Rieke, F., Warland, D., Van Steveninck, R. d. R., & Bialek, W. (1999). *Spikes: Exploring the neural code*. MIT Press.

Roberts, J. A., Friston, K. J., & Breakspear, M. (2017). Clinical applications of stochastic dynamic models of the brain, part i: A primer. *Biological psychiatry: cognitive neuroscience and neuroimaging*, *2*(3), 216–224.

Rolls, E. T., & Deco, G. (2010). *The noisy brain: stochastic dynamics as a principle of brain function*. Oxford academic.

Rueckauer, B., & van Gerven, M. (2023). An in-silico framework for modeling optimal control of neural systems. *Frontiers in Neuroscience*, *17*, 1141884.

Rutherford, S., Kia, S. M., Wolfers, T., Fraza, C., Zabihi, M., Dinga, R., . . . others (2022). The normative modeling framework for computational psychiatry. *Nature protocols*, *17*(7), 1711–1734.

Ryder, T., Golightly, A., McGough, A. S., & Prangle, D. (2018). Black-Box Variational Inference for Stochastic Differential Equations. In *International conference on machine learning* (pp. 4423–4432).

Sakmann, B. (2013). *Single-channel recording*. Springer Science & Business Media.

Särkkä, S., & Solin, A. (2019). *Applied stochastic differential equations* (Vol. 10). Cambridge University Press.

Schaeffer, R., Khona, M., & Fiete, I. (2022). No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit. *Advances in Neural Information Processing Systems*, *35*, 16052–16067.

Schiff, S. J. (2011). *Neural control engineering: The emerging intersection between control theory and neuroscience*. MIT Press.

Schimel, M., Kao, T.-C., Jensen, K. T., & Hennequin, G. (2021). ilqr-vae: control-based learning of input-driven dynamics with applications to neural data. *bioRxiv*, 2021–10.

Sedler, A. R., Versteeg, C., & Pandarinath, C. (2022). Expressive architectures enhance interpretability of dynamics-based neural population models. *arXiv preprint arXiv:2212.03771*.

Seeliger, K., Güçlü, U., Ambrogioni, L., Güçlütürk, Y., & van Gerven, M. A. (2018). Generative adversarial networks for reconstructing natural images from brain activity. *NeuroImage*, *181*, 775–785.

Shen, C., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, *4*(8), 552–567.

Steinmetz, N. A., et al. (2021). Neuropixels 2.0: A Miniaturized High-Density Probe for Stable, Long-Term Brain Recordings. *Science*, *372*, eabf4588. doi: 10.1126/science.abf4588

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, *576*(7786), 266–273.

Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology*, *25*, 156–163.

Sussillo, D., & Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, *25*(3), 626–649.

Tang, E., & Bassett, D. S. (2018). Colloquium: Control of dynamics in brain networks. *Reviews of Modern Physics*, *90*(3), 031003.

Tuckwell, H. C. (1988). *Introduction to theoretical neurobiology: linear cable theory and dendritic structure* (Vol. 1). Cambridge University Press.

Tzen, B., & Raginsky, M. (2019a). Neural Stochastic Differential Equations: Deep Latent Gaussian Models in the Diffusion Limit. *arXiv preprint arXiv:1905.09883*.

Tzen, B., & Raginsky, M. (2019b). Theoretical Guarantees for Sampling and Inference in Generative Models with Latent Diffusions. In *Conference on learning theory* (pp. 3084–3114).

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

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and brain sciences*, *21*(5), 615–628.

Versteeg, C., Sedler, A. R., McCart, J. D., & Pandarinath, C. (2023). Expressive dynamics models with nonlinear injective readouts enable reliable recovery of latent features from neural activity. *arXiv preprint arXiv:2309.06402*.

Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. *Annual Review of Neuroscience*, *43*, 249–275.

Wang, H., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60.

Warland, D. K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, *78*(5), 2336–2350.

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. *ACM Computing Surveys*, *55*(4), 1–37.

Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., . . . others (2023). A high-performance speech neuroprosthesis. *Nature*, 1–6.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*(1), 1–24.

Yang, Y., Connolly, A. T., & Shanechi, M. M. (2018). A control-theoretic system identification framework and a real-time closed-loop clinical simulation testbed for electrical brain stimulation. *Journal of Neural Engineering*, *15*(6), 066007.

Appendix

Case Study: Predicting Stimulus-Invoked Neural and Behavioral Responses during Visual Decision-Making using the UDE Framework

To illustrate the practical benefits and flexibility of the Universal Differential Equation (UDE) framework outlined in this paper, we apply it to the challenging problem of modeling multi-regional neural activity and behavioral responses during a complex cognitive task. This case study demonstrates how the UDE approach can integrate prior knowledge, handle system stochasticity, and provide interpretable insights while achieving strong predictive performance.

Objective We aimed to develop a generative model capable of predicting both multi-regional neural spiking activity and continuous behavioral output (wheel movement) in mice performing a visual decision-making task. The model takes only the presented visual stimuli and task timing cues as input. This serves as a concrete application for evaluating different modeling choices within the UDE framework, particularly the trade-offs between purely data-driven and knowledge-informed approaches.

Data We utilized publicly available data from (Steinmetz, Zatka-Haas, Carandini, & Harris, 2019), involving mice trained on a two-alternative contrast discrimination task. Mice indicate the side with the higher contrast stimulus by turning a wheel after a go cue. The dataset includes simultaneous recordings of spiking activity from hundreds of neurons across multiple brain regions (including visual cortex, motor cortex, thalamus, and striatum) via Neuropixels probes, as well as the corresponding wheel velocity. Model inputs consist of the left/right stimulus contrast levels and the timing of stimulus presentation and the go cue. We focus our analysis on data from a single session from one mouse, comprising 327 trials and recordings from 879 neurons across 5 distinct brain regions.

Model Implementation using the UDE Framework We employed the neural system identification method described in the main text. The core idea is to model the latent neural state's evolution as the solution to a UDE, specifically a latent Stochastic Differential Equation (SDE), and infer its parameters and states using variational inference.

- **Stimulus Encoder (α_θ, π):** In this setting, low-dimensional task inputs are already provided as categorical variables describing the contrast level for each side. We applied one-hot encoding to these inputs and concatenated the stimulus presentation timing and the timing of the go cue. We then applied zero-order interpolation to obtain a continuous-time input signal u.
- **Recognition Model (ζ_φ):** We used a 2-layer bi-directional LSTM as the recognition model ζ_φ. It processes an initial segment of the observed neural and behavioral data along with the corresponding encoded stimulus ($u_{c:0}$) to estimate the parameters (mean and variance) of the Gaussian approximate posterior distribution for the initial latent state x_0, approximating $p(x_0 \mid y, u)$.
- **Process Model (Latent UDE):** This module captures the latent neural dynamics $p(x \mid u)$ via an SDE (Eq. 8). To explore the UDE spectrum (Fig. 1b), we implemented and compared two distinct variants:
  - **Neural SDE:** Representing a highly flexible, data-driven approach, both the drift function (μ_θ) and the diffusion function (σ_θ) are parameterized by a 2-layer fully connected neural networks. This formulation aligns with the Neural Differential Equation category representing a black-box dynamics model.
  - **Coupled Oscillator SDE (CO-SDE):** To incorporate domain knowledge, we structured the drift term based on the dynamics of coupled limit-cycle oscillators, a model class frequently used to study neural rhythms and synchronization (Breakspear, Heitmann, & Daffertshofer, 2010). Specifically, we adapted the formulation from (Matthews, Mirollo, & Strogatz, 1991) described by the following complex-valued ODE to describe the drift function of the UDE:

    $$\alpha_\theta = ((a_\theta + i\omega_\theta)x + |x|^2 x + K_\theta(u)x) \tag{13}$$

    where x represents the position of oscillators in the complex plane, a the bifurcation parameters, ω denotes the natural frequency, and K represents all-to-all coupling strength. This formulation captures the essential dynamics of coupled limit-cycle oscillators near a supercritical Hopf bifurcation. Leveraging this model in the UDE formulation, we used an MLP to map the inputs to the coupling strength K. We used another MLP to define the diffusion function. This formulation represents a grey-box dynamics model, where some of the the parameters of the model are learned via neural networks.
- **Observation Model (λ_θ):** Following the principles in Section 4.1.4, separate decoders map the latent state $x(t)$ to the different observation modalities:
  - *Neural Data (Spikes):* An MLP transforms the latent state, and its output is exponentiated to yield the rate parameter for a Poisson distribution, predicting spike counts for each neuron. Distinct MLPs are used for neurons in different recorded brain regions, allowing for region-specific mappings.

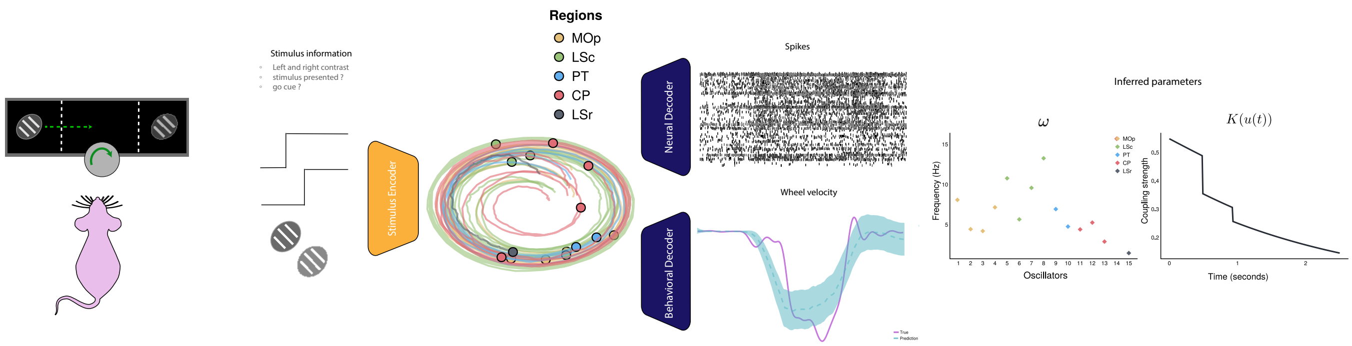

Figure 4: **Generative modeling of neural and behavioral data via latent coupled oscillators across three different datasets.** The generative model takes as input the contrast levels of both the right and left images, wether the stimulus is presented or not, and wether the go cue is issued or not. The model is trained to predict both neural activity across multiple brain regions and wheel velocity. Here latent neural dynamics is represented a latent UDE representing stochastic coupled oscillators. The frequencies of the oscillators after training is displayed along with the dynamic coupling strength inferred for an example trail. Abbreviations: MOp: primary motor area , LSc: lateral sensory cortex, PT: posterior thalamus, CP: caudoputamen, LSr: Lateral sensory rostral area, PMd: dorsal pre-motor cortex, M1: primary motor cortex.

  - *Behavioral Data (Wheel Velocity):* Two parallel linear layers map $x(t)$ to the mean and variance parameters of a Gaussian distribution predicting the continuous wheel velocity.

- **Approximate Posterior SDE:** Training employs variational inference (Section 4.2). An auxiliary SDE (Eq. 10) defines the approximate posterior path measure Q. Its drift μ_ϕ is parameterized by an MLP conditioned on the latent state $\tilde{x}$, observations y, and encoded stimulus u. Crucially, it shares the same diffusion function σ_θ as the generative (prior) SDE to ensure a tractable KL divergence ($D_{\text{KL}}(Q \parallel P)$).

Training All model parameters (θ for the generative model, ϕ for the recognition model) are trained jointly by maximizing the Evidence Lower Bound (ELBO) via stochastic gradient descent. Gradients are computed using backpropagation through the entire computational graph, including the numerical SDE solver (Euler-Maruyama).

Results and Benefits of the UDE Framework We evaluated the predictive performance of the trained models on held-out neural and behavioral data from the Visual Decision-Making dataset. Specifically, we compared stochastic UDE-based models (Neural SDE, CO-SDE) to their deterministic counterparts (Neural ODE, CO-ODE, where $\sigma_\theta = 0$) and to a standard LSTM baseline. Neural activity prediction was quantified using the bits-per-spike (bps) metric (Pei et al., 2021), while behavioral prediction was assessed via the R^2 score. The results reveal a number of key benefits of adopting UDE-based models.

Table 1: Performance comparison of latent sequential models on the Visual Decision-Making dataset. Neural responses are evaluated via bits per spike (bps) and behavioral responses via R^2. The number of parameters of the generative dynamics for each model is also shown. Encoders and decoder architectures are shared across all models. Arrows (↑/↓) indicate whether higher or lower values are better. Best performances are highlighted in **bold**. Results are the mean and std. of 5-fold cross-validation.

Latent Dynamics	**# Params** ↓	**Neural (bps)** ↑	**Behavior (R^2)** ↑
LSTM	33280	0.16 ± 0.006	0.66 ± 0.015
Neural ODE	4880	0.15 ± 0.010	0.59 ± 0.025
CO-ODE	**465**	0.21 ± 0.009	0.61 ± 0.02
Neural SDE	4944	0.18 ± 0.012	**0.70 ± 0.02**
CO-SDE	526	**0.22 ± 0.011**	0.67 ± 0.025

- **Stochasticity Improves Neural Prediction.** The SDE-based models (Neural SDE and CO-SDE) outperformed their ODE-based counterparts in predicting neural activity, achieving higher bits-per-spike values. This highlights the benefit of explicitly modeling intrinsic stochasticity in neural processes, rather than attributing variability solely to uncertainty in initial conditions. The UDE framework, when realized through SDEs, thus better captures trial-to-trial variability in neural dynamics.

- **Structural Priors Enable Parameter-Efficient Learning.** The CO-SDE model, which embeds oscillatory prior structure into the latent dynamics, matched or slightly exceeded the performance of the purely data-driven Neural SDE in neural prediction and performed comparably in behavioral prediction. Importantly, it achieved this with an order of magnitude fewer parameters. Compared to LSTM models, CO-SDE required over 60× fewer parameters. This result demonstrates that incorporating domain knowledge via structured priors can act as an effective inductive bias, leading to models that are both compact and also highly performant.
- **Unified Framework for Model Comparison.** By situating both black-box (Neural SDE) and grey-box (CO-SDE) models within the same UDE framework and training them using identical variational inference procedures, we enabled a direct and fair comparison. This flexibility facilitates rigorous investigation into how different structural assumptions affect model performance and generalization, supporting systematic exploration along the spectrum from data-driven to theory-driven modeling.
- **Interpretability Through Structured Dynamics.** The structured design of CO-SDE provides avenues for mechanistic interpretation. Analysis of the learned time-varying coupling strength parameter $K(u(t))$ revealed systematic modulations aligned with task events (e.g., stimulus onset, go cue), suggesting dynamic shifts in network-level interactions during decision-making as illustrated in Figure 4. Such insights are difficult to extract from black-box models, underscoring how embedding known structure within UDEs can provide anchor points for scientific interpretation, even if full interpretability remains challenging.
- **Bridging Modeling Traditions.** This study illustrates how the UDE framework naturally bridges different modeling approaches: the Neural SDE represents a data-driven, deep learning-based model, while the CO-SDE integrates classical, phenomenological structure from computational neuroscience. Both models can be trained, compared, and interpreted within the same modern machine learning infrastructure, demonstrating UDEs' potential to unify traditionally distinct modeling paradigms.

Conclusion This case study, applying the UDE framework to model complex neurobehavioral data from a decision-making task, showcases its potential. By leveraging latent UDEs, we demonstrated the framework's ability to effectively handle stochasticity, integrate prior knowledge for improved efficiency, facilitate systematic model comparison, and offer avenues for interpretation. These capabilities position UDEs as a versatile and powerful common language for building, testing, and understanding models of neural systems.