

Neural Operators for Power Systems: A Physics-Informed Framework for Modeling Power System Components

Ioannis Karampinis*, Petros Ellinas, Johanna Vorwerk and Spyros Chatzivasileiadis
Technical University of Denmark, Wind and Energy Systems, Kongens Lyngby, Denmark
iokar@dtu.dk

Abstract—Modern power systems require fast and accurate dynamic simulations for stability assessment, digital twins, and real-time control, but classical ODE solvers are often too slow for large-scale or online applications. We propose a neural-operator framework for surrogate modeling of power system components, using Deep Operator Networks (DeepONets) to learn mappings from system states and time-varying inputs to full trajectories without step-by-step integration. To enhance generalization and data efficiency, we introduce Physics-Informed DeepONets (PI-DeepONets), which embed the residuals of governing equations into the training loss. Our results show that DeepONets, and especially PI-DeepONets, achieve accurate predictions under diverse scenarios, providing over 30 times speedup compared to high-order ODE solvers. Benchmarking against Physics-Informed Neural Networks (PINNs) highlights superior stability and scalability. Our results demonstrate neural operators as a promising path toward real-time, physics-aware simulation of power system dynamics.

Index Terms—DeepONet, Dynamical systems, Neural operators, Physics-informed machine learning, Surrogate modeling

I. INTRODUCTION

Modern power systems are increasingly stressed by high penetrations of converter-interfaced resources, fluctuating loads, and more aggressive market dispatch, all of which demand faster and yet accurate dynamic simulations [1]. Traditional time-domain simulations for ordinary differential equations (ODEs) that describe power systems, e.g. swing equations for synchronous machines or dynamics of power electronic converters, provide reliable accuracy but become prohibitively slow [2] when applied to large networks or real-time tasks including online stability assessment, digital-twin operation, and closed-loop control design.

To address these challenges, the power systems community has turned to machine learning (ML) surrogates. Recurrent networks such as Long Short-Term Memory(LSTM) Networks [3] and neural ODEs [4] have been used to regress trajectories directly, but their reliance on large datasets generated by costly numerical solvers limits practicality. Physics-Informed Neural Networks (PINNs) integrate system equations into the training loss, reducing data requirements [5]. However, vanilla PINNs,

composed of a single Neural Network, are generally formulated for specific initial conditions and parametric evolution equations [6]. Moreover, they may accumulate inaccuracies over long integration horizons [7], which could limit their suitability for real-time operation.

Neural operator architectures offer a promising alternative, bypassing the need for sequential time-stepping in solving the governing ODEs. The Deep Operator Network (DeepONet) [8] learns mappings between function spaces, acting as a continuous-time surrogate that can evaluate the response of a dynamical system at arbitrary time points without re-integration. Once trained, inference is nearly instantaneous. The learned operator can act as a rapid simulator—predicting future states from candidate control actions for real-time control loops, or as a digital twin for simulation, diagnostics, and analysis. Even though recent studies have demonstrated operator learning for power system components [9] and for Multi-Input-Multi-Output control frameworks [10], these efforts are mainly data-driven, with limited robustness and generalization capabilities..

To address this gap, this work explores Physics-Informed DeepONets (PI-DeepONets)[11] for non-autonomous power system components with time-varying inputs. By embedding governing physics into operator learning, we aim to combine data-driven speed with improved generalization, offering accurate and efficient surrogates for real-time simulation, analysis, and control.

To the best of our knowledge, this is the first application of Physics-Informed Neural Operators to power system components subject to time-varying inputs. The main contributions of the presented work are:

- **Neural-Operator Framework for Power System Dynamics.** We propose one of the first applications of data-driven operator learning surrogate modeling framework to directly map initial conditions and time-varying inputs (e.g., mechanical torque, voltage set-points, fault events) directly to system states. The framework encodes time-varying inputs through sensor sampling, enabling trajectory evaluation at any query time without the need for step-by-step numerical integration.
- **Physics-Informed Operator Priors.** By embedding the governing differential-equation residuals into the loss function, our PI-DeepONets [11] enforce physical constraints (e.g. swing-equation and voltage-dynamics), im-

This work was supported by the ERC Starting Grant VeriPhIED, Grant Agreement 949899, and the ERC Proof of Concept PINNSim, Grant Agreement 101248667, both funded by the European Research Council. Submitted to the 24th Power Systems Computation Conference (PSCC 2026).

proving generalization to unseen operating points and reducing required training data. To the best of our knowledge, this is the first application of Physics-Informed Neural Operators in power systems.

- **Speedup and Generalization Study.** We demonstrate that our operator models achieve at least a 30× speedup in inference time compared to high-order ODE solvers across diverse component types, all while maintaining high accuracy.

Furthermore, we provide the implementation of our framework, including dataset generation, training routines, and evaluation scripts, to support adoption and ensure reproducibility of our results. The code is publicly available on GitHub: <https://github.com/radiakos/PowerDeepONet>

II. DEEP NEURAL OPERATORS FOR TIME-VARYING INPUT DYNAMICAL SYSTEMS

A. Problem Formulation

Power system components exhibit dynamic behavior that can be mathematically represented using ODEs. These equations describe the evolution of state variables over time and are typically framed as initial value problems (IVPs). An IVP refers to finding a function $x(t)$ solution to a differential equation subject to given initial conditions and its external inputs. The uniqueness of the solution is ensured by the Picard-Lindelof theorem, assuming that the function f governing the ODE is Lipschitz continuous. The IVP and its solution $x(t)$ are formulated as:

$$\frac{d}{dt} \mathbf{x}(\mathbf{t}) = f(t, \mathbf{x}(\mathbf{t}), \mathbf{u}(\mathbf{t}), \boldsymbol{\mu}), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad (1a)$$

$$\mathbf{x}(\mathbf{t}) = \mathbf{x}_0 + \int_0^t f(\tau, \mathbf{x}(\tau), \mathbf{u}(\tau), \boldsymbol{\mu}) d\tau, \quad (1b)$$

where $\mathbf{x}(t)$ represents the state vector that evolves over time t , f is a given function that maps the system parameters to the states, vector $\mathbf{u}(t)$ represents the time-varying inputs of the system, vector $\boldsymbol{\mu}$ includes the system parameters that are to be determined, and \mathbf{x}_0 represents the initial conditions. For readability, the time argument (t) is dropped from $\mathbf{u}(t)$, with the understanding that \mathbf{u} is time-dependent.

The objective is to determine the function $\mathbf{x}(t; \mathbf{x}_0, \mathbf{u}, \boldsymbol{\mu})$ that fulfills (1a) and describes the system's time-domain response. This function represents the unique evolution of the system over a time horizon T for the given initial conditions and time-varying input, referred to as a trajectory. Yet, because the problem is nonlinear, obtaining a closed-form expression for $\mathbf{x}(t)$ is not feasible, so we must rely on approximations:

$$\hat{\mathbf{x}}(t; \mathbf{x}_0, \mathbf{u}, \boldsymbol{\mu}) \approx \mathbf{x}(t; \mathbf{x}_0, \mathbf{u}, \boldsymbol{\mu}). \quad (2)$$

In practice the input function \mathbf{u} is often unknown or partially known, and we rely on an approximation $\hat{\mathbf{u}}$. Since our objective is to approximate the future state of a non-autonomous, i.e. interconnected, dynamical system, knowledge of the external input is required. The proposed multi-step prediction framework can be combined with model predictive control (MPC) strategies, either leveraging existing established approaches or employing an ML-based predictor for the evolution of

external inputs. In the presented work, we stress the ability of the operator to approximate the behavior of the system under various inputs, having access to the same knowledge as traditional ODE solvers.

Since no analytical solution is available for the ODEs, the system response must be approximated through discretization, i.e., by evaluating the dynamics at specific time instances.

B. Classical Ordinary Differential Equations Solvers

Traditional numerical solvers, such as Runge–Kutta schemes, are commonly used for solving ODEs. These methods approximate the solution trajectory over a specified time interval T by iteratively computing discrete updates based on derivative evaluations for small time steps in specific time instances. Increasing the number of data points, i.e., using smaller step sizes, generally improves the accuracy of the approximation but also increases the number of iterations required and subsequently the computational cost. This iterative process can be computationally expensive, especially for complex systems with many state variables, stiff equations.

These limitations motivate the exploration of operator-learning approaches, such as DeepONets, which directly approximate the solution operator.

C. Deep Operator Network

DeepONet is composed of two sub-networks: the *Branch Net* and the *Trunk Net*, which are both introduced in the following.

1) *Branch Net*: encodes the input function(s) at a fixed number of sensors $x_i \in [1, \dots, m]$ and outputs a set of coefficients $b_k(x)$. In this context, sensors correspond to discrete sampling locations of the input signal. Similarly to the traditional ODE solvers, the operator requires discretization of the time-varying input function \mathbf{u} at $m+1$ sensor locations $\{t_i\}_{i=0}^m \subset [0, T]$. The resulting vector $(u(t_0), u(t_1), \dots, u(t_m))$ time traces are provided as input to the Branch Net. The number and placement of sensors-samples within the time horizon must be chosen such that the discretized time traces preserve sufficient information about \mathbf{u} . In particular, m should be large enough to ensure that the induced sampling entropy of \mathbf{u} is sufficient for the Branch Net to capture its core characteristics.

The Branch Net can consist of either a single Neural Network(NN), called *Unstacked DeepONet* as presented in Fig. 1, or multiple networks, called *Stacked DeepONet*. In the latter case, each sensor value is processed by its own Neural Network. Both architectures have distinct advantages in terms of approximation accuracy and generalization, since they encode the information from the input functions in different ways. In this work, we evaluate the performance of the Unstacked DeepONet and an alternative variant of the Stacked DeepONet, which we refer to as the *Stacked-N DeepONet* and shown in Fig. 2. In the Stacked-N architecture, we employ two dedicated branch networks, each corresponding to a different type-nature of input: one to encode the initial conditions and another one to encode the time-varying input. This modular design is particularly promising for modeling multi-stage or

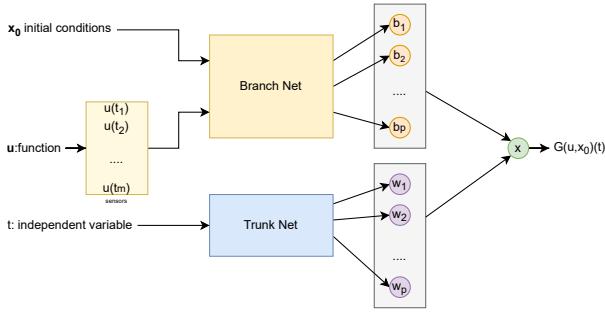


Fig. 1. Traditional Unstacked DeepONet architecture

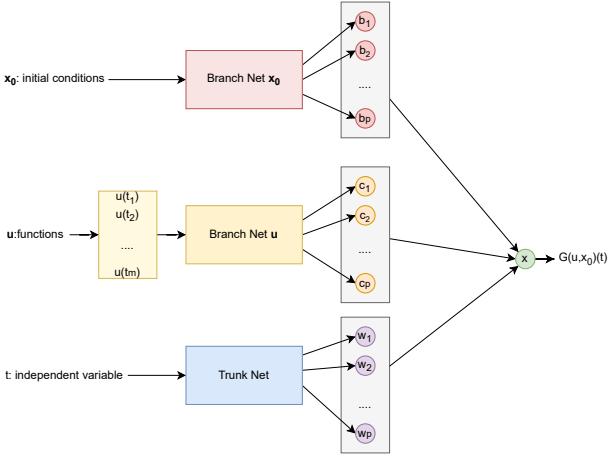


Fig. 2. Stacked-N DeepONet architecture

hierarchical operators, where the external input \mathbf{u} can be handled by a specialized encoder or even by an entire surrogate approximator.

2) *Trunk Net*: encodes the coordinates of the output function, typically the temporal or spatio-temporal evaluation points, and outputs a set of basis functions $\tau_k(\mathbf{t})$.

The output of the DeepONet is obtained by combining the latent representations of the two sub-networks as follows: let G denote the operator mapping the input functions \mathbf{x} and the query point \mathbf{t} , so that the desired solution is $G(\mathbf{x})(\mathbf{t})$. In practice, DeepONet approximates this operator through an expansion of the form

$$G(\mathbf{x})(\mathbf{t}) \approx \sum_{k=1}^p b_k(\mathbf{x}) \tau_k(\mathbf{t}), \quad (3)$$

where $b_k(\mathbf{x})$ is the output of the Branch Net and $\tau_k(\mathbf{t})$ is the output of the Trunk Net. Based on the universal approximation theorem for operators [8], DeepONet architectures with a sufficient number of sensors and basis functions can approximate nonlinear continuous operators to an arbitrary degree of accuracy.

In our setting, DeepONet G aims to directly map the initial system state \mathbf{x}_0 , the time-varying input function(s) \mathbf{u} (e.g., mechanical torque, voltage set-points, fault events), and the query time \mathbf{t} to the resulting system state at \mathbf{t} . Thus, DeepONet provides a surrogate for the solution operator of a non-autonomous dynamical system, by mapping the initial

conditions, the discretized time-varying input, and the independent variable \mathbf{t} . Importantly, it should satisfy the underlying dynamical relation without the need for explicit numerical re-integration of the governing equations. Consequently, it is expected to fulfill equation (1b) as follows:

$$G(\mathbf{u}, \mathbf{x}_0)(\mathbf{t}) = \mathbf{x}_0 + \int_0^{\mathbf{t}} f(\tau, G(\mathbf{u}, \mathbf{x}_0)(\tau), \mathbf{u}(\tau), \boldsymbol{\mu}) d\tau. \quad (4)$$

D. Training the Solution Operator DeepONet

DeepONets can be trained on labeled simulated data of input-output pairs, allowing them to generalize to unseen inputs when the training set sufficiently represents the target domain. Training involves iteratively updating the model parameters to minimize the discrepancy between predictions and desired outputs. In our setting, such data can be generated using ODE solvers. Trajectories are simulated and transformed into the required format. The resulting dataset, defined over the input domain Ω_d with N_d samples has the following format: $\{(x_0^{(i)}, \mathbf{u}^{(i)}, t^{(i)}), x^{(i)}\}_{i=1}^{N_d}$. In case of using the mean squared error (MSE) as the loss function between the ground truth state $x^{(i)}$ and the DeepONet output $\hat{x}^{(i)}$, the trainable parameters are updated according to:

$$\mathcal{L}_{\text{data}} = \frac{1}{N_d} \sum_{i=1}^{N_d} \|\hat{x}^{(i)} - x^{(i)}\|^2. \quad (5)$$

Physics-Informed DeepONet: The structure of the proposed DeepONet allows for Physics-Informed (PI) training when the underlying system of ODEs is available. As proposed in [12] and implemented in [5], we exploit automatic differentiation (AD) to incorporate physical constraints directly into the loss function. In more detail, AD provides the derivative of the DeepONet's output with respect to the query point \mathbf{t} (given as input), corresponding to the left-hand side of (1a), while the right-hand side can be evaluated either using the ground-truth $x^{(i)}$ from the simulated dataset or the DeepONet prediction itself. The resulting PI loss based on simulated data $\mathbf{x}^{(i)}$ is expressed as:

$$\mathcal{L}_{\text{phy data}} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left| \underbrace{\frac{d\hat{x}^{(i)}}{dt}}_{\text{Derivative of DeepONet output}} - \underbrace{-f(t^{(i)}, \mathbf{x}^{(i)}, \mathbf{u}, \boldsymbol{\mu})}_{\text{Source term evaluation}} \right|^2. \quad (6)$$

DeepONet serves as a solution operator and should satisfy the governing physics equations at any arbitrary point within the training domain. When the physics loss is formulated using only the DeepONet approximation, both sides of (1a) depend solely on the DeepONet's predictions $\hat{x}^{(i)}$. This allows us to introduce additional evaluation points, referred to as collocation points, where the physics residual is enforced using as input to f the DeepONet outputs $\hat{x}^{(i)}$ and not $x^{(i)}$ like in (6), as follows:

$$\mathcal{L}_{\text{phy col}} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left| \underbrace{\frac{d\hat{x}^{(i)}}{dt}}_{\text{Derivative of DeepONet output}} - \underbrace{-f(t^{(i)}, \hat{x}^{(i)}, \mathbf{u}, \boldsymbol{\mu})}_{\text{DeepONet approximation term evaluation}} \right|^2. \quad (7)$$

III. A TRAINING PIPELINE FOR DEEPONETS

The main goal of this work is to create a reliable and efficient pipeline for training DeepONets to model non-autonomous dynamic components and demonstrate their performance.. The pipeline is built to be fully parameterizable and highly adaptable at each stage. The following sections present the core steps of the methodology and explain the reasoning behind the choices made at each step.

A. Setting up the ODEs with the Time-Varying Input

Initially, we define the set of ODEs in (1a) that govern the desired non-autonomous power system dynamic component without replacing any value. The system parameters μ in (1a) are considered static and are stored separately. This approach allows testing components over a range of parameters, such as varying the inertia of a synchronous machine. The resulting ODEs are then employed for the dataset generation by the ODE solver and for the PI training.

B. Dataset Generation

The second phase includes the generation of the datasets where one dataset contains the simulated trajectories and the other the collocation points. During the training, we expose the ML model to a limited training region, while for the testing we can assess its ability to capture the behavior of the examined system both within the same input domain, and also stress its generalization ability. This results into a limited domain, i.e. limited samples within a range of values, for the initial conditions, a specific time horizon and a family of external input signals with specified characteristics:

- **Sampling the initial conditions:** We seek to approximate the system's behavior over a defined input domain Ω_d . To do so, we first specify the boundaries of this domain and determine how many distinct samples to take for each state. Different sampling methods can be employed, such as uniform, random, Latin Hypercube sampling (LHS) [13] and others. LHS is preferred as a stratified sampling technique to ensure that all portions of the input space are sampled, allowing for comprehensive coverage of the input space with fewer samples.
- **Time horizon:** The surrogate model must be trained over a finite time window $[0, T]$, where T is chosen according to the needs of the study. Additionally, the time intervals within the time window are to be defined based on the user's needs.
- **External input signals:** In addition to the initial conditions and the time horizon, the dataset must incorporate a representative family of time-varying input, whose characteristics are assumed to be known or predefined. These signals can be polynomial, sinusoidal, stochastic, or designed to mimic realistic operating conditions. Their characteristics (e.g., amplitude, frequency, bandwidth) are selected within predefined bounds to ensure that the model learns to generalize across a sufficiently diverse set of operating scenarios.

Having defined all the above, the ODE solver produces a number of simulated trajectories. These are then further

processed to obtain the needed input-output format. The resulting labeled data is exploited during training through the labeled losses (5) and (6). In contrast, for the collocation points dataset, only the construction of the input format is necessary, and therefore the use of the ODE solver is not required.

C. Deep Operator Neural Network (DeepONet) Architecture

Depending on the desired approach, we must define the architecture of DeepONet to be implemented. The complexity of the examined power system component directly affects the number of hidden layers and the neurons per layer in the Branch and Trunk Nets. The more complex the examined system is, the more neurons and/or layers must be included. The width of the input layer of the Branch Net is determined by the number of the initial conditions and the sensors used, while for the Trunk Net has a single scalar input corresponding to the evaluation time t . The width of the output layer of both the Branch Net and the Trunk Net is a shared latent dimension p . The Branch Net maps the initial conditions and external inputs to $B(u) \in \mathbb{R}^p$, while the Trunk Net maps the evaluation time t to $T(t) \in \mathbb{R}^p$. Their inner product forms the system response. The choice of p depends on the complexity of the examined component: higher nonlinearity or richer dynamics require larger p , while simpler systems can be modeled with smaller values. This selection is typically treated as part of the hyperparameter tuning process.

D. Training of DeepONet

Having defined both the training datasets and the DeepONet's architecture, the next phase includes the training of the DeepONet, which is the optimization of the trainable parameters (i.e. the weights of neurons). The goal is to minimize the losses described in (5), (6) and (7). Since the loss term consists of different terms, a balance between the data points discrepancy and the physics loss must be found. Different strategies can be employed to optimize the total performance of the trained ML model, to provide reliable results with the minimal error. The total loss is a weighted sum of all the losses available, where the weights can be static or dynamically adaptable. The total loss will then be used by the selected optimizer to adjust the trainable parameters of the ML model over a repetitive process.

E. Testing: Performance Assessment of DeepONet

The developed surrogate models are evaluated along two key dimensions: accuracy and computational efficiency.

1) *Accuracy:* Depending on the intended application, the required performance of the DeepONet may vary, e.g. higher accuracy may be needed near the steady state, during the initial transients, or uniformly across the entire time horizon. Accordingly, different sampling strategies for dataset generation and weighting schemes in the loss function can be employed to emphasize specific regions of interest.

In this work, we aim to train and evaluate a DeepONet surrogate model that performs uniformly well within the interval $[0, T]$. For testing, a number of labeled trajectories from the same input domain can be used. The model's performance is

TABLE I
SYNCHRONOUS MACHINE PARAMETERS

Parameter	Value
Damping factor, D	2
Inertia constant, H [s]	5.06
Stator resistance, R_s [p.u.]	0
Direct-axis transient time constant, T_d' [s]	4.75
Quadrature-axis transient time constant, T_q' [s]	1.6
Direct-axis reactances, X_d , X_d' [p.u.]	1.25, 0.232
Quadrature-axis reactances, X_q , X_q' [p.u.]	1.22, 0.715
Line reactance, X_{ep} [p.u.]	0.1
Line resistance, R_e [p.u.]	0
System angular frequency, Ω_b [rad/s]	314.159

evaluated using the Mean Absolute Error (MAE), the Mean Squared Error (MSE), and the Maximum Absolute Error (MaxAE) that captures the upper error bound. These metrics are computed both globally over the full time horizon and locally at each time step for all state variables.

While MAE provides a linear measure of the average magnitude of prediction errors, offering a direct interpretation of model's deviation, the MaxAE quantifies the largest observed absolute error, defining the model's upper error bound.

2) *Time*: The computational efficiency of the developed DeepONet is assessed by comparing its inference time with the traditional numerical ODE solver. Unlike solvers that iteratively compute all intermediate steps, the DeepONet directly approximates system states at arbitrary time points, bypassing the integration process. This allows rapid evaluation of single points or entire trajectories in one forward pass. Moreover, the parallelism nature of Neural Networks allows efficient GPU utilization and larger batch sizes, allowing simultaneous evaluation of multiple trajectories. The resulting speedup of DeepONet inference time is measured against solver runtime, in both single- and multi-trajectory simulations, and also in single point evaluation.

IV. CASE STUDY

This section demonstrates DeepONets ability to approximate the behavior of a power system's component with time-varying external input. Specifically, we present results for the two different aforementioned architectures of DeepONet, Unstacked and Stacked-N, as a surrogate for a 4th-order synchronous machine (SM) connected to a bus with varying voltage. This study considers the following representation found in [14] and the parameters used are shown in Table I :

$$\begin{bmatrix} \frac{1}{\Omega_B} \\ \frac{2H}{\Omega_B} \\ \frac{d}{dt} \\ T_{do}' \\ T_{qo}' \end{bmatrix} = \begin{bmatrix} \delta \\ \omega \\ P_m - E_d' I_d + E_q' I_q + (X_q' - X_d') I_d I_q - D\omega \\ -E_q' - (X_d - X_d') I_d + E_{fd} \\ -E_d' + (X_q - X_q') I_q \end{bmatrix} \quad (8)$$

and

$$\begin{bmatrix} (R_s + R_e) & -(X_q + X_e) \\ (X_d' + X_e) & (R_s + R_e) \end{bmatrix} \begin{bmatrix} I_d \\ I_q \end{bmatrix} = \begin{bmatrix} E_d' - V_s \sin(\delta - \theta_{vs}) \\ E_q' - V_s \cos(\delta - \theta_{vs}) \end{bmatrix}. \quad (9)$$

A. Datasets of the Input Domain

1) *Domain of Initial Conditions*: Different initial conditions for the simulations are sampled from the domain of initial conditions Ω_{ic} . The selected ranges are as follow: $\theta \in [-2, 2]$ rad, $\omega \in [-1, 1]$ rad/s, $E_d' = 0$ pu, $E_q' \in [0.9, 1.1]$ pu, $R_F = 1$ pu, $V_r = 1.105$ pu, $E_{fd} = 1.08$ pu, $P_{sv} = 0.7048$ pu, and $P_m = 0.7048$ pu. In total, 2200 different sets of initial conditions were sampled within the state space using LHS, consisting of 1000 for the labeled training data, 1000 for the collocation points, and 100 for each of the labeled validation and test sets.

2) *Time-Varying External Input*: The external voltage inputs, defined by the magnitude V_s and the phase angle θ_{vs} , are modeled using either first or second-order ODEs. An alternative modeling approach was proposed in [15] and employed in [6], where the evolution of the external input is represented in polar form. The two general approaches used in this work are formulated as follows:

$$\frac{dy}{dt} = -k(y - y_{ref}) + A_{noise} \sin(\omega_{noise} t), \quad (10)$$

or

$$\frac{d^2y}{dt^2} + 2\zeta\omega_n \frac{dy}{dt} + \omega_n^2(y - y_{ref}) = 0, \quad (11)$$

where k is the system gain, y_{ref} the steady-state reference, A_{noise} and ω_{noise} denote the amplitude and frequency of perturbations, and ζ and ω_n are the damping ratio and natural frequency, respectively.

Two configurations of external voltage dynamics are considered for the presented case study:

- 1) **Slow response**: Both V_s and θ_{vs} follow first-order dynamics defined in (10) with low-frequency perturbations.
 - *Voltage magnitude*: $k \in [1.0, 2.0]$, $A_{noise} \in [0.0, 0.03]$, $\omega_{noise} \in [0.5, 1.5]$, $V_s(0) \in [0.7, 1.2]$
 - *Phase angle*: $k \in [1.0, 3.0]$, $A_{noise} \in [0.0, 0.05]$, $\omega_{noise} \in [0.1, 0.5]$, $\theta_{vs}(0) \in [-0.5, 0.5]$
- 2) **Fast response**: Both quantities are modeled as second-order systems as in (11) with higher frequencies.
 - *Voltage magnitude*: $\zeta \in [0.6, 0.8]$, $\omega_n \in [8.0, 12.0]$, $V_s(0) \in [0.7, 1.2]$, $\dot{V}_s(0) \in [-0.2, 0.2]$
 - *Phase angle*: $\zeta \in [0.2, 0.4]$, $\omega_n \in [10.0, 18.0]$, $\theta_{vs}(0) \in [-0.5, 0.5]$, $\dot{\theta}_{vs}(0) \in [-0.8, 0.8]$
- 3) *Generating Labeled and Collocation Datasets*: The numerical solver RK45 from the `scipy.integrate` library is used to generate the simulated trajectories representing the transient behavior of the SM. The solver inputs the time-varying external signals, the initial conditions, and the time horizon. The latter is set to 1 s with an evaluation time step of 1 ms. The external input functions were sampled at the same evaluation points, resulting in 1000 sensors in total. The outputs correspond to the simulated trajectories of each SM state variable, which were subsequently processed to form the input-output pairs required for DeepONet training.

In contrast, the collocation dataset consists only of inputs. Specifically, it includes the evaluation points for the trunk network, the initial conditions, and the sampled external input vectors.

The final datasets have the following format:

$$\mathcal{X}_{data} = \{(\mathbf{x}_0^{(i)}, \mathbf{u}^{(i)}, t^{(i)}), \mathbf{x}^{(i)}\}_{i=1}^{N_l} \quad (12a)$$

$$\mathcal{X}_{collocation} = \{(\mathbf{x}_0^{(i)}, \mathbf{u}^{(i)}, t^{(i)})\}_{i=1}^{N_c} \quad (12b)$$

The training dataset comprises 1 000 labeled trajectories, each containing 1 000 data points (in total, 1 000 000 data points), along with $N_c = 1 000 000$ collocation points. The validation and testing datasets each contain 100 trajectories of identical dimensionality in time.

B. DeepONet Architecture and Hyperparameter Training

For both implementations, Unstacked and Stacked-N, the Branch and Trunk networks are fully connected neural networks consisting of three hidden layers with 64 neurons each and a hyperbolic tangent (\tanh) activation function. The output layer of each network produces 64 latent features, whose inner product approximates the response of the dynamical component. The architecture was determined empirically by testing different hyperparameters, with the final configuration providing a favorable trade-off between accuracy and computational cost & efficiency.

Training is performed using the SOAP optimizer [16] with a learning rate of 3×10^{-3} and a weight decay of 0.01. The models are trained for 10 000 epochs using the Mean Squared Error (MSE) loss function. Early stopping is employed to prevent overfitting, with a patience of 200 epochs and a minimum improvement threshold of 1×10^{-6} . The static loss weights are set to $\lambda_d = 1$, $\lambda_{pd} = 1 \times 10^{-3}$, and $\lambda_{pc} = 1 \times 10^{-4}$.

All implementations were developed in PyTorch [17]. Weights & Biases (WandB) [18] was used for experiment tracking and hyperparameter optimization, providing detailed insights into model training and testing performance. All experiments were conducted on the High-Performance Computing (HPC) cluster at the Technical University of Denmark, utilizing a 16-core Intel Xeon 6226R processor with 256 GB of RAM and an NVIDIA V100 GPU with 16 GB GPU memory.

C. Results

This section compares the performance of the proposed DeepONet solutions in approximating the dynamics of a 4th-order SM subject to time-varying voltage inputs. The objective is to assess their accuracy, generalization capability across diverse operating conditions, and computational efficiency. We benchmarked the performance against a different ML surrogate, a PINN with identical inputs, outputs, and training data, differing only in its internal network architecture. All training hyperparameters were kept identical and the number of trainable parameters was of comparable magnitude. Each experiment is repeated three times to ensure fairness, with a different random seed used for every run. The average results are reported below.

1) Accuracy: The evaluation was conducted on the 100 unseen trajectories of the test dataset, where each trajectory includes the full dynamic evolution of the four state variables. Figure. 3 highlights the accurate trajectory reconstruction of the operator-learning surrogates across different initial conditions and external inputs.

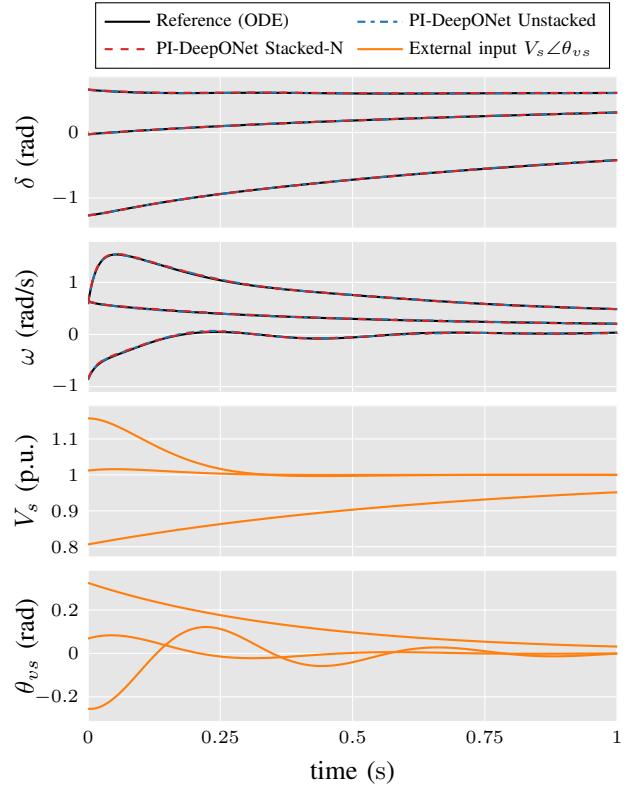


Fig. 3. Illustration of three trajectories from the test dataset showing the rotor angle δ and speed ω obtained from the reference ODE solver, the PI-DeepONet surrogates, and the single PINN model. The corresponding time-varying external voltage magnitude V_s and phase θ_{vs} are shown in the lower panels.

Based on the accuracy metrics summarized in Table II, among all variations, the Unstacked PI-DeepONet achieved the best overall performance, followed by the Stacked-N PI-DeepONet, while all DeepONet-based models significantly outperformed the baseline PINN. The temporal evolution of the AE is illustrated in Figure 4. Both PI-DeepONet architectures exhibit slightly higher errors during the initial transient period but remain stable and low throughout the remainder of the time horizon. In contrast, the PINN exhibits consistently higher errors and demonstrates weaker performance, particularly in the transient region. Overall, these results confirm the superior capability of PI-DeepONets to learn functional mappings between input and output trajectories with high accuracy and robustness. It is worth noting that, although not presented in this work, additional experiments indicate that increasing the number of training samples consistently enhances the accuracy of all implementations.

2) Time: In addition to accuracy, computational efficiency is a critical factor for enabling real-time applications such as online stability assessment and digital twins. Once trained within a specified input domain, with an upfront training cost of up to 15 minutes, the DeepONets can infer complete system trajectories without requiring step-by-step numerical integration. Table III summarizes the runtime comparison: for a single trajectory, the DeepONet achieves more than a $30\times$ speedup compared to a high-precision Runge–Kutta solver, while the

TABLE II
AVERAGE ACCURACY METRICS OF ML SURROGATES.

ML Surrogate	MSE	MAE	Max AE
PINN	38.67×10^{-6}	3.87×10^{-3}	172.93×10^{-3}
Unstacked DeepONet	38.25×10^{-6}	3.66×10^{-3}	117.23×10^{-3}
Unstacked PI-DeepONet	26.87×10^{-6}	3.26×10^{-3}	73.03×10^{-3}
Stacked-N DeepONet	38.6×10^{-6}	3.68×10^{-3}	157.44×10^{-3}
Stacked-N PI-DeepONet	36.14×10^{-6}	3.36×10^{-3}	104.69×10^{-3}

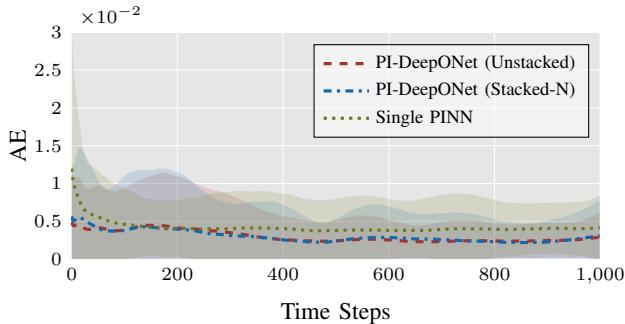


Fig. 4. Comparison of the absolute error (AE) over time among the three surrogate models, illustrating their relative accuracy and temporal stability.

advantage becomes even more pronounced when evaluating multiple trajectories in parallel, reaching a speedup exceeding $6000\times$ for 1000 trajectories. It should be noted that the ODE solver runs sequentially on a single CPU core, whereas DeepONet inference leverages parallel GPU execution through a single forward evaluation. In contrast, PINNs remain slower due to their dense architecture and large number of neuron connections, which increase computational and memory costs. Overall, these results highlight the scalability and suitability of the proposed approach for real-time applications.

V. CONCLUSION AND FUTURE WORK

This paper introduced physics-informed neural operators for modeling of power system dynamic components. By combining the capabilities of the DeepONet's architecture with physics-informed training, the proposed approach enables accurate and efficient trajectory prediction for systems subject to time-varying inputs. Once trained, the operator can evaluate entire trajectories in a single forward pass, bypassing the need for iterative integration and achieving substantial speedups compared to conventional solvers. The results demonstrate that embedding physics into the learning process enhances generalization to previously unseen operating conditions.

Future research will focus on improving the scalability and adaptability of neural operator models. This includes investigating sensor placement strategies and search for optimal operator hyperparameters, developing uncertainty quantification and adaptive loss weighting techniques, and extending the framework to encompass inverter-based resources and multi-machine systems. Integrating neural operators into real-time simulation and control frameworks, together with advancing

TABLE III
INFERENCE TIME IN MS (AND SPEEDUP)

	Single Trajectory	10 Trajectories	1000 Trajectories
ODE solver	15.517	155.169	15516.834
PINN	0.189 (80 \times)	0.665 (230 \times)	6.861 (2260 \times)
Unstacked DeepONet	0.358 (40 \times)	0.425 (360 \times)	2.311 (6720 \times)
Stacked-N DeepONet	0.485 (30 \times)	0.539 (290 \times)	2.415 (6430 \times)

online learning from measurement data, constitutes a significant step toward developing fully data- and physics-informed digital twins of the power grid.

REFERENCES

- [1] S. Zhang, A. Pandey, D. Pahuja, T. Nudell, J. Zhao, R. Banerji, X. Luo, and S. Liu, "Grid Modernization in the Age of Artificial Intelligence and Cloud: Synergies Between Grid Modernization, Artificial Intelligence, and Cloud," *IEEE Power and Energy Magazine*, vol. 23, no. 5, pp. 102–116, 2025.
- [2] F. Milano, *Power system modelling and scripting*. Springer Science & Business Media, 2010.
- [3] J. Li, M. Yue, Y. Zhao, and G. Lin, "Machine-learning-based online transient analysis via iterative computation of generator dynamics," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2020, pp. 1–6.
- [4] T. Xiao, Y. Chen, S. Huang, T. He, and H. Guan, "Feasibility study of neural ODE and DAE modules for power system dynamic component modeling," *IEEE Transactions on Power Systems*, vol. 38, no. 3, pp. 2666–2678, 2022.
- [5] I. Karampinis, P. Ellinas, I. V. Nadal, R. Nellikkath, and S. Chatzivasileiadis, "Toolbox for Developing Physics Informed Neural Networks for Power Systems Components," *arXiv preprint arXiv:2502.06412*, 2025.
- [6] J. Stiasny, B. Zhang, and S. Chatzivasileiadis, "PINNSim: A simulator for power system dynamics based on physics-informed neural networks," *Electric Power Systems Research*, vol. 235, p. 110796, 2024.
- [7] S. Wang and P. Perdikaris, "Long-time integration of parametric evolution equations with physics-informed deepnets," *Journal of Computational Physics*, vol. 475, p. 111855, 2023.
- [8] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," *Nature Machine Intelligence*, vol. 3, no. 3, 10 2021. [Online]. Available: <https://www.osti.gov/biblio/2281727>
- [9] O. Arowolo, J. Stiasny, and J. Cremer, "Exploring extrapolation of machine learning models for power system time domain simulation," *Sustainable Energy, Grids and Networks*, p. 101908, 2025.
- [10] T. O. de Jong, K. Shukla, and M. Lazar, "Deep Operator Neural Network Model Predictive Control," *arXiv preprint arXiv:2505.18008*, 2025.
- [11] S. Goswami, A. Bora, Y. Yu, and G. E. Karniadakis, "Physics-informed deep neural operator networks," in *Machine learning in modeling and simulation: methods and applications*. Springer, 2023, pp. 219–254.
- [12] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.
- [13] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 42, no. 1, pp. 55–61, 2000.
- [14] P. W. Sauer, M. A. Pai, and J. H. Chow, *Power system dynamics and stability: with synchrophasor measurement and power system toolbox*. John Wiley & Sons, 2017.
- [15] B. Wang, N. Duan, and K. Sun, "A time-power series-based semi-analytical approach for power system simulation," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 841–851, 2018.
- [16] N. Vyas, D. Morwani, R. Zhao, M. Kwun, I. Shapira, D. Brandfonbrener, L. Janson, and S. Kakade, "Soap: Improving and stabilizing shampoo using adam," *arXiv preprint arXiv:2409.11321*, 2024.
- [17] PyTorch Foundation, "PyTorch." [Online]. Available: www.pytorch.org
- [18] L. Biewald and others, "Experiment tracking with weights and biases," 2020.