

Received 10 July 2024; revised 5 November 2024; accepted 1 January 2025.  
Date of publication 15 January 2025; date of current version 4 February 2025.

Digital Object Identifier 10.1109/OAJPE.2025.3529928

# Physics-Informed Kolmogorov-Arnold Networks for Power System Dynamics

HANG SHUAI<sup>ID</sup> (Member, IEEE) AND FANGXING LI<sup>ID</sup> (Fellow, IEEE)

Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996 USA

CORRESPONDING AUTHOR: F. LI (fli6@utk.edu)

This work was supported in part by the Center for Ultra-Wide-Area Resilient Electric Energy Transmission Networks (CURENT), which is a graduated U.S. National Science Foundation (NSF) Engineering Research Center and was jointly supported by NSF and the Department of Energy (DoE) from 2011 to 2021.

**ABSTRACT** This paper presents, for the first time, a framework for Kolmogorov-Arnold Networks (KANs) in power system applications. Inspired by the recently proposed KAN architecture, this paper proposes physics-informed Kolmogorov-Arnold Networks (PIKANs), a novel KAN-based physics-informed neural network (PINN) tailored to efficiently and accurately learn dynamics within power systems. PIKANs offer a promising alternative to conventional Multi-Layer Perceptrons (MLPs) based PINNs, achieving superior accuracy in predicting power system dynamics while employing a smaller network size. Simulation results on test power systems underscore the accuracy of the PIKANs in predicting rotor angle and frequency with fewer learnable parameters than conventional PINNs. Specifically, PIKANs can achieve higher accuracy while utilizing only 50% of the network size required by conventional PINNs. Furthermore, simulation results demonstrate PIKANs' capability to accurately identify uncertain inertia and damping coefficients. This work opens up a range of opportunities for the application of KANs in power systems, enabling efficient dynamic analysis and precise parameter identification.

**INDEX TERMS** Kolmogorov-Arnold Networks (KANs), power system dynamics, deep learning, swing equation, physics-informed neural network (PINN).

## I. INTRODUCTION

DEEP learning (DL) has demonstrated remarkable success in addressing complex tasks, particularly in fields where precise mathematical models are difficult to establish, such as computer vision, natural language processing [1], protein structure prediction [2], and medical image analysis [3]. In the power sector, DL has also increasingly been investigated for applications such as renewable energy forecasting [4], power system stability assessment [5], microgrid energy management [6], [7], reflecting its growing influence and great application potentials for power grids as projected in [8].

Power system dynamics are governed by Differential-Algebraic Equations (DAEs), which impose substantial computational demands, particularly in tasks like transient stability analysis. Reducing this computational burden while retaining high accuracy is critical across many power system applications. Because analytical solutions for DAEs are generally unavailable, numerical methods—such as the Euler method or the Runge-Kutta method—are typically employed to approximate the dynamic processes of power

systems [9]. However, the integration of variable renewable energy (VRE) sources and inverter-based resources (IBRs) has introduced significant uncertainty and nonlinearity into power system dynamics, requiring the assessment of a vast number of scenarios to ensure system security. Traditional numerical methods for solving these DAEs are often computationally intensive and may struggle with the extensive simulation times required for such exhaustive assessments. This highlights the urgent need for advanced methods that can significantly accelerate these computations, which is crucial for the evolving needs of the power industry.

With advancements in artificial intelligence (AI), significant efforts have been made to develop various data-driven algorithms for the modeling of power system dynamics and online identification of power system dynamics [10], [11], [12]. Among these, DL techniques have been increasingly utilized [13]. However, these DL-based approaches often lacked integration with the underlying power system model. As a result, they relied heavily on the quality and quantity of training data, necessitating large datasets and complex neural network architectures. Considering this,

researchers further proposed physics-informed neural network (PINN) based algorithms for power system dynamic identification. For example, in [14], a PINN approach was developed to learn the rotor angle and frequency dynamics of a single-machine infinite bus (SMIB) power system. The PINN based method leverages the underlying physical model, resulting in significantly reduced computation times and a lesser need for training data. Researchers further proposed a practical framework for identifying essential features of non-linear voltage dynamics. This approach converts PINNs into a mixed-integer linear program [15]. It enables adjustment of the neural network's output conservativeness concerning stability boundaries, eliminating the necessity for exhaustive time-domain simulations.

Despite promising results in designing PINNs for power system dynamics, there remains significant room for improvement in the accuracy of the learned dynamic models. For instance, the PINNs agent developed in [14], exhibits relative  $L^2$  errors of 2.37% between the exact and predicted solutions for the rotor angle of a SMIB power system. When used to identify the generator inertia constant and the damping coefficient of a power system, the mean error of parameter identification of PINNs would reach around 50% when only limited measurements (such as rotor angle) are available [16]. Furthermore, the aforementioned PINNs struggle to effectively learn both the stable and unstable dynamics of the same power system. This limitation necessitates the use of distinct, trained PINNs to achieve high accuracy in stable and unstable regimes [14]. In this work, inspired by recently proposed Kolmogorov-Arnold Networks (KANs) in [17], we propose physics-informed KANs (PIKANs) algorithm for more accurately predicting power system dynamics that reduces dependency on training data and enables a more smaller number of learnable parameter in neural networks.

Current DL architectures such as deep neural networks (DNNs), recurrent neural networks (RNNs), convolutional neural networks (CNNs) and PINNs, largely rely on Multi-Layer Perceptrons (MLPs) [18]. MLPs are fully connected neural networks featuring fixed activation functions in neurons, while weights associated with network connections are adjusted using backpropagation techniques during training. Further, universal approximation theorems [19], [20] imply that MLPs are universal approximators, which can represent a wide variety of interesting functions with appropriate weights and activation functions. However, MLPs based DL techniques face challenges such as the requirement for large training datasets, catastrophic forgetting [21], and a lack of interpretability [22] (often referred to as the "black box" problem). KANs [17], promising alternatives to MLPs, also feature fully-connected network structures. Unlike MLPs, KANs place learnable activation functions on the edges, which usually allow much smaller computation graphs than MLPs and could reach more accurate learning results at the same time. While MLPs have the potential to learn generalized additive structures, they often struggle

with efficiently approximating exponential and sine functions using traditional activation functions, such as ReLU. In contrast, KANs excel in learning both compositional structures and univariate functions effectively, thereby significantly outperforming MLPs [17]. Considering sine and cosine functions are fundamental functions in power system dynamics models, so KANs would have great potential to be more effective at representing dynamics in power systems than MLPs. In summary, KANs are mathematically sound, accurate and interpretable, which offer a range of opportunities in power systems by precisely and adaptively capturing grid dynamics as described by DAEs. Compared to PINNs, the proposed PIKANs have the potential to enhance learning capabilities for power system dynamics by incorporating physical models within the KAN framework.

This is the first work to propose the use of KANs for power system applications. Specifically, we utilize the swing equation in power systems as an example to demonstrate their potential. We also demonstrate the proposed method can be used to estimate uncertain inertia and damping coefficients. The main contributions of this work can be summarized as follows:

- For the first time, we present a framework that integrates KANs with the PINNs architecture for power system applications, and PIKAN algorithms for power system dynamics are developed. We propose a PIKAN training procedure that leverages the power system swing equation model to reduce data dependency and achieve higher accuracy. Compared to traditional MLP-based PINNs, the proposed PIKANs provide a highly accurate alternative for modeling power system dynamics.
- The performance of the proposed method is demonstrated on three test systems: a single-machine infinite bus (SMIB) system, a 4-bus 2-generator system, and the IEEE New England power system. Simulation results show that PIKANs achieve higher accuracy in solving the DAEs of power systems while requiring only 50% of the neural network size compared to traditional MLP-based PINNs.

This paper is organized as follows: Section II introduces the power system dynamic model investigated in this work. Section III presents KANs and the framework designation that integrates KANs with the PINNs architecture for the power system dynamic application. Section IV shows numerical simulation results on the three testing systems demonstrating the performance of the proposed algorithm. Section V discusses the findings and limitations of this work. Section VI concludes the paper.

## II. POWER SYSTEM DYNAMIC MODEL

Power system dynamics are described by swing equations. By assuming the bus voltage magnitudes to be 1 per unit (p.u.), and neglecting the reactive power flows, the frequency dynamics of each generator  $i$  can be described by the

following equation [23]:

$$\begin{aligned} \frac{d\theta_i}{dt} &= \omega_i \\ M_i \cdot \frac{d\omega_i}{dt} &= P_{m_i} - P_{e_i} - D_i \cdot \omega_i \end{aligned} \quad (1)$$

where  $\theta_i$  and  $\omega_i$  are the voltage angle and angular frequency of the generator  $i$  (also connected to bus  $i$ ), respectively.  $t$  is the time index.  $M_i$  and  $D_i$  are the inertia and damping constant of the generator  $i$ , respectively.  $P_{m_i}$  is the net power injection.  $P_{e_i}$  is the electrical power output (p.u.) of the  $i$ th generator, which can be calculated by the following equation [24]:

$$P_{e_i} = \sum_{j=1}^n V_i V_j [B_{ij} \cdot \sin(\theta_i - \theta_j) + G_{ij} \cdot \cos(\theta_i - \theta_j)] \quad (2)$$

where  $B_{ij}$  and  $G_{ij}$  are the susceptance and conductance of the transmission line between bus  $i$  and  $j$ , respectively.  $V_i$  and  $V_j$  represent the voltage magnitudes at bus  $i$  and bus  $j$ , respectively.

For transmission systems, when the line reactance  $X$  greatly exceeds the resistance  $R$ , and assuming the bus voltage is 1 p.u., equation (2) can be simplified to:

$$P_{e_i} = \sum_{j=1}^n B_{ij} \cdot \sin(\theta_i - \theta_j) \quad (3)$$

For the frequency dependent load  $i$ , the frequency dynamics in equation (1) can be simplified to:

$$P_{m_i} - P_{e_i} - D_i \cdot \omega_i = 0 \quad (4)$$

where  $\omega_i = \frac{d\theta_i}{dt}$ .

Therefore, the system dynamics can be described by equations (1) and (4), which can be expressed in the form of a DAE system:

$$\begin{aligned} \dot{\mathbf{x}}_{\text{sys}} &= \mathbf{h}(\mathbf{x}_{\text{sys}}, \mathbf{y}, \mathbf{p}; \lambda) \\ \mathbf{0} &= \mathbf{g}(\mathbf{x}_{\text{sys}}, \mathbf{y}, \mathbf{p}; \lambda) \\ \mathbf{P}_m &\in [\mathbf{P}_m^{\min}, \mathbf{P}_m^{\max}], t \in [0, T] \end{aligned} \quad (5)$$

where  $\mathbf{x}_{\text{sys}} = [\theta; \omega]$  is the power system state variables vector.  $\mathbf{y} = [\mathbf{P}_e]$  is the algebraic variables vector.  $\mathbf{p} = [\mathbf{P}_m]$  represents the power system input variables.  $\lambda = [\mathbf{M}; \mathbf{D}; \mathbf{B}]$  is the parameter of the power system.

In this work, we focus on using PIKANs to learn dynamics described by equation (5), and identify uncertain inertia and damping parameters in  $\lambda$ .

### III. PHYSICS-INFORMED KOLMOGOROV-ARNOLD NETWORKS FOR POWER SYSTEM DYNAMICS

#### A. KOLMOGOROV-ARNOLD NETWORKS

Based on Kolmogorov-Arnold Representation theorem [25], for a multivariate continuous function  $f$  on a bounded domain, it can be represented by a finite composition of continuous functions of a single variable and the binary operation

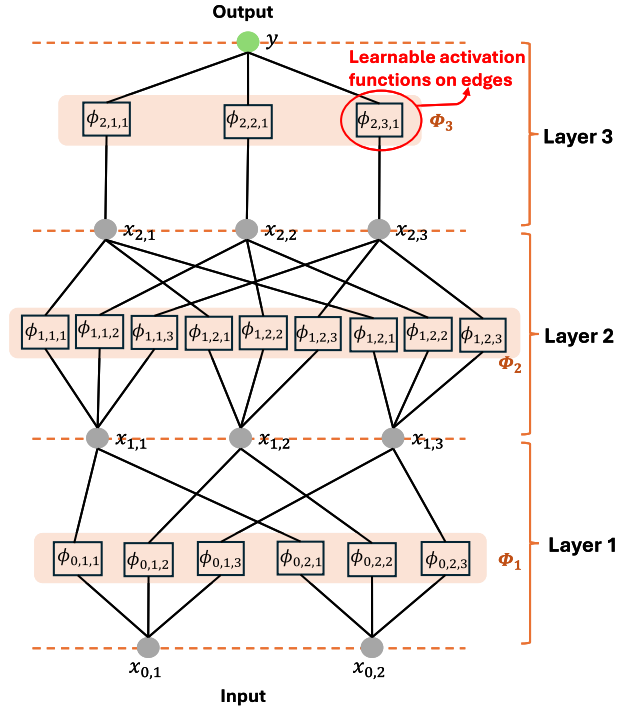


FIGURE 1. Illustration of a 3-layer KAN having a shape of [2, 3, 3, 1].

of addition, as given by the following equation:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_k) = \sum_{q=1}^{2k+1} \Phi_q \left( \sum_{p=1}^k \phi_{q,p}(x_p) \right) \quad (6)$$

where  $\phi_{q,p}: [0, 1] \rightarrow \mathbb{R}$  and  $\Phi_q: \mathbb{R} \rightarrow \mathbb{R}$ .  $\mathbf{x}$  is the input vector. The theorem shows that learning a high-dimensional function  $f$  can be boiled down to learning a polynomial number of 1D univariate functions  $\phi_{q,p}$  and  $\Phi_q$ . Actually, equation (6) can be treated as a two layer network having a shape of  $[n, 2n + 1, 1]$ , which has two-layer nonlinearities. However, these 1D functions can be non-smooth and even fractal, so they may not be learnable in practice [26]. Thus, the theorem was thought to be practically useless in machine learning.

To enable the Kolmogorov-Arnold theorem for machine learning, [17] innovatively proposed the KANs architecture, as illustrated in Fig. 1. In KANs, each 1D function of equation (6) are parametrized as a B-spline curve. Each B-spline curve is with learnable coefficients of local B-spline basis functions. It is worth noting that the activation functions are placed on edges instead of nodes in Fig. 1. To generalize the network described by equation (6) to arbitrary widths and depths, [17] further defined a KAN layer and stacking more KAN layers as needed. A KAN layer with  $n_l$ -dimensional inputs and  $n_{l+1}$ -dimensional outputs is defined as a matrix of 1D functions:

$$\Phi_l = \{\phi_{l,j,i}\}, i = 1, 2, \dots, n_l, j = 1, 2, \dots, n_{l+1} \quad (7)$$

where function  $\phi_{l,j,i}$  has trainable parameters, which is the activation function that connects the  $i^{th}$  neuron in the  $l^{th}$  layer and the  $j^{th}$  neuron in the  $l + 1^{th}$  layer.  $l$  is the index of the layer. Therefore, the output of the  $l^{th}$  layer of the KAN is

$$\begin{aligned} \mathbf{x}_{l+1} &= \Phi_l \mathbf{x}_l \\ &= \begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \cdots & \phi_{l,2,n_l}(\cdot) \\ \vdots & \vdots & & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix} \mathbf{x}_l, \end{aligned} \quad (8)$$

In this way, the output of a KAN network composed of  $L$  layers can be written as

$$KAN(\mathbf{x}) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \cdots \circ \Phi_1 \circ \Phi_0) \mathbf{x} \quad (9)$$

where  $\mathbf{x} \in \mathbb{R}^{n_0}$  is the input vector of the network. Also, the operator  $\circ$  in (9) denotes function composition, indicating that the operations are applied sequentially from right to left. Considering all the above operations are differentiable, KANs can be trained with back propagation techniques.

To make the KAN easy to train, we can design activation functions as given below:

$$\phi(x_{l,i}) = w \cdot (b(x_{l,i}) + \text{spline}(x_{l,i})) \quad (10)$$

where  $w$  is a factor to control the overall magnitude of the activation function.  $b(x)$  is a basis function which can be set to

$$b(x_{l,i}) = \text{silu}(x_{l,i}) = \frac{x_{l,i}}{1 + e^{-x_{l,i}}} \quad (11)$$

$\text{spline}(x_{l,i})$  is the spline function which can be parametrized as a linear combination of B-splines:

$$\text{spline}(x_{l,i}) = \sum_s c_s \cdot B_s(x_{l,i}) \quad (12)$$

where  $B_s(x_{l,i})$  is the B-spline function. During the training process,  $\text{spline}(\cdot)$  and  $w$  are trainable, and we can initialize  $\text{spline}(\cdot)$  by drawing B-spline coefficients  $c_s \sim \mathcal{N}(0, 0.1^2)$  and  $w$  initialized according to the Xavier initialization. It worth noting that other activation functions other than B-spline can be also utilized. For instance, to address computational cost problem caused by training learnable B-splines, [27] developed a wavelet KAN architecture based on the work in [17].

For a  $L$ -layer KAN with layers of equal width  $N$  (which means each layer has  $N$  neurons), there are in total  $O(N^2 L(G + k_b)) \sim O(N^2 LG)$  parameters, where  $k_b$  and  $G$  are the order and intervals of the spline. Contrarily, an MLP with depth  $L$  and width  $N$  typically requires  $O(N^2 L)$  parameters, suggesting it might be more parameter-efficient than a KAN. However, KANs often operate effectively with much smaller  $N$  than MLPs. This not only reduces parameter count but also enhances generalization and facilitates interpretability.

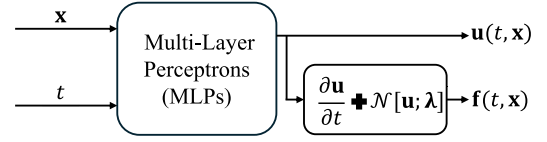


FIGURE 2. General structure of a PINN [14], [28]: it predicts the output  $\mathbf{u}(t, \mathbf{x})$  given inputs  $\mathbf{x}$  and  $t$ .

## B. PHYSICS-INFORMED KANs FOR POWER SYSTEM DYNAMICS

PINNs are universal function approximators that incorporate the knowledge of physical laws governing a given dataset into the neural network training process [28]. This approach mitigates the need for large amounts of training data and the large network sizes typically required by traditional DNNs. In traditional PINNs, the architecture consists of a MLP with an input layer, several fully connected hidden layers featuring fixed nonlinear activation functions at each neuron, and an output layer. Each layer transition involves the application of a weight matrix  $\mathbf{W}_l$  and an activation function  $\sigma_l$ :

$$\begin{aligned} \text{MLP}(\mathbf{x}) &= (\mathbf{W}_{L-1} \circ \sigma_{L-1} \circ \mathbf{W}_{L-2} \circ \sigma_{L-2} \circ \cdots \circ \mathbf{W}_1 \circ \sigma_1 \circ \mathbf{W}_0) \mathbf{x} \end{aligned} \quad (13)$$

During the training process, these weights are adjusted to minimize an objective function, which typically penalizes the difference between the neural network's predictions and the actual labels of the training data.

Based on [28], the dynamics of a physical system governed by parametrized and nonlinear partial differential equations (PDEs), as shown in equation (14), can be effectively learned using PINNs.

$$\frac{\partial \mathbf{u}}{\partial t} + \mathcal{N}[\mathbf{u}; \boldsymbol{\lambda}] = \mathbf{0}, \mathbf{x} \in \Omega, t \in [0, T] \quad (14)$$

where  $\mathbf{u}(t, \mathbf{x})$  is the solution of the PDE, depending on time  $t$  and system input  $\mathbf{x}$ .  $\mathcal{N}[\cdot; \boldsymbol{\lambda}]$  is a nonlinear operator parametrized by  $\boldsymbol{\lambda}$ .  $\Omega$  is a subset of  $\mathbb{R}^D$ .  $[0, T]$  is the time interval within which the system evolves.

For the traditional PINNs, we can define a physics informed neural network  $\mathbf{f}(t, \mathbf{x})$  as equation (15) and proceed by approximating  $\mathbf{u}(t, \mathbf{x})$  by a MLP, as illustrated in Fig. 2.

$$\mathbf{f}(t, \mathbf{x}) = \frac{\partial \mathbf{u}}{\partial t} + \mathcal{N}[\mathbf{u}; \boldsymbol{\lambda}] \quad (15)$$

As shown in Fig. 2, the MLPs used for predicting  $\mathbf{f}(t, \mathbf{x})$  shares the same parameters as the MLPs used for predicting  $\mathbf{u}(t, \mathbf{x})$ , with the distinction lying in their activation functions. The parameters common to both neural networks are optimized by minimizing the following loss function:

$$\begin{aligned} \text{loss}_l &= \text{MSE}_u + \text{MSE}_f \\ &= \frac{1}{N_u} \sum_{n=1}^{N_u} |\mathbf{u}(t_u^n, \mathbf{x}_u^n) - \mathbf{u}^n|^2 + \frac{1}{N_f} \sum_{n=1}^{N_f} |\mathbf{f}(t_f^n, \mathbf{x}_f^n)|^2 \end{aligned} \quad (16)$$



where loss  $MSE_u$  corresponds to the initial and boundary data, while  $MSE_f$  enforces the structure imposed by equation (14) at a finite set of collocation data points. The loss  $MSE_u$  is calculated over  $N_u$  initial and boundary training data points, and  $MSE_f$  is calculated over  $N_f$  collocation points.  $\mathbf{u}(t_u^n, \mathbf{x}_u^n)$  and  $\mathbf{f}(t_f^n, \mathbf{x}_f^n)$  are outputs of the PINN, while  $\mathbf{u}^n$  is the label value of the  $n$ th data point.

Considering we can usually obtain the measurement of derivatives of  $\mathbf{u}(t, \mathbf{x})$  with respect to the input  $t$  (i.e., angular frequency  $\omega$  in this work), we can also use the following loss function to train the PINN network [16]:

$$\begin{aligned} loss_{II} &= MSE_u + MSE_f \\ &= \frac{1}{N_u} \sum_{n=1}^{N_u} |\mathbf{u}(t_u^n, \mathbf{x}_u^n) - \mathbf{u}^n|^2 + |\dot{\mathbf{u}}(t_u^n, \mathbf{x}_u^n) - \dot{\mathbf{u}}^n|^2 \\ &\quad + \frac{1}{N_f} \sum_{n=1}^{N_f} |\mathbf{f}(t_f^n, \mathbf{x}_f^n)|^2 \end{aligned} \quad (17)$$

By using automatic differentiation in PyTorch, we can easily obtain the derivatives of  $\mathbf{u}(t_u^n, \mathbf{x}_u^n)$  with respect to the input  $t$ .

To improve the accuracy of the PINN-based power system dynamic model, we designed the PIKAN architecture, illustrated in Fig. 3. The primary difference from the traditional PINN is that we utilize KAN to predict  $\mathbf{u}(t, \mathbf{x})$  based on the input state  $\mathbf{x}$  and time  $t$ . In other words, the PIKAN replaces the conventional MLP with the KAN, as defined in equation (8). This PIKAN offers two advantages: 1) increased model learning accuracy, and 2) reduced network size without sacrificing accuracy, which will be demonstrated in Section IV.

#### 1) PIKAN FOR CAPTURING POWER SYSTEM DYNAMICS

When the PIKAN is used for capturing power system dynamics, we assume system parameters  $\lambda = [\mathbf{M}; \mathbf{D}; \mathbf{B}]$  in equation (5) are known. Therefore, the input of KAN is defined as  $\mathbf{x} := \mathbf{P}_m$ . By inputting  $\mathbf{P}_m$  and time period of interest to the PIKAN in Fig. 3, it can predict the voltage angle of each bus, i.e.,  $\mathbf{u} = \theta(t, \mathbf{P}_m)$ . The output of KAN is fed into the DAE module of the PIKAN to incorporate the power system dynamics model, as described by equation (5), into the neural network architecture. The training objective is to optimize the activation function  $\Phi$  to minimize loss function in equation (16) or (17). Thus, by minimising the total loss function over the KAN parameters, we can obtain the optimal KAN:

$$\Phi^* = \arg \min_{\Phi} (MSE_u + MSE_f) \quad (18)$$

Solving the above highly non-convex and multi-parameter optimization problem is challenge. We can use the LBFGS or Adam optimiser to get a solution. We refer to the PIKAN using the loss function in equation (16) as PIKAN-I, and the PIKAN using the loss function in equation (17) as PIKAN-II. In other words, PIKAN-I uses only the measurements of the voltage angle  $\theta$  to train the KAN, while PIKAN-II uses both

#### Algorithm 1 PIKAN for Capturing Power System Dynamics

---

**Data:** Power system training and test dataset generated by time domain simulation; Power system parameters (e.g.,  $\mathbf{M}$ ,  $\mathbf{D}$ , and  $\mathbf{B}$ )

**Result:** KAN parameters

Initialize KAN parameters:  $\{\Phi_l\}_{l=1}^L$ ,  $G$ , and  $k_b$ ;

Specify the loss function as equation (16) or (17);

Specify the initial & boundary training data points:  $\{(t_u^n, \mathbf{x}_u^n), \mathbf{u}^n\}_{n=1}^{N_u}$ , and specify collocation training points:  $\{(t_f^n, \mathbf{x}_f^n)\}_{n=1}^{N_f}$ ;

Specify the test points:  $\{(t_{test}^n, \mathbf{x}_{test}^n), \mathbf{u}_{test}^n\}_{n=1}^{N_{test}}$ ;

Set the maximum number of training steps  $N$ , and learning rate;

**while**  $n_{iter} < N$  **do**

Forward pass of KAN to calculate all  $\mathbf{u}(t_u^n, \mathbf{x}_u^n)$ .

If loss function (17) is adopted, further calculate  $\dot{\mathbf{u}}(t_u^n, \mathbf{x}_u^n)$  using automatic differentiation;

Calculate  $MSE_u$  based on the output of KAN and the measurements;

Calculate  $MSE_f$  based on the output of KAN and the power system dynamics given in equation (5);

Find the best KAN parameters to minimize the loss function using the LBFGS optimizer;

**if**  $n_{iter} \% 10 == 0$  **then**

Evaluate the performance of the PIKAN agent over the test points based on equation (20);

**end**

**end**

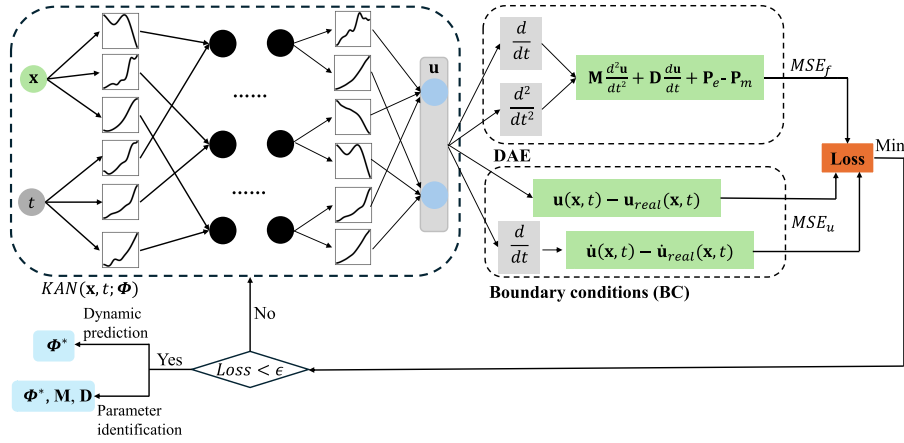
---

the voltage angle  $\theta$  and the angular frequency  $\omega$  measurements to train the KAN. The proposed PIKAN for power system dynamics can be summarized in Algorithm 1.

#### 2) PIKAN FOR POWER SYSTEM PARAMETER IDENTIFICATION

Estimating power system inertia and damping coefficients is crucial for maintaining frequency stability. With the increased installation of inverter-based resources (IBRs) in modern power systems, the inertia and damping constants can vary with the control strategies employed, potentially affecting system stability and dynamic performance. Therefore, it is necessary to frequently estimate these parameters. When the PIKAN is used for parameter identification,  $\mathbf{M}$  and  $\mathbf{D}$  in  $\lambda$  will be unknown in equation (5). The structure of the KAN remains unchanged, except that the  $\mathbf{M}$  and  $\mathbf{D}$  parameters are now considered as additional variables during the minimization of the loss function in the network training process. So, by minimising the total loss function over the KAN parameters and power system uncertain parameters, we can obtain the optimal KAN:

$$\Phi^*, \mathbf{M}^*, \mathbf{D}^* = \arg \min_{\Phi, \mathbf{M}, \mathbf{D}} (MSE_u + MSE_f) \quad (19)$$



**FIGURE 3. Physics-Informed Kolmogorov-Arnold Network (PIKAN) for power system dynamics.**

The proposed PIKAN for power system parameter identification can be summarized in **Algorithm 2** (see Appendix).

To measure the performance during the training, we defined the mean squared error (MSE) of the predictions on the test dataset as

$$MSE_t = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} |\theta_{pred,n} - \theta_n|^2 \quad (20)$$

where  $n$  is the index of the sampled test data point.  $\theta_{pred,n}$  and  $\theta_n$  are the predicted and real voltage angle vector of all the buses in the system, respectively.  $N_{test}$  is the total points of the test dataset.

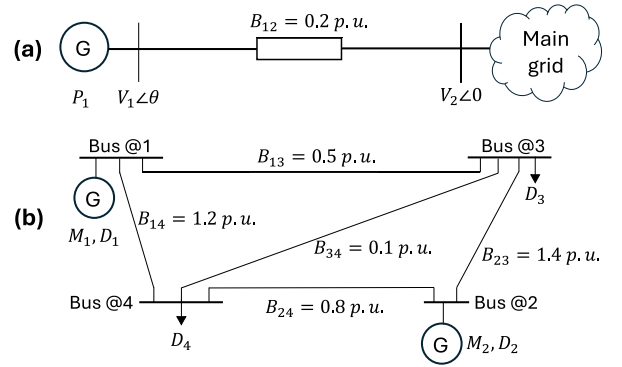
To evaluate the predictive performance of the well-trained PIKANs, we defined the relative prediction error of the voltage angle as:

$$\begin{aligned} e_\theta &= \frac{\|\theta^{0:T} - \theta_{pred}^{0:T}\|_2}{\|\theta^{0:T}\|_2} \\ &= \frac{\sqrt{\sum_{i=1}^{n_b} \sum_{t=0}^T (\theta_i^t - \theta_{pred,i}^t)^2}}{\sqrt{\sum_{i=1}^{n_b} \sum_{t=0}^T (\theta_i^t)^2}} \end{aligned} \quad (21)$$

where  $\theta^{0:T}$  and  $\theta_{pred}^{0:T}$  represent the actual and predicted voltage angles of all buses from time 0 to  $T$ , respectively.  $\theta_i^t$  and  $\theta_{pred,i}^t$  are the actual and predicted voltage angle of bus  $i$  at time  $t$ , respectively.  $\|\cdot\|_2$  is the  $l^2$  norm for finite-dimensional vectors. For the inertia and damping coefficients identification performance, we defined the relative estimation error as:

$$e_{M_i} = \frac{|M_i - M_{pred,i}|}{M_i}, \quad e_{D_i} = \frac{|D_i - D_{pred,i}|}{D_i} \quad (22)$$

where  $M_i$  and  $M_{pred,i}$  represent the actual and predicted inertia coefficients of the generator connected to bus  $i$ , respectively.  $D_i$  and  $D_{pred,i}$  represent the actual and predicted damping coefficients of bus  $i$ , respectively.



**FIGURE 4. Testing systems: (a) SMIB power system, (b) 4-bus system with two generator.**

#### IV. SIMULATION AND RESULTS

The performance of the proposed PIKANs for frequency dynamics was firstly demonstrated on a SMIB power system and a 4-bus 2-generator system, as shown in Fig. 4. The testing power system parameters are presented in Table 1 and Fig. 4. To generate the training and test datasets, we utilized time domain simulations implemented with SciPy in Python. The generated frequency dynamic data is with a time step of 0.1s over time window  $[0, T]$  for each trajectory. Each data trajectory represents the response of the testing system to a random disturbance. In the SMIB system, we assume initial values for  $\theta_1$  and  $\omega_1$  to be 0.1 rad and 0.1 rad/s, respectively. The value of  $P_{m1}$  ranges between 0.08 p.u. and 0.18 p.u., within which the SMIB system remains stable. In this case setting, we generated 100 trajectories under different  $P_{m1}$  in the range of  $[0.08, 0.18]$  p.u., using increments of  $\Delta P = 0.001$  p.u. For each trajectory, the training and test datasets consist of time intervals from 0 to 20 seconds with a 0.1-second step, including the corresponding  $\theta$  and  $\omega$  values at each time step and the corresponding power injection value  $P_{m1}$ . As a result, our entire training and test datasets consist

**TABLE 1. Parameters of the case studies.**

Parameters	SMIB system		4-bus 2-generator system	
	$M$ (p.u.)	$D$ (p.u.)	$M$ (p.u.)	$D$ (p.u.)
Bus @1	0.4	0.15	0.3	0.15
Bus @2	—	—	0.2	0.3
Bus @3	—	—	0	0.25
Bus @4	—	—	0	0.2

Note: The line parameters of the testing systems can be found in Fig. 4.

of 20,100 samples. For the 4-bus 2-generator system, similar to the setup in reference [16], we assume the system is in equilibrium at  $t = 0$ . We then perturb the system with a constant input signal  $\mathbf{P}_m = a \times [0.1, 0.2, -0.1, -0.2]$  p.u. for  $t > 0$  in each trajectory. We generated 19 trajectories, with  $a$  ranging from 0.5 to 9.5 in increments of 0.5. For each trajectory, the training and test datasets consist of time intervals from 0 to 5 seconds with a 0.1-second step, including the corresponding  $[\theta_1, \theta_2, \theta_3, \theta_4]$  values at each time step and the corresponding input signal  $\mathbf{P}_m$ . All the datasets were normalized to the range of  $[-1, 1]$  before feeding into the neural networks. We conducted PIKANs training and performance testing in PyTorch on an Intel Xeon(R) Gold 6248R CPU @3.00 GHz  $\times$  48 Windows based server with 64 GB RAM.

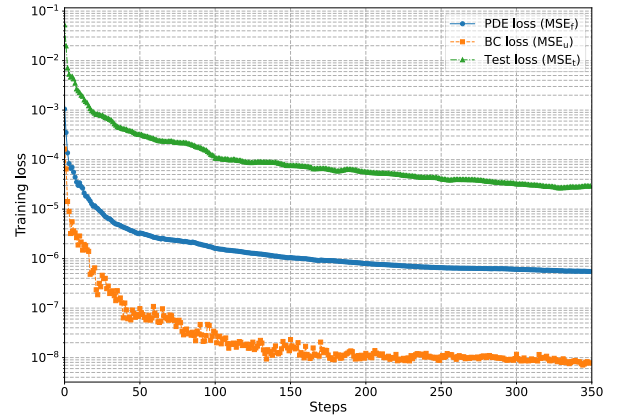
#### A. DATA-DRIVEN SOLUTION OF FREQUENCY DYNAMICS

In the study of capturing frequency dynamics, the inertia and damping coefficients of the testing systems are known parameters. We evaluated the capability of the PIKANs to accurately predict trajectories of  $\theta$  and  $\omega$  for uncertain power injections.

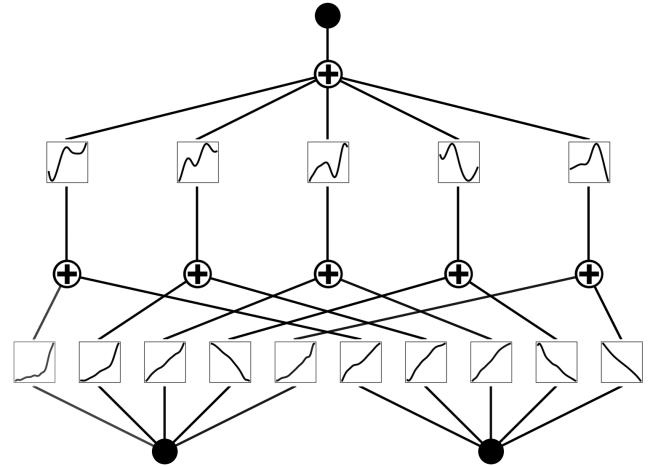
##### 1) SMIB SYSTEM

For the SMIB system, we used a 2-layer KAN with a shape of  $[2, 5, 1]$ . In each training step, the randomly sampled time  $t$  and power injection  $P_{m1}$  were fed into the KAN, and trained to minimize the loss function in equation (16) for the PIKAN-I algorithm (or equation (17) for the PIKAN-II algorithm). For both PIKAN-I and PIKAN-II algorithms, the intervals of the B-spline were set to  $G = 10$ , and the order of the B-spline was set to  $k_b = 3$ . We set  $N_u = 40$ ,  $N_f = 800$ , and  $N_{test} = 20, 100$ . The training convergence process of the PIKAN-I algorithm is depicted in Fig. 5. It shows that the PIKAN-I converges quickly and achieves lower losses within hundreds of training steps. Fig. 6 illustrates the architecture of the well-trained PIKAN-I agent. It can be observed that the activation functions for each edge are uniquely tailored, having been adjusted based on the training data to achieve optimal estimation performance.

Fig. 7 depicts the comparison between the PIKAN-I predicted and the actual trajectory of the angle  $\theta$  and the angular frequency  $\omega$  of bus 1 in the SMIB system. The angular frequency  $\omega$  in the figure was calculated by differentiating the signal associated with the voltage angle  $\theta$ . In the left



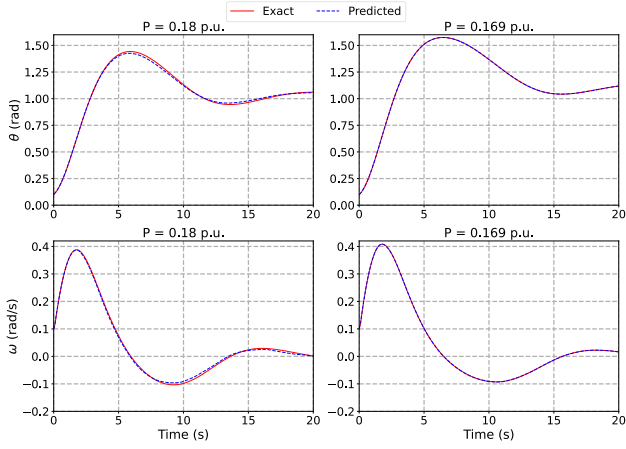
**FIGURE 5. Training convergence process of the PIKAN-I algorithm for capturing SMIB system frequency dynamics. The LBFGS optimizer was employed, with parameter maximum iteration set to 20. Thus, each optimization step in the figure contains 20 iterations.**



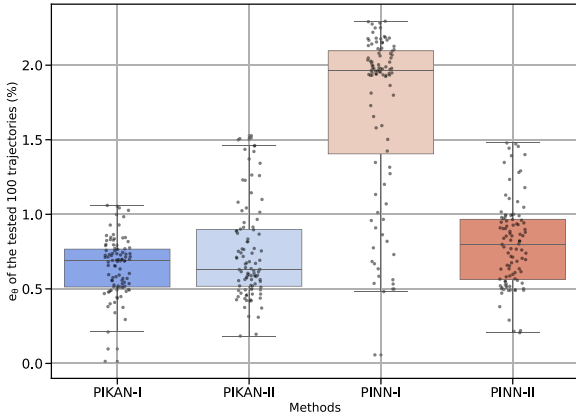
**FIGURE 6. Illustration of the architecture of the well-trained PIKAN-I for the SMIB power system.**

figures of Fig. 7, we present the least accurate estimation of the voltage angle and frequency trajectory, yielding a relative prediction error ( $e_\theta$ ) of 1.06%. Conversely, in the right figures, we demonstrate the most accurate estimation of the voltage angle and frequency trajectory, achieving a relative prediction error ( $e_\theta$ ) of 0.014%. The median value of the prediction error on voltage angle over the 100 trajectories is 0.688%, which indicate that the PIKAN-I is able to predict the trajectory of the angle with high accuracy.

If we use measurements of both  $\theta$  and  $\omega$  to train the PIKAN, denoted as the PIKAN-II algorithm, the accuracy of the agent can be further improved, with the median value of the prediction error on the voltage angle decreasing to 0.633% (see Table 2). We also compared the performance of the proposed method with the MLP-based PINNs for power systems proposed in [14] and [16]. The prediction errors for the 100 tested trajectories are presented in Fig. 8 and



**FIGURE 7.** Comparison of the predicted and exact solution for the voltage angle and frequency with the PIKAN-I for SMIB power system dynamics.

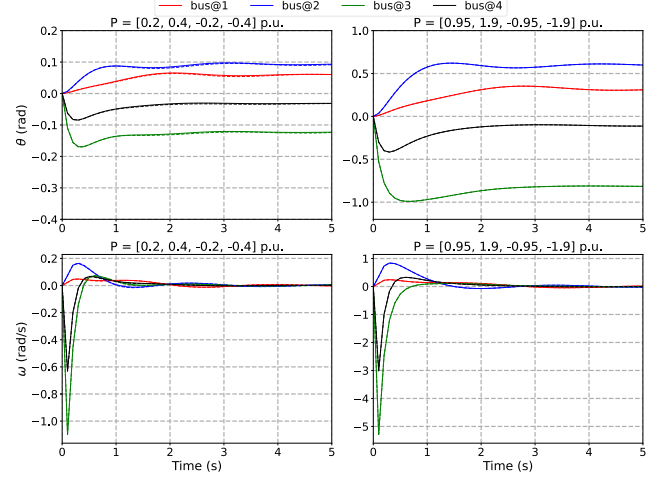


**FIGURE 8.** Performance of the proposed method and the MLPs based PINNs for the SMIB system. The parameters and hyperparameters setting is same with reference [14]. For the PINN-I and PINN-II, we set  $N_u = 40$  and  $N_f = 8,000$ .

Table 2. The proposed method outperforms the traditional PINNs, demonstrating the effectiveness of the PIKANs in learning the dynamics of the SMIB system. From the results in Fig. 8, we can observe that incorporating measurements of  $\omega$  (i.e., using the loss function defined in equation (17)) during training improves the performance of the agent for both the PIKAN and traditional PINN methods.

## 2) 4-BUS 2-GENERATOR SYSTEM

To further test the performance of the proposed method in capturing the dynamics of multi-machine power systems, we evaluated it on a 4-bus 2-generator system as shown in Fig. 4(b). For this case study, we employed a 2-layer KAN with a structure of [5, 10, 4]. In each training step, the randomly sampled time  $t$  and power injection  $[P_{m1}, P_{m2}, P_{m3}, P_{m4}]$  are fed into the KAN, which is then trained to minimize the loss function in equation (16) or (17), ultimately outputting the voltage angles of the four buses at time  $t$ . For



**FIGURE 9.** Comparison of the predicted and exact solutions for the voltage angle and frequency using PIKAN-II for the 4-bus 2-generator power system dynamics. Solid lines represent the exact trajectory, while dashed lines represent the predicted trajectory by PIKAN-II. The  $\theta$  and  $\omega$  in the figure represent the changes relative to the initial voltage angle and frequency values, respectively.

both PIKAN-I and PIKAN-II, the intervals of the B-spline were set to  $G = 5$ , and the order of the B-spline was set to  $k_b = 3$ . We set  $N_u = 80$ ,  $N_f = 4000$ , and  $N_{test} = 969$ .

Fig. 9 depicts the comparison between the predicted and the actual trajectory of the angle  $[\theta_1, \theta_2, \theta_3, \theta_4]$  and the frequency  $[\omega_1, \omega_2, \omega_3, \omega_4]$  of the four buses in the system. In the left figures of Fig. 9, we present the least accurate estimation of the voltage angle and frequency trajectory, yielding a relative prediction error ( $e_\theta$ ) of 1.94%. Conversely, in the right figures, we demonstrate the most accurate estimation of the voltage angle and frequency trajectory, achieving a relative prediction error ( $e_\theta$ ) of 0.04%. The median value of the estimation error on voltage angle over the 19 trajectories is 0.538%, indicating that PIKAN-II can predict the trajectory of the angle with high accuracy. In contrast, the traditional PINN-I and PINN-II algorithms performed much worse, with median estimation errors on the voltage angle of 5.03% and 2.59%, respectively. The performance comparisons between PIKANs and PINNs are summarized in Table 2. The results on the 4-bus 2-generator system also demonstrate that the proposed methods outperform traditional PINN-based approaches.

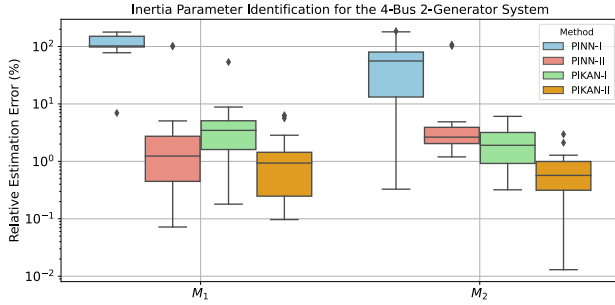
## B. DATA-DRIVEN INERTIA AND DAMPING COEFFICIENTS IDENTIFICATION

In the parameter identification study, the inertia and damping coefficients of the testing systems are unknown parameters—reflecting situations where system parameters may be inaccurately known or subject to variations due to control strategies of IBRs. We assessed the capability of PIKANs to accurately estimate these unknown parameters from observed trajectories.

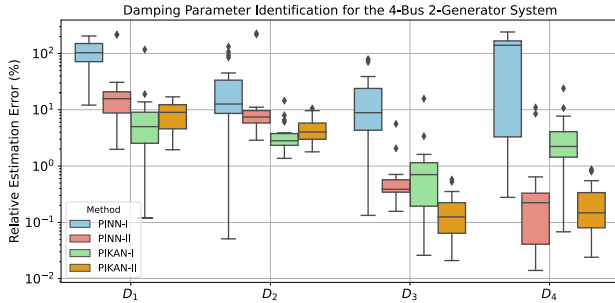


**TABLE 2. Dynamic capturing study results: Estimation error of the trajectory of  $\theta(t)$ .**

Estimation error	SMIB system			4-bus 2-generator system		
	Max (%)	Min (%)	Median (%)	Max (%)	Min (%)	Median (%)
PIKAN-I	1.06	0.014	0.688	4.85	0.043	4.64
PIKAN-II	1.53	0.184	0.633	1.94	0.040	0.538
PINN-I ([14])	2.30	0.057	1.96	6.35	0.151	5.03
PINN-II ([16])	1.48	0.206	0.800	5.98	0.076	2.59

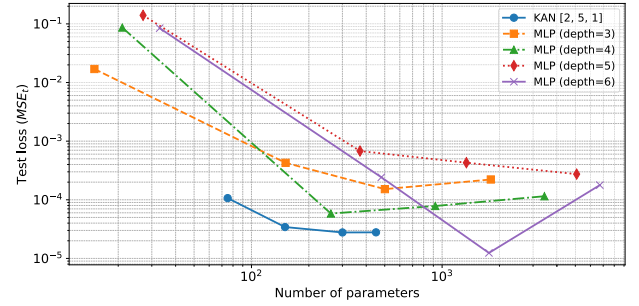


**FIGURE 10. Inertia coefficients estimation errors of PIKANs and PINNs.  $M_1$  and  $M_2$  are inertia coefficients of the two generators in the 4-bus 2-generator system, respectively.**



**FIGURE 11. Damping coefficients estimation errors of PIKANs and PINNs.  $D_1$ ,  $D_2$ ,  $D_3$  and  $D_4$  are damping coefficients of the four buses in the 4-bus 2-generator system, respectively.**

The parameters and hyperparameters of the PIKANs for assessing inertia and damping coefficients are the same as those of the PIKANs agents in Section IV-A. Since the neural network's weights are initialized randomly, we run each estimation of the four algorithms 20 times. Figs. 10 and 11 show the distribution of parameter estimation errors on the 4-bus 2-generator system for the proposed methods and the comparison methods. PIKAN-I achieves a median relative error below 10% for evaluating the inertia and damping coefficients of the system. In contrast, PIKAN-II demonstrates significantly better performance, achieving a median relative error of around 1% for inertia coefficients and 0.1% for several damping coefficients. The traditional PINNs, however, perform much worse than the proposed methods. For instance, the traditional PINNs can result in estimation errors as high as 110%, as shown in the figures. Additionally, we observed that incorporating measurements of  $\omega$  during



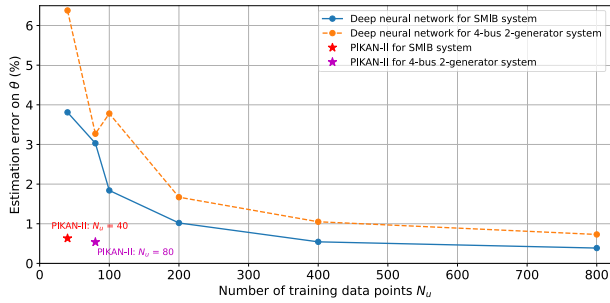
**FIGURE 12. Scaling laws of losses against the number of parameters for different physics-informed neural networks applied to the SMIB system.**

training can significantly improve the parameter estimation accuracy of the agent for both the PIKAN and traditional PINN methods.

### C. NUMBER OF NETWORK PARAMETERS VS. PIKAN PERFORMANCE

Results in Tables 2 and 3 indicate that, for the SMIB case, PIKANs achieved greater accuracy in grid dynamic learning while using only 41% of the network size of the PINNs. Similarly, for the 4-bus 2-generator case, PIKANs achieved higher accuracy while utilizing only 58% of the network size of the PINNs.

For a  $L$ -layer KAN with layers of equal width  $N$ , there are in total  $O(N^2LG)$  parameters, and an MLP only needs  $O(N^2L)$  parameters for the same number of layers and width. However, KANs typically achieve similar performance with a much smaller  $N$  compared to MLPs. Fig. 12 illustrates the scaling laws of losses as a function of the number of parameters in both PIKANs and PINNs. The results demonstrate that KANs exhibit steeper scaling laws than MLPs. This implies that PIKANs can achieve comparable or even superior accuracy in power system dynamic learning with fewer parameters than PINNs. The implications of these findings are significant. They suggest that while KANs may initially seem to require more parameters due to the  $O(N^2LG)$  scaling, their ability to use a smaller  $N$  effectively reduces the overall parameter count needed for high performance. Consequently, PIKANs offer a more efficient and scalable solution for complex learning tasks in power systems, surpassing traditional MLP-based PINNs in terms of both accuracy and parameter efficiency.



**FIGURE 13.** Performance comparison of traditional DNNs and PIKANs with varying numbers of training data points. For the SMIB case, a [2, 10, 10, 10, 10, 1] DNN architecture was employed. For the 4-bus system, a [5, 30, 30, 4] DNN architecture was utilized, and the Adam optimizer was employed.

#### D. REDUCED DATA DEPENDENCY COMPARED TO DNN

PIKANs introduce a KAN training framework designed to leverage the inherent dynamics of power systems. Consequently, compared to traditional DNNs without a physics-informed architecture, PIKANs can significantly reduce the required size of the training dataset. For the two testing systems in Table 2, the performance of traditional DNNs, employing identical architecture and parameters as the PINNs, varies with the number of training data points, as illustrated in Fig. 13. From the results, it is observed that PIKANs can achieve similar or even better performance while requiring only 10% of the training data points compared to traditional DNNs.

#### E. PERFORMANCE ON IEEE NEW ENGLAND POWER SYSTEM

To further demonstrate the effectiveness of the proposed methods on larger power systems, case studies were conducted on the IEEE New England 10-machine 39-bus power system. The system parameters are referenced from [29]. The system is represented in its Kron-reduced form, with dynamics described by equation (1). The initial bus angles,  $\theta_i(0)$ , were randomly generated, uniformly distributed in the range of  $[-0.01, 0.01]$  rad [29].

A total of 80 trajectories were generated under different initial system states. For each trajectory, the training and test datasets were constructed over a 5-second interval with a time step of 0.05 seconds. A 3-layer KAN with architecture [11, 5, 10] was used, and the PIKAN hyperparameters were set as the same with those used in the 4-bus 2-generator case. The initial bus angles of the 10 buses and the time  $t$  are fed into the neural networks as inputs. The PIKAN model has 840 trainable parameters. When applied to capture the system's frequency dynamics following disturbances (training convergence shown in Fig. 14), the median estimation errors for voltage angle achieved by PIKAN-I and PIKAN-II were 2.72% and 1.55%, respectively. For comparison, a traditional PINN with architecture [11, 32, 32, 10] was tested after evaluating performance across hidden layers with 32, 64, and

128 neurons. This PINN model has 1,770 trainable parameters. The median estimation errors for voltage angle with PINN-I and PINN-II were 4.83% and 4.68%, respectively. These results indicate that the proposed method achieves lower estimation error while utilizing only 47.5% of the network size of the traditional PINNs.

#### V. DISCUSSION

This paper introduces, for the first time in power systems, a KAN-based PINN (i.e., PIKAN) architecture that explicitly considers the physical laws describing the frequency dynamics of grids. As a promising alternative to traditional MLP-based PINNs, the proposed PIKANs for power system dynamics can achieve higher accuracy with fewer neural network parameters compared to MLP-based PINNs. The advantage of PIKANs is particularly significant given the challenges of training large neural networks, such as large language models (LLMs), which are resource-intensive and consume substantial amounts of energy. This opens up numerous opportunities in power systems, as PIKANs can potentially be used to accurately and efficiently solve DAEs in power grids. In addition, PIKANs require only a very limited amount of training data. For instance, for the SMIB system, PIKANs need only  $N_u = 40$  points  $\{(t_u^n, \mathbf{x}_u^n), \mathbf{u}^n\}_{n=1}^{N_u}$  to train the agent. Even for a larger power system, such as the 4-bus 2-generator system, PIKANs still need only  $N_u = 80$  training data points. Although we require significantly more collocation points (for example,  $N_f = 800$  for the SMIB case) to evaluate the  $MSE_f$  term in the loss function given in equation (16), it is important to note that this evaluation is not dependent on measured voltage angle and angular frequency data. This means we can generate any number of collocation points without needing to produce labels for those data points.

Similar with traditional PINNs, PIKANs have the capability to directly compute the voltage angle at any given time step  $t_1$ . In contrast, numerical methods must integrate starting from the initial conditions at  $t = t_0$  and proceed sequentially to reach  $t = t_1$ . This provides significant advantages over traditional numerical integration methods.

For inertia and damping coefficient identification using the PIKAN-based method, the inputs of the KAN network are  $t$  and  $\mathbf{P}_m$ . In real-world applications, PIKAN models would be trained using historical data from the power system. However, in cases where measurements of  $\mathbf{P}_m$  are not readily available due to the limitations in the current measurement infrastructure, practical approaches must be employed to approximate  $\mathbf{P}_m$  using observable system data and state estimation techniques.

In this study, our primary focus was on exploring how PIKANs could achieve higher accuracy in learning power system dynamics while maintaining a smaller network size. Theoretically, KANs outperform MLPs in terms of accuracy, interpretability, and reduction of catastrophic forgetting. Nevertheless, to fully harness the potential of PIKANs, several challenges must be addressed.

**TABLE 3. Training time for results given in Table 2.**

System	Methods	Network layers	Order of B-spline ( $k_b$ )	Intervals of B-spline ( $G$ )	No. of parameters	Training iterations	Training time (ms/iter.)
SMIB system	PIKAN-I	[2, 5, 1]	3	10	195	7000	87.5
	PIKAN-II	[2, 5, 1]	3	10	195	7000	130
	PINN-I ([14])	[2, 10, 10, 10, 10, 10, 1]	–	–	481	50000	0.54
	PINN-II ([16])	[2, 10, 10, 10, 10, 10, 1]	–	–	481	10000	3.41
4-Bus 2-Generator System	PIKAN-I	[5, 10, 4]	3	5	720	3000	1225
	PIKAN-II	[5, 10, 4]	3	5	720	3000	1390
	PINN-I ([14])	[5, 30, 30, 4]	–	–	1234	50000	2.89
	PINN-II ([16])	[5, 30, 30, 4]	–	–	1234	50000	3.78

- 1) *Training and computing time*: From the results presented in Table 3, it is evident that the training of PIKANs requires considerably more time compared to PINNs. Liu et al. [17] attribute this slower training to the inefficiency of current activation functions in batch computations. Despite the extended training duration, the superior performance and accuracy of PIKANs may justify the additional time investment, especially in scenarios requiring high precision. After offline training, we evaluated the PIKAN's performance based on its online computational speed required to solve the DAE defined by equation (5). For 19 different initial conditions of the 4-bus 2-generator system, the ode45 solver averages 0.017 seconds to solve the swing equations across the time interval from 0 to 5 seconds, whereas PIKAN averages 0.024 seconds. In future research, we aim to explore techniques utilizing more efficient activation functions, such as Jacobi polynomials proposed by [30], to substantially enhance training speeds. And primary investigation in [31] demonstrates that Jacobi polynomials can reduce training times by two orders of magnitude compared to KANs using B-spline activation functions in the context of solving specific PDEs.
- 2) *Accuracy*: Our simulation results demonstrated that KAN-based PINNs exhibit higher accuracy compared to MLP-based PINNs in modeling power system dynamics. Researchers have also found that KANs generally achieve greater accuracy than MLPs in most PDE problems [32]. However, whether KANs consistently outperform MLPs in various power dynamic problems requires further investigation. Additionally, understanding why PIKANs have higher accuracy than conventional PINNs warrants further exploration. One possible reason could be that KANs employ learnable activation functions, allowing for more complex learned activations compared to the fixed activation functions (such as ReLU) used in MLPs.
- 3) *Interpretability*: KANs have the potential to serve as foundational models for AI + Science due to their accuracy and interpretability [17]. With KANs,

humans can interactively obtain the symbolic formula of the model's output, which significantly facilitates the analysis of complex physical systems, such as the dynamics of bulk power systems. However, in the case of the swing equations examined in this study, we observed that the symbolic formula provided by the well-trained PIKAN does not accurately capture the frequency dynamics of the testing power systems, despite the PIKAN model itself precisely predicting the grid dynamics. This discrepancy may stem from the limited library of symbolic formulas available in the current version of KAN package in [33], or perhaps the formula for the grid dynamics is not inherently symbolic.

- 4) *Continual learning*: One drawback of MLPs is their tendency to forget previously learned tasks when transitioning from one task to another. Liu et al. [17] demonstrate that for a 1D regression task, KANs exhibit local plasticity and can prevent catastrophic forgetting by leveraging the locality of splines. However, the extent to which KANs can avoid catastrophic forgetting in more complex learning tasks, such as power system dynamics as explored in this study, remains unclear. In our investigation, we observed that a well-trained PIKAN, initially trained on data from stable scenarios (e.g.,  $P_{m1} \in [0.08, 0.18]$  p.u. for the SMIB case), tends to forget previously learned dynamics when further trained on dynamics from unstable scenarios (e.g.,  $P_{m1} \in [0.20, 0.25]$  p.u. for the SMIB case). Therefore, further investigation into the continual learning capabilities of the proposed PIKANs is warranted in future research.

## VI. CONCLUSION

This is the first paper to propose physics-informed KANs for power system applications. By integrating KAN with PINN, we achieve higher accuracy in solving the differential-algebraic equations of power systems with smaller neural network size compared to traditional MLP-based PINNs. In our case studies, we showcased the effectiveness of the proposed PIKANs in accurately capturing

**Algorithm 2** PIKAN for Grid Parameter Identification

**Data:** Power system training and test dataset generated by time domain simulation

**Result:** KAN parameters and estimated inertia  $\mathbf{M}$  and damping  $\mathbf{D}$  parameters

Initialize KAN parameters:  $\{\Phi_l\}_{l=1}^L$ ,  $G$ , and  $k_b$ ;

Initialize inertia  $\mathbf{M}$  and damping  $\mathbf{D}$  parameters;

Specify the loss function as equation (16) or (17);

Specify the initial & boundary training data points:

$\{(t_u^n, \mathbf{x}_u^n), \mathbf{u}_u^n\}_{n=1}^{N_u}$ , and specify collocation training points:

$\{(t_f^n, \mathbf{x}_f^n)\}_{n=1}^{N_f}$ ;

Specify the test points:  $\{(t_{test}^n, \mathbf{x}_{test}^n), \mathbf{u}_{test}^n\}_{n=1}^{N_{test}}$ ;

Set the maximum number of training steps  $N$ , and learning rate;

**while**  $n_{iter} < N$  **do**

Forward pass of KAN to calculate all  $\mathbf{u}(t_u^n, \mathbf{x}_u^n)$ .

If loss function (17) is adopted, further calculate  $\dot{\mathbf{u}}(t_u^n, \mathbf{x}_u^n)$  using automatic differentiation;

Calculate  $MSE_u$  based on the output of KAN and the measurements;

Calculate  $MSE_f$  based on the output of KAN and the power system dynamics given in equation (5);

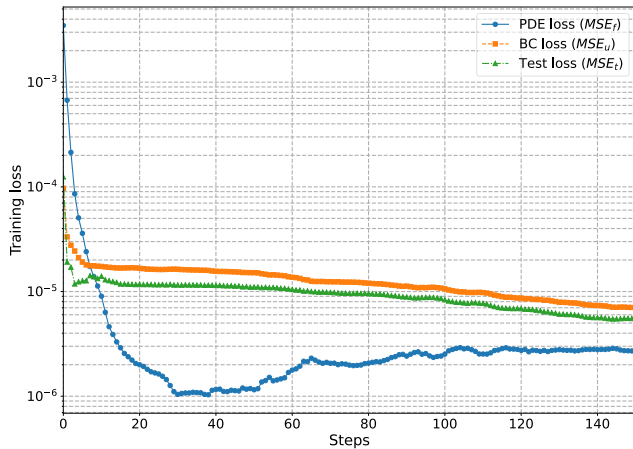
Find the best KAN parameters and inertia  $\mathbf{M}$  and damping  $\mathbf{D}$  parameters to minimize the loss function using the LBFGS optimizer;

**if**  $n_{iter} \% 10 == 0$  **then**

Evaluate the performance of the PIKAN agent over the test points based on equation (20);

**end**

**end**



**FIGURE 14.** Training convergence process of the PIKAN-I algorithm for capturing IEEE New England power system frequency dynamics. The LBFGS optimizer was employed, with parameter maximum iteration set to 20. Thus, each optimization step in the figure contains 20 iterations.

the dynamics of power systems. Furthermore, we demonstrated their capability to identify uncertain system inertia and

damping parameters, with high accuracy using a limited set of training data points. Using the proposed method reduces the median value of the inertia coefficient estimation error from approximately 100% to 5% compared to traditional PINNs. These results underscore the promising potential of the PIKANs for practical applications in power systems, opening up new avenues for their use.

**APPENDIX I.**

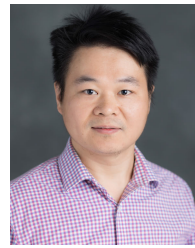
See Algorithm 2 and Fig. 14.

**REFERENCES**

- [1] H. Li, "Deep learning for natural language processing: Advantages and challenges," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 24–26, Jan. 2018.
- [2] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, Aug. 2021.
- [3] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps: Automation of Decision Making*. Cham, Switzerland: Springer, pp. 323–350, doi: 10.1007/978-3-319-65981-7\_12.
- [4] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access J. Power Energy*, vol. 7, pp. 376–388, 2020.
- [5] C. Ren, Y. Xu, and R. Zhang, "An interpretable deep learning method for power system transient stability assessment via tree regularization," *IEEE Trans. Power Syst.*, vol. 37, no. 5, pp. 3359–3369, Sep. 2022.
- [6] Y. Du and F. Li, "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1066–1076, Mar. 2019.
- [7] H. Shuai and H. He, "Online scheduling of a residential microgrid via Monte-Carlo tree search and a learned model," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1073–1087, Mar. 2021.
- [8] F. Li and Y. Du, "From AlphaGo to power system AI: What engineers can learn from solving the most complex board game," *IEEE Power Energy Mag.*, vol. 16, no. 2, pp. 76–84, Mar. 2018.
- [9] G. Gurrula, A. Dimitrovski, S. Pannala, S. Simunovic, and M. Starke, "Parareal in time for fast power system dynamic simulations," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1820–1830, May 2016.
- [10] S. Sinha, S. P. Nandanoori, and E. Yeung, "Data driven online learning of power system dynamics," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2020, pp. 1–5.
- [11] R. Satheesh, N. Chakkungal, S. Rajan, M. Madhavan, and H. H. Alhelou, "Identification of oscillatory modes in power system using deep learning approach," *IEEE Access*, vol. 10, pp. 16556–16565, 2022.
- [12] N. Bhusal, R. M. Shukla, M. Gautam, M. Benidris, and S. Sengupta, "Deep ensemble learning-based approach to real-time power system state estimation," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106806.
- [13] Y. Zhang, X. Shi, H. Zhang, Y. Cao, and V. Terzija, "Review on deep learning applications in frequency analysis and control of modern power system," *Int. J. Electr. Power Energy Syst.*, vol. 136, Mar. 2022, Art. no. 107744.
- [14] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, "Physics-informed neural networks for power systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2020, pp. 1–5.
- [15] G. S. Misyris, J. Stiasny, and S. Chatzivasileiadis, "Capturing power system dynamics by physics-informed neural networks and optimization," in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, Dec. 2021, pp. 4418–4423.
- [16] J. Stiasny, G. S. Misyris, and S. Chatzivasileiadis, "Physics-informed neural networks for non-linear system identification for power system dynamics," in *Proc. IEEE Madrid PowerTech*, Jun. 2021, pp. 1–6.
- [17] Z. Liu et al., "KAN: Kolmogorov–Arnold networks," 2024, *arXiv:2404.19756*.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [19] T. Chen and H. Chen, "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems," *IEEE Trans. Neural Netw.*, vol. 6, no. 4, pp. 911–917, Jul. 1995.



- [20] Y. Lu and J. Lu, "A universal approximation theorem of deep neural networks for expressing probability distributions," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2020, pp. 3094–3105.
- [21] R. Kemker, M. McClure, A. Abitino, T. L. Hayes, and C. Kanan, "Measuring catastrophic forgetting in neural networks," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2018, vol. 32, no. 1, pp. 3390–3398.
- [22] F.-L. Fan, J. Xiong, M. Li, and G. Wang, "On interpretability of artificial neural networks: A survey," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 6, pp. 741–760, Nov. 2021.
- [23] H. Shuai, B. She, J. Wang, and F. Li, "Safe reinforcement learning for grid-forming inverter based frequency regulation with stability guarantee," *J. Mod. Power Syst. Clean Energy*, vol. 2024, pp. 1–8, Apr. 2024.
- [24] V. Vittal, J. D. McCalley, P. M. Anderson, and A. Fouad, *Power System Control and Stability*, vol. 3, 3rd ed., Hoboken, NJ, USA: Wiley, 2019.
- [25] A. N. Kolmogorov, *On the Representation of Continuous Functions of a Smaller Number of Variables by Superpositions of Continuous Functions of a Smaller Number of Variables*. Providence, RI, USA: American Mathematical Society, 1961.
- [26] T. Poggio, A. Banburski, and Q. Liao, "Theoretical issues in deep networks," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 48, pp. 30039–30045, Jun. 2020.
- [27] Z. Bozorgasl and H. Chen, "Wav-KAN: Wavelet Kolmogorov–Arnold networks," 2024, *arXiv:2405.12832*.
- [28] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (Part I): Data-driven solutions of nonlinear partial differential equations," 2017, *arXiv:1711.10561*.
- [29] W. Cui, Y. Jiang, and B. Zhang, "Reinforcement learning for optimal primary frequency control: A Lyapunov approach," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1676–1688, Mar. 2023.
- [30] SynodicMonth. (2024). *ChebyKAN*. [Online]. Available: <https://github.com/SynodicMonth/ChebyKAN>
- [31] K. Shukla, J. D. Toscano, Z. Wang, Z. Zou, and G. E. Karniadakis, "A comprehensive and FAIR comparison between MLP and KAN representations for differential equations and operator networks," 2024, *arXiv:2406.02917*.
- [32] Y. Wang et al., "Kolmogorov Arnold informed neural network: A physics-informed deep learning framework for solving forward and inverse problems based on Kolmogorov Arnold networks," 2024, *arXiv:2406.11045*.
- [33] KindXiaoming. (2024). *PyKAN*. Accessed: Jun. 2024. [Online]. Available: <https://github.com/KindXiaoming/pykan>



**HANG SHUAI** (Member, IEEE) received the B.Eng. degree from Wuhan Institute of Technology (WIT), Wuhan, China, in 2013, and the Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, in 2019. He was a Visiting Student Researcher with the University of Rhode Island (URI), Kingston, RI, USA, from 2018 to 2019. He was a Postdoctoral Researcher with URI, and The University of Tennessee (UTK), Knoxville, TN, USA, from 2019 to 2022. Currently, he is a Research Assistant Professor with UTK. His research interests include reinforcement learning for power systems, microgrid operation and control, and bulk power system resilience.



**FANGXING LI** (Fellow, IEEE) received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree from Virginia Tech, Blacksburg, VA, USA, in 2001. Currently, he is the John W. Fisher Professor in electrical engineering with The University of Tennessee, Knoxville, TN, USA. He has been the Director of CURENT, since 2024. His research interests include resilience, artificial intelligence in power, distributed energy resources (DERs) and microgrid, and electricity markets. He has received several awards and honors including R&D 100 Award in 2020, the IEEE PES Technical Committee Prize Paper Award (twice) in 2019 and 2024, six best or prize paper awards at international journals, and seven best papers/posters at international conferences. From 2020 to 2021, he served as the Chair for the IEEE PES Power System Operation, Planning and Economics (PSOPE) Committee. Currently, he is serving as the Chair for IEEE WG on Machine Learning for Power Systems. He is the Editor-in-Chief of IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY. He is also known as Fran Li.

...