

---

# Stable Port-Hamiltonian Neural Networks

---

Fabian J. Roth, Dominik K. Klein, Maximilian Kannapinn, Jan Peters, Oliver Weeger

Technical University of Darmstadt, Darmstadt, Germany

{roth,klein,kannapinn,weeger}@cps.tu-darmstadt.de,  
peters@ias.tu-darmstadt.de

## Abstract

In recent years, nonlinear dynamic system identification using artificial neural networks has garnered attention due to its broad potential applications across science and engineering. However, purely data-driven approaches often struggle with extrapolation and may yield physically implausible forecasts. Furthermore, the learned dynamics can exhibit instabilities, making it difficult to apply such models safely and robustly. This article introduces stable port-Hamiltonian neural networks, a machine learning architecture that incorporates physical biases of energy conservation and dissipation while ensuring global Lyapunov stability of the learned dynamics. Through illustrative and real-world examples, we demonstrate that these strong inductive biases facilitate robust learning of stable dynamics from sparse data, while avoiding instability and surpassing purely data-driven approaches in accuracy and physically meaningful generalization. Furthermore, the model’s applicability and potential for data-driven surrogate modeling are showcased on multiphysics simulation data.

## 1 Introduction

Recent years have seen a strongly growing interest in using machine learning (ML) to identify the dynamics of physical systems directly from observations [5, 28, 31, 32, 40, 42, 53]. This approach has diverse applications, ranging from modeling phenomena with partially or entirely unknown underlying physics to creating surrogate and reduced-order models for rapid simulations and digital twins [23, 53, 54]. In the latter application, data-driven methods can address the drawbacks of traditional numerical methods, which often require tremendous computational resources and expertise [1, 53]. However, in turn, classical ML algorithms typically require large amounts of training data and lack robustness, as they often suffer from poor extrapolation and are prone to making physically implausible forecasts [1, 33, 53, 54]. To overcome these limitations, a general consensus is to avoid discarding centuries’ worth of established knowledge in physics, which is an implicit consequence of relying solely on data-driven methods. Instead, there is a growing emphasis on incorporating physical priors and inductive biases into ML approaches [33, 54]. This *physics-guided* ML promises improved accuracy, robustness, and interpretability while reducing data requirements [33, 53].

Several physics-guided architectures have been proposed for learning dynamical systems. These approaches often parameterize an ordinary differential equation (ODE) to identify continuous-time dynamics. To ensure adherence to the laws of thermodynamics, many architectures build upon frameworks such as Hamiltonian mechanics [9, 16, 20], Lagrangian mechanics [10, 34], Poisson systems [19], GENERIC [18, 56], port-Hamiltonian systems (PHS) [11, 13, 37, 38, 57], or the generalized Onsager principle [55]. The latter three allow for the modeling of dissipative dynamics, while the former are generally limited to energy-conserving systems. Additionally, PHS explicitly incorporate external excitations, making them particularly suited for modeling controlled systems.

The mentioned physical modeling frameworks have implications for the stability of the learned dynamics. For example, in Hamiltonian neural networks, stable equilibria correspond to extrema in

the Hamiltonian energy function. However, these effects are often not discussed in the referenced works. Nevertheless, stability is an important and fundamental property of any dynamic system. Guarantees on the stability of the dynamics may improve the model’s robustness, help trust their predictions, and can themselves be viewed as physically motivated biases [14].

Given the importance of the stability property, multiple other works have proposed methods to enhance the stability of learned dynamics, often using neural ordinary differential equations (NODEs) [7] as a basis. Massaroli et al. [36] propose a provably stable NODE variant that guarantees the existence of locally asymptotically stable equilibria. Multiple studies have combined Lyapunov’s theory with neural networks (NNs) to guarantee the stability of discrete-time dynamics [14, 29] or to estimate the region of attraction for given dynamical systems [3, 44]. Kolter and Manek [27] describe another Lyapunov-stable architecture based on NODEs. They concurrently learn unconstrained dynamics with a NODE and a convex Lyapunov function. Global stability around an equilibrium point is then ensured by projecting the NODE-dynamics onto the space of stable dynamics as given by the Lyapunov function. Takeishi and Kawahara [50] extend this approach from a stable equilibrium to stable invariant sets. Similar projection-based techniques have also been proposed to learn input-output stable dynamics [26] and dissipative systems [39]. However, these projection methods can introduce discontinuities, hindering training strategies that rely on integrating trajectories rather than matching state-derivative pairs [46]. Other approaches employ physical modeling frameworks to formulate the learning problem directly in the subspace of stable dynamics. Rettberg et al. [43] propose identifying a linear latent PHS, which makes enforcing global stability relatively straightforward. Yu et al. [55] propose a model based on the generalized Onsager principle that can guarantee bounded solutions and the existence of locally asymptotically stable equilibria. While their evolution equation is similar to PHS, their approach does not consider arbitrary, time-dependent external input signals.

This work exploits the connections between stability and physical frameworks and proposes stable port-Hamiltonian neural network (sPHNN) for learning *nonlinear* and *globally stable* dynamical systems. The method is projection-free and can incorporate input signals. We show that the approach is capable of learning equilibrium points, though our numerical evaluations focus on practical applications where the equilibrium is known a priori. In such cases, this prior knowledge can be seamlessly integrated into the proposed architecture, improving prediction accuracy while simultaneously reducing data requirements.

## 2 Background

### 2.1 Stability of dynamical systems

We consider autonomous dynamical systems as first-order ODEs, where the state  $\mathbf{x}(t) \in \mathbb{R}^n$  evolves according to a vector-valued function  $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ :

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)). \quad (1)$$

Suppose  $\mathbf{f}$  is sufficiently smooth to ensure a unique solution  $\mathbf{x}(t)$  for any initial condition  $\mathbf{x}(t_0)$ . A solution  $\hat{\mathbf{x}}(t)$  is *stable* in the sense of Lyapunov if, for any small perturbation in the initial condition, the perturbed solution  $\bar{\mathbf{x}}(t)$  remains close for all times, that is, if for any  $\epsilon > 0$ , there exists a  $\delta > 0$  such that  $\|\hat{\mathbf{x}}(t_0) - \bar{\mathbf{x}}(t_0)\| < \delta$  implies  $\|\hat{\mathbf{x}}(t) - \bar{\mathbf{x}}(t)\| < \epsilon$  for all  $t > t_0$  [35, 52]. A stronger version of stability, namely *asymptotic stability*, is obtained if additionally the perturbed solution  $\bar{\mathbf{x}}(t)$  converges to  $\hat{\mathbf{x}}(t)$  for  $t \rightarrow \infty$  for initial conditions  $\|\hat{\mathbf{x}}(t_0) - \bar{\mathbf{x}}(t_0)\| < \Delta$  with some  $\Delta > 0$  [52].

Lyapunov stability is a *local* property, in the sense that for sufficiently large perturbations, the perturbed solution may stray arbitrarily far even from asymptotically stable solutions [49]. Quantifying the permissible extent of perturbations on the initial condition such that the perturbed solution still converges to the unperturbed solution leads to the concept of a *basin of attraction* [30]. Asymptotically stable equilibria, for which the basin of attraction extends to the entire phase space, are called *globally asymptotically stable* [30]. A system with a globally stable equilibrium cannot have any other equilibria, as their existence would imply that there are solutions (namely, the other equilibria) that do not converge to the globally stable equilibrium.

In the following, we consider the case of an equilibrium solution  $\mathbf{x}(t) = \mathbf{0}$ , i.e.,  $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ . Placing the equilibrium at the origin is done for ease of notation and is without loss of generality, as arbitrary equilibria can be translated to the origin with a time-independent transformation. The stability of the

equilibrium can be shown through the use of Lyapunov's stability theory by finding a suitable scalar Lyapunov function  $V$  [24, 52]. Let  $V : D \rightarrow \mathbb{R}$  be continuously differentiable and positive definite, that is,  $V(\mathbf{0}) = 0$  and  $V(\mathbf{x}) > 0$  in  $D \setminus \{\mathbf{0}\}$ , where  $D \subseteq \mathbb{R}^n$  is a neighborhood of the origin. If the value of  $V$  is nonincreasing along trajectories of the system in  $D$ , then the equilibrium is stable in the sense of Lyapunov. Mathematically, this is the case if  $\dot{V}(\mathbf{x})$  is negative semi-definite:

$$\dot{V}(\mathbf{x}) = \frac{\partial V^\top}{\partial \mathbf{x}} \dot{\mathbf{x}} = \frac{\partial V^\top}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}) \leq 0 \quad \forall \mathbf{x} \in D \setminus \{\mathbf{0}\}. \quad (2)$$

In the case of a negative definiteness ( $\dot{V}(\mathbf{x}) < 0$  in  $D \setminus \{\mathbf{0}\}$ ), *asymptotic* stability can be concluded. To obtain *global* asymptotic stability, these conditions must be satisfied *globally*, that is, for all  $\mathbf{x} \in D = \mathbb{R}^n$ , and the Lyapunov function must be radially unbounded, i.e.,  $V(\mathbf{x}) \rightarrow \infty$  for  $\|\mathbf{x}\| \rightarrow \infty$  [24]. The latter condition avoids the existence of solutions that do not converge to the equilibrium by diverging to infinity without violating the decrease condition in Equation (2).

## 2.2 Port-Hamiltonian systems

PHS theory provides a framework for modeling physical systems that generalizes the underlying mathematical structure of Hamiltonian dynamics. It incorporates the modeling of energy-dissipating elements, which are typically absent in classical Hamiltonian systems [51]. Additionally, PHS theory has roots in control theory, emphasizing the interactions of dynamic systems with their environment. In this work, we consider PHS of the form

$$\dot{\mathbf{x}} = [\mathbf{J}(\mathbf{x}) - \mathbf{R}(\mathbf{x})] \frac{\partial \mathcal{H}}{\partial \mathbf{x}}(\mathbf{x}) + \mathbf{G}(\mathbf{x}) \mathbf{u}(t). \quad (3)$$

Here,  $\mathbf{x}(t) \in \mathbb{R}^n$  describes the system's state,  $\mathbf{u}(t) \in \mathbb{R}^m$  its input, and  $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}$  denotes the Hamiltonian, which typically represents the total energy of the system [4, 51]. The skew-symmetric structure matrix  $\mathbf{J} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ ,  $\mathbf{J}(\mathbf{x}) = -\mathbf{J}^\top(\mathbf{x})$  describes the conservative energy flux within the system. The symmetric positive semi-definite matrix  $\mathbf{R} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ ,  $\mathbf{R}(\mathbf{x}) = \mathbf{R}^\top(\mathbf{x}) \succeq 0$  is called *dissipation matrix* and describes the energy losses. Finally, the *input matrix*  $\mathbf{G} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$  describes how energy enters and exits the system via the inputs. PHS extend the energy conservation of Hamiltonian systems to the energy dissipation inequality:

$$\dot{\mathcal{H}} = \underbrace{\frac{\partial \mathcal{H}^\top}{\partial \mathbf{x}} \mathbf{J} \frac{\partial \mathcal{H}}{\partial \mathbf{x}}}_{=0} - \underbrace{\frac{\partial \mathcal{H}^\top}{\partial \mathbf{x}} \mathbf{R} \frac{\partial \mathcal{H}}{\partial \mathbf{x}}}_{\geq 0} + \underbrace{\frac{\partial \mathcal{H}^\top}{\partial \mathbf{x}} \mathbf{G} \mathbf{u}(t)}_{s(\mathbf{x}, \mathbf{u})} \leq s(\mathbf{x}, \mathbf{u}). \quad (4)$$

In the language of system theory, the system is dissipative with respect to the supply rate  $s(\mathbf{x}, \mathbf{u})$ . The formal definition of dissipativity requires  $\mathcal{H} \geq 0$ , which will be ensured later. Since the supply rate  $s(\mathbf{x}, \mathbf{u})$  is linear in the input  $\mathbf{u}$ , it vanishes for zero-input  $s(\mathbf{x}, \mathbf{0}) = 0$ . The inequality thus ensures that the system cannot gain energy in the absence of an external excitation ( $\mathbf{u} = \mathbf{0}$ ). Besides this physical motivation, the dissipativity of PHS plays a crucial role in stability analysis. For the unforced system with  $s = 0$ , the Hamiltonian inherently fulfills the decrease condition  $\dot{\mathcal{H}} \leq 0$ . Thus, by using  $\mathcal{H}$  as a Lyapunov function, it can be shown that any strict minimum in the Hamiltonian implies a stable equilibrium of the unforced dynamics [51]. Furthermore, the equilibrium is asymptotically stable if the system continuously dissipates energy ( $\mathbf{R} \succ 0$ ) for states near the minimum.

The use of the port-Hamiltonian formulation as a basis for an ML model has multiple favorable consequences. When learning system dynamics from data, the structure of Equation (3) provides both a physical bias and the opportunity for interpreting the resulting trained model, while enabling the modeling of a wide range of relevant systems. Equation (4) highlights the distinctive roles played by the three matrix-valued functions. The structure matrix  $\mathbf{J}(\mathbf{x})$  results in conservative dynamics, the dissipation matrix  $\mathbf{R}(\mathbf{x})$  contributes only to the dissipative dynamics, and the input matrix  $\mathbf{G}(\mathbf{x})$  captures the energy flow over the system boundaries due to the inputs. This knowledge improves the interpretability of the trained model components. Furthermore, it allows new ways of incorporating prior knowledge into the system. For example, the learned system can be constrained to be conservative by fixing  $\mathbf{R} = \mathbf{0}$ . Furthermore, the connection of PHS to the decrease condition from Lyapunov's theory provides the opportunity to introduce further desirable stability constraints.

To illustrate the above concepts with a simple example, consider the dynamics of a three-dimensional spinning rigid body described by Euler's rotation equations. Expressing the angular velocity vector

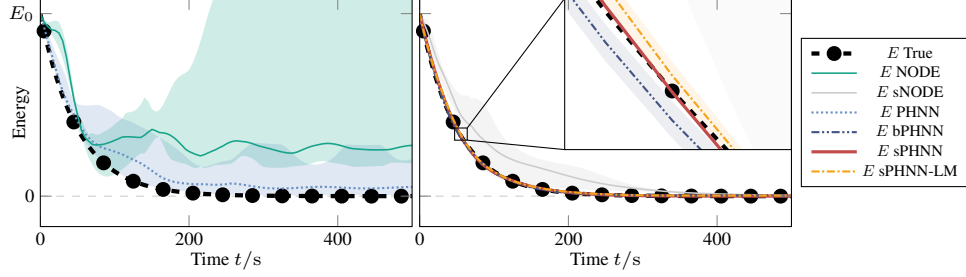


Figure 1. *Spinning rigid body*: Interquartile mean (lines) and range (shaded regions) of the energy  $E$  computed from the predicted states. Left: Models without stability bias; Right: Models with stability bias.

$\omega(t) \in \mathbb{R}^3$  in the principal-axis coordinate frame, the evolution equation is given by

$$\mathbf{I}\dot{\omega} + \omega \times (\mathbf{I}\omega) = -\mu\omega \quad \forall t > 0. \quad (5)$$

Here,  $\mathbf{I} = \text{diag}(I_1, I_2, I_3)$  denotes the inertia matrix and  $\mu$  is a damping coefficient. By introducing

$$\mathcal{H}(\omega) = \frac{1}{2}I_1\omega_1^2 + \frac{1}{2}I_2\omega_2^2 + \frac{1}{2}I_3\omega_3^2, \quad (6)$$

which describes the kinetic energy  $E$ , the system can be rewritten in the form of Equation (3), with state-dependent structure matrix  $\mathbf{J}$ , constant dissipation matrix  $\mathbf{R}$ , and vanishing input matrix  $\mathbf{G}$ :

$$\begin{bmatrix} \dot{\omega}_1 \\ \dot{\omega}_2 \\ \dot{\omega}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & -\frac{I_3}{I_1 I_2} \omega_3 & \frac{I_2}{I_3 I_1} \omega_2 \\ \frac{I_3}{I_1 I_2} \omega_3 & 0 & -\frac{I_1}{I_2 I_3} \omega_1 \\ -\frac{I_2}{I_3 I_1} \omega_2 & \frac{I_1}{I_2 I_3} \omega_1 & 0 \end{bmatrix}}_{\mathbf{J}(\omega)} - \underbrace{\mu \begin{bmatrix} \frac{1}{I_1^2} & 0 & 0 \\ 0 & \frac{1}{I_2^2} & 0 \\ 0 & 0 & \frac{1}{I_3^2} \end{bmatrix}}_{\mathbf{R}} \frac{\partial \mathcal{H}}{\partial \omega}. \quad (7)$$

In anticipation of the numerical experiment in Section 4.1, Figure 1 shows the predicted energy  $E$  from various models trained on trajectories from the spinning body system. Those incorporating both, physical biases from the PHS-formulation and stability (bPHNN, sPHNN, and sPHNN-LM) perform best, highlighting the importance of these constraints for learning system dynamics.

### 3 Stable port-Hamiltonian neural networks

Our goal is to learn the dynamics of physical systems in the form of Equation (3) with desirable stability properties from observations. To this end, we parameterize the components  $\mathcal{H}$ ,  $\mathbf{J}$ ,  $\mathbf{R}$  and  $\mathbf{G}$  of PHS with NNs.

#### 3.1 Stability of port-Hamiltonian dynamics

To guarantee the existence of stable equilibria of port-Hamiltonian systems, it is sufficient to enforce the existence of strict local minima in the Hamiltonian. However, these stability implications are only *local*, meaning the size and shape of the corresponding basin of attraction are not constrained by Equation (3). Small perturbations might be enough to stray far from the stable equilibria. Furthermore, the general PHS can represent dynamics such as unbounded trajectories, which must be considered unstable from a practical standpoint. The following describes the necessary requirements for achieving *global* stability. Essentially, the Hamiltonian  $\mathcal{H}$  is required to be a suitable Lyapunov function, which includes positive definiteness and radial unboundedness. These properties are achieved by constraining the Hamiltonian to be convex and ensuring the existence of a strict minimum. This approach is similar to the construction of a Lyapunov function by Kolter and Manek [27], but differs in the way the minimum is enforced.

**Theorem 3.1.** *Consider the PHS Equation (3) in the unforced case  $\mathbf{u}(t) = \mathbf{0}$ :*

$$\dot{\mathbf{x}} = [\mathbf{J}(\mathbf{x}) - \mathbf{R}(\mathbf{x})] \frac{\partial \mathcal{H}}{\partial \mathbf{x}}(\mathbf{x}), \quad \text{with} \quad \mathbf{J} = -\mathbf{J}^\top, \mathbf{R} = \mathbf{R}^\top \succeq 0. \quad (8)$$

*Suppose the Hamiltonian  $\mathcal{H}(\mathbf{x})$  is convex, twice continuously differentiable, and fulfills:*

$$\mathcal{H}(\mathbf{0}) = 0, \quad \left. \frac{\partial \mathcal{H}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{0}} = \mathbf{0}, \quad \left. \frac{\partial^2 \mathcal{H}}{\partial \mathbf{x} \partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{0}} \succ 0. \quad (9)$$



Then,  $\mathcal{H}$  is a suitable Lyapunov function for showing stability of the equilibrium at  $\mathbf{x}(t) = \mathbf{0}$ , and all solutions are bounded. Furthermore, the equilibrium is globally asymptotically stable if  $\mathbf{R}(\mathbf{x}) \succ 0$ .

A proof of Theorem 3.1 is provided in Appendix A. However, its implications also follow a clear physical intuition: If a dynamical system continuously loses energy via internal dissipation ( $\mathbf{R} \succ 0$ ), it will eventually settle into an energy minimum corresponding to an asymptotically stable equilibrium. The convexity requirement, together with Equation (9), guarantees the existence of a unique strict and global minimum at the origin. Additionally, it ensures that the energy for states at infinity is unbounded, i.e.,  $\mathcal{H}(\mathbf{x}) \rightarrow \infty$  as  $\|\mathbf{x}\| \rightarrow \infty$ . Thus, no finite state has more energy than a state at infinity, and all trajectories must be bounded. As a result, all solutions eventually converge to the global energy minimum. While convexity is sufficient to ensure global asymptotic stability, it is not a necessary condition. To enhance model expressiveness when needed, techniques such as input warping [27] can be employed, but were not required in the applications considered in this work.

For conservative systems with  $\mathbf{R} = \mathbf{0}$ , energy is not dissipated, and attractive equilibria cannot exist. Therefore, a conservative system cannot possess asymptotically stable equilibria. Nevertheless, the minimum in the energy still represents a stable equilibrium, and all trajectories stay bounded. The mixed case, where  $\mathbf{R} \succeq 0$ , only allows the same conclusions as the conservative case. However, unless the dissipation matrix vanishes in a large region in phase space, it is unlikely that a system with  $\mathbf{R} \succeq 0$  is not globally asymptotically stable.

### 3.2 Neural network model architecture

Having established the theoretical requirements for stability of PHS, we now outline how these conditions can be met with a dedicated NN model. We call the resulting architecture sPHNN. Its computation graph is illustrated in Figure 2a. Figure 2b shows the decomposition of a sPHNN with randomly initialized NNs into conservative and dissipative dynamics, highlighting the model's physical bias and inherent interpretability.

**Hamiltonian** To fulfill the constraints of Theorem 3.1, the Hamiltonian  $\mathcal{H}(\mathbf{x})$  is represented by a fully input convex neural network (FICNN) [2]. FICNNs can approximate any convex function [8]. They are defined by the recurrence relation

$$\begin{aligned} z_1 &= \sigma_0(\mathbf{W}_0 \mathbf{x} + \mathbf{b}_0), \\ z_{i+1} &= \sigma_i(\mathbf{U}_i z_i + \mathbf{W}_i \mathbf{x} + \mathbf{b}_i), \quad i = 1, 2, \dots, k-1, \\ f(\mathbf{x}) &= z_k, \end{aligned} \tag{10}$$

with the pass-through weight matrices  $\mathbf{W}_i$  that map the input  $\mathbf{x}$  directly to the activation of the  $k-1$  hidden layers, bias vectors  $\mathbf{b}_i$ , and layer outputs  $z_i$ . If the weight matrices  $\mathbf{U}_i$  have non-negative

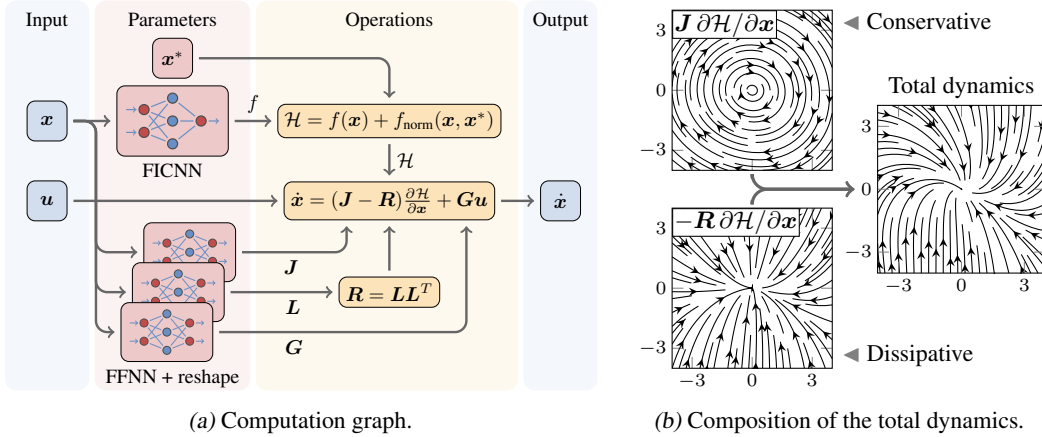


Figure 2. *sPHNN model architecture*: (a) Computation graph of sPHNN. The FICNN parameterizing  $f$  is normalized to obtain  $\mathcal{H}$  and the outputs of the FFNNs for  $\mathbf{J}$ ,  $\mathbf{L}$ , and  $\mathbf{G}$  are reshaped to be skew-symmetric, lower triangular, and rectangular matrices, respectively. (b) Dynamics of a randomly initialized sPHNN for  $n = 2$ , showing the built-in interpretability through the separation into conservative and dissipative dynamics.

components, the first activation function  $\sigma_0$  is convex, and all subsequent  $\sigma_i$  are convex and non-decreasing, then the output  $f(\mathbf{x})$  is convex in the input  $\mathbf{x}$ . Since the Hamiltonian should be twice continuously differentiable, the same must hold for the activation functions. One suitable choice is the softplus activation. The conditions for stability in Equation (9) are then fulfilled by normalizing the FICNN  $f$ :

$$\mathcal{H}(\mathbf{x}) = \underbrace{f(\mathbf{x}) - f(\mathbf{x}^*) - \frac{\partial f}{\partial \mathbf{x}} \bigg|_{\mathbf{x}^*}^\top (\mathbf{x} - \mathbf{x}^*)}_{f_{\text{norm}}(\mathbf{x}, \mathbf{x}^*)} + \underbrace{\epsilon \|\mathbf{x} - \mathbf{x}^*\|^2}_{f_{\text{reg}}(\mathbf{x}, \mathbf{x}^*)}. \quad (11)$$

Consequently, the Hamiltonian  $\mathcal{H}$  and its gradient vanish at  $\mathbf{x}^*$  due to the normalization term  $f_{\text{norm}}$ . Moreover, the regularization term  $f_{\text{reg}}$  with an arbitrarily small  $\epsilon > 0$  ensures positive definiteness of the Hamiltonian’s Hessian. However, in practice,  $f_{\text{reg}}$  can be omitted since the FICNN provides enough bias towards local strict convexity. After training, the stability guarantee can then be recovered by verifying the Hamiltonian’s positive definiteness at  $\mathbf{x}^*$ . Choosing  $\mathbf{x}^* = \mathbf{0}$  fulfills all requirements of Theorem 3.1. However, the presented approach is more general. As the assumption that the equilibrium is located at the origin is only made for ease of notation, the normalization approach in Equation (11) can position the stable equilibrium anywhere in phase space and is not restricted to the origin. If the true equilibrium position is known beforehand, fixing  $\mathbf{x}^*$  introduces an additional inductive bias. While this information can often be obtained from prior physical knowledge, it may not always be available. In such cases, the equilibrium can be inferred directly from data by treating  $\mathbf{x}^*$  as a trainable parameter during optimization.

**Structure, dissipation and input matrices** The matrix-valued functions  $\mathbf{J}(\mathbf{x})$ ,  $\mathbf{R}(\mathbf{x})$  and  $\mathbf{G}(\mathbf{x})$  are parameterized by feedforward neural networks (FFNNs), where the outputs are suitably reshaped into matrices. As  $\mathbf{J}$  is constrained to be skew-symmetric, the respective FFNN’s output vector is mapped to the space of skew-symmetric matrices via a linear transformation. Similarly, the positive definiteness (or semi-definiteness) of  $\mathbf{R}$  is ensured via the Cholesky factorization  $\mathbf{R} = \mathbf{L}\mathbf{L}^\top$ , where the reshaped FFNN output  $\mathbf{L}$  is a lower triangular matrix with positive (or non-negative) elements along the main diagonal. For some applications,  $\mathbf{J}$ ,  $\mathbf{R}$  and  $\mathbf{G}$  are independent of  $\mathbf{x}$ , and thus, learning constant matrices is sufficient. In particular, this is the case when the conservative dynamics of the system under consideration are Hamiltonian in the given coordinates. Furthermore, by fixing  $\mathbf{J}$  to the symplectic matrix (and choosing  $\mathbf{R} = \mathbf{G} = \mathbf{0}$ ), the sPHNN can be reduced to a Hamiltonian NN.

### 3.3 Training stable port-Hamiltonian neural networks

There are two fundamentally different approaches for training time-continuous dynamic models such as NODEs and sPHNNs: *derivative* and *trajectory fitting*. The former directly compares the predicted state derivatives  $\dot{\mathbf{x}}$  from Equation (3) to the true ones. While this approach is efficient, it requires the true time derivatives to be available as training data, which is typically not the case for real-world measurement data. Furthermore, derivative fitting is not applicable when augmented states are used [12], i.e., when  $\mathbf{x}$  is padded with additional dimensions compared to the observable states to enrich the representable dynamics. Further, a model proficient in predicting  $\dot{\mathbf{x}}$  does not necessarily produce accurate predictions for trajectories  $\mathbf{x}(t)$ , as small derivative errors can accumulate exponentially when integrating an ODE [17]. On the other hand, trajectory fitting directly compares integrated model trajectories  $\mathbf{x}(t)$  with ground-truth trajectories, ensuring long-term accuracy. Here, the optimization process necessitates either the propagation of the loss gradients through the ODE solver or the use of the adjoint sensitivity method [7], which computes the gradients by solving a modified ODE. In both cases, trajectory fitting incurs significant computational costs due to gradient propagation and model integration, making trajectory fitting much slower than derivative fitting.

A distinct advantage of sPHNNs is the favorable influence the stability constraint exerts on the training process when using trajectory fitting. NODEs and similar architectures can suffer from exploding gradients, where small parameter changes lead to large loss gradients. This can slow the training process or prevent it from converging. A potential cause is the crossing between the boundaries of local basins of attraction during training, leading to significant differences in long-term behavior [41]. Since sPHNNs have only one global basin of attraction, this source of training difficulties is eliminated. In our numerical experiments, we observe good convergence of the training error of sPHNNs where NODEs struggle without the use of regularization terms or gradient clipping.

## 4 Experiments

We evaluate the proposed model on multiple example problems, including measurement data from a real physical system and multiphysics simulation data. The performance of sPHNNs is compared to that of NODEs, PHNNs, and bounded PHNNs (bPHNNs). The latter two are PHS-based NN models that do not enforce or relax the convexity constraint of sPHNNs, and thus do not guarantee global stability. In PHNNs, the Hamiltonian is modeled by an unconstrained FFNN, whereas in bPHNNs, it is constructed to be positive and radially unbounded, which ensures bounded trajectories. The latter architecture is intentionally designed to resemble OnsagerNet [55], but with the key difference that it supports arbitrary time-dependent inputs, making it suitable for our experimental setting. For examples that purely rely on derivative fitting, we further include stable NODEs (sNODEs) [27]. Using trajectory fitting, we could not achieve convergence of sNODEs, excluding them from most experiments. This is likely due to difficulties caused by discontinuous dynamics stemming from the projection-based stability enforcement, as also mentioned by [46]. Finally, we include a variant of sPHNNs with learnable minima (sPHNN-LM). In contrast to sPHNNs, which have  $\mathbf{x}^*$  fixed to an a-priori known equilibrium position, sPHNN-LMs have to infer the equilibrium position from the training data and start out with a randomly initialized  $\mathbf{x}^*$ . All models are trained by minimizing the mean squared error using the ADAM optimizer [25]. Model hyperparameters and computational costs are listed in Appendix B. Predictions are generated by rollout with a numerical integrator using the Runge-Kutta scheme Tsit5, with adaptive step size and a known input function  $\mathbf{u}(t)$ . Code and model weights are available on GitHub<sup>1</sup>.

### 4.1 Spinning rigid body

We generate data of a damped spinning rigid body by numerically integrating Euler’s rotation equations, which were already introduced in Section 2.2, Equation (5). Here, accurately capturing the dynamics of the angular velocity vector  $\mathbf{x} = \boldsymbol{\omega}$  with a PHS requires a state-dependent structure matrix  $\mathbf{J}(\mathbf{x})$ . We train ten instances per model type using derivative fitting with  $(\boldsymbol{\omega}, \dot{\boldsymbol{\omega}})$  pairs sampled from the trajectories. Figure 1 shows the rotational energy  $E(\boldsymbol{\omega})$  computed from model predictions  $\boldsymbol{\omega}$  via Equation (6). The PHNN, NODE, and sNODE models show deviations from ground truth, with the former two not converging to zero. Furthermore, the NODEs show unphysical behavior with at times increasing energy and exhibit high variance between instances. In contrast, sPHNNs and bPHNNs match the true energy well, highlighting the importance of providing both stability and physical biases. The sPHNN-LM variant recovered the actual equilibrium position accurately, with an interquartile mean distance between the true and inferred minimum across all instances of 0.010. As a result, its predictive performance is nearly identical to that of the sPHNN with fixed  $\mathbf{x}^*$ .

### 4.2 Cascaded tanks

Next, the sPHNN model is evaluated using the cascaded tanks dataset [47, 48]. It contains measurements of a physical fluid level control system consisting of a pump and two tanks, see Appendix D for details. The training data consists of a single trajectory with 1024 pairs of pump voltages  $u$  and water level measurements  $y$  of the lower tank. We model the system using a two-dimensional state vector  $\mathbf{x} = [x_1, x_2]^\top$ , where  $y = x_2$ . Assuming the augmented state  $x_1$  describes the water level in the upper tank, it is straightforward to determine the equilibrium of the autonomous system. With zero input, the tanks will eventually drain; therefore, the stable equilibrium is given by  $\mathbf{x}^* = \mathbf{0}$ . We thus fix the minimum of the sPHNN’s Hamiltonian to the origin. To investigate the effects of a potential lack of this knowledge, we also train the sPHNN-LM variant. Its minimum location  $\mathbf{x}^*$  is randomly initialized and optimized during training. Apart from this, both sPHNN versions are identical and use a constant fixed symplectic matrix as the structure matrix, as well as constant but learnable dissipation and input matrices. The same structure applies to PHNN and bPHNN.

We train 20 instances per model type and show a statistical evaluation of the root mean squared errors (RMSEs) in Figure 3a. Overall, sPHNN, sPHNN-LM, and bPHNN achieve comparable accuracy, while NODE and PHNN perform worse on average across both trajectories. In addition to their better performance, the models with stability bias exhibit lower variance across model instances, indicating more consistent results. To evaluate the models’ extrapolation capabilities, we extend the evaluation data with 400 s of zero-input. The resulting predictions are depicted in

<sup>1</sup><https://github.com/CPShub/sphnn-publication>

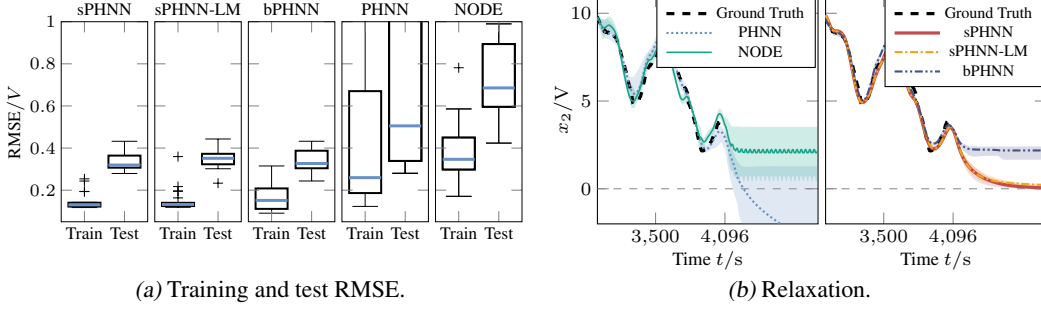


Figure 3. *Cascaded tanks*: (a) RMSE of the models on training and test trajectory. (b): Predictions for the extended test trajectory. At  $t = 4096$  s, the pump is turned off. Lines correspond to the interquartile mean and shaded areas represent the interquartile range of the predictions from the 20 model instances.

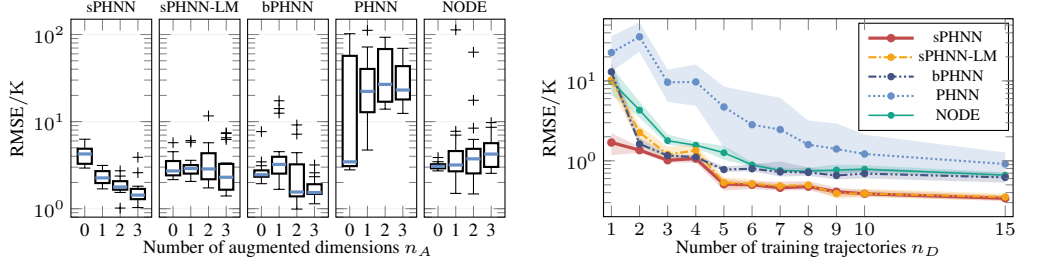
Figure 3b. Assuming perfectly calibrated sensors, the true system response is an eventual zero output as both tanks drain. By construction, this behavior is guaranteed for sPHNNs. The 20 sPHNN-LM instances reliably identified the systems equilibrium to lie between  $-0.196$  V to  $0.432$  V, which aligns with physical intuition. In contrast the bPHNN, NODE and PHNN models exhibit poor physical extrapolation, predicting equilibria far from zero, diverging water levels, or oscillatory behavior. In an additional experiment presented in Appendix D, the sPHNN models are also shown to have successfully learned the hard saturation effect that occurs when a water tank overflows — a task at which both NODE and PHNN fail.

### 4.3 Thermal food processing surrogate

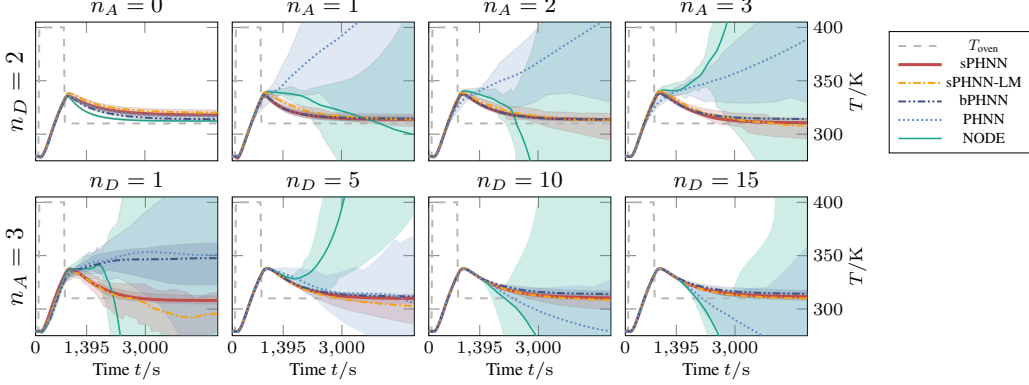
With the applicability of sPHNNs to real-world data established, this section focuses on one of the promising applications of data-driven dynamic system identification: the efficient construction of surrogate models. Here, data from a conjugate heat transfer simulation of a convection oven coupled with a soft matter model for meats from Kannapinn et al. [21, 23] is employed. The data consists of trajectories representing the temperature histories  $T_A$  and  $T_B$  at two probe points within the meat. These result from a predefined excitation signal  $T_{\text{oven}}$ , which controls the oven temperature. The surrogate’s task is to predict the probe temperatures given the oven temperature as input.

We fit the models to initially  $n_D = 2$  trajectories consisting of 280 samples each (see Appendix E). Per model, we vary the dimensionality of the state  $\mathbf{x} \in \mathbb{R}^n$  with  $n = 2 + n_A$  by adding 0 to 3 augmented dimensions  $n_A$ . The minimum  $\mathbf{x}^*$  of the sPHNNs’ Hamiltonian is set to the point in phase space corresponding to the ambient temperature, as motivated by physical intuition about the stable thermal equilibrium. The initial values of the augmented states are fixed to zero. In total, 20 instances are trained per model type and augmented dimension.

The RMSEs evaluated on 15 test trajectories (see Appendix E) are depicted in Figure 4a. The sPHNN predictions tend to improve with more augmented dimensions, achieving errors comparable to those obtained by Kannapinn et al. [21] using commercial software. This appears intuitive, as additional hidden state variables may be required to accurately describe the evolution of  $T_A$  and  $T_B$  that results from the multiphysics, conjugate heat transfer process, see also [12]. However, for bPHNNs and sPHNN-LMs no clear trend emerges. The NODEs and PHNNs even tend to deteriorate in performance with more augmented dimensions, likely due to instabilities in the learned dynamics. This becomes apparent in Figure 4c (top row), which depicts the models’ predictions for an extended test case. While the sPHNNs consistently converge to temperatures close to thermal equilibrium, the predictions of NODEs and PHNNs tend to become unstable when one or more augmented dimensions are used. Although all models improve with more training data (Figure 4b), these instabilities persist when extrapolating in time (Figure 4c, bottom row  $t > 1395$  s). In contrast, the stability constraint of the sPHNNs and sPHNN-LMs allows safe use of the additional flexibility provided by a higher-dimensional state space, avoiding the pitfalls observed in the unconstrained models. Given sufficient training data, the sPHNN-LM variant performs on par with the sPHNN, whereas in the scarce-data regime its errors increase and are comparable to those of the bPHNNs. Although the bPHNNs also remain stable, they are consistently outperformed by the sPHNNs for  $n_A = 3$ , regardless of dataset size, and especially when only a single training trajectory is available  $n_D = 1$ , see Figure 4b and Figure 4c (bottom left).



(a) Test RMSEs for  $n_D = 2$  training trajectories. (b) Test RMSEs for  $n_A = 3$  augmented dimensions.



(c) Model predictions for different combinations of augmented dimensions and training data.

Figure 4. Thermal food processing surrogate: (a) and (b) RMSEs evaluated on 15 test trajectories for various numbers of augmented dimensions  $n_A$  and training trajectories  $n_D$ . (c) Interquartile mean and range of the  $T_A$  predictions for a custom test case. *Top row*: Varying number of augmented dimensions with fixed number  $n_D = 2$  of training trajectories. *Bottom row*: Varying number of training trajectories with fixed number  $n_A = 3$  of augmented dimensions. The time  $t = 1395$  s marks the length of the training trajectories.

Table 1. Median RMSE of 20 instances per models using  $n_A = 3$  augmented dimensions, trained on  $n_D = 2$  noisy trajectories and evaluated on noise-free training and test sets.

Model	Training			Test		
	5 % noise	15 % noise	25 % noise	5 % noise	15 % noise	25 % noise
sPHNN	<b>0.332</b>	<b>0.854</b>	<b>1.354</b>	<b>1.170</b>	<b>2.595</b>	<b>3.177</b>
sPHNN-LM	0.638	1.096	1.622	2.420	2.888	3.677
bPHNN	0.500	1.049	2.207	1.572	2.828	5.231
PHNN	2.828	6.750	5.779	28.233	25.631	24.224
NODE	0.993	1.974	2.423	3.577	3.482	4.362

To assess the models' robustness to noise, we conducted an additional experiment using  $n_A = 3$  augmented dimensions and  $n_D = 2$  training trajectories, introducing zero-mean Gaussian noise for both the input  $T_{\text{oven}}$  and the outputs  $T_A$  and  $T_B$ . The noise amplitudes were varied between 5 % and 25 % of the original signals' standard deviations. Table 1 reports the average RMSE over 20 model instances per model type, evaluated on noise-free test data. To isolate the effect of noise and eliminate generalization error, we also evaluated the RMSE on the noise-free training data. All models demonstrate a degree of robustness to noise, since the integration step involved in trajectory fitting acts as a low-pass filter, effectively smoothing the learned dynamics. As a result, high-frequency Gaussian noise is attenuated, which helps prevent overfitting and improves generalization. Nevertheless, the sPHNN consistently outperforms the baselines on both the training and test sets.

#### 4.4 Additive manufacturing surrogate

Finally, we evaluate the applicability of sPHNNs to higher-dimensional state spaces and build a reduced order model (ROM) as a surrogate of a 3D partial differential equation problem. To this end, we generate field data by numerically solving the heat conduction equation for a moving heat

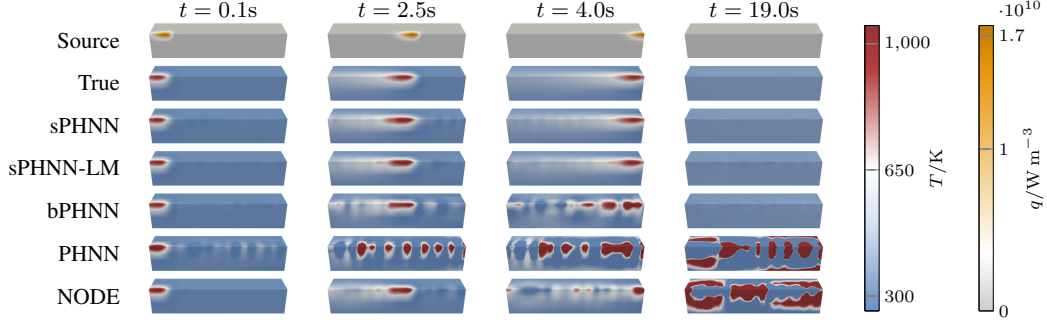


Figure 5. Heat source field and temperature field predictions on the cuboid domain for a test case with  $v = 12.5 \text{ mm s}^{-1}$ ,  $Q = 400 \text{ W}$ . The instances selected for this evaluation resulted in the median test error for the corresponding model type. Colors are clipped to remain in the legend’s range.

source on a cuboid domain with convective and radiative thermal boundary conditions. The finite element simulation is designed to model the evolution of the temperature field of a metal additive manufacturing process, see [22] for details. In total, 25 trajectories are obtained by varying the heat source speed  $v$  from  $10 \text{ mm s}^{-1}$  to  $20 \text{ mm s}^{-1}$  and power  $Q$  from  $300 \text{ W}$  to  $500 \text{ W}$ . All trajectories span  $20 \text{ s}$ , with a non-zero heat source lasting for the initial  $2.5 \text{ s}$  to  $6 \text{ s}$ , depending on  $v$ . However, directly using the discretized field data would lead to very high-dimensional models and render the training and inference inefficient. We thus apply a proper orthogonal decomposition (POD) to map the temperature and heat source fields onto 40-dimensional latent spaces, respectively (see Appendix F for details). The models are trained on the latent representations of the two trajectories with  $(v, Q) = (10 \text{ mm s}^{-1}, 300 \text{ W})$  and  $(20 \text{ mm s}^{-1}, 500 \text{ W})$  using derivative fitting and subsequent fine-tuning with trajectory fitting.

Figure 5 depicts the heat source field (top row) and true and predicted temperature fields from the different model types. While the sPHNN’s total test RMSE peaks below  $30 \text{ K}$  (see Appendix F, Figure F.2) and decreases as the system approaches thermal equilibrium, the errors for NODE and PHNN increase rapidly, driven by instabilities in the learned dynamics. These instabilities are likely caused by the sparsity of the training data, which leaves gaps in the phase space, allowing unstable dynamics to emerge. The stability constraint of sPHNN and sPHNN-LM mitigates these issues, ensuring robust performance even with scarce training data, while outperforming the bPHNN thanks to their stronger inductive bias.

## 5 Conclusion

This work proposes stable port-Hamiltonian neural networks (sPHNNs) for accurate, robust, and reliable identification of stable nonlinear dynamics. They leverage the port-Hamiltonian framework, which provides a basis for model interpretation and ensures adherence to the laws of thermodynamics in terms of energy conservation or dissipation balances with respect to a learned Hamiltonian energy function. Additionally, the approach ensures global asymptotic stability of the identified dynamics, without requiring projection, by constraining the Hamiltonian to be a convex, positive definite Lyapunov function. While this limits sPHNNs to modeling systems that appear globally stable within the region of interest, it provides a strong inductive bias that enables robust learning and reasonable generalization from sparse data. We demonstrate the approach’s viability using data from synthetic examples, real-world measurements, and complex multiphysics simulation models. The resulting models, built with small neural network architectures, exhibit low variance across instances, accurate prediction, and robust generalization. Furthermore, the proposed method can successfully exploit augmented dimensions where unconstrained alternative models struggle with instabilities or data scarcity. While the approach can identify the equilibrium point from data, we find that providing the equilibrium as a prior can further enhance prediction accuracy and reduce data requirements. Future work will extend the approach to multistable systems by allowing Hamiltonians with multiple minima, while retaining favorable stability properties.

## Acknowledgments and Disclosure of Funding

The authors acknowledge the financial support provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, research grant number 492770117 and the CRC/TRR 361 CREATOR), the Hessian Center for Artificial Intelligence (Hessian.AI, project EnDyRo), and the Graduate School Computational Engineering at TU Darmstadt. Furthermore, we would like to thank An Thai Le, Daniel Palenicek, and Joe Watson from TU Darmstadt (IAS) for their constructive feedback on the manuscript.

## References

- [1] A. Abbasi, P. N. Kambali, P. Shahidi, and C. Nataraj. “Physics-Informed Machine Learning for Modeling Multidimensional Dynamics”. In: *Nonlinear Dynamics* (2024). DOI: 10.1007/s11071-024-10163-3.
- [2] B. Amos, L. Xu, and J. Z. Kolter. “Input Convex Neural Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. Proceedings of Machine Learning Research. PMLR, 2017, pp. 146–155.
- [3] M. Barreau and N. Bastianello. “Learning and Verifying Maximal Taylor-neural Lyapunov Functions”. In: *arXiv e-prints* (2024). DOI: 10.48550/arXiv.2408.17246.
- [4] C. Beattie, V. Mehrmann, H. Xu, and H. Zwart. “Linear Port-Hamiltonian Descriptor Systems”. In: *Mathematics of Control, Signals, and Systems* 30.4 (2018), p. 17. DOI: 10.1007/s00498-018-0223-3.
- [5] L. Böttcher. “Gradient-Free Training of Neural ODEs for System Identification and Control Using Ensemble Kalman Inversion”. In: *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*. 2023.
- [6] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Version 29. Cambridge New York Melbourne New Delhi Singapore: Cambridge University Press, 2023. 716 pp.
- [7] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. “Neural Ordinary Differential Equations”. In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018.
- [8] Y. Chen, Y. Shi, and B. Zhang. “Optimal Control via Neural Networks: A Convex Approach”. In: *International Conference on Learning Representations*. 2019.
- [9] A. Choudhary, J. F. Lindner, E. G. Holliday, S. T. Miller, S. Sinha, and W. L. Ditto. “Forecasting Hamiltonian Dynamics without Canonical Coordinates”. In: *Nonlinear Dynamics* 103.2 (2021), pp. 1553–1562. DOI: 10.1007/s11071-020-06185-2.
- [10] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho. “Lagrangian Neural Networks”. In: *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*. 2019.
- [11] S. Desai, M. Mattheakis, D. Sondak, P. Protopapas, and S. Roberts. “Port-Hamiltonian Neural Networks for Learning Explicit Time-Dependent Dynamical Systems”. In: *Physical Review E* 104.3 (2021). DOI: 10.1103/PhysRevE.104.034312.
- [12] E. Dupont, A. Doucet, and Y. W. Teh. “Augmented Neural Odes”. In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019.
- [13] S. Eidnes, A. J. Stasik, C. Sterud, E. Bøhn, and S. Riemer-Sørensen. “Pseudo-Hamiltonian Neural Networks with State-Dependent External Forces”. In: *Physica D: Nonlinear Phenomena* 446 (2023), p. 133673. DOI: 10.1016/j.physd.2023.133673.
- [14] N. B. Erichson, M. Muehlebach, and M. W. Mahoney. “Physics-Informed Autoencoders for Lyapunov-stable Fluid Flow Prediction”. In: *Machine Learning and the Physical Sciences (NeurIPS 2019)*. Neural Information Processing Systems. Vancouver, Canada, 2019.
- [15] X. Glorot and Y. Bengio. “Understanding the Difficulty of Training Deep Feedforward Neural Networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Vol. 9. Sardinia, Italy: PMLR, 2010, pp. 249–256.
- [16] S. Greydanus, M. Dzamba, and J. Yosinski. “Hamiltonian Neural Networks”. In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019.
- [17] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. 2nd rev. ed. Springer Series in Computational Mathematics 8. Heidelberg ; London: Springer, 2009. 528 pp.



- [18] Q. Hernández, A. Badias, D. Gonzalez, F. Chinesta, and E. Cueto. “Structure-Preserving Neural Networks”. In: *Journal of Computational Physics* 426 (2021), p. 109950. DOI: 10.1016/j.jcp.2020.109950.
- [19] P. Jin, Z. Zhang, I. G. Kevrekidis, and G. E. Karniadakis. “Learning Poisson Systems and Trajectories of Autonomous Systems via Poisson Neural Networks”. In: *IEEE Transactions on Neural Networks and Learning Systems* 34.11 (2023), pp. 8271–8283. DOI: 10.1109/TNNLS.2022.3148734.
- [20] P. Jin, Z. Zhang, A. Zhu, Y. Tang, and G. E. Karniadakis. “SympNets: Intrinsic Structure-Preserving Symplectic Networks for Identifying Hamiltonian Systems”. In: *Neural Networks* 132 (2020), pp. 166–179. DOI: 10.1016/j.neunet.2020.08.017.
- [21] M. Kannapinn, M. K. Pham, and M. Schäfer. “Physics-Based Digital Twins for Autonomous Thermal Food Processing: Efficient, Non-Intrusive Reduced-Order Modeling”. In: *Innovative Food Science & Emerging Technologies* 81 (2022), p. 103143. DOI: 10.1016/j.ifset.2022.103143.
- [22] M. Kannapinn, F. Roth, and O. Weeger. “Digital Twin Inference from Multi-Physical Simulation Data of DED Additive Manufacturing Processes with Neural ODEs”. 2024.
- [23] M. Kannapinn, M. Schäfer, and O. Weeger. “TwinLab: A Framework for Data-Efficient Training of Non-Intrusive Reduced-Order Models for Digital Twins”. In: *Engineering Computations* (2024). DOI: 10.1108/EC23-0855.
- [24] H. K. Khalil. *Nonlinear Systems*. 3rd ed. Upper Saddle River, N.J: Prentice Hall, 2002. 750 pp.
- [25] D. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *International Conference on Learning Representations*. 2015.
- [26] R. Kojima and Y. Okamoto. “Learning Deep Input-Output Stable Dynamics”. In: *Advances in Neural Information Processing Systems*. Vol. 35. Curran Associates, Inc., 2022, pp. 8187–8198.
- [27] J. Z. Kolter and G. Manek. “Learning Stable Deep Dynamics Models”. In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019.
- [28] Z. Lai, C. Mylonas, S. Nagarajaiah, and E. Chatzi. “Structural Identification with Physics-Informed Neural Ordinary Differential Equations”. In: *Journal of Sound and Vibration* 508 (2021), p. 116196. DOI: 10.1016/j.jsv.2021.116196.
- [29] N. Lawrence, P. Loewen, M. Forbes, J. Backstrom, and B. Gopaluni. “Almost Surely Stable Deep Dynamics”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 18942–18953.
- [30] G. C. Layek. *An Introduction to Dynamical Systems and Chaos*. 1st ed. 2015. New Delhi: Springer India: Imprint: Springer, 2015. 1 p. DOI: 10.1007/978-81-322-2556-0.
- [31] K. Lee and E. J. Parish. “Parameterized Neural Ordinary Differential Equations: Applications to Computational Physics Problems”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 477.2253 (2021), p. 20210162. DOI: 10.1098/rspa.2021.0162.
- [32] C. Legaard, T. Schranz, G. Schweiger, J. Drgoňa, B. Falay, C. Gomes, A. Iosifidis, M. Abkar, and P. Larsen. “Constructing Neural Network Based Models for Simulating Dynamical Systems”. In: *Acm Computing Surveys* 55.11 (2023). DOI: 10.1145/3567591.
- [33] D. Liu and Y. Wang. “Multi-Fidelity Physics-Constrained Neural Network and Its Application in Materials Modeling”. In: *Journal of Mechanical Design* 141.12 (2019), p. 121403. DOI: 10.1115/1.4044400.
- [34] M. Lutter, C. Ritter, and J. Peters. “Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning”. In: *International Conference on Learning Representations*. 2019.
- [35] J. G. Malkin. *Theorie der Stabilität einer Bewegung*. München: R. Oldenbourg, 1959. 400 pp.
- [36] S. Massaroli, M. Poli, M. Bin, J. Park, A. Yamashita, and H. Asama. “Stable Neural Flows”. In: *arXiv e-prints* (2020), arXiv:2003.08063. DOI: 10.48550/arXiv.2003.08063.
- [37] H. Nakano, R. Ariizumi, T. Asai, and S.-i. Azuma. “Model Estimation Ensuring Passivity by Using Port-Hamiltonian Model and Deep Learning”. In: 61st Annual Conference of the Society of Instrument and Control Engineers (SICE). Kumamoto, Japan: IEEE, 2022, pp. 886–891. DOI: 10.23919/SICE56594.2022.9905855.
- [38] C. Neary and U. Topcu. “Compositional Learning of Dynamical System Models Using Port-Hamiltonian Neural Networks”. In: *Proceedings of the 5th Annual Learning for Dynamics and Control Conference*. Vol. 211. PMLR, 2023, pp. 679–691.



- [39] Y. Okamoto and R. Kojima. “Learning Deep Dissipative Dynamics”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 39. 2025, pp. 19749–19757. DOI: 10.1609/aaai.v39i18.34175.
- [40] J. S. R. Park, S. W. Cheung, Y. Choi, and Y. Shin. “tLaSDI: Thermodynamics-Informed Latent Space Dynamics Identification”. In: *Computer Methods in Applied Mechanics and Engineering* 429 (2024), p. 117144. DOI: 10.1016/j.cma.2024.117144.
- [41] R. Pascanu, T. Mikolov, and Y. Bengio. “On the Difficulty of Training Recurrent Neural Networks”. In: *Proceedings of the 30th International Conference on Machine Learning*. Vol. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA: PMLR, 2013, pp. 1310–1318.
- [42] S. Ramasinghe, H. Saratchandran, V. Shevchenko, and S. Lucey. “On the Effectiveness of Neural Priors in Modeling Dynamical Systems”. In: *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*. 2023.
- [43] J. Rettberg, J. Kneifl, J. Herb, P. Buchfink, J. Fehr, and B. Haasdonk. “Data-Driven Identification of Latent Port-Hamiltonian Systems”. In: *arXiv e-prints* (2024). DOI: 10.48550/arXiv.2408.08185.
- [44] S. M. Richards, F. Berkenkamp, and A. Krause. “The Lyapunov Neural Network: Adaptive Stability Certification for Safe Learning of Dynamical Systems”. In: *Proceedings of the 2nd Conference on Robot Learning*. Vol. 87. PMLR, 2018, pp. 466–476.
- [45] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [46] A. Schlaginhaufen, P. Wenk, A. Krause, and F. Dorfler. “Learning Stable Deep Dynamics Models for Partially Observed or Delayed Dynamical Systems”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 11870–11882.
- [47] M. Schoukens, P. Mattson, T. Wigren, and J.-P. Noël. *Cascaded Tanks Benchmark Combining Soft and Hard Nonlinearities*. .mat .csv. Version 1. (License: CC BY-SA 4.0). 4TU.ResearchData, 2020. DOI: 10.4121/12960104.v1.
- [48] M. Schoukens, P. Mattsson, T. Wigren, and J.-P. Noel. “Cascaded Tanks Benchmark Combining Soft and Hard Nonlinearities”. In: *Workshop on Nonlinear System Identification Benchmarks 2016*. Brussels, Belgium, 2016.
- [49] R. Seydel. *Practical Bifurcation and Stability Analysis*. 3rd ed. Interdisciplinary Applied Mathematics 5. New York: Springer, 2010. 483 pp.
- [50] N. Takeishi and Y. Kawahara. “Learning Dynamics Models with Stable Invariant Sets”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.11 (2021), pp. 9782–9790. DOI: 10.1609/aaai.v35i11.17176.
- [51] A. J. van der Schaft and D. Jeltsema. *Port-Hamiltonian Systems Theory: An Introductory Overview*. Foundations and Trends in Systems and Control 1, 2/3 (2014). Boston Delft: Now, 2014. 216 pp.
- [52] F. Verhulst. *Nonlinear Differential Equations and Dynamical Systems*. Universitext. Berlin New York Paris [etc.]: Springer, 1990.
- [53] R. Wang and R. Yu. “Physics-Guided Deep Learning for Dynamical Systems: A Survey”. In: *arXiv e-prints* (2021). DOI: 10.48550/arXiv.2107.01272.
- [54] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar. “Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems”. In: *Acm Computing Surveys* 55.4 (2022). DOI: 10.1145/3514228.
- [55] H. Yu, X. Tian, W. E, and Q. Li. “OnsagerNet: Learning Stable and Interpretable Dynamics Using a Generalized Onsager Principle”. In: *Physical Review Fluids* 6.11 (2021), p. 114402. DOI: 10.1103/PhysRevFluids.6.114402.
- [56] Z. Zhang, Y. Shin, and G. Em Karniadakis. “GFINNs: GENERIC Formalism Informed Neural Networks for Deterministic and Stochastic Dynamical Systems”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380.2229 (2022), p. 20210207. DOI: 10.1098/rsta.2021.0207.
- [57] Y. D. Zhong, B. Dey, and A. Chakraborty. “Dissipative SymODEN: Encoding Hamiltonian Dynamics with Dissipation and Control into Deep Learning”. In: *ICLR Workshop on Integration of Deep Neural Models and Differential Equations*. 2020. DOI: 10.48550/arXiv.2002.08860.

## A Proof of stability

This section provides a proof of Theorem 3.1, which describes the essential requirements for achieving stable dynamics with port-Hamiltonian systems. Due to the relevance of convexity to the following discussion, the following Lemma first provides some widely known equivalent formulations of convexity for scalar-valued multivariate functions.

**Lemma A.1.** *Consider a twice continuously differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\mathbf{x} \mapsto f(\mathbf{x})$ . If  $f$  is convex, the following statements hold:*

1.  $\lambda f(\mathbf{x}_1) + (1 - \lambda)f(\mathbf{x}_2) \geq f(\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n, \lambda \in [0, 1],$
2.  $\frac{\partial f}{\partial \mathbf{x}} \Big|_{\mathbf{x}_1}^\top (\mathbf{x}_2 - \mathbf{x}_1) \leq f(\mathbf{x}_2) - f(\mathbf{x}_1) \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n,$
3.  $\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}}(\mathbf{x}) \succeq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n.$

For a proof, the reader is referred to [6]. Note that for strict convexity, the inequalities in statements 1 and 2 must be strict for all  $\mathbf{x}_1 \neq \mathbf{x}_2$ , and the Hessian in statement 3 must be positive definite. With the notion of convexity established, attention can now be directed to the proof of Theorem 3.1. For convenience, the statement of the theorem is repeated verbatim below:

**Theorem A.2.** *Consider the PHS Equation (3) in the unforced case  $\mathbf{u}(t) = \mathbf{0}$ :*

$$\dot{\mathbf{x}} = [\mathbf{J}(\mathbf{x}) - \mathbf{R}(\mathbf{x})] \frac{\partial \mathcal{H}}{\partial \mathbf{x}}(\mathbf{x}), \quad \text{with} \quad \mathbf{J} = -\mathbf{J}^\top, \mathbf{R} = \mathbf{R}^\top \succeq 0. \quad (\text{A.1})$$

*Suppose the Hamiltonian  $\mathcal{H}(\mathbf{x})$  is convex, twice continuously differentiable, and fulfills:*

$$\mathcal{H}(\mathbf{0}) = 0, \quad \frac{\partial \mathcal{H}}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{0}} = \mathbf{0}, \quad \frac{\partial^2 \mathcal{H}}{\partial \mathbf{x} \partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{0}} \succ 0. \quad (\text{A.2})$$

*Then, the system in Equation (A.1) has a stable equilibrium at  $\mathbf{x}(t) = \mathbf{0}$ , and all solutions are bounded. Furthermore, the equilibrium is globally asymptotically stable if  $\mathbf{R}(\mathbf{x}) \succ 0$ .*

*Proof.* Since the gradient of the Hamiltonian vanishes and its Hessian is positive definite at  $\mathbf{x} = \mathbf{0}$ , it has a strict local minimum at the origin. This implies the existence of an  $R > 0$  such that

$$\mathcal{H}(\mathbf{x}) > \mathcal{H}(\mathbf{0}) = 0 \quad \forall \mathbf{x} \in \{\|\mathbf{x}\| \leq R, \mathbf{x} \neq \mathbf{0}\}. \quad (\text{A.3})$$

Consider the function  $g(\mathbf{x}) = \frac{a}{r} \|\mathbf{x}\|$  for some  $r \in (0, R]$  with  $a = \inf_{\|\mathbf{x}\|=r} \mathcal{H}(\mathbf{x}) > 0$ . We claim that  $g$  is a lower bound for  $\mathcal{H}$ , i.e.  $g(\mathbf{x}) \leq \mathcal{H}(\mathbf{x})$  for all  $\|\mathbf{x}\| > r$ .

Suppose there existed an  $\tilde{\mathbf{x}} \in \mathbb{R}^n$  that violated the claim, that is  $\|\tilde{\mathbf{x}}\| > r$  and  $g(\tilde{\mathbf{x}}) > \mathcal{H}(\tilde{\mathbf{x}})$ . Due to the convexity of  $\mathcal{H}$ , it holds

$$\lambda \mathcal{H}(\tilde{\mathbf{x}}) + (1 - \lambda)\mathcal{H}(\mathbf{0}) \geq \mathcal{H}(\lambda \tilde{\mathbf{x}} + (1 - \lambda) \cdot \mathbf{0}) \quad \forall \lambda \in [0, 1] \quad (\text{A.4})$$

and thus  $\lambda g(\tilde{\mathbf{x}}) > \lambda \mathcal{H}(\tilde{\mathbf{x}}) \geq \mathcal{H}(\lambda \tilde{\mathbf{x}})$ . Choosing  $\lambda = \frac{r}{\|\tilde{\mathbf{x}}\|}$  and applying the definition of  $g$  we obtain

$$\frac{r}{\|\tilde{\mathbf{x}}\|} g(\tilde{\mathbf{x}}) = a > \mathcal{H}\left(\frac{r}{\|\tilde{\mathbf{x}}\|} \tilde{\mathbf{x}}\right) \geq a. \quad (\text{A.5})$$

This is a contradiction, and therefore  $g$  is a lower bound.

With the bound established, showing that  $\mathcal{H}$  is radially unbounded is straightforward. For any path  $\mathbf{x}(t)$  with  $\lim_{t \rightarrow \infty} \|\mathbf{x}(t)\| = \infty$ , once  $\|\mathbf{x}(t)\| > r$  we have

$$\lim_{t \rightarrow \infty} \mathcal{H}(\mathbf{x}(t)) \geq \lim_{t \rightarrow \infty} g(\mathbf{x}(t)) = \infty. \quad (\text{A.6})$$

We would like to remark that the same conclusion could alternatively be reached by showing that  $\mathcal{H}$  has at least one bounded sublevel set due to its strict minimum and then concluding radial unboundedness (i.e., coercivity) using Corollary 8.7.1 in [45] together with the standard equivalence for proper closed convex functions. However, the presented approach using a lower bound can directly be extended to show global positive definiteness of  $\mathcal{H}$ . For this, note that for all  $\|\mathbf{x}\| > r$  the lower bound implies  $0 < g(\mathbf{x}) \leq \mathcal{H}(\mathbf{x})$ . Together with Equation (A.3), this implies that  $\mathcal{H}$  is indeed positive definite for all  $\mathbf{x} \in \mathbb{R}^n$ .

From the port-Hamiltonian structure of Equation (A.1) we immediately obtain the decrease condition  $\dot{\mathcal{H}} \leq 0$  (see Equation (4)). Using the  $\mathcal{H}$  as a Lyapunov function, we obtain all requirements to conclude *local* stability of the equilibrium. Furthermore, the decrease condition implies that a solution starting at  $\mathbf{x}^*$  cannot leave the sublevel set  $\Omega_{\mathcal{H}} = \{\mathbf{x} \in \mathbb{R}^n | \mathcal{H}(\mathbf{x}) \leq \mathcal{H}(\mathbf{x}^*)\}$ . Since  $\Omega_{\mathcal{H}} \subseteq \Omega_g = \{\mathbf{x} \in \mathbb{R}^n | g(\mathbf{x}) \leq \mathcal{H}(\mathbf{x}^*)\}$  and  $\Omega_g$  is clearly bounded, all solutions of Equation (A.1) must be bounded as well.

Finally, to show that *global* asymptotic stability follows from  $\mathbf{R} \succ 0$ , it remains to be demonstrated that  $\dot{\mathcal{H}} < 0$  for all  $\mathbf{x} \neq \mathbf{0}$ . This is implied by Equation (4) if  $\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \neq \mathbf{0}$  for all  $\mathbf{x} \neq \mathbf{0}$ . Proving that the gradient of  $\mathcal{H}$  vanishes only at the origin is done by contradiction. Assume that there exists an  $\tilde{\mathbf{x}} \neq \mathbf{0}$  such that  $\frac{\partial \mathcal{H}}{\partial \mathbf{x}}|_{\tilde{\mathbf{x}}} = \mathbf{0}$ . Furthermore, let  $\mathbf{x}^* = \mathbf{0}$ . Since  $\mathcal{H}$  is convex it follows (see Lemma A.1) that

$$\mathcal{H}(\mathbf{x}^*) - \mathcal{H}(\tilde{\mathbf{x}}) \geq \frac{\partial \mathcal{H}}{\partial \mathbf{x}}|_{\tilde{\mathbf{x}}}^\top (\mathbf{x}^* - \tilde{\mathbf{x}}) = 0, \quad (\text{A.7})$$

$$\mathcal{H}(\tilde{\mathbf{x}}) - \mathcal{H}(\mathbf{x}^*) \geq \frac{\partial \mathcal{H}}{\partial \mathbf{x}}|_{\mathbf{x}^*}^\top (\tilde{\mathbf{x}} - \mathbf{x}^*) = 0. \quad (\text{A.8})$$

This implies  $\mathcal{H}(\mathbf{x}^*) = \mathcal{H}(\tilde{\mathbf{x}}) = 0$ . However, since  $\mathcal{H}$  is globally positive definite and thus only vanishes at the origin, we have  $\mathbf{x}^* = \tilde{\mathbf{x}}$ . This is a contradiction as  $\tilde{\mathbf{x}} \neq \mathbf{0}$ , which concludes the proof.  $\square$

## B Experiment hyperparameters

In each experiment presented within this manuscript, the different models were trained for the same number of steps using identical learning rates. The mean squared error loss function and the ADAM [25] optimizer were used throughout. The model and training hyperparameters, along with the number of model instances trained for statistical evaluation, are provided in Table 2. Throughout the experiments presented in Sections 4.1 to 4.3, all NNs consisted of two hidden layers, each containing 16 neurons. The network width was increased to 32 neurons for the additive manufacturing surrogate in Section 4.4. All network weights were initialized using the Glorot uniform [15] scheme.

The PHS-based models (sPHNN, sPHNN-LM, PHNN, bPHNN) share the same structure, differing only in how the Hamiltonian  $\mathcal{H}$  is parametrized. Specifically, the sPHNNs use an FICNN with a normalization term, as described in Section 3. The only difference for the sPHNN-LM variant is that the minimum  $\mathbf{x}^*$  of the Hamiltonian is learned as an additional parameter. The PHNNs, by contrast, use an unconstrained FFNN without any normalization. Finally, the bPHNNs ensures positivity and radially unboundedness of the Hamiltonian by adopting the method described by Yu et al. [55] for the OnsagerNets Energy function. Here, the output of an unconstrained network is squared and combined with a quadratic regularization term. However, unlike OnsagerNet, we used standard FFNNs instead of ResNets.

The models were trained on a Windows machine with Intel i7-13700K (13th Generation) and 32 GB of memory. Table 3 provides representative computational costs for training and prediction. Here, *Derivative evaluation* refers to the average time taken by each model to compute the state derivative  $\dot{\mathbf{x}}$

Table 2. Hyperparameters per experiment. TF and DF refer to trajectory and derivative fitting, respectively. The number of instances trained per model type is listed in the "# inst." column, and  $n$  and  $m$  denote the number of dimensions of state and input. The final three columns list the parametrization of the matrices for sPHNN, sPHNN-LM, bPHNN, and PHNN as either state-dependent matrices (s.) or constant matrices (c.).

Experiment	# training steps	learning rate	# inst.	$n$	$m$	matrices		
						$\mathbf{J}$	$\mathbf{R}$	$\mathbf{G}$
Randomly initialized model	-	-	1	2	-	s.	s.	-
Spinning rigid body	50 000 DF	$1 \times 10^{-3}$	10	3	-	s.	c.	-
Cascaded tanks	50 000 TF	$5 \times 10^{-4}$	20	2	1	c.	c.	c.
Food processing surrogate	30 000 TF	$1 \times 10^{-4}$	20	2 to 5	1	c.	c.	c.
Additive manufacturing surrogate	20 000 DF and 10 000 TF	$1 \times 10^{-3}$ $1 \times 10^{-5}$	20	40	40	c.	c.	c.

Table 3. Representative computational costs for training and prediction. The top two rows report the training times for derivative and trajectory fitting, respectively. The bottom two rows show the time for a single derivative evaluation and a full trajectory integration using data and models from Section 4.3.

	sPHNN	sPHNN-LM	sNODE	bPHNN	PHNN	NODE
Derivative fit Section 4.1 (min)	0.577	0.644	0.672	0.635	0.539	0.346
Trajectory fit Section 4.3 (min)	37.663	39.160	-	38.329	19.111	17.951
Derivative evaluation (ms)	0.007	0.007	-	0.007	0.007	0.005
Model integration (ms)	0.365	0.393	-	0.422	0.237	0.221

for a given state  $x$  and input  $u$ , which needs to be evaluated for time integration. In contrast, *Model integration* denotes the time required to generate a single trajectory prediction via integration. While the latter better reflects practical use cases, the reported integration time also depends on the integration scheme and chosen step size. Since we use an adaptive step size controller, the step size in turn depends on the learned dynamics.

## C Spinning rigid body system

The governing equations of the three-dimensional spinning rigid body used in the experiment in Section 4.1 are described in Section 2.2. We generated  $(\omega, \dot{\omega})$  pairs for training by evaluating Equation (5) for varying  $\omega$ . To simulate a measurement process, we selected  $\omega$  values from 50 s long numerically integrated trajectories, using 10 randomly chosen initial conditions  $\omega(0) \in [0, 1]^3$  with parameters  $I = \text{diag}(1, 2, 3)$  and  $\mu = 0.01$ . A representative trajectory of  $\omega$  is shown in Figure C.1.

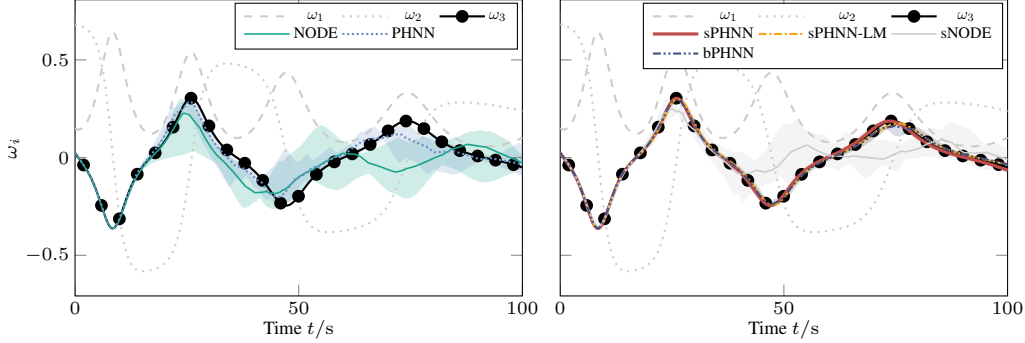


Figure C.1. Representative, extended training trajectory for spinning rigid body, including interquartile mean and range of model predictions for  $\omega_3$ . Left: Models without stability bias; Right: Models with stability bias.

## D Cascaded tanks data

The cascaded tanks dataset [47, 48] used in Section 4.2 consists of measurements of a fluid level control system. Liquid flows from an upper tank through an outlet into a lower tank, exits via a similar outlet into a reservoir, and is pumped back to the upper tank, see Figure D.1a. The system’s input  $u$  is the voltage applied to the pump, with higher values resulting in a faster volume flow. When the pump speed is large enough, an overflow can occur in the upper tank. A capacitive sensor measures the fluid level in the lower tank, and the resulting voltage represents the system’s output  $y$ . The dataset consists of two trajectories comprising 1024 pairs  $(u, y)$  each and Figure D.1b depicts the training trajectory. The initial water level in the upper tank is unknown, but it is the same for both trajectories.

Besides the experiments presented in Section 4.2, we further investigated the models’ capabilities to learn the hard saturation that the cascaded tanks data system exhibits at an output value of 10 V. This occurs when the tanks overflow in response to a large input voltage over an extended period of time (see Figure D.1b) [48]. We explore the models’ behavior in this regime by constructing a pump voltage signal that linearly increases from 1 V initially to 8 V at 2800 s. The resulting predictions

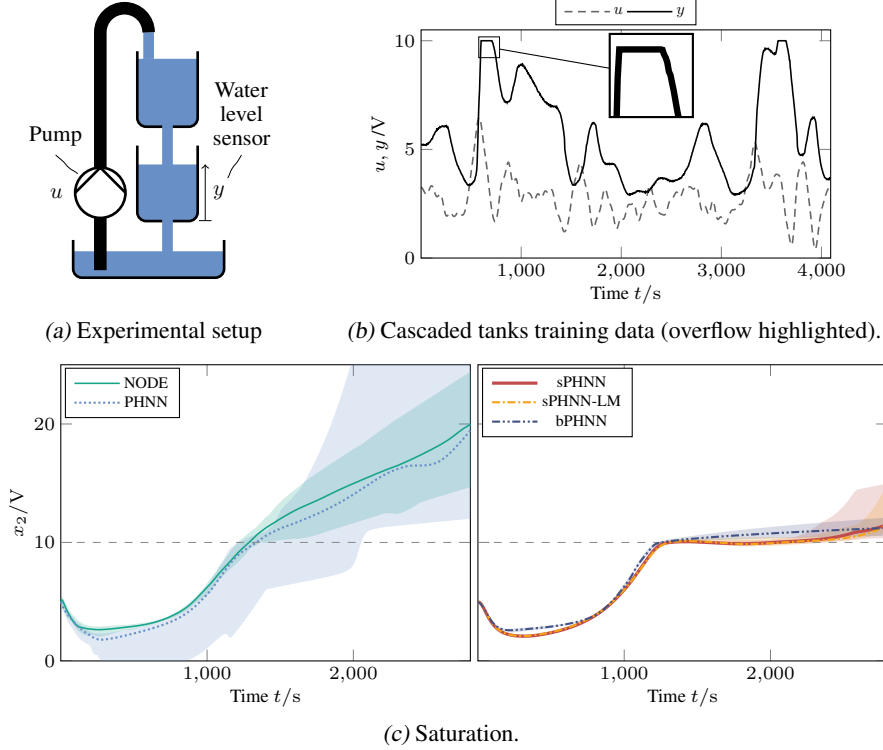


Figure D.1. *Cascaded tanks*: (a) System setup, (b) Training trajectory. (c) Predictions for a linearly increasing pump voltage. Due to an overflow, the true system exhibits a hard saturation at  $x_2 = 10$  V. Lines correspond to the interquartile mean, whereas shaded areas represent the interquartile range of the predictions from the 20 model instances.

are shown in Figure D.1c. For pump speeds up to  $\sim 7$  V ( $\sim 2400$  s), the sPHNN, sPHNN-LM, and bPHNN instances capture the saturation effect, while the NODE and PHNN instances fail to do so.

## E Thermal food processing surrogate data

This section provides additional details about the data used in Section 4.3, which is a subset of the data generated and used by Kannapinn et al. [21] in the creation of ROMs. The dataset describes the thermal processing of chicken breasts in a convection oven and was generated via the simulation of a soft-matter model for meats implemented in COMSOL Multiphysics. Each trajectory in the dataset describes the temperature history at two points inside the full order model. These result from a predefined excitation signal that controls the oven temperature. The probe points are situated in the center and close to the surface of the food item, with the corresponding temperatures labeled as  $T_A$  and  $T_B$ , respectively. To facilitate cross-referencing, the alphanumeric identifiers used by Kannapinn et al. [21] for the individual trajectories are retained. The training data consists of trajectories 745 and 795. They use an amplitude-modulated pseudo-random binary sequence (APRBS) and a multi-sine as the excitation signal  $T_{\text{oven}}$ . Figure E.1 (left and middle) illustrates  $T_{\text{oven}}$  along with the core and surface temperatures for both trajectories. This selection provides a fair evaluation of the models across the relevant temperature spectrum. For studying the impact of the number of training trajectories on the models, we additionally included 13 trajectories with identifiers 835, 851, 853, 861, 868, 873, 874, 888, 899, 902, 905, 906 and 917. In the experiment, the test group AP15 [21] is utilized as test data. It comprises 15 trajectories with APRBS excitations, specifically selected to ensure an even distribution of amplitudes and median values. Figure E.1 (right) shows an exemplary test trajectory, including the predictions for  $T_A$  from all models using 3 augmented dimensions. All trajectories span 1395 s and consist of 280 samples. The data is normalized for training to ensure the numerical values are of magnitude one. For this, an affine transformation is applied to  $T_{\text{oven}}$ ,  $T_A$ , and  $T_B$  individually. It shifts the ambient temperature to a value of zero and rescales the signal to have unit variance. The scaling and shifting factors are calculated using only the training data to ensure that the test data does not influence the training procedure.

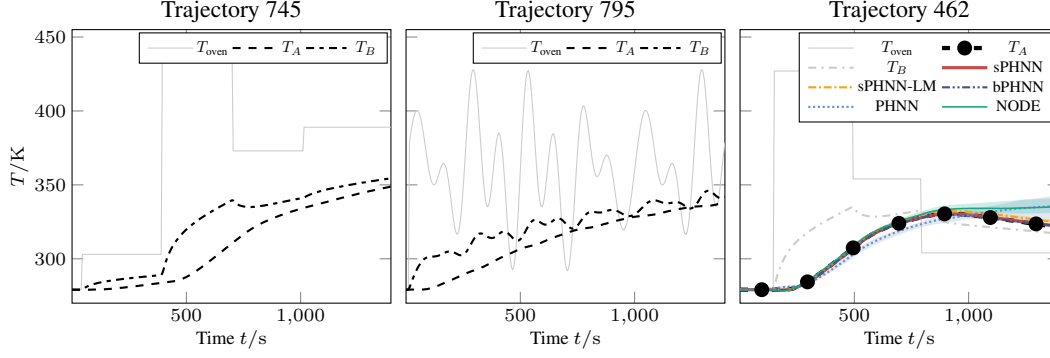


Figure E.1. *Left and middle:* Training trajectories. *Right:* Exemplary test trajectory with interquartile mean and range of the model predictions for  $T_A$ .

## F Thermal field data

### F.1 Dimensionality reduction

The following describes the POD method for dimensionality reduction of the field data applied in Section 4.4. The process is first described for general field data given as a matrix  $\mathbf{A} = [\mathbf{a}_0, \dots, \mathbf{a}_M]^\top$  of  $M$  snapshots  $\mathbf{a}_i \in \mathbb{R}^N$ , where  $M$  is the number of snapshots and  $N$  is the number of finite element nodes. To identify the dominant modes of the data, we perform a singular value decomposition (SVD) of  $\mathbf{A}$ :

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \in \mathbb{R}^{M \times N}, \quad (\text{F.1})$$

where  $\mathbf{U} \in \mathbb{R}^{M \times M}$  and  $\mathbf{V} \in \mathbb{R}^{N \times N}$  are unitary matrices and  $\mathbf{\Sigma} \in \mathbb{R}^{M \times N}$  is a rectangular diagonal matrix containing the singular values  $\sigma_1 \geq \dots \geq \sigma_M$  on its diagonal. Dimensionality reduction is achieved by truncating the SVD  $\mathbf{A} \approx \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^\top$ , where  $\tilde{\mathbf{U}} \in \mathbb{R}^{M \times n}$  and  $\tilde{\mathbf{V}} \in \mathbb{R}^{N \times n}$  contain the first  $n$  columns of  $\mathbf{U}$  and  $\mathbf{V}$ , and  $\tilde{\mathbf{\Sigma}} = \text{diag}(\sigma_1, \dots, \sigma_n)$  is the truncated matrix of the dominant singular values. The dominant modes in the data are then given by  $\mathbf{M} = c\tilde{\mathbf{V}}$ . Here, the scaling factor  $c$  is the standard deviation of the elements in  $\tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}$ . This scaling ensures that the latent trajectories will have unit variance. To map a latent state  $\mathbf{x} \in \mathbb{R}^n$  to the full-order field  $\mathbf{X} \in \mathbb{R}^N$  we compute:

$$\mathbf{X} = \mathbf{M}\mathbf{x}. \quad (\text{F.2})$$

Conversely, to map from the full-order field to the latent space, we solve the least-squares regression problem:

$$\mathbf{x} = \arg \min_{\mathbf{x}'} \|\mathbf{X} - \mathbf{M}\mathbf{x}'\| = \frac{1}{c^2} \mathbf{M}^\top \mathbf{X}. \quad (\text{F.3})$$

In Section 4.4, we consider two fields: The thermal field describing the state of the system and the excitation given by the heat source field. For each, a separate SVD is performed to compute  $\mathbf{M}_T$  and  $\mathbf{M}_q$  using the respective snapshots from the two training trajectories. We thus obtain two pairs of encoders  $\psi$  and decoders  $\phi$ :

$$\psi_T(\mathbf{T}) = \arg \min_{\mathbf{x}} \|\mathbf{T} - \mathbf{M}_T \mathbf{x}\|, \quad \phi_T(\mathbf{x}) = \mathbf{M}_T \mathbf{x}, \quad (\text{F.4})$$

$$\psi_q(\mathbf{q}) = \arg \min_{\mathbf{u}} \|\mathbf{q} - \mathbf{M}_q \mathbf{u}\|, \quad \phi_q(\mathbf{u}) = \mathbf{M}_q \mathbf{u}. \quad (\text{F.5})$$

To fully leverage the benefits of sPHNNs, it is essential to ensure that the absence of a heat source is encoded as the zero-vector in the corresponding latent space. Since the mapping from the latent space Equation (F.2) is linear, and  $\mathbf{q} = \mathbf{0}$  corresponds to the absence of a heat source in the full-order space, this requirement is automatically fulfilled. However, for applications where a nonzero state  $\tilde{\mathbf{x}}$  describes the absence of an external energy source or for general nonlinear latent mappings (such as autoencoders), the condition can be met by applying a constant translation in the latent space, which shifts the encoded state  $\tilde{\mathbf{x}} = \psi(\tilde{\mathbf{x}})$  to the origin. In Section 4.4, we use this method to ensure that the thermal equilibrium corresponds to the origin of the temperature latent space. While this is not strictly necessary, knowledge of the equilibrium location in the latent space enables its integration

into the sPHNN’s architecture by fixing the stable equilibrium position. Consequently, we arrive at the representation of the latent state

$$\mathbf{x} = \psi_T(\mathbf{T}) - \psi_T(\mathbf{T}_{\text{eq}}), \quad \mathbf{T} \approx \phi_T(\mathbf{x} + \psi_T(\mathbf{T}_{\text{eq}})). \quad (\text{F.6})$$

where  $\mathbf{T}_{\text{eq}}$  is the full-order state vector describing thermal equilibrium. The complete model architecture is represented in Figure F.1. There, the function  $\mathbf{f}$  is used as a placeholder for an arbitrary evolution equation, to generalize to all model architectures used in Section 4.4.

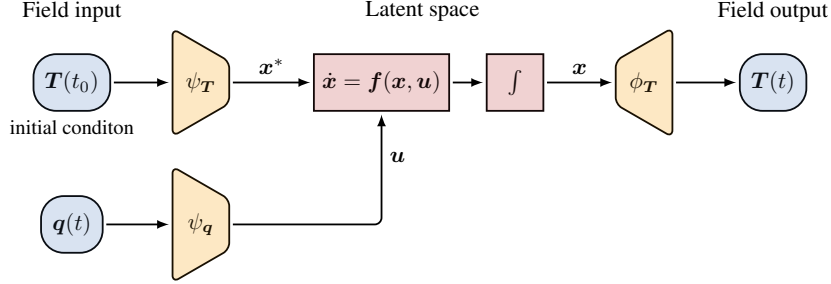


Figure F.1. Model setup used in Section 4.4. The initial temperature field  $\mathbf{T}(t_0)$  and input field  $\mathbf{q}(t)$  are mapped via encoders  $\psi_T, \psi_q$  into the latent space, where the latent dynamics  $\mathbf{f}$  are integrated. Here,  $\mathbf{f}$  is used as a placeholder for a specific model, e.g., NODE or sPHNN. Finally, the predicted latent trajectory  $\mathbf{x}(t)$  is mapped to the full-order field via the decoder  $\phi_T$ .

Finally, we note that alternative dimensionality reduction techniques could be applied. In preliminary experiments, we explored the use of autoencoders for this purpose. Although autoencoders offer advantages over the linear SVD regarding achievable compression rates, they exhibited poor extrapolation performance when trained on sparse data. This issue could potentially be mitigated by training the autoencoders with additional synthetic data, which would not necessitate costly simulations; however, further research is needed to explore this approach.

## F.2 Additional results

We briefly present additional results from Section 4.4 to demonstrate that the differences in predictions across the various model architectures remain consistent across multiple instances. Figure F.2 depicts the RMSEs of each model from Section 4.4 computed over all 25 trajectories and the spatial domain. The behavior of the models with stability bias can be distinguished from the NODE and PHNN, which suffer from instabilities that result in rapidly increasing errors. While both sPHNN variants and bPHNNs exhibit decreasing errors as the system returns to thermal equilibrium, sPHNNs achieve a peak RMSE nearly an order of magnitude lower, due to their strong inductive bias from enforcing global stability. The sPHNN-LM performs almost identically to the sPHNN initially, while showing slightly larger errors in the convergence to equilibrium. This is due to slight errors in the learned equilibrium position, which can be addressed by increasing the amount of training data, as demonstrated in Section 4.3.

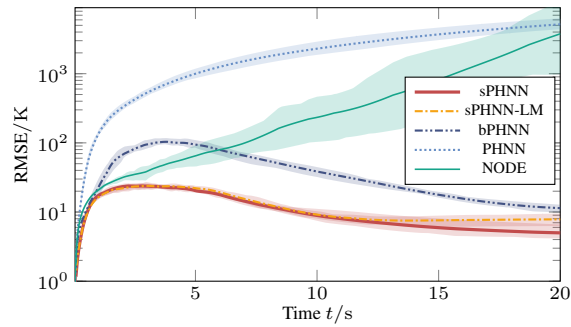


Figure F.2. Interquartile mean and range of the RMSEs evaluated with all 25 trajectories of the thermal field data.

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: Our claim, that the enforcement of global stability and physics-based conditions results in a strong inductive bias and increased performance in applicable cases, is highlighted in the abstract and introduction, and demonstrated by our empirical results.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: Although we do not include a dedicated "Limitations" section, the paper explicitly discusses the assumptions and limitations of our approach, such as the global stability requirement and its implications for model expressiveness, at relevant points throughout the text.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory assumptions and proofs



Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: Our theoretical result provided in Theorem 3.1 are accompanied by a intuitive discussion in Section 3 and a rigorous proof in the appendix Appendix A.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: We provide an extensive description of the proposed model architecture as well as the experimental setup. This includes the model hyperparameters, which are listed in Appendix B.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

## 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Code and model weights are provided in this GitHub repository: <https://github.com/CPSHub/sphnn-publication>. The repository includes instructions on how to set up the environment, access the data, and run the experiments to reproduce the results of the paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We supply the essential training and test details necessary to understand the results in the main paper, and provide additional information about the data and model hyperparameters in the appendix and code repository.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: In the discussion of our experiments, we provide suitable measures on the variance between individual initializations of the different models.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

#### 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Given that the target application involves small neural networks trained on sparse data, the computational requirements are relatively modest. We have specified the system used to generate the reported results, along with approximate training times, in Appendix B.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

#### 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: The present research conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

#### 10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: This work focuses on foundational research involving small-scale, physically structured neural networks trained on sparse data for modeling dynamical systems. It is not tied to specific applications or deployments, and we do not foresee any direct societal impacts arising from its current form.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

## 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This paper does not involve the release of models or datasets that pose a high risk of misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All external assets used in this work are properly cited in the bibliography, including relevant version information, URLs, and licenses where applicable. We ensured

that the license terms and conditions of use were respected and named the license explicitly in the bibliography.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

### 13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: This work does not release any new datasets, models, or other assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This work does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

### 15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: This work does not involve human subjects, user studies, or any form of data collection from individuals.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

#### 16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This work does not involve large language models in any part of the core research methodology.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.