Nonlinear Dyn (2025) 113:29909–29926
https://doi.org/10.1007/s11071-025-11634-x

RESEARCH

# Universal-basis neural ODE modeling of the discrete sine-gordon system

**Mingkun Li · Zaijiu Shang · Peng Wang · Hong-Kun Zhang · Junjie Fan**

Received: 29 April 2025 / Accepted: 17 July 2025 / Published online: 6 August 2025


**Abstract** We propose a data-driven framework for learning and predicting solutions of the discrete sine-Gordon equation by combining universal-basis expansions with neural ODE architectures. In this approach, polynomial and trigonometric basis functions are embedded into the network's representation of the Hamiltonian, enabling efficient approximation of non-polynomial interactions in the lattice. We further incorporate *symmetry-informed* penalties, which enforce invariants such as reflection symmetry and energy conservation, thereby enhancing stability and long-horizon accuracy. Numerical experiments on both sine-wave and breather soliton initializations demonstrate that our *universal-basis neural ordinary differential equations* (UB-NODEs) yield accurate particle trajectories and maintain essential soliton properties over extended times. Moreover, empirical comparisons reveal the advantages of adding symmetry-based constraints, including faster convergence and reduced overfitting. This methodology is broadly applicable to other Hamiltonian lattice systems and paves the way for deeper machine-learning investigations of complex nonlinear dynamics.

**Keywords** Discrete sine-Gordon equation · Machine learning · Neural ordinary differential equations · Universal basis neural networks · Universal basis neural ordinary differential equations

M. Li, Z. Shang, P. Wang, H.-K. Zhang, J. Fan: These authors contributed equally to this work.

M. Li · P. Wang
Great Bay Institute for Advanced Study, Dongguan 523000, China

M. Li
School of Mathematical Sciences, University of Science and Technology of China, Hefei 230026, China

M. Li · H.-K. Zhang
School of Sciences, Great Bay University, Dongguan 523000, China

Z. Shang
Center for Mathematics and Interdisciplinary Sciences, Fudan University, Shanghai 200433, China

Z. Shang
Shanghai Institute for Mathematics and Interdisciplinary Sciences, Shanghai 200438, China

P. Wang
Department of Mathematics, Sun Yat-sen University, Guangzhou 510275, China

J. Fan (✉)
School of Science, Inner Mongolia University of Technology, Hohhot 010051, China
e-mail: tracyfan2432@126.com

## 1 Introduction

The discrete sine-Gordon equation is a significant mathematical model with broad applications in physics and applied mathematics. It emerges as a discrete analogue of the continuous sine-Gordon equation [1], a classical integrable field theory with well-documented relevance in condensed matter, nonlinear optics, and

string theory [2]. Within its lattice-based form, the sine-Gordon model describes systems possessing nonlinear particle interactions, appearing in settings such as wave propagation [3–6], crystal dislocations [7–9], coupled pendulum arrays [10–13], and Josephson junction networks [14,15]. The discrete sine-Gordon equation also underpins physical phenomena in superconductivity [16,17] and DNA base-pair dynamics [18,19], making it both challenging and representative of more extensive classes of complex lattice models.

Other nonlinear lattice systems, including the discrete nonlinear Schrödinger model [20], the discrete $\phi^4$ model [21], the Fermi–Pasta–Ulam–Tsingou system [22], and the discrete complex Ginzburg–Landau equation [23], typically employ polynomial Hamiltonians, which simplifies certain analytical and numerical steps. In contrast, the discrete sine-Gordon system involves a non-polynomial Hamiltonian, substantially raising the level of complexity in studying its soliton-type solutions and long-term behaviors.

Earlier research pursued explicit formulations of the discrete sine-Gordon model as a Hamiltonian ODE [24] and applied classical solution methods. Meanwhile, machine learning strategies like physics-informed neural networks (PINNs) [25] have emerged as powerful new tools for parameter estimation and solution approximation in lattice systems [26–28]. However, two inherent difficulties arise in classical or existing ML-based methods. (1) In many practical scenarios, only the system's positions and velocities are observed, with no guarantee it follows a discrete sine-Gordon equation. How can one infer correct governing dynamics under partial or incomplete knowledge? No comprehensive data-driven machine-learning scheme specifically tailored to the discrete sine-Gordon equation has yet been fully explored. (2) The discrete sine-Gordon system's sensitive dependence on initial states can lead to exponential trajectory divergence. Is there a robust method to maintain accurate solutions for extended times, despite potential chaotic drifts?

To construct a data-driven model for the machine learning of the discrete sine-Gordon equation, we develop a universal basis neural network (UBNN) architecture, embedding both polynomial and trigonometric basis functions into the Hamiltonian framework of the discrete sine-Gordon equation. We then formulate *universal basis neural ordinary differential equations* (UB-NODEs) by placing UBNNs within the neural ODE paradigm (see Fig. 2). By incorporating *symmetry-informed loss functions*, our approach preserves crucial physical properties of the discrete sine-Gordon lattice, particularly those salient for soliton solutions.

We additionally study how symmetry-loss weighting and lattice spacing influence energy conservation and stability over long simulation times. Our work makes three principal contributions:

1) **Automated Discovery of Non-Polynomial Dynamics.** UB-NODEs embed trigonometric and polynomial basis expansions within neural ODEs to capture the nonlinear terms of discrete sine-Gordon systems, even without prior knowledge of the model form.
2) **Symmetry-Informed Learning for Soliton Dynamics.** By integrating loss functions that encode invariants, such as energy and momentum conservation, UB-NODEs effectively mitigate trajectory divergence, resulting in improved stability over long horizons.
3) **Soliton Prediction and Extended Robustness.** The synergy between universal basis expansions and Hamiltonian constraints enables precise approximations of soliton solutions while supporting stable large-time integration, even under conditions where classical methods become overly sensitive.

Hence, the UB-NODE methodology constitutes our main contribution: a flexible neural network platform designed to learn non-polynomial Hamiltonian structures, augmented by symmetry-preserving techniques for faster training and improved stability. By combining Hamiltonian mechanics with UBNNs, we provide a robust, versatile solution for data-driven investigations of the discrete sine-Gordon equation and its soliton phenomena, setting the stage for future advancements in data-driven modeling of complex lattice dynamics.

The remainder of this paper is organized as follows. In Sect. 2, we introduce the discrete sine-Gordon model and its Hamiltonian formulation, highlighting the challenges of non-polynomial dynamics and soliton behavior. Section 3 describes our methodology in detail: universal-basis neural networks (UBNNs), their integration into a neural ODE framework (UB-NODEs), and the symmetry-informed loss function. Section 4 presents a series of numerical experiments, including data generation, a comparative study of models with and without symmetry constraints, and extended tests on sine-wave and breather solitons. Section 5 offers a

discussion of our findings, underscoring the method's strengths, limitations, and broader implications for Hamiltonian lattice systems. Finally, Section 6 concludes the paper by summarizing key takeaways and outlining directions for future research.

## 2 Background on the discrete sine-Gordon model

In this section, we formally state the problem to be solved and provide a concise yet comprehensive overview of the discrete sine-Gordon equation along with its Hamiltonian formulation.

### 2.1 Problem statement

In this article, we aim to develop a data-driven model for the discrete sine-Gordon equation. Specifically, we are provided with a collection of $S$ discrete solution trajectories, each representing the evolution of the system over time as governed by the true dynamics of the discrete sine-Gordon equation:

$$\mathcal{D} = \left\{ \left( \mathbf{u}_0^{(\alpha)}, \mathbf{v}_0^{(\alpha)} \right), \left( \mathbf{u}_1^{(\alpha)}, \mathbf{v}_1^{(\alpha)} \right), \ldots, \left( \mathbf{u}_K^{(\alpha)}, \mathbf{v}_K^{(\alpha)} \right) \;\middle|\; 1 \le \alpha \le S \right\},$$

where each trajectory $\left\{ (\mathbf{u}_k^{(\alpha)}, \mathbf{v}_k^{(\alpha)}) \right\}_{k=0}^{K}$ corresponds to a sequence of time steps originating from the initial state $(\mathbf{u}_0^{(\alpha)}, \mathbf{v}_0^{(\alpha)})$.

Our objective is to construct a predictive model that, given a new initial condition, can accurately recover the future states of the system, consistent with the discrete sine-Gordon dynamics.

### 2.2 From continuous to discrete sine-Gordon

To establish a clear foundation, we recall the continuous sine-Gordon equation, highlighting its fundamental properties and structural characteristics. We then describe how these properties extend naturally to the discrete case, preserving essential dynamical and Hamiltonian structures while adapting to the constraints of a discrete spatial domain.

A classical sine-Gordon system in continuous space-time $(x, t) \in \mathbb{R} \times \mathbb{R}^+$ is governed by the nonlinear hyperbolic partial differential equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - \sin(u). \tag{2.1}$$

Since many physical media are naturally discrete, one often replaces the continuous coordinate $x$ by a uniform lattice $\{x_k = x_0 + k\,h,\ k = 1, \ldots, N\}$, where $h$ denotes the lattice spacing [29]. On this discrete set of sites, the field $u_k(t)$ approximates $u(x_k, t)$. The discrete sine-Gordon equation then takes the form

$$\frac{d^2 u_k}{dt^2} = \frac{1}{h^2}(u_{k+1} - u_k) - \frac{1}{h^2}(u_k - u_{k-1}) - \sin(u_k), \quad k = 1, \ldots, N. \tag{2.2}$$

Boundary conditions for (2.2) may be specified in various ways, such as $u_{N+1} = u_N,\ u_1 = u_0$ or periodic identifications $u_{N+1} = u_1,\ u_N = u_0$.

The discrete sine-Gordon model retains many of the integrable properties of its continuous counterpart, such as the existence of *soliton solutions* [30], i.e., a traveling wave solution consisting of a single peak or trough that propagates in isolation without change in size, shape or speed. Figure 1 shows a soliton solution of discrete sine-Gordon equation called breather.

### 2.3 Hamiltonian formulation and challenges

An important perspective on (2.2) is that it arises from a non-polynomial Hamiltonian system. Specifically, let

$$(\mathbf{u}, \mathbf{v}) = (u_1, \ldots, u_N,\ v_1, \ldots, v_N)$$

be the position and velocity coordinates in a $2N$-dimensional phase space $\mathbb{R}^{2N}$. The discrete sine-Gordon Hamiltonian function

$$H(\mathbf{u}, \mathbf{v}) = T(\mathbf{v}) + V(\mathbf{u}), \tag{2.3}$$

splits naturally into kinetic and potential parts,

$$T(\mathbf{v}) = \sum_{k=1}^{N} \frac{1}{2} v_k^2,$$

$$V(\mathbf{u}) = \sum_{k=1}^{N} \left[ \tfrac{1}{2h^2} (u_k - u_{k-1})^2 + 1 - \cos(u_k) \right].$$

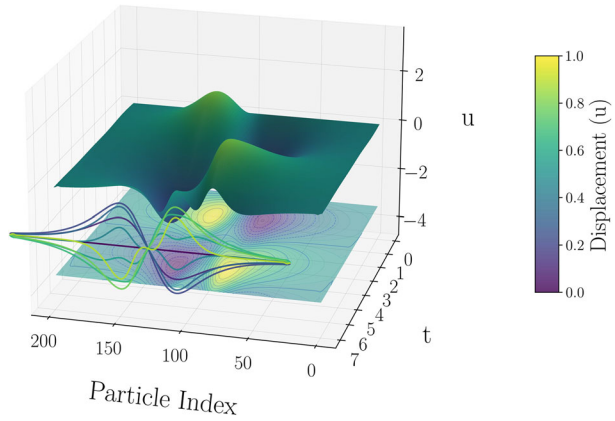



**Fig. 1** A discrete sine-Gordon breather solution visualized in three dimensions. The horizontal axes represent the particle index and time, while the vertical axis (and color map) indicates the displacement field $u$. Several trajectories are superimposed in the foreground to highlight how individual sites evolve in time. The contour projection beneath the surface further reveals local peaks and troughs. Such a breather remains localized and oscillatory without dispersing, reflecting the integrable nature of the sine-Gordon structure

The induced Hamiltonian ODE system is

$$\dot{u}_k = \frac{\partial H}{\partial v_k}, \quad \dot{v}_k = -\frac{\partial H}{\partial u_k}, \quad k = 1, \ldots, N, \tag{2.4}$$

which, upon substitution of (2.3) into (2.4), yields

$$\begin{aligned} \frac{du_k}{dt} &= v_k, \\ \frac{dv_k}{dt} &= \frac{1}{h^2}\left(u_{k+1} - u_k\right) \\ &\quad -\frac{1}{h^2}\left(u_k - u_{k-1}\right) - \sin(u_k), \end{aligned} \tag{2.5}$$

equivalent to the discrete sine-Gordon form (2.2).

Compared to many simpler lattice equations whose Hamiltonians are polynomial, the sine-Gordon system's potential contains a non-polynomial term $\cos(u_k)$. This introduces additional complications in analytic treatment and in capturing long-term soliton-like behaviors. Numerical integrators must also handle the high sensitivity to initial data, as small perturbations can evolve into large phase errors over extended times [24,31,32]. These features underline why the discrete sine-Gordon equation is viewed as a challenging and representative model for nonlinear lattice dynamics. In the context of data-driven neural networks that incorporate numerical algorithms, the network generates approximations at successive time steps, which are in turn used to guide further training and refine the predictions. As a result, stability in such models refers to the algorithm's capacity to reliably forecast the system's future behavior. In this work, the stability of our proposed method is verified through numerical experiments conducted under various initial conditions, demonstrating its robustness across different dynamical scenarios.

## 3 Methodology

In this section, we present a systematic framework for learning the discrete sine-Gordon equation in a data-driven manner. We begin by establishing a universal approximation theorem (Theorem 3.1) for vector-valued functions whose Jacobian is symmetric, which is relevant for gradients of scalar potentials such as Hamiltonian functions. We then introduce our *universal basis neural networks* (UBNNs), which leverage these approximation results to learn the gradient of the sine-Gordon Hamiltonian. Lastly, we describe how these UBNNs are embedded into the *neural ODE* framework, yielding what we term *UB-NODEs*, and we discuss the role of *symmetry-informed loss* in preserving the discrete sine-Gordon system's essential physical properties.

### 3.1 Universal approximation for symmetric gradients

A central goal in learning a nonlinear Hamiltonian lattice system is to approximate its Hamiltonian gradient from observed data.

***Basis definitions***

Let $I \subset \mathbb{R}$ be an open interval and $\overline{I}$ its closure. Denote by $\mathcal{F}(\overline{I}, \mathbb{R})$ the space of $C^2$ functions on $\overline{I}$. A *basis* for $\mathcal{F}(\overline{I}, \mathbb{R})$ is a set of smooth functions $\{e_0, e_1, e_2, \dots\}$ whose finite linear combinations can approximate any $C^2$ function arbitrarily well. By the Weierstrass approximation theorem, polynomials form such a basis for continuous functions on a closed interval, while Dirichlet-type theorems ensure that trigonometric functions can serve as bases for periodic functions.

We begin by stating a universal approximation theorem showing that, under suitable smoothness and symmetry assumptions, one can approximate any vector field whose Jacobian is symmetric by finite sums of certain *multiplicative basis* functions (e.g., polynomials or trigonometric functions). This result underlies the design of UBNNs.

**Theorem 3.1** (Universal approximate theorem) *Let $I$ be an interval of $\mathbb{R}$, and let $\mathcal{F}(\overline{I}, \mathbb{R})$ denote the space of $C^2$ functions defined on the closure $\overline{I}$. Suppose $\mathbf{f} : \overline{I}^N \to \mathbb{R}^N$ is a $C^1$ function whose Jacobian, $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$, is symmetric for all $\mathbf{x} \in \overline{I}^N$. Then, for any $\epsilon > 0$, there exist integers $M, N_1 \in \mathbb{Z}^+$, a diagonal matrix $\Lambda_i = \mathrm{diag}(\lambda_{i;1}, \dots, \lambda_{i;N_1})$, and a matrix*

$$P_i = \begin{pmatrix} p_{i;11} & \cdots & p_{i;1N} \\ \vdots & \ddots & \vdots \\ p_{i;N_1 1} & \cdots & p_{i;N_1 N} \end{pmatrix} \in \mathbb{R}^{N_1 \times N},$$

*such that*

$$\left\| \mathbf{f}(\mathbf{x}) - \sum_{i=0}^{M} P_i^\top \circ \Lambda_i \circ e_i' \circ P_i \circ \mathbf{x} \right\| < \epsilon, \quad (3.1)$$

*where $\{e_0, e_1, e_2, \dots\}$ is a* multiplicative basis *in $\mathcal{F}(\overline{I}, \mathbb{R})$, and $e_i'$ denotes the derivative of $e_i$, applied element-wise. In other words, $\mathbf{f}$ can be uniformly approximated by a finite sum of suitably transformed and element-wise applied basis derivatives.*

*Proof* Consider the differential 1-form $\alpha = \mathbf{f} d\mathbf{x}$. $\alpha$ is closed since $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ is symmetric. Then by Poincaré's lemma, there exists a $C^2$ function $F : \overline{I}^N \to \mathbb{R}$, such that $\frac{\partial F}{\partial \mathbf{x}} = \mathbf{f}$. Multi-variable function $F$ can be approximated by linear combination of basis functions, i.e.,

$$F(x_1, \dots, x_N) = \sum_{i_1, \dots, i_N = 0}^{\infty} a_{i_1, \dots, i_N} e_{i_1}(x_1) \dots e_{i_N}(x_N) \quad (3.2)$$

By a multiplicative basis, we mean a basis $\{e_0, e_1, e_2, \dots\}$ for $\mathcal{F}(\overline{I}, \mathbb{R})$, such that for any product $e_{i_1}(x_1) \dots e_{i_N}(x_N)$ evaluated at $(x_1, \dots, x_N) \in \overline{I}^N$, one can rewrite

$$e_{i_1}(x_1) \dots e_{i_N}(x_N) = \sum_{j=1}^{N_1} \mu_j \, e_K \left( \sum_{k=1}^{N} p_{jk} \, x_k \right),$$

for some integer $N_1$ and suitable real parameters $(\mu_j)$ and $(p_{jk})$. In other words, the product of basis elements remains in the span of the basis under suitable re-indexing. Since $\{e_0, e_1, e_2, \dots\}$ is a multiplicative basis of $\mathcal{F}(\overline{I}, \mathbb{R})$, there exists $N_{i_1, \dots, i_N} \in \mathbb{Z}^+$, a vector $\boldsymbol{\mu} = \left( \mu_{i_1, \dots, i_N; 1} \cdots \mu_{i_1, \dots, i_N; N_{i_1, \dots, i_N}} \right) \in \mathbb{R}^{N_{i_1, \dots, i_N}}$ and matrix

$$P = \begin{pmatrix} p_{i_1, \dots, i_N; 11} & \cdots & p_{i_1, \dots, i_N; 1N} \\ \vdots & \ddots & \vdots \\ p_{i_1, \dots, i_N; N_{i_1, \dots, i_N} 1} & \cdots & p_{i_1, \dots, i_N; N_{i_1, \dots, i_N} N} \end{pmatrix} \in \mathbb{R}^{N_{i_1, \dots, i_N} \times N},$$

such that the term $a_{i_1, \dots, i_N} e_{i_1}(x_1) \dots e_{i_N}(x_N)$ can be written as

$$a_{i_1, \dots, i_N} e_{i_1}(x_1) \dots e_{i_N}(x_N) = \sum_{j=1}^{N_{i_1, \dots, i_N}} a_{i_1, \dots, i_N} \mu_{i_1, \dots, i_N; j} e_{K_{i_1, \dots, i_N}} \left( \sum_{k=1}^{N} p_{i_1, \dots, i_N; jk} x_k \right).$$

Denote $a_{i_1, \dots, i_N} \mu_{i_1, \dots, i_N; j}$ by $\lambda_{i_1, \dots, i_N; j}$, then we write

$$a_{i_1, \dots, i_N} e_{i_1}(x_1) \dots e_{i_N}(x_N)$$

$$= \sum_{j=1}^{N_{i_1,\ldots,i_N}} \lambda_{i_1,\ldots,i_N;j} e_{K_{i_1,\ldots,i_N}} \left( \sum_{k=1}^{N} p_{i_1,\ldots,i_N;jk} x_k \right).$$

Hence, multi-variable function $F$ is approximated by

$$F(x_1, \ldots, x_N) = \sum_{i=0}^{\infty} \sum_{K_{i_1,\ldots,i_N}=i} \sum_{j=1}^{N_{i_1,\ldots,i_N}} \lambda_{i_1,\ldots,i_N;j} e_{K_{i_1,\ldots,i_N}} \left( \sum_{k=1}^{N} p_{i_1,\ldots,i_N;jk} x_k \right).$$

By combining like terms, we can simplify the equation into

$$F(x_1, \ldots, x_N) = \sum_{i=0}^{\infty} \sum_{j=1}^{N_i} \lambda_{i;j} e_i \left( \sum_{k=1}^{N} p_{i;jk} x_k \right). \tag{3.3}$$

The derivative $\mathbf{f}$ of $F$ can then be approximated by the derivative of (3.3), i.e.,

$$\mathbf{f}(\mathbf{x}) = \sum_{i=0}^{\infty} \sum_{j=1}^{N_i} \lambda_{i;j} e_i' \left( \sum_{k=1}^{N} p_{i;jk} x_k \right) (p_{i;j1}, \ldots, p_{i;jN})^T,$$

which means for $\epsilon > 0$, there exists $M, N_1 \in \mathbb{Z}^+$ such that

$$\left\| \mathbf{f}(\mathbf{x}) - \sum_{i=0}^{M} \sum_{j=1}^{N_1} \lambda_{i;j} e_i' \left( \sum_{k=1}^{N} p_{i;jk} x_k \right) (p_{i;j1}, \ldots, p_{i;jN})^T \right\| < \epsilon. \tag{3.4}$$

We see $e_i'$ as an element-wise function, then (3.4) can be written in an equivalent form

$$\left\| \mathbf{f}(\mathbf{x}) - \sum_{i=0}^{M} P_i^T \circ \Lambda_i \circ e_i' \circ P_i \circ \mathbf{x} \right\| < \epsilon.$$

□

Intuitively, Theorem 3.1 shows that for any smooth function $\mathbf{f}$ arising as a gradient field (hence a symmetric Jacobian), there is a finite linear combination of basis elements (and their derivatives) that approximates $\mathbf{f}$ to arbitrary accuracy. This sets the stage for constructing neural networks that approximate gradients of Hamiltonians, including those of sine-Gordon type.

### 3.2 Universal basis neural networks (UBNNs)

Using Theorem 3.1 as a foundation, we introduce **universal basis neural networks** (UBNNs), which approximate vector fields of interest by sums of basis expansions. Specifically, we aim to learn the partial derivatives of a separable Hamiltonian

$$H(\mathbf{u}, \mathbf{v}) = T(\mathbf{v}) + V(\mathbf{u}).$$

In the discrete sine-Gordon model [6,7], $H$ is not polynomial, but it can still be suitably approximated by polynomials, trigonometric expansions, or a mixture thereof.

***Overview of basis expansions***

Let $\{e_0, e_1, \ldots, e_M\}$ be a finite subset of a multiplicative basis for $\mathcal{F}(I, \mathbb{R})$. In the polynomial case, for instance, $e_i(x) = x^i$. For each $e_i$, we also consider its derivative $e_i'$, applied element-wise to vector arguments. The UBNN is then constructed by combining these basis expansions with learned linear transformations, as indicated in Theorem 3.1.

***Approximation of $T_{\mathbf{v}}$ and $V_{\mathbf{u}}$***

Given that $H(\mathbf{u}, \mathbf{v}) = T(\mathbf{v}) + V(\mathbf{u})$ separates into a velocity part and a position part, we define two networks

$$\boldsymbol{\Phi}(\mathbf{v}; P_\Phi, Q_\Phi) \quad \text{and} \quad \boldsymbol{\Psi}(\mathbf{u}; P_\Psi, Q_\Psi),$$

where each one approximates the partial gradient needed for Hamilton's equations:

$$\nabla_{\mathbf{v}} T(\mathbf{v}) \quad \text{and} \quad \nabla_{\mathbf{u}} V(\mathbf{u}),$$

respectively. Symbolically,

$$\begin{aligned} \boldsymbol{\Phi}(\mathbf{v}; P_\Phi, Q_\Phi) &= \sum_{i=0}^{M} \tau_{e_i}\left(P_{\Phi,i}, Q_{\Phi,i}\right) \circ \mathbf{v}, \\ \boldsymbol{\Psi}(\mathbf{u}; P_\Psi, Q_\Psi) &= \sum_{i=0}^{M} \tau_{e_i}\left(P_{\Psi,i}, Q_{\Psi,i}\right) \circ \mathbf{u}, \end{aligned} \tag{3.5}$$

where each $\tau_{e_i}$ denotes a transformation by $e_i'$ and appropriate parameters $P_{\Phi,i}, Q_{\Phi,i}$ (or $P_{\Psi,i}, Q_{\Psi,i}$), in

accordance with the finite-sum approximation in (3.1). These parameters are learned through gradient-based training.

***Choice of basis***

For continuous or analytic Hamiltonians (e.g., polynomial-like expansions), one might favor polynomials $x^i$. For periodic or near-periodic cases, sines/cosines are frequently beneficial [33]. Indeed, the discrete sine-Gordon equation can exhibit periodic boundary conditions and soliton solutions akin to traveling waves, so trigonometric expansions can be advantageous. In any scenario, the theorem ensures that there exists a suitable finite expansion for approximating the Hamiltonian gradient to arbitrary precision.

### 3.3 Embedding UBNNs in neural ODEs (UB-NODEs)

Having built UBNNs that approximate the partial derivatives of $H$, we now place them in a *Neural ODE* framework to learn an ODE-based solver for the discrete sine-Gordon system. Concretely, from (2.5), we recall that

$$\dot{\mathbf{u}}(t) \;=\; \nabla_{\mathbf{v}}\, T\big(\mathbf{v}(t)\big), \quad \dot{\mathbf{v}}(t) \;=\; -\,\nabla_{\mathbf{u}}\, V\big(\mathbf{u}(t)\big).$$

Using the UBNNs $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ from (3.5), we replace these partial derivatives by learned expansions:

$$\dot{\mathbf{u}}(t) = \boldsymbol{\Phi}\big(\mathbf{v}(t), \boldsymbol{\theta}\big), \quad \dot{\mathbf{v}}(t) = -\,\boldsymbol{\Psi}\big(\mathbf{u}(t), \boldsymbol{\theta}\big),$$

where $\boldsymbol{\theta} = (P_\Phi, Q_\Phi, P_\Psi, Q_\Psi)$ is the parameter to be trained.

Because these right-hand sides are differentiable in the parameters, one can integrate them using standard neural ODE solvers and backpropagate to train $\boldsymbol{\theta} = (P_\Phi, Q_\Phi, P_\Psi, Q_\Psi)$ based on observed data. We refer to this combination of a universal basis expansion with a neural ODE solver as a **UB-NODE** (Fig. 2).

In the optimization of the parameters, we begin by computing the predicted solution $(\hat{\mathbf{u}}_{t_1}, \hat{\mathbf{v}}_{t_1})$ at time $t_1$ using the Hamiltonian vector field $\hat{X}_{\text{Hamilton}}(\boldsymbol{\theta})$, parameterized by $\boldsymbol{\theta}$. This predicted state is then compared with the ground-truth solution $(\mathbf{u}_{t_1}, \mathbf{v}_{t_1})$ at the same time point. The discrepancy between these two values is quantified using a loss function, which incorporates both data fidelity and symmetry-preserving constraints, and will be formally introduced in the next subsection. Based on this loss, the parameters $\boldsymbol{\theta}$ are updated using gradient descent. This procedure is iterated, wherein each step updates the Hamiltonian vector field to more accurately reflect the system's true dynamics while preserving physical symmetries.

### 3.4 Symmetry-informed loss for preserving physical properties

Even though a UB-NODE adequately approximates the sine-Gordon equations on short trajectories, naive training can lead to numerical drift or a lack of long-term stability. This issue arises because typical objectives, such as matching short-horizon trajectories via mean squared error (MSE) or mean absolute error (MAE), do not necessarily enforce global invariants like energy or momentum. These traditional loss functions focus purely on minimizing discrepancies between predicted and observed data points, often neglecting the inherent symmetries and conservation laws that govern the system's dynamics. Several methods have recently been proposed to enforce symmetries when learning dynamical systems, leveraging prior physical invariances to improve model accuracy and generalization. Notable examples include the works in [34–36], where symmetry constraints are explicitly embedded into the learning architectures or loss functions, demonstrating improved performance in modeling complex physical dynamics. .

In physical systems described by the sine-Gordon equation, certain fundamental properties—such as energy conservation, momentum conservation, and symmetry invariants—must be preserved throughout the evolution of the system. The discrete sine-Gordon equation, in particular, is governed by a non-polynomial Hamiltonian that exhibits symmetries such as reflection symmetry and energy conservation. If these invariants are not enforced, the learned model may fail to capture the true underlying dynamics, especially for long-time predictions, leading to a drift or instability in the system's behavior.

This issue arises because traditional training methods, using objective functions such as MSE or MAE, focus only on reducing the short-term discrepancy between model predictions and observed data. However, they do not inherently preserve the structural properties that are crucial for long-term stability, such as the

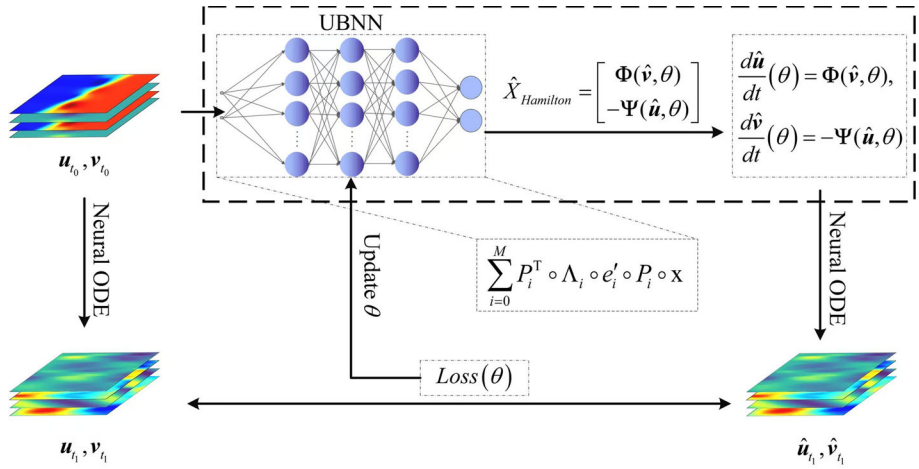



**Fig. 2** In universal basis neural ODEs (UB-NODEs), initial values $(\mathbf{u}_{t_0}, \mathbf{v}_{t_0})$ at $t_0$ and target values $(\mathbf{u}_{t_1}, \mathbf{v}_{t_1})$ at $t_1$ are given for training. A UBNN produces outputs $\boldsymbol{\Phi}(\hat{\mathbf{v}}, \boldsymbol{\theta})$ and $\boldsymbol{\Psi}(\hat{\mathbf{u}}, \boldsymbol{\theta})$ to approximate $T_{\mathbf{v}}(\hat{\mathbf{v}})$ and $V_{\mathbf{u}}(\hat{\mathbf{u}})$. These yield an approximation $\hat{X}_{\text{Hamilton}}$ of the Hamiltonian vector field. Then, following the neural ODE idea, $\hat{X}_{\text{Hamilton}}$ is integrated from $t_0$ to $t_1$. Comparing predictions $(\hat{\mathbf{u}}_{t_1}, \hat{\mathbf{v}}_{t_1})$ to the true values refines the parameters $\boldsymbol{\theta}$ via a symmetry-informed loss

conservation of energy or momentum. Without incorporating these fundamental physical invariants into the learning process, the model may yield solutions that are not consistent with the underlying dynamics of the system, especially when extrapolated over longer time periods.

To address these challenges and promote physically consistent solutions, we introduce *symmetry-informed* constraints in the loss function. These constraints incorporate known invariants or symmetries of the discrete sine-Gordon system into the training process. Specifically, by enforcing symmetry-based penalties, we ensure that the learned model adheres to the reflection symmetry of the system's potential and conserves energy in a physically plausible manner. This integration of physical knowledge into the loss function helps the neural network learn more stable and accurate long-term trajectories by aligning its predictions with the conserved quantities that govern the system.

Furthermore, symmetry-informed penalties guide the model to respect the intrinsic structure of the underlying dynamical system. For instance, enforcing reflection symmetry ensures that the model remains invariant under certain transformations of the state variables, which is crucial for maintaining the fidelity of soliton solutions. Similarly, incorporating energy conservation into the loss function helps mitigate the numerical drift that often occurs in standard neural network training, where the model might otherwise produce unphysical trajectories over extended time horizons.

Certain Hamiltonian systems exhibit symmetry properties that can be leveraged to improve generalization and physical fidelity during training. Denote by $H(\mathbf{u}, \mathbf{v})$ the Hamiltonian function [24]. A symmetry of the Hamiltonian can be compactly encoded by a matrix $G \in \mathbb{R}^{2N\times 2N}$, acting on the concatenated state vector $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^{2N}$. We say that $H$ is *G-symmetric* if

$$H((\mathbf{u}, \mathbf{v}) \cdot G) = H(\mathbf{u}, \mathbf{v}). \tag{3.6}$$

To incorporate this symmetry into a learning-based framework, we introduce a *symmetry loss* that penalizes violations of the $G$-invariance in either the learned Hamiltonian itself or its associated vector field. Specifically, given a Hamiltonian model $\mathcal{H}_{\boldsymbol{\theta}}$ parameterized by neural network weights $\boldsymbol{\theta}$, we may define:

$$\mathcal{L}_{\text{symmetry}}(\mathcal{D}, \boldsymbol{\theta}) = \sum_i \Big| \mathcal{H}_{\boldsymbol{\theta}}\big((\hat{\mathbf{u}}(t_i), \hat{\mathbf{v}}(t_i)) \cdot G\big) - \mathcal{H}_{\boldsymbol{\theta}}\big(\hat{\mathbf{u}}(t_i), \hat{\mathbf{v}}(t_i)\big)\Big|, \tag{3.7}$$

which enforces $G$-symmetry directly at the level of energy. Alternatively, we may constrain the *vector field* generated by the Hamiltonian. Let $\mathcal{V}_{\boldsymbol{\theta}} := \nabla_{(\mathbf{u},\mathbf{v})}\mathcal{H}_{\boldsymbol{\theta}}$ denote the learned Hamiltonian vector field. Then an

equivalent symmetry loss can be defined as:

$$\mathcal{L}_{\text{symmetry}}(\mathcal{D},\boldsymbol{\theta}) = \sum_i \Big| \mathcal{V}_{\boldsymbol{\theta}}\big((\hat{\mathbf{u}}(t_i),\hat{\mathbf{v}}(t_i))\cdot G\big)\cdot G^{\top} - \mathcal{V}_{\boldsymbol{\theta}}\big(\hat{\mathbf{u}}(t_i),\hat{\mathbf{v}}(t_i)\big)\Big|. \tag{3.8}$$

These constraints ensure that the network learns a physically plausible vector field consistent with the symmetry structure of the true dynamics.

The total training loss is therefore augmented to include the symmetry constraint:

$$\mathcal{L}_{\text{total}}(\mathcal{D},\boldsymbol{\theta}) = \underbrace{\mathcal{L}_{\text{standard}}(\mathcal{D},\boldsymbol{\theta})}_{\text{e.g., MSE or MAE}} + \lambda\, \mathcal{L}_{\text{symmetry}}(\mathcal{D},\boldsymbol{\theta}), \tag{3.9}$$

where $\mathcal{L}_{\text{standard}}$ penalizes deviations from observed trajectories, and the hyperparameter $\lambda > 0$ balances trajectory fidelity against the enforcement of physical symmetry.

In the case of the discrete sine-Gordon system, the Hamiltonian satisfies a $G$-symmetry for

$$G = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}.$$

To enforce this symmetry, we may define a penalty on the gradient of the Hamiltonian, which is computed as follows:

$$\mathcal{L}_{\text{symmetry}}(\mathcal{D},\boldsymbol{\theta}) = \sum_i \left| \boldsymbol{\Psi}(-\hat{\mathbf{u}}(t_i),\boldsymbol{\theta}) + \boldsymbol{\Psi}(\hat{\mathbf{u}}(t_i),\boldsymbol{\theta}) \right|. \tag{3.10}$$

This formulation enforces symmetry in the network's output consistent with the $G$-symmetry of the discrete sine-Gordon Hamiltonian. Such physics-informed losses encourage the model to generalize beyond the training data and preserve structural properties of the underlying dynamical system.

Figure 3 illustrates the impact of including such symmetry-informed penalties. Models trained purely with standard losses tend to reduce training errors but may exhibit larger testing fluctuations, possibly indicating overfitting or drift in long-term predictions. By contrast, models incorporating energy or symmetry constraints often converge more reliably and maintain stability over extended time. Similarly, Fig. 4 shows that symmetry-informed UB-NODE solutions track the true sine-Gordon orbit closely, underscoring how these additional terms guide the learned system to remain faithful to the underlying lattice physics.

Overall, the symmetry-informed approach helps the network learn beyond mere trajectory fitting, embedding deeper physical consistency into the training objective. This results in solutions that more robustly capture long-time soliton behaviors, adhere to conservation laws, and better reflect the essential structure of the discrete sine-Gordon model. By incorporating these symmetry-informed constraints, we improve the model's ability to generalize beyond the training data and maintain stability over long time periods. This results in a model that not only fits short-term data more accurately but also preserves the essential physical properties of the system, ensuring that the learned dynamics remain consistent with the true behavior of the discrete sine-Gordon equation.

## 4 Experiment

In this section, we present numerical experiments that demonstrate the efficacy of our proposed UB-NODE framework for learning the discrete sine-Gordon equation. We first detail the data generation process using a fourth-order Runge–Kutta (RK4) solver and outline the general training setup. Next, we compare models trained with and without symmetry-informed components, highlighting the effect of symmetry-based constraints on accuracy and stability. Finally, we show the performance of our method on two types of soliton solutions—the sine-wave soliton and the breather soliton—focusing on long-term accuracy and preservation of energy-like invariants.

### 4.1 Data generation and training setup

We employ an RK4 solver [24,29] to numerically integrate the discrete sine-Gordon system over a prescribed time interval. Specifically, we begin with initial conditions sampled uniformly from $[-0.5, 0.5]^{2N}$, where $\mathbf{u}$ and $\mathbf{v}$ represent particle positions and velocities, respectively. We denote by $S$ the total number of sampled initial conditions and by $K$ the number of RK4 time steps, each of size $\Delta t$. The dataset thus takes the form

$$\mathcal{D} = \left\{ \left(\mathbf{u}_0^{(\alpha)}, \mathbf{v}_0^{(\alpha)}\right), \left(\mathbf{u}_1^{(\alpha)}, \mathbf{v}_1^{(\alpha)}\right), \ldots, \right.$$

$$\left(\mathbf{u}_K^{(\alpha)}, \mathbf{v}_K^{(\alpha)}\right) \mid 1 \le \alpha \le S\Big\},$$

where $\mathbf{u}_k^{(\alpha)}$ and $\mathbf{v}_k^{(\alpha)}$ are the numerical solutions at time $k\Delta t$. We then interpret $\left(\mathbf{u}_k^{(\alpha)}, \mathbf{v}_k^{(\alpha)}\right)$ as input features and $\left(\mathbf{u}_{k+1}^{(\alpha)}, \mathbf{v}_{k+1}^{(\alpha)}\right)$ as training targets.

Throughout our experiments, we partition the data into training and test splits: $S_1$ sets of initial conditions for training (each with $K$ time steps) and $S_2$ sets for testing. For concreteness, we set $S_1 = 50$, $S_2 = 100$, with each trajectory having $K = 2500$ points (each advanced by $\Delta t = 4 \times 10^{-3}$). The model parameters $\left\{P_\Phi, Q_\Phi, P_\Psi, Q_\Psi\right\}$ (see Sect. 3) are initialized with a normal distribution and updated via Adam (lr $= 10^{-3}$) plus an RK4 ODE solver inside the UB-NODE structure and $N_1$ takes the value of 32.

### 4.2 Comparative analysis: with vs. without symmetry-informed components

We first investigate how adding symmetry-related constraints affects training and generalization. In many Hamiltonian lattice systems, including the discrete sine-Gordon equation, potentials satisfy symmetries such as $V_{\mathbf{u}}(\mathbf{u}) + V_{\mathbf{u}}(-\mathbf{u}) = 0$. To exploit these symmetries, we modify the usual MSE or MAE loss by incorporating a penalty term that enforces consistency under reflection [27].

Concretely, our baseline models minimize a standard loss:

$$\mathcal{L}_{\text{standard}}(\mathcal{D}, \boldsymbol{\theta}) = \begin{cases} \text{MSE:} & \frac{1}{K}\sum_{i=1}^{K} \left\|\hat{\mathbf{v}}_i - \mathbf{v}_{\text{true},i}\right\|^2 + \left\|\hat{\mathbf{u}}_i - \mathbf{u}_{\text{true},i}\right\|^2, \\ \text{MAE:} & \frac{1}{K}\sum_{i=1}^{K} \left\|\hat{\mathbf{v}}_i - \mathbf{v}_{\text{true},i}\right\| + \left\|\hat{\mathbf{u}}_i - \mathbf{u}_{\text{true},i}\right\|, \end{cases}$$

where $\hat{\mathbf{u}}_i, \hat{\mathbf{v}}_i$ are predictions at the $i$-th step. In the *symmetry-informed* variant, we add

$$\mathcal{L}_{\text{Symmetry}}(\mathcal{D}, \boldsymbol{\theta}) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} \left\|\boldsymbol{\Psi}(\hat{\mathbf{u}}_i) + \boldsymbol{\Psi}(-\hat{\mathbf{u}}_i)\right\|^2 & \text{(for MSE-based training)}, \\ \frac{1}{N}\sum_{i=1}^{N} \left\|\boldsymbol{\Psi}(\hat{\mathbf{u}}_i) + \boldsymbol{\Psi}(-\hat{\mathbf{u}}_i)\right\| & \text{(for MAE-based training)}. \end{cases}$$

Hence, the final objective combines these terms:

$$\mathcal{L}_{\text{total}}(\mathcal{D}, \boldsymbol{\theta}) = \mathcal{L}_{\text{standard}}(\mathcal{D}, \boldsymbol{\theta}) + \lambda\, \mathcal{L}_{\text{symmetry}}(\mathcal{D}, \boldsymbol{\theta}). \tag{4.1}$$

where $\lambda > 0$ balances fidelity to data against enforcement of the system's reflection invariance.

Empirically, we observe faster training convergence and more stable test performance when $\lambda$ is suitably tuned. See Fig. 3, comparing test losses with and without symmetry constraints. Figure 3 compares the training and testing losses for two UB-NODE variants: one incorporating symmetry-informed learning (green lines) and one trained in a purely data-driven manner (blue lines). As shown in the left panel, both methods reduce the training loss rapidly and converge to comparable values; however, adding symmetry constraints ($\lambda = 0.01$) accelerates learning and lowers the final training error slightly. In the right panel, the symmetry-free model exhibits larger fluctuations in the test loss—an indication of possible overfitting—while the symmetry-informed model's test loss remains stable and converges more quickly to an equally low plateau.

Figure 4 further demonstrates the predictive accuracy of our symmetry-informed *Universal-Basis Neural Ordinary Differential Equation* (UB-NODE) model. The predicted trajectory, represented by colored spheres, aligns closely with the true solution (solid line) of the discrete sine-Gordon equation. The minimal discrepancy between the predicted and actual trajectories highlights the model's capability to preserve the underlying Hamiltonian dynamics, even over extended time horizons.

For comparison, we also include the predictions of standard *Hamiltonian Neural Networks* (HNNs), shown in the same figure. Despite having comparable network complexity, HNNs exhibit significant deviation from the true orbit, particularly in capturing the fine-grained phase structure and amplitude of oscillations. This contrast emphasizes the advantage of embedding physical symmetries directly into the learning framework when modeling discrete Hamiltonian systems.

***Discrete Sine-Gordon Model: Predictive Performance Comparison***

Overall, these results illustrate the advantages of incorporating symmetry-informed learning in both training and testing. The symmetry-based constraints lead to faster convergence and greater predictive stability, reinforcing the importance of physical invariants in preserving the long-term fidelity of the learned model (Fig. 4). In the example of the discrete sine-Gordon Eq. (2.2), we set $C = \frac{1}{h^2}$, a parameter of considerable physical significance that can represent the torsion con-

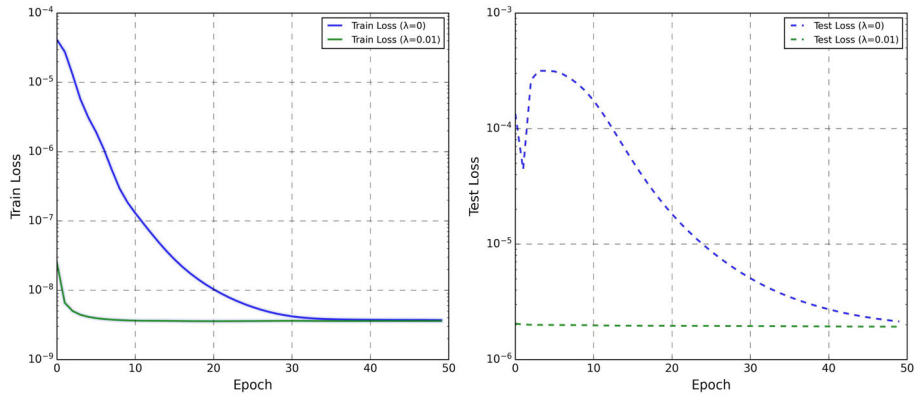



**Fig. 3** Comparison of training (left panel) and test (right panel) losses for models trained *without* symmetry-informed constraints ($\lambda = 0$, blue curves) versus *with* symmetry information ($\lambda = 0.01$, green curves). Although both approaches converge to similarly low training losses, the symmetry-informed UB-NODE learns more efficiently (left), while also achieving smoother and more stable test loss trajectories (right). The non-symmetry-informed model exhibits fluctuations and potential overfitting on the test set, whereas the symmetry-informed variant steadily decreases to comparable final accuracy

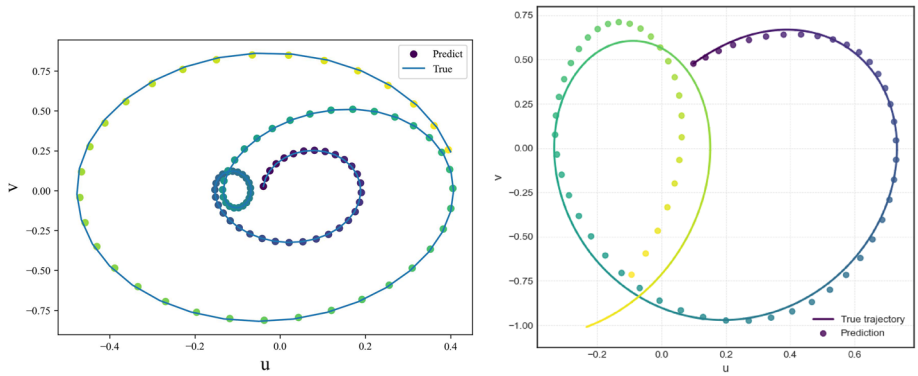



**Fig. 4** Representative trajectory for a discrete sine-Gordon orbit (solid line) compared against the prediction given by neural networks (colored spheres). In the left panel, we observe a close match between the true trajectory and the predictions made by our proposed *Universal-Basis Neural ODEs* (UB-NODEs), demonstrating its ability to accurately capture the long-term Hamiltonian dynamics of the system. The model effectively preserves the qualitative and quantitative features of the discrete sine-Gordon dynamics, highlighting the benefits of incorporating physical structure and conservation laws into the learning process. In contrast, the right panel displays the results produced by a standard *Hamiltonian Neural Network* (HNN) of comparable complexity. The discrepancy between the predicted and true trajectories is noticeably larger, indicating that HNNs struggle to maintain both accuracy and stability over long time horizons, even when applied to an orbit that is comparatively simpler than the one shown in the left panel

stant in coupled pendulums, the effective mass in crystal dislocations, or the inductance in Josephson junction arrays. A larger $C$ also pushes the discrete sine-Gordon system closer to its continuous counterpart.

Table 1 summarizes the model's convergent values for $C = 2$ under varying symmetry-loss coefficients $\lambda$. Reported metrics include MSE and MAE for training and testing, indicating that $\lambda = 0.1$ yields the best training performance while $\lambda = 1$ excels in the testing phase, albeit by a relatively small margin in each case. This finding suggests that modest symmetry-informed penalties can substantially enhance stability and accuracy, but the precise choice of $\lambda$ may be problem- or data-dependent (Table 2).

Table 3 presents the mean squared error (MSE) of our method as the number $N$ of lattice points increases.

**Table 1** Training and test losses for the discrete sine-Gordon system at $C = 2$, computed under various symmetry-informed loss coefficients $\lambda$. Here, $\lambda = 0.1$ yields the smallest MSE and MAE for training (bold) while $\lambda = 1$ does so for testing (bold)

| $\lambda$ | Train MSE Loss | Train MAE Loss | Test MSE Loss | Test MAE Loss |
|---|---|---|---|---|
| 0 | 4.8401E-08 | 1.7685E-04 | 9.6893E-05 | 7.6732E-03 |
| 0.001 | 6.8231E-09 | 6.3888E-05 | 4.6370E-06 | 1.4936E-03 |
| 0.01 | 5.5082E-09 | 5.7157E-05 | 3.5260E-06 | 1.2820E-03 |
| 0.1 | **4.8432E-09** | **5.5508E-05** | 3.2996E-06 | 1.2309E-03 |
| 1 | 4.8752E-09 | 7.2923E-05 | **3.1368E-06** | **1.1838E-03** |

**Table 2** This table reports the mean squared error (MSE) of our method as the number $N$ of lattice points varies. We observe that the MSE remains relatively stable as $N$ increases, indicating the robustness of our approach with respect to system size

| N | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| **Test Loss** | $2.2167 \times 10^{-6}$ | $1.7710 \times 10^{-6}$ | $1.8091 \times 10^{-6}$ | $1.8307 \times 10^{-6}$ |

We observe that the MSE remains relatively stable across varying values of $N$, demonstrating the robustness of our approach with respect to system dimensionality. These results indicate that our model effectively scales to high-dimensional settings and can reliably approximate the solutions of discrete sine-Gordon equations even when the number of lattice points becomes large.

Although the primary focus of this paper is on the discrete sine-Gordon equation, we emphasize that the proposed *Universal-Basis Neural ODE* (UB-ODE) framework is not limited to this specific system. The methodology is general and can be readily applied to a wide range of lattice-based Hamiltonian systems.

For instance, in preliminary experiments on the Toda lattice—a well-known integrable system characterized by exponential inter-particle interactions—the UB-ODE model demonstrates similarly strong performance in accurately capturing the underlying dynamics and preserving long-term stability (Fig. 5). This generalizability suggests that our approach is broadly applicable to other structured physical systems governed by lattice dynamics, particularly those with known conservation laws and symmetries.

### 4.3 Performance on sine wave and breather solitons

We now demonstrate the capability of UB-NODEs (with symmetry-informed constraints) to accurately capture soliton solutions of the discrete sine-Gordon equation. Our focus is on two canonical soliton profiles:

1. **Sine-wave soliton:** The initial condition is prescribed as
$$u_k(0) = \sin\left(\frac{2\pi k}{N}\right), \quad v_k(0) = 0.1\cos\left(\frac{2\pi k}{N}\right), \quad k = 0, \ldots, N-1.$$
This yields an oscillatory wave across the lattice.
2. **Breather soliton:** A localized, pulsating wave with
$$u_k(0) = 0, \quad v_k(0) = \frac{4}{\sqrt{2}} \frac{\sinh\left(\frac{2\pi k}{\sqrt{2}N}\right)}{\cosh^2\left(\frac{2\pi k}{\sqrt{2}N}\right)}, \quad k = 0, \ldots, N-1,$$
which forms a single, smoothly oscillating structure often referred to as a "breather" [30].

Throughout these experiments, we fix $N = 30$ and $C = 1$.

Figure 6a and b illustrate the time evolution of two distinct types of solitons on the $(k, u)$ plane. For each lattice site $k$, the model tracks its displacement $u$ over time, comparing the ground-truth numerical solution (blue) to the predictions produced by the proposed UB-NODEs (green). The instantaneous discrepancies between the predicted and true solutions are highlighted by purple lines. While minor deviations appear at later times, both waveforms maintain their characteristic shapes for extended durations, demonstrating the long-term accuracy and robustness of the UB-NODEs.

In contrast, Fig. 7a and b show the results obtained by applying HNNs to the discrete sine-Gordon model.

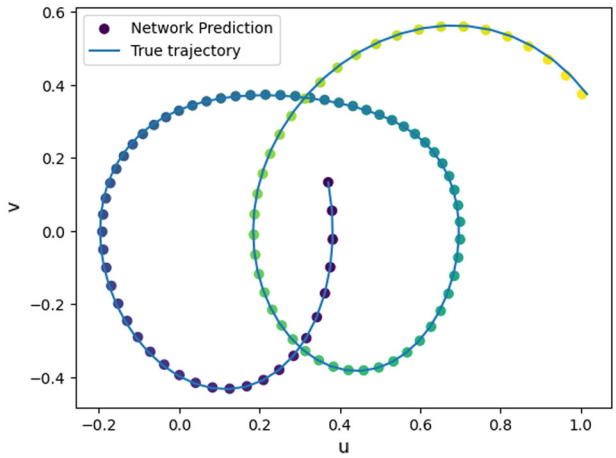
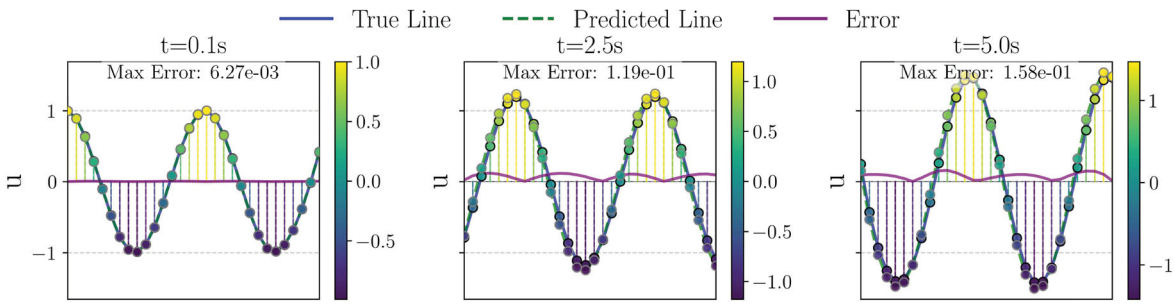
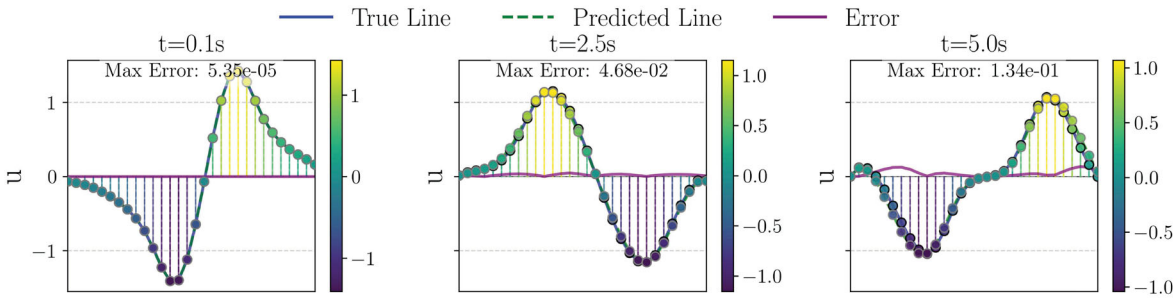



**Fig. 5** Representative trajectories for the discrete sine-Gordon system are shown, where the solid line denotes the true orbit and the colored spheres represent the predictions given by neural networks. The close agreement highlights the effectiveness of our proposed *Universal-Basis Neural ODEs* (UB-NODEs) in accurately capturing the underlying systems. Furthermore, we remark that UB-NODEs also demonstrate strong performance on other lattice systems, such as the Toda lattice, indicating the generality and robustness of the method across different classes of integrable Hamiltonian systems

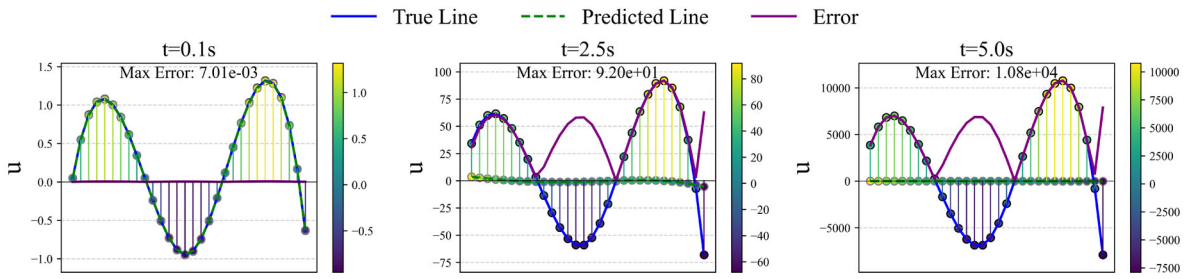



(a) True and predicted wave evolution of a sine-wave soliton obtained using UB-NODEs

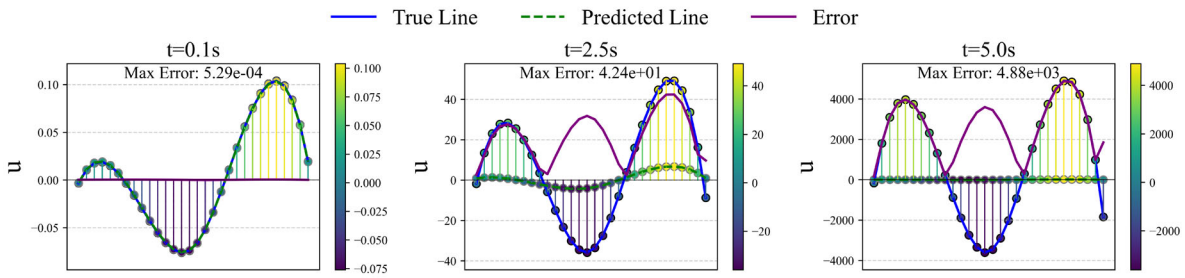



(b) True and predicted wave evolution of a breather soliton obtained using UB-NODEs

**Fig. 6 Comparisons of true vs. predicted wave evolutions given by UB-NODEs.** Each subplot shows the lattice index $k$ on the horizontal axis and the displacement $u$ on the vertical axis. Blue solid lines represent the exact solution from RK4, green dashed lines show the UB-NODE prediction, and purple lines highlight the error at each time slice. Colored markers denote the particle states, with color indicating the magnitude of $u$




(a) True and predicted wave evolution of a sine-wave soliton obtained using HNNs




(b) True and predicted wave evolution of a breather soliton obtained using HNNs

**Fig. 7 Comparisons of true vs. predicted wave evolutions given by HNNs.** Each subplot shows the lattice index $k$ on the horizontal axis and the displacement $u$ on the vertical axis. Blue solid lines represent the exact solution from RK4, green dashed lines show the HNN prediction, and purple lines highlight the error at each time slice. Colored markers denote the particle states, with color indicating the magnitude of $u$

Although the HNNs align with the true solution at early times (e.g., $t = 0.1$), significant errors quickly accumulate, and the predicted waveforms deteriorate as time progresses. These results underscore the superior predictability and stability of UB-NODEs when compared to HNNs, particularly in long-term simulations.

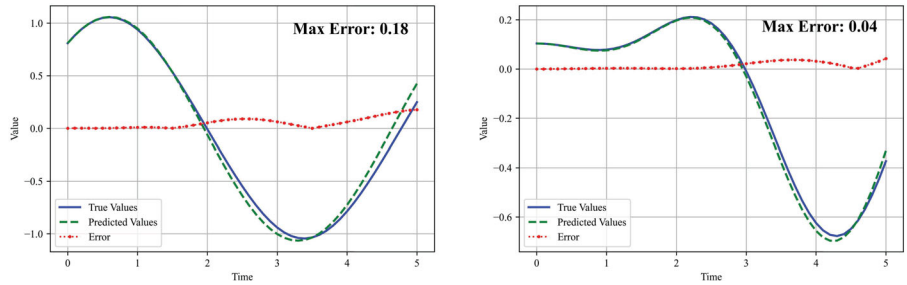
To isolate specific site-level dynamics, Fig. 8a and b plot the displacement–time curves for the single lattice site $k = 19$. Again, we see close alignment between the UB-NODE outputs and the reference solution, with the instantaneous error (red) remaining relatively small, especially for the breather case until around $t = 5$.

Finally, Fig. 9a and b present contour plots of the true vs. predicted orbits and their difference. Even over multiple oscillations, the UB-NODE approach maintains fidelity to the main wave patterns. In particular, the breather's localized "pulse" is successfully captured until minor amplitude drift appears for large $t$. In summary, these results confirm that symmetry-informed UB-NODEs achieve robust modeling of both extended and localized soliton solutions in the discrete sine-Gordon equation, demonstrating a balanced mix of short-term accuracy and long-term physical consistency.

In summary, these numerical results confirm that UB-NODEs accurately reconstruct both the extended sine-wave solution and the localized breather in the discrete sine-Gordon equation. Even under chaotic-like lattice dynamics or for extended time horizons, the model's predictions track the principal soliton shapes with only moderate error, especially when the symmetry-informed constraints are employed to maintain the underlying Hamiltonian structure.

### 4.4 Summary/observations of experimental results

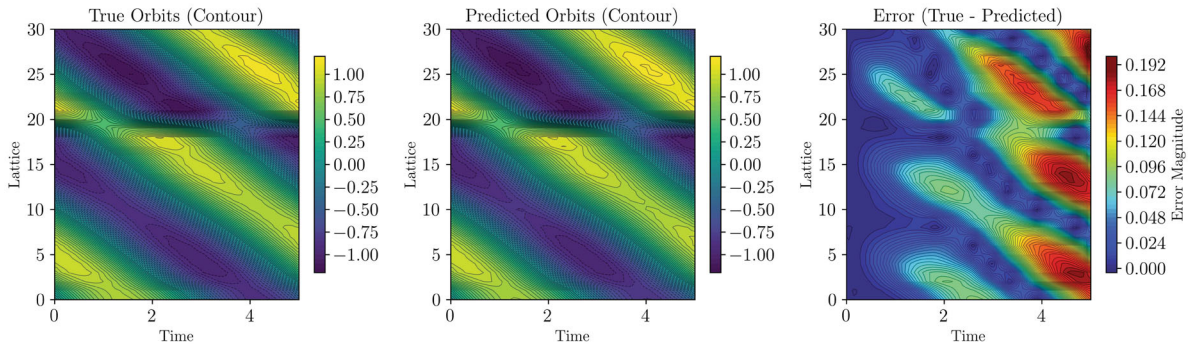
Our numerical experiments confirm that the proposed UB-NODE framework—especially when supplemented with symmetry-informed constraints—achieves both accurate short-term trajectory matching and improved



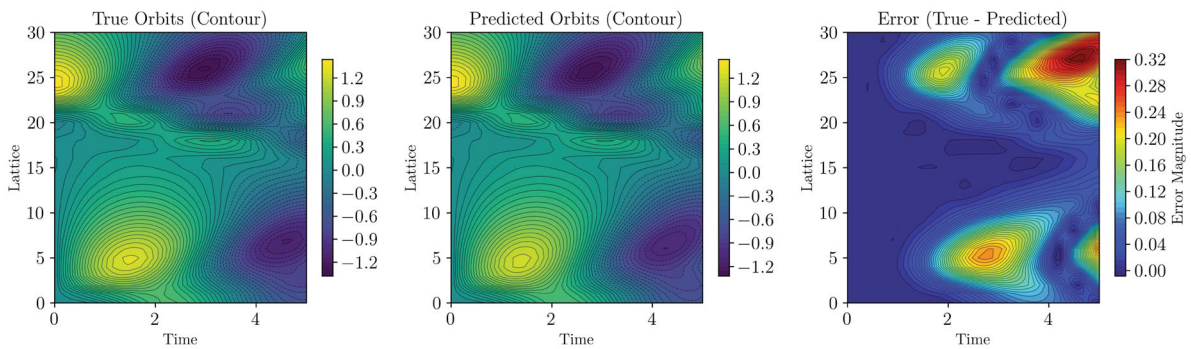

(a) Sine-wave soliton at particle $k = 19$

(b) Breather soliton at particle $k = 19$

**Fig. 8 Individual particle trajectories in time.** True evolution (blue solid) is juxtaposed with UB-NODE predictions (green dashed), and the red dotted line highlights the instantaneous error. Although the breather case exhibits some drift for $t > 5.0$, both solitons remain well-approximated for earlier times




(a) Sine-wave soliton: true, predicted, and error contours




(b) Breather soliton: true, predicted, and error contours

**Fig. 9 Contour views of full soliton evolution**. Each row shows (left) the true orbit, (middle) the UB-NODE-based prediction, and (right) the resulting error map. Warmer or cooler hues in the error map reflect higher or lower discrepancy in displacement. Both soliton configurations—extended sinusoid and localized breather—are reproduced accurately

**Table 3** Test Loss under Different Particle Numbers $n$

| **n** | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| **Test Loss** | $2.2167 \times 10^{-6}$ | $1.7710 \times 10^{-6}$ | $1.8091 \times 10^{-6}$ | $1.8307 \times 10^{-6}$ |

long-term stability for the discrete sine-Gordon equation. Below, we summarize the main findings:

- **Efficiency of Symmetry-Informed Learning.** As shown in Fig. 3, incorporating a symmetry-based penalty ($\lambda > 0$) into the loss function not only speeds up convergence but also diminishes overfitting in test data. Table 1 further demonstrates that a modestly tuned $\lambda$ can produce lower MSE and MAE, indicating that physical invariants (e.g., reflection symmetry in the potential) help guide the network to more stable solutions.
- **Robustness Over Extended Time Horizons.** Figures 6a–b show that the UB-NODE solution tracks the true lattice orbits closely across multiple periods of a soliton's motion. Although small drift emerges in the breather soliton beyond $t \approx 5$, the model retains the principal wave shape better than a purely data-driven baseline without symmetry terms.

The framework accurately recovers both extended "sine-wave" solitons and localized "breather" solitons (see Fig. 8a–b). In each case, key waveform features—e.g., crest amplitude or soliton width—remain well captured. Even in the presence of relatively large lattice spacing parameters, the method continues to predict major wave characteristics, highlighting its adaptability to various discrete parameter regimes.

While the choice of $\lambda$, the polynomial/trigonometric basis order, and the training rate does influence final accuracy, the method converges reliably over a practical range of these settings. For real-world applications, small-scale hyperparameter tuning typically suffices.

Taken together, these observations indicate that universal basis expansions combined with a neural ODE formulation provide a robust, physically interpretable approach to learning dynamics of nonlinear lattice equations, such as the discrete sine-Gordon system.

## 5 Conclusion

In this work, we introduced universal-basis neural ordinary differential equations (UB-NODEs) for learning and predicting solutions of the discrete sine-Gordon equation. Our approach uses basis-function expansions to approximate the Hamiltonian's derivatives and integrates symmetry-informed constraints to uphold fundamental invariants, such as reflection symmetry.

The key contributions of our work are:

- **Universal-Basis Approximation.** By embedding polynomial or trigonometric bases into a neural ODE architecture, UB-NODEs effectively capture the non-polynomial structure of the discrete sine-Gordon system. Our experiments show this approach performs well for both extended (sine-wave) and localized (breather) soliton solutions.
- **Symmetry-Informed Loss.** Introducing a symmetry-based penalty ($\lambda$) in the loss function proved to reduce long-horizon drift and enhance model stability during training. Both training and test performance benefited from this additional constraint, which enforces the discrete sine-Gordon system's built-in reflection symmetry.
- **Long-Term Accuracy.** Numerical results show that UB-NODEs can maintain good predictive accuracy across multiple soliton oscillations. In particular, for breather solitons, the approach preserves the wave's localized structure, with only minor deviations emerging after extended simulation times.
- **Robustness and Efficiency.** Compared to standard neural ODEs lacking physical constraints, the symmetry-enhanced UB-NODE converges faster, exhibits fewer test-loss fluctuations, and preserves key energy-like properties more effectively. Its training overhead remains moderate, relying on widely used optimizers and ODE solvers.

Although demonstrated here on the discrete sine-Gordon problem, our UB-NODE approach naturally extends to other Hamiltonian lattice models (e.g., $\phi^4$ theories, FPUT models), provided a suitable basis can approximate each system's potential.

Our findings highlight the pitfalls of purely data-driven networks in overfitting and losing long-time fidelity. By contrast, incorporating physics-based con-

straints (e.g., symmetry or energy conservation) promotes stable extrapolation in realistic scenarios.

Possible next steps include scaling the UB-NODE framework to larger lattice sizes, integrating multi-scale or operator-learning techniques for improved efficiency, and exploring domain-decomposition or streaming-data extensions for real-time monitoring.

In conclusion, combining universal-basis expansions, neural ODE solvers, and symmetry-based penalties offers a robust, physically grounded method for modeling complex discrete soliton dynamics over extended time horizons—a powerful foundation for future advances in machine learning for nonlinear lattice systems.

**Acknowledgements** This material is based upon work supported by National Natural Science Foundation of China (NSFC 12071471), and by the Natural Science Foundation Key Project of Inner Mongolia Autonomous Region (Nos. 2022ZD05), the Key Laboratory of Infinite-dimensional Hamiltonian System and Its Algorithm Application (Inner Mongolia Normal University), Ministry of Education (Nos. 2023KFZD01). Hong-kun Zhang is partially supported by Guangdong Basic and Applied Basic Research Foundation 2025A1515011952.

**Author contributions** Mingkun Li wrote the main manuscript draft and contributed to the overall structure of the paper. Zaijiu Shang provided theoretical guidance and assisted in the formulation of the modeling framework. Peng Wang curated the references and contributed to manuscript editing and polishing. Hong-Kun Zhang provided theoretical support and assisted in revising the manuscript. Junjie Fan implemented the code, generated the experimental data and figures, and validated the model performance. All authors reviewed and approved the final manuscript.

**Data Availability** No datasets were generated or analysed during the current study.

**Declarations**

**Conflict of interest** The authors declare no Conflict of interest.

## References

1. Cuevas-Maraver, J., Kevrekidis, P.G., Williams, F.: The sine-Gordon Model and Its Applications, vol. 10. Springer, Cham, Switzerland (2014)
2. Wallace, D.J.: Solitons and instantons: An introduction to solitons and instantons in quantum field theory. Phys. Bull. **34**(1), 29 (1983)
3. Dutykh, D., Caputo, J.-G.: Wave dynamics on networks: Method and application to the Sine-Gordon equation. Appl. Numer. Math. **131**, 54–71 (2018)
4. Miller, M.D.: Nonlinear wave propagation in periodic systems: The driven Sine-Gordon chain. Phys. Rev. B **33**, 1641–1656 (1986)
5. Cisneros, L.A., Minzoni, A.A.: Asymptotics for kink propagation in the discrete Sine-Gordon equation. Physica D **237**(1), 50–65 (2008)
6. Macías-Díaz, J.E.: Nonlinear wave transmission in harmonically driven Hamiltonian Sine-Gordon regimes with memory effects. Chaos, Solitons Fractals **142**, 110362 (2021)
7. Braun, O.M., Kivshar, Y.S.: The Frenkel-Kontorova Model: Concepts, Methods, and Applications. Springer, Berlin (2004)
8. Nikan, O., Avazzadeh, Z., Rasoulizadeh, M.N.: Soliton solutions of the nonlinear sine-Gordon model with Neumann boundary conditions arising in crystal dislocation theory. Nonlinear Dyn. **106**(1), 783–813 (2021)
9. Parker, R., Kevrekidis, P.G., Aceves, A.: Stationary multikinks in the discrete Sine-Gordon equation. Nonlinearity **35**(2), 1036 (2021)
10. Semenov, M.E., Reshetova, O.O., Tolkachev, A.V., Solovyov, A.M., Meleshenko, P.A.: Oscillations under hysteretic conditions: From simple oscillator to discrete sine-gordon model. In: Topics in Nonlinear Mechanics and Physics, pp. 229–253. Springer, Singapore (2019)
11. Goedde, C.G., Lichtenberg, A.J., Lieberman, M.A.: Chaos and the approach to equilibrium in a discrete Sine-Gordon equation. Physica D **59**(1–3), 200–225 (1992)
12. Ezawa, M.: Nonlinear topological phase transitions in the dimerized Sine-Gordon model. Phys. Rev. B **105**(16), 165418 (2022)
13. Hanif, Y., Saleem, U.: Coupled Sine-Gordon systems in living cellular structure. Mod. Phys. Lett. B **34**(05), 2050070 (2020)
14. Watanabe, S., van der Zant, H.S.J., Strogatz, S.H., Orlando, T.P.: Dynamics of circular arrays of Josephson junctions and the discrete Sine-Gordon equation. Physica D **97**(4), 429–470 (1996)
15. Hu, W., Wang, Z., Wang, G., Wazwaz, A.-M.: Local dynamic behaviors of long 0-$\pi$ Josephson junction. Phys. Scr. **95**(8), 085221 (2020)
16. Trías, E., Mazo, J.J., Orlando, T.P.: Discrete breathers in nonlinear lattices: Experimental detection in a Josephson array. Phys. Rev. Lett. **84**, 741–744 (2000)
17. Binder, P., Abraimov, D., Ustinov, A.V., Flach, S., Zolotaryuk, Y.: Observation of breathers in Josephson ladders. Phys. Rev. Lett. **84**, 745–748 (2000)
18. Yomosa, S.: Soliton excitations in deoxyribonucleic acid (DNA) double helices. Phys. Rev. A **27**, 2120–2125 (1983)
19. Peyrard, M.: Nonlinear dynamics and statistical physics of DNA. Nonlinearity **17**(2), 1 (2004)
20. Kevrekidis, P.G.: The Discrete Nonlinear Schrödinger Equation: Mathematical Analysis. Numerical Computations and Physical Perspectives. Springer, Berlin (2009)
21. Kevrekidis, P.G., Cuevas-Maraver, J. (eds.): A Dynamical Perspective on the $\phi^4$ Model. Past, Present and Future. Nonlinear Systems and Complexity, vol. 26. Springer, Cham, Switzerland (2019)
22. Gallavotti, G.: The Fermi-Pasta-Ulam Problem: A Status Report. Springer, Berlin, Heidelberg (2007)
23. Thomson, S.J., Durey, M., Rosales, R.R.: Discrete and periodic complex Ginzburg-landau equation for a hydrodynamic active lattice. Phys. Rev. E **103**, 062215 (2021)
24. Arnold, V.I.: Mathmatical Methods of Classical Mechanics. Springer, New York (1978)

25. Saqlain, S., Zhu, W., Charalampidis, E.G., Kevrekidis, P.G.: Discovering governing equations in discrete systems using PINNS. Commun. Nonlinear Sci. Numer. Simul. **126**, 107498 (2023)
26. Gao, Y., Geng, R., Kevrekidis, P.G., Zhang, H.-K., Zu, J.: $\alpha$-SGHN: A Robust Model for Learning Particle Interactions in Lattice Systems (2024). arxiv:2407.11684
27. Geng, R., Zu, J., Gao, Y., Zhang, H.-K.: Separable graph Hamiltonian network: A graph deep learning model for lattice systems. Phys. Rev. Res. **6**, 013176 (2024)
28. Lin, Z., Chen, Y.: Pseudo grid-based physics-informed convolutional-recurrent network solving the integrable nonlinear lattice equations. Physica D **468**, 134304 (2024)
29. Malomed, B.A.: The sine-Gordon Model: General Background, Physical Motivations, Inverse Scattering, and Solitons, pp. 1–30. Springer, Cham (2014)
30. Manukure, S., Booker, T.: A short overview of solitons and applications. Partial Differ. Equ. Appl. Math. **4**, 100140 (2021)
31. Melnikov, V.K.: On the stability of the center for time-periodic perturbations. Trans. Moscow Math. Soc. **12**, 1–57 (1963)
32. Ott, E.: Chaos in Dynamical Systems. Cambridge University Press, Cambridge (2002)
33. Ngom, M., Marin, O.: Fourier neural networks as function approximators and differential equation solvers. Statist. Anal. Data Mining **14**, 647–661 (2021)
34. Dierkes, E., Flaßkamp, K.: Learning hamiltonian systems considering system symmetries in neural networks. IFAC-PapersOnLine **54**(19), 210–216 (2021)
35. Zhang, Z.-Y., Zhang, H., Zhang, L.-S., Guo, L.-L.: Enforcing continuous symmetries in physics-informed neural network for solving forward and inverse problems of partial differential equations. J. Comput. Phys. **492**, 112415 (2023)
36. Vaquero, M., Cortés, J., Diego, D.M.: Symmetry preservation in Hamiltonian systems: Simulation and learning. J. Nonlinear Sci. **34**, 115 (2024)