Separable graph Hamiltonian network: A graph deep learning model for lattice systems

Ru Geng, Jian Zu, and Yixian Gao*
School of Mathematics and Statistics, Center for Mathematics and Interdisciplinary Sciences, Northeast Normal University, Changchun 130024, People's Republic of China

Hong-Kun Zhang†
Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003, USA

(Received 10 August 2023; accepted 14 December 2023; published 16 February 2024)

Addressing the challenges posed by nonlinear lattice models, which are vital across diverse scientific disciplines, we present a new deep learning approach that harnesses the power of graph neural networks. By representing the lattice system as a graph and leveraging the graph structures to identify complex nonlinear relationships, we have developed a flexible solution that outperforms traditional techniques. Our model not only offers precise trajectory predictions and energy conservation properties by incorporating separable Hamiltonians but also proves superior to existing top-tier models when tested on classic nonlinear oscillator lattice problems: a mixed Fermi-Pasta-Ulam Klein-Gordon, a Klein-Gordon system with long-range interactions, and a two-dimensional Frenkel-Kontorova, highlighting its potential for wide-reaching applications.

DOI: 10.1103/PhysRevResearch.6.013176

I. INTRODUCTION

Nonlinear lattice systems have garnered significant attention across a vast array of disciplines, from condensed matter physics [1,2] and high-energy physics [3,4] to cosmology [5,6], materials science [7–9], chemistry [10], and biology [11–14]. Their importance stems from their ability to exhibit a multitude of complex behaviors like chaotic dynamics, phase transitions, and pattern formations. As such, they are invaluable in comprehending the behavior of complex systems and engineering materials and equipment with specific attributes.

In this research, we mainly focus on the classical coupled nonlinear lattice system, which is governed by

$$\frac{d}{dt}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \mathbf{J}\begin{pmatrix}\frac{\partial \mathcal{H}}{\partial \mathbf{q}}\\ \frac{\partial \mathcal{H}}{\partial \mathbf{p}}\end{pmatrix}, \quad \mathbf{J} = \begin{pmatrix} O & \mathbf{I} \\ -\mathbf{I} & O \end{pmatrix}, \tag{1}$$

where $\mathbf{q} = (q_\alpha, \alpha \in \mathbb{Z}^d)$ and $\mathbf{p} = (p_\alpha, \alpha \in \mathbb{Z}^d)$ are the generalized position and momentum, respectively; d is the dimension of the system. $\mathbf{I} \in \mathbb{R}^{d\times d}$ is the identity matrix. Here $\mathcal{H} = \mathcal{T} + \mathcal{V}$ is the Hamiltonian function, with kinetic energy

$$\mathcal{T}(\mathbf{p}) = \sum_{\alpha\in\mathbb{Z}^d} \frac{|p_\alpha|^2}{2m_\alpha}$$

and potential energy

$$\mathcal{V}(\mathbf{q}) = \sum_{\alpha\in\mathbb{Z}^d} (V_1(q_{N_k(\alpha)}) + V_2(q_\alpha)). \tag{2}$$

The intersite potential V_1 encapsulates particle interactions within the kth order neighbor $N_k(\alpha)$, while the on-site potential V_2 accounts for potential interactions with an external environment, such as a substrate. Moreover, m_α denotes the mass of the αth pointlike particle, and we set $m_\alpha = 1$ in this research for simplicity. This model is applicable to a variety of physical problems, including heat conductivity, atomic vibrations in crystals and molecules, and field modes in optics or acoustics [15,16].

The exceptional capabilities of deep neural networks have been increasingly leveraged in recent years to address scientific challenges in physical system modeling [17–19]. These cutting-edge neural network technologies have emerged as some of the most promising tools for analyzing nonlinear lattice systems. This includes solving nonlinear equations with physics informed neural networks (PINNs) [20], identifying phase transitions via multilayer perceptrons (MLPs) [21], and uncovering governing equations through a synergistic use of PINNs and symbolic regression [22]. Additionally, building lattice models based on PINNs [23] and exploring Poisson lattice systems through Poisson neural networks (PNNs) [24], a technology derived from symplectic neural networks (SympNets) [25], are burgeoning areas of interest.

Despite their utility, the aforementioned methods dependent on conventional neural networks present challenges. Some require an understanding of system equations [20,23], while some necessitate the creation of an overcomplete operator library [22]. These techniques are primarily tailored toward simpler lattice systems existing on one-dimensional

*Corresponding author: gaoyx643@nenu.edu.cn
†Corresponding author: hongkun@math.umass.edu

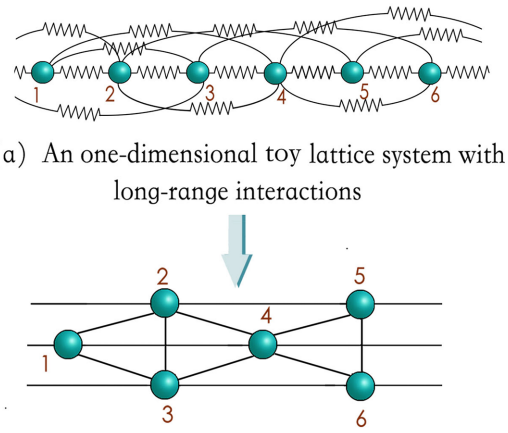


FIG. 1. A one-dimensional lattice system with long-range interactions.

lattice points, with a predominant focus on short-range interactions. However, it's well understood that lattice systems typically encompass myriad degrees of freedom and complex interactions, giving rise to diverse and intricate lattice structures. These structures could exhibit irregularities [26] or even demonstrate long-range interactions [27,28], as depicted in Fig. 1(a). Some lattice systems might present even more complex phenomena, further complicating the simulation of trajectories. Therefore, the demand for more accurate and efficient deep learning methodologies for lattice system modeling is both immediate and vital.

To address these challenges, we have designed an innovative deep learning methodology, utilizing graph neural networks for the exploration of lattice systems. Our primary step entails constructing a graph structure, indicative of pointlike particle interactions. In this configuration, the nodes of the graph correspond to the particles or field variables within the system, while the edges mirror the interactions between these pointlike particles, as visualized in Fig. 1(b). Utilizing this graph-based approach not only provides us with powerful tools to analyze irregularly structured lattices and those presenting long-range interactions, but also offers a significant advantage over traditional methods. Traditional techniques often struggle with the complex topology and multiscale nature of lattice systems, whereas our graph-based method naturally integrates these aspects into the model. This inherent flexibility and adaptability enable us to accurately capture and learn the unique features of each lattice system, regardless of its complexity.

Proceeding from this, we combine graph neural networks (GNN) to achieve lattice system learning. Contrasting with conventional neural networks constructed for vector data, GNNs are purposefully designed to process graphical data. Whereas conventional neural networks analyze each element of the input data individually, GNNs exploit the global structure of the graph and interactions between nodes to capture expansive features and patterns with higher efficacy. As a result, GNNs outshine in recognizing complex nonlinear relationships and topological structures. Additionally, GNNs can adeptly handle graphs of various sizes, thus obviating the need for preprocessing or fixed dimensions, a necessity in conventional network technologies. In addition to this, GNNs utilize the graph structure and message-passing mechanisms to incorporate global information, leading to notably accurate predictions.

In the heart of our model lies the integration of GNNs. This integration is accomplished by setting up neural network parameterized update functions that operate on both node and edge features. In doing so, we harness the rich information encapsulated in the graph formulated from the lattice system for effective learning of system representation. The model, aptly named the separable graph Hamiltonian network (SGHN), is a method for learning the kinetic and potential energies of systems based on graph neural networks. It not only assimilates generalized coordinates and momentum as part of its input data but also incorporates the graph structure resulting from the lattice system. The architecture of SGHN is illustrated in Fig. 2.

To evaluate the performance of SGHN, we subjected it to benchmark tests against three state-of-the-art baseline models in the context of one-dimensional, two-dimensional, and long-range interacting lattice systems. The results indicate that SGHN displays an in-depth understanding of Hamiltonian dynamics, showing significant superiority over the baseline models in preserving system energy and accurately predicting system trajectory. This superiority is particularly evident in systems with multidimensional attributes or notable nonlinearity. Given its versatility and adaptability to any lattice structure irrespective of its size or shape, our approach offers a promising avenue for examining complex lattice physics models that have yet to be fully explored.

Our contributions can be summarized as follows:

(i) We introduce an innovative method of graph-based representation of nonlinear lattice systems with short and/or long-range interactions. This paradigm shift conceptualizes lattice pointlike particles as nodes and their interactions as edges, paving the way for a fresh perspective in the study of lattice systems.

(ii) We pioneer the integration of GNNs into nonlinear Hamiltonian lattice system learning. This approach creates a synergistic link between traditional neural networks and graph theory techniques, proving to be markedly superior.

(iii) Our third major contribution is the extraordinary versatility of our model, which demonstrates its ability to effectively handle a variety of intricate lattice models. This includes, but is not restricted to, high-dimensional and irregular systems, as well as those with long-range interactions. Such adaptability positions our model at the forefront of new explorations in the complex and diverse realm of lattice physics.

II. REPRESENTING LATTICE SYSTEMS ON GRAPHS

Consider a given lattice system, defined by Eqs. (1) and (2), with N pointlike particles. This system can be represented as a directed (or an undirected) graph, symbolized by $\mathcal{G} = (V, E)$, consisting of N nodes. The node set $V = \{v_1, \cdots, v_N\}$ signifies the particles inhabiting our lattice system. The presence of interaction between particles is depicted through linking corresponding nodes with directed edges $E = \{e_{i,j}, i, j = 1, \cdots, N\}$. Each $e_{i,j}$ embodies a directed edge originating

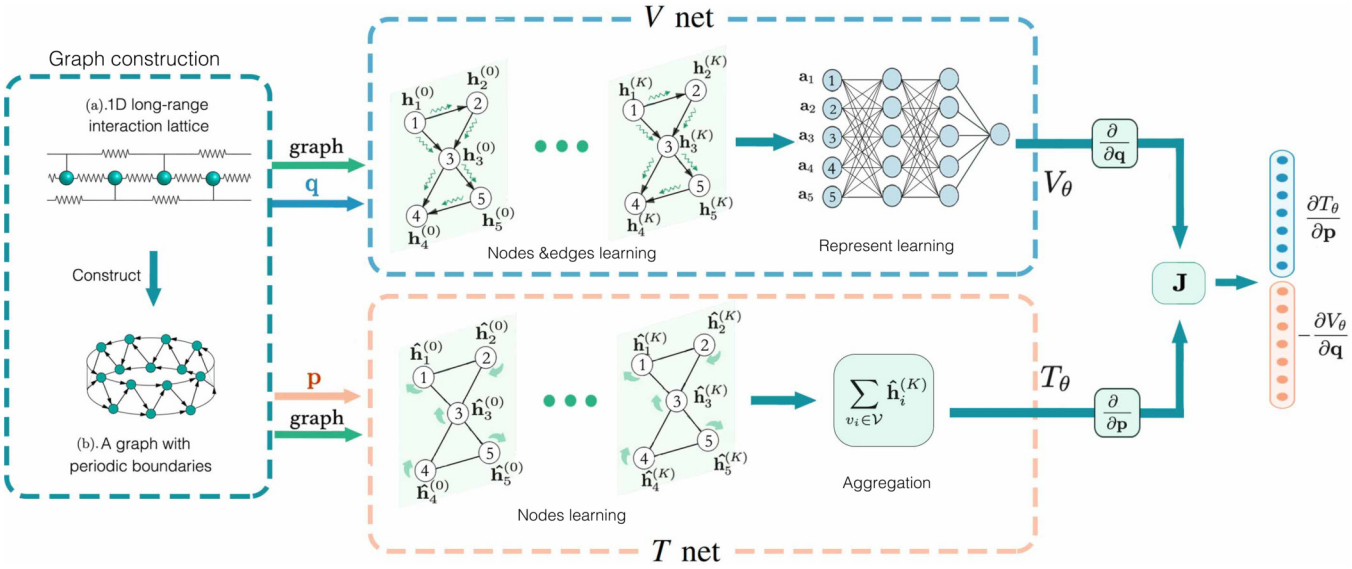


FIG. 2. The architecture of the separable graph Hamiltonian neural network (SGHN). For the definition about $\mathbf{h}_i^{(k)}$, $\hat{\mathbf{h}}_i^{(k)}$, and $\mathbf{a}_i$ see the "Neural Networks Models" section.

from node v_i and ending at node v_j. If such an edge is present, we designate v_i as the in-neighbor of v_j and reciprocally, v_j is referred to as the out-neighbor of v_i.

In this research, we consider three lattice systems. The first one is the 1D Fermi-Pasta-Ulam Klein-Gordon (FPU-KG) chain model. The system is assumed to be periodic, implying that $(q_{1+N}, p_{1+N}) = (q_1, p_1)$. When constructing the graph, we connect the first and last nodes to form a ring graph to maintain the periodicity of the model.

The second one is the 1D Klein-Gordon lattice system with long-range interactions (KG-LRI). The physical interaction structure of the KG-LRI system is depicted in Fig. 2(a). We again extend our investigation to the periodic boundary conditions. Given the inclusion of long-range interactions, the resulting graph is exhibited in Fig. 2(b).

We also consider the 2D Frenkel-Kontorova (FK) model. We assume that the model is located in the quadrilateral lattice. In accordance with the periodic boundary condition, we link nodes located on the parallel edges of the quadrilateral grid graph, so that the graph structure is a torus.

For a detailed introduction to these three lattice models, please refer to the Appendix.

III. NEURAL NETWORK MODELS

Our neural network leverages two neural network components, V net and T net, to model lattice systems with separable Hamiltonian functions, as shown in Fig. 2.

The V net is the parametrization of the potential energy, V_θ, within the lattice system. This parametrization is built upon a GNN that leverages the position, $\mathbf{q}$, and graph structure as inputs.

In the model, each node and edge is tied to a node feature vector, $\mathbf{h}_i^{(0)}$, and an edge feature vector, $\mathbf{e}_{i,j}$, respectively. To kickstart the process, the position q_i is utilized as the initial feature of node v_i within the lattice graph, thereby defining $\mathbf{h}_i^{(0)} = q_i$. In an adaptive fashion, edge features are learned through neural networks as follows:

$$\mathbf{e}_{i,j}^{(0)} = \mathcal{F}_e(q_i - q_j). \tag{3}$$

Here, $\mathcal{F}_e$ signifies functions parameterized by deep neural networks.

Further, we prescribe K-layer node update and edge update operations as such:

$$\mathbf{h}_i^{(k+1)} = \sum_{v_j \in \mathcal{N}_{(i)}} \mathcal{F}_{un}^{(k)}\left(\mathbf{e}_{i,j}^{(k)}\right), \tag{4}$$

$$\mathbf{e}_{i,j}^{(k+1)} = 1/S \sum_{s=1}^{S} \left(\mathbf{e}_{i,j}^{(k)} + \mathcal{F}_{ue}^{(k,s)}\left(\mathbf{h}_i^{(k)} \odot \mathbf{h}_j^{(k)}\right)\right). \tag{5}$$

In these expressions, $\mathcal{F}_{un}^{(k)}$ and $\mathcal{F}_{ue}^{(k,s)}$, with $0 \leqslant k \leqslant K-1$ are functions parametrized by deep neural networks. $\mathcal{N}_{(i)}$ indicates the out-neighbor of v_i, and $\odot$ signifies the Hadamard product used to assess node interactions.

Subsequently, we amalgamate the node and edge information to depict the final node features:

$$\mathbf{a}_i = \mathcal{F}_a\left(\mathcal{F}_{v_1}\left(\mathbf{h}_i^{(0)}\right) \parallel \mathcal{F}_{v_2}\left(\mathbf{h}_i^{(K)}\right) \parallel \sum_{v_j \in \mathcal{N}_{(i)}} \mathcal{F}_{v_3}\left(\mathbf{e}_{ij}^{(K)}\right)\right). \tag{6}$$

In this context, $\mathcal{F}_\cdot$ is a function parametrized by deep neural networks, and $\parallel$ represents the concatenation operation.

As a final step, we employ the function parametrized by deep neural networks to learn the definitive graph representation learning,

$$y_v = \mathcal{F}_v\{\mathbf{a}_i | v_i \in \mathcal{V}\}, \tag{7}$$

where $y_v \in \mathbb{R}$ is the approximate value of potential energy by the neural network model V_θ.

The T net involves the parametrization of the kinetic energy, T_θ. We define the node feature vector as $\hat{\mathbf{h}}_i^{(0)} = p_i$. Then, we use K-layer node update operations as follows:

$$\hat{\mathbf{h}}_i^{(k+1)} = \mathcal{F}_{ut}^{(k)}\left(\hat{\mathbf{h}}_i^{(k)}\right). \tag{8}$$

TABLE I. Specify the parameters of the baseline models. For the LA-SympNet, the last column represents the number of sublayers.

Networks	Hidden layers	Hidden neurons
NODE	2	256
HNN	2	200
LA-SympNet	20	4 (sublayers)
G-SympNet	20	50

Finally, we obtain the kinetic energy of the system by aggregation function as follows:

$$y_t = \sum_{v_i \in \mathcal{V}} \hat{\mathbf{h}}_i^{(K)}. \tag{9}$$

We learned the system Hamiltonian parametrization function $V_\theta + T_\theta$ through neural networks, and defined the following loss function to ensure the ability to learn precise conserved quantities from the data:

$$\mathcal{L} = \left\| \frac{\partial T_\theta}{\partial \mathbf{p}} - \frac{d\mathbf{q}}{dt} \right\|_2 + \left\| -\frac{\partial V_\theta}{\partial \mathbf{q}} - \frac{d\mathbf{p}}{dt} \right\|_2. \tag{10}$$

IV. EMPIRICAL EVALUATION

To evaluate the efficacy of our SGHN, we made a comparative analysis with three established models: neural ordinary differential equations (NODEs) [29], Hamiltonian neural networks (HNNs) [30], and symplectic networks (SympNets) [25]. The latter two models, HNNs and SympNets, are widely recognized as the leading-edge gray-box modeling tools for learning Hamiltonian systems.

For the SGHN model, Eq. (8) is the fully connected neural network that has two hidden layers, each with ten hidden units. Equation (7) is the fully connected neural network that has two hidden layers, each with sixty hidden units. The neural network parametrization functions in Eqs. (3)–(6) are all represented by the fully connected neural network, with two hidden layers and five units per layer. In Eqs. (7) and (8), the layers for node and edge updates are $K = 1$. $S = 2$ in Eq. (7).

The specific parameters of the NODE, HNN, LA-SympNet, and G-SympNet are shown in Table I. The total parameters of the network model for 1D lattice systems and 2D lattice systems are shown in Table II.

TABLE II. Network model parameters.

	NODE	HNN	LA-SympNet	G-SympNet	SGHN
1D	98 816	53 400	83 808	34 000	6390
2D	213 504	98 200	1 667 376	146 000	13 110

For dataset creation, training details, and initial value settings, as well as some supplementary experiments, please refer to the Appendix.

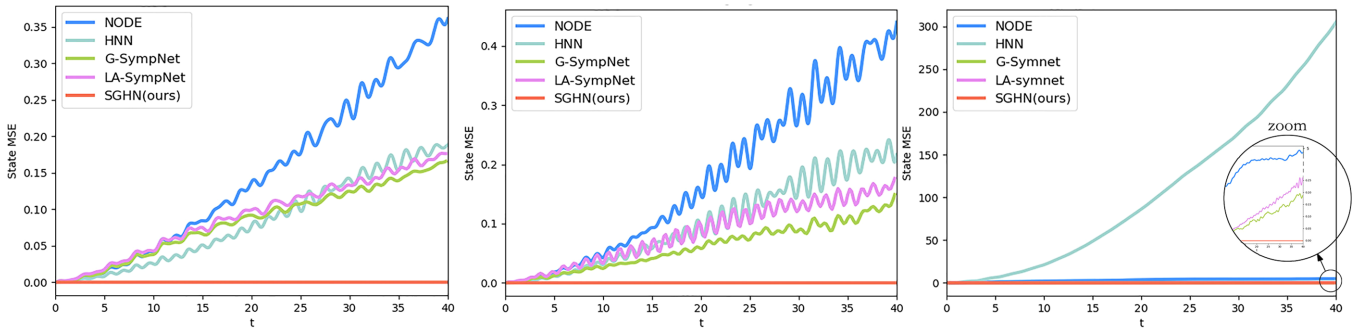
V. RESULTS

To appraise the performance of our model, we relied on the mean squared error (MSE) of the predicted energies ($\mathcal{H}$), and the MSE of the predicted state variables ($\mathbf{q}_\theta, \mathbf{p}_\theta$). As presented in Table III, the MSE of the predicted energies for 20 test samples distinctly underlines the superiority of the SGHN over other models, a point further reinforced by the highest performing results highlighted in bold. Figure 3 provides a compelling visual comparison of the MSE of the predicted state variables over time, which reveals the exceptional long-term predictive stability of the SGHN. In addition, our dynamic demonstration (refer to README.md in [31]). comparing predicted and ground truth positions given an arbitrary initial value further testifies to SGHN's proficiency, with the predicted trajectory mirroring the ground truth so closely that discerning any deviation is beyond the capabilities of the unaided eye.

Analysis of the two key metrics clearly shows that SGHN significantly excels over the other models in all aspects. In addition, compared to the baseline model, SGHN uses the least number of parameters and has the best performance in maintaining system energy over the long term and accurately predicting system trajectories. To investigate the impact of graph neural networks on our model's performance, we substituted the V net and T net components with a three-layer fully connected neural network. Each layer of this network contains 200 hidden units. The energy prediction MSE for this configuration is presented in Table IV. This comparison reveals a significant decline in the network's performance when graph neural networks are not utilized, underscoring their importance in our model. These outcomes underscore the instrumental role of the graph neural network in learning lattice systems.

TABLE III. The MSE of the predicted energies, which measures whether the network model conforms to the property of system energy conservation when predicting over long timespans. The best results are emphasized by bold fonts.

	FPU-KG	KG-LRI	2D FK
NODE	$5.36 \times 10^{-1} \pm 1.22 \times 10^{0}$	$3.54 \times 10^{-1} \pm 3.73 \times 10^{-1}$	$9.85 \times 10^{4} \pm 2.14 \times 10^{5}$
HNN	$5.98 \times 10^{-3} \pm 6.21 \times 10^{-3}$	$4.48 \times 10^{-3} \pm 5.64 \times 10^{-3}$	$1.26 \times 10^{8} \pm 5.22 \times 10^{7}$
LA-SympNet	$5.01 \times 10^{-2} \pm 4.06 \times 10^{-2}$	$4.55 \times 10^{-2} \pm 4.11 \times 10^{-2}$	$3.89 \times 10^{-1} \pm 1.74 \times 10^{-1}$
G-SympNet	$4.63 \times 10^{-2} \pm 3.96 \times 10^{-2}$	$8.27 \times 10^{-2} \pm 5.90 \times 10^{-2}$	$2.13 \times 10^{-1} \pm 3.46 \times 10^{-2}$
SGHN(ours)	$\mathbf{3.18 \times 10^{-7} \pm 4.13 \times 10^{-7}}$	$\mathbf{1.28 \times 10^{-6} \pm 1.95 \times 10^{-6}}$	$\mathbf{2.72 \times 10^{-6} \pm 1.19 \times 10^{-5}}$



(a) The $\mathrm{MSE}_{\mathrm{state}}$ of FPU-KG model

(b) The $\mathrm{MSE}_{\mathrm{state}}$ of KG-LRI model

(c) The $\mathrm{MSE}_{\mathrm{state}}$ of 2D FK model

FIG. 3. MSE of the predicted state variables, which measures the stability of the network model in predicting generalized momentum and generalized position over long timespans. The color line represents the MSE of the predicted state over time, which is the average value of 20 test samples. See README.md in [31], for a dynamic demonstration of comparing the predicted position with the ground truth position given a random initial value.

VI. CONCLUSION

In conclusion, our work presents a groundbreaking methodology for deciphering and understanding lattice systems, substantially transforming our comprehension of these structures. We effectively merge graph structures with the intricate topology and varied interaction ranges of lattice particles, culminating in a robust model that excels in deciphering complex lattice systems, particularly those exhibiting high dimensionality or notable nonlinearity.

By leveraging the inherent structure of these systems, our graph-centric approach effectively encapsulates the potential energy generated from particle interactions or external influences, as well as the kinetic energy of the system. This model proficiency is encapsulated in the adoption of GNNs to house this potential energy and kinetic energy. Our model continually outperforms base models in predictive accuracy and generalization across all evaluated scenarios, notably in systems with heightened dimensional properties or significant nonlinearity. Furthermore, our simulation results highlight the crucial role of GNNs in bolstering neural network performance.

This study initiates a paradigm shift in the exploration and understanding of lattice systems, paving the way for new research directions within this field. We are confident that this innovative approach will propel further inquiries, leading to a plethora of potential applications and advancements in this domain.

ACKNOWLEDGMENTS

Y.X.G. was supported by NSFC grants (Grants No. 12071065 and No. 11871140). H.K.Z. was partially supported by Simons Foundation Collaboration Grants for Mathematicians (Grant No. 706383), as well as NSF-DMS (Grant No. 2220211).

APPENDIX: SUPPLEMENTAL MATERIAL

In this Appendix, we will introduce the lattice system used for testing, the acquisition of datasets, training settings, and some supplementary experiments.

1. Lattice systems

(i) Fermi-Pasta-Ulam Klein-Gordon (FPU-KG) model.

The FPU-KG model, a derivative of the extended discrete nonlinear Schrödinger equation, provides profound insights into mixed-type lattices. Especially achievable through rotational wave type approximation, this model serves as a touchstone for the exploration of breather-forming mechanisms [32,33].

The potential energy of the model can be expressed as follows:

$$V_1(q_i, q_{i+1}) = a\frac{(q_{i+1}-q_i)^2}{2} + b\frac{(q_{i+1}-q_i)^4}{4}, \quad \text{(A1)}$$

$$V_2(q_i) = \frac{q_i^2}{2} + \frac{q_i^4}{4}, \quad \text{(A2)}$$

where $i = 1, \cdots, N$, $a > 0$, and $b \geqslant 0$. In our experimental setting, we assign the values $a = 1$, $b = 0.25$, and $N = 32$.

(ii) The Klein-Gordon (KG) Lattice System.

The KG lattice system stands as a paradigmatic model within theoretical and applied physics, particularly instrumental in studying nonlinear phenomena like localized excitations in ionic crystals [34], and thermal denaturation of DNA [13].

TABLE IV. The MSE of the predicted energies, where V net and T net are fully connected neural networks.

	FPU-KGE	KG-LRI	2D FK
Model parameters	94 000	94 000	138 800
MSE	$1.02 \times 10^{-4} \pm 6.98 \times 10^{-5}$	$3.26 \times 10^{-4} \pm 5.17 \times 10^{-4}$	$3.36 \times 10^{-3} \pm 2.989 \times 10^{-3}$

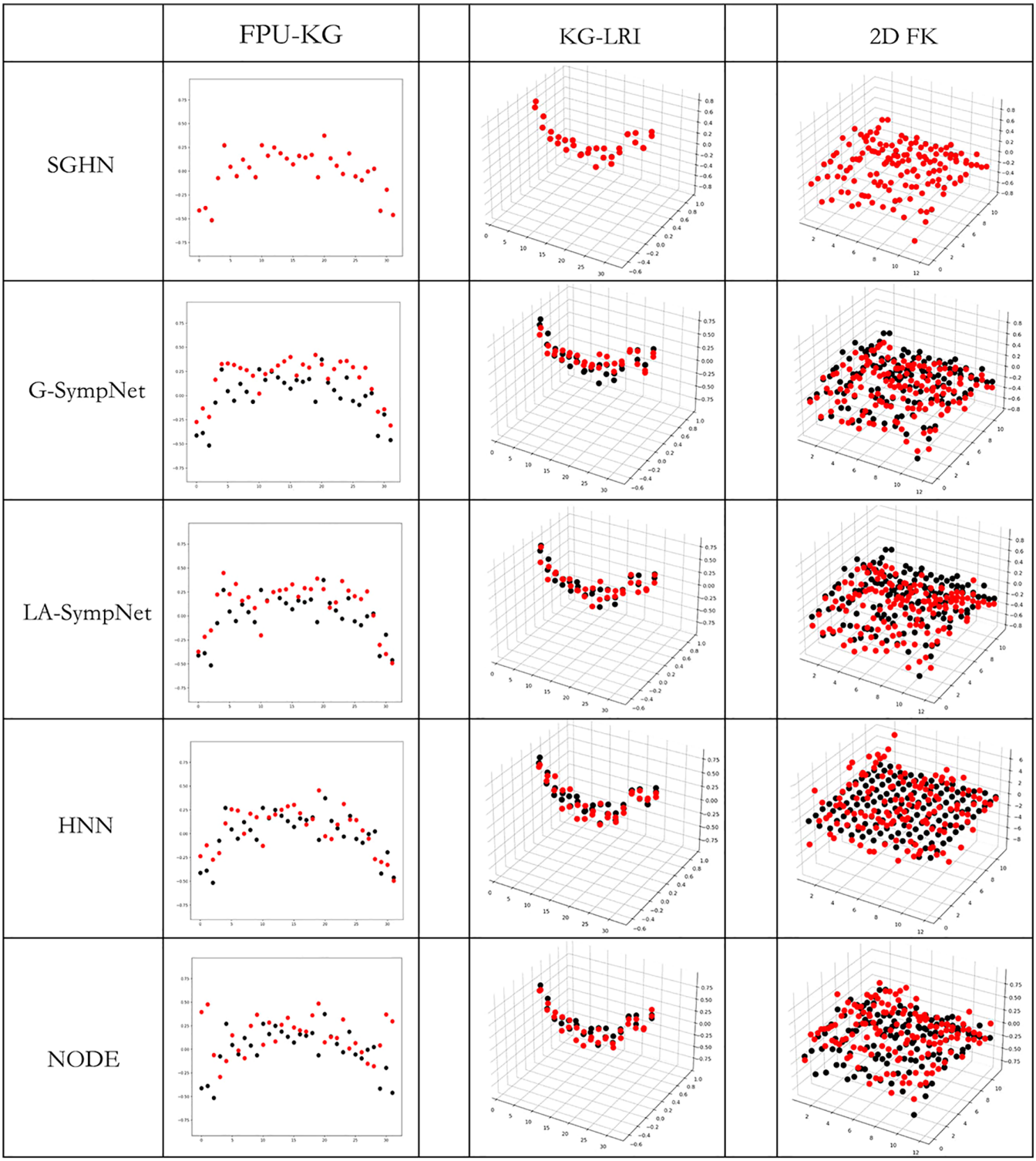


FIG. 4. When $t = 40s$, the comparison between the predicted position and the ground truth position with a given random initial value. See Ref. [31] for dynamic demonstrations.

The potential energy of the KG system with long-range interactions (KG-LRI) is delineated as follows [35]:

$$V_1(q_i, q_{i+1}, q_{i+2}) = a\frac{(q_{i+1} - q_i)^2}{2} + b\frac{(q_{i+2} - q_i)^2}{2},$$

where $a > 0$, $b \geqslant 0$, and $V_2(q_i)$ is as defined in Eq. (A2), for $i = 1, \ldots, N$.

We extend our investigation to the periodic KG-LRI model, presuming $a = b = 1$ and a total of $N = 32$ nodes.

(*iii*) *The 2D Frenkel-Kontorova (FK) Model.*

The FK model holds significant theoretical value within the realms of solid-state physics and nonlinear dynamics [36]. It encapsulates the intriguing interplay between local particle interactions and a periodic potential enforcing long-range order

TABLE V. Specify the parameters of the network structure. For the LA-SympNet, the last column represents the number of sublayers.

Networks	Hidden layers	Hidden neurons
NODE′	5	200
NODE″	2	600
HNN′	5	200
HNN″	2	600
LA-SympNet′	50	8 (sublayers)
LA-SympNet″	10	2 (sublayers)
G-SympNet′	10	100
G-SympNet″	20	100

[37–41]. This interaction results in an intriguing variety of dynamical behaviors, including the manifestation of topological solitons, which are localized particle displacements moving through the lattice without altering shape. In this work, we focus on a two-dimensional variant of the FK model [42].

For $i = 1, \ldots, N$ and $j = 1, \ldots, M$, the potential energy $V_2(q_{i,j}) = -\cos(q_{i,j})$ and $V_1(q_{i,j}, q_{i+1,j}, q_{i,j+1})$ is given by

$$a\frac{(q_{i+1,j} - q_{i,j} - \rho)^2}{2} + b\frac{(q_{i,j+1} - q_{i,j})^2}{2},$$

where $a > 0$, $b > 0$, and ρ denotes the average particle distance absent an external potential. Within our 2D FK model framework, we set $a = b = \rho = 1$. Assuming $M = N = 12$, this model portrays a quadrilateral grid with a total of 144 particles.

2. Dataset acquisition

For the FPU-KG model and KG-LRI model, we set $N = 32$, and for 2D FK, we set $M = N = 12$. We used the fourth-order Runge-Kutta integrator by scipy.integrate.solve_ivp on the time interval $[0, 40]$ sec with time step 0.002 and error tolerance 10^{-12} to find 100 trajectories. The initial conditions (ICs) of FPU-KG and KG-LRI were

$$\mathbf{q}_0(i) = \lambda_i \sin\left(\frac{(i-1)\pi}{N-1}\right), \tag{A3}$$

$$\mathbf{p}_0(i) = 0, \quad i = 1, \cdots, N, \tag{A4}$$

where $\lambda_i \sim \mathcal{N}(0, 1)$ and $\mathcal{N}$ represents the normal distribution. Their boundary conditions (BCs) were $q_{i+N} = q_i$ and $p_{i+N} = p_i$. The ICs of 2D FK were

$$\mathbf{q}_0(i, j) = \lambda_{i,j} \sin\left(\frac{\big(M(i-1) + (j-1)\big)\pi}{MN - 1}\right), \tag{A5}$$

$$\mathbf{p}_0(i, j) = 0, \quad i = 1, \cdots, N, j = 1, \cdots, M. \tag{A6}$$

The BCs were $q_{i+N,j+M} = q_{i,j}$ and $p_{i+N,j+M} = p_{i,j}$.

We subsampled the trajectories at a fixed timestep of 0.2 as a dataset and then performed a 50/30/20% train/validation/test set split over trajectories.

3. Training settings

We adopt a learning rate piecewise constant decay strategy [43]. The learning rate was initialized to 10^{-4}, and after 3000 epochs, the learning rate was 10^{-5}. We also applied an early stop strategy [44] for the validation set's loss, with the patience set to 100 epochs. The total epochs were set to 100 000. The optimizer was Adam and the batch size was 256. The activation functions of all models were taken tanh.

4. Trajectory prediction

The prediction of trajectory was accomplished by integrating neural network models as per the following equation:

$$(\mathbf{q}_{\theta,t}, \mathbf{p}_{\theta,t}) = (\mathbf{q}_0, \mathbf{p}_0) + \int_{t_0}^{t} \mathcal{N}_\theta dt. \tag{A7}$$

We utilized a fourth-order Runge-Kutta integrator for this purpose, where $(\mathbf{q}_0, \mathbf{p}_0)$ represents the initial values from the test set. In our study, $t_0 = 0$ and $t = 40$, with a time step size of 0.002. Here, $\mathcal{N}_\theta$ refers to the neural network models, which include NODE, HNN, and SGHN. For the SympNet model, the predicted state was obtained using the subsequent equation:

$$(\mathbf{q}_{\theta,t+1}, \mathbf{p}_{\theta,t+1}) = \mathcal{N}_{\text{SympNet}}(\mathbf{q}_{\theta,t}, \mathbf{p}_{\theta,t}). \tag{A8}$$

In this context, the time step size was kept consistent with the training step size at 0.2. $\mathcal{N}_{\text{SympNet}}$ symbolizes LA type and G

TABLE VI. The MSE of the predicted energies, which measures whether the network model conforms to the property of system energy conservation when predicting over long timespans. The best results are emphasized by bold fonts. The parameters required for the network model are in parentheses, and the parameters used for the FPU-KG and KG-LRI network models are the same.

	FPU-KG	KG-LRI (parameters)	2D FK (parameters)
NODE′	$1.36 \times 10^{4} \pm 7.12 \times 10^{3}$	$1.07 \times 10^{3} \pm 4.04 \times 10^{2}$ (174 000)	$1.70 \times 10^{5} \pm 3.32 \times 10^{4}$ (218 800)
NODE″	$1.83 \times 10^{-2} \pm 1.34 \times 10^{-2}$	$1.09 \times 10^{-2} \pm 1.27 \times 10^{-2}$ (438 000)	$4.13 \times 10^{-1} \pm 6.89 \times 10^{-1}$ (706 800)
HNN′	$7.18 \times 10^{-4} \pm 5.12 \times 10^{-4}$	$3.76 \times 10^{-2} \pm 2.68 \times 10^{-2}$ (174 000)	$2.33 \times 10^{8} \pm 1.08 \times 10^{8}$ (218 800)
HNN″	$1.47 \times 10^{-4} \pm 1.31 \times 10^{-4}$	$2.35 \times 10^{-4} \pm 1.82 \times 10^{-4}$ (400 200)	$1.61 \times 10^{-2} \pm 8.25 \times 10^{-3}$ (534 600)
LA-SympNet′	$2.04 \times 10^{-2} \pm 1.78 \times 10^{-2}$	$4.87 \times 10^{-2} \pm 4.10 \times 10^{-2}$ (663 008)	$6.04 \times 10^{-1} \pm 1.79 \times 10^{-1}$ (8 315 856)
LA-SympNet″	$3.05 \times 10^{-2} \pm 2.86 \times 10^{-2}$	$5.50 \times 10^{-2} \pm 4.36 \times 10^{-2}$ (21 408)	$4.12 \times 10^{-2} \pm 1.20 \times 10^{-2}$ (418 896)
G-SympNet′	$4.89 \times 10^{-2} \pm 3.99 \times 10^{-2}$	$8.78 \times 10^{-2} \pm 6.04 \times 10^{-2}$ (34 000)	$2.35 \times 10^{-1} \pm 3.39 \times 10^{-2}$ (146 000)
G-SympNet″	$4.75 \times 10^{-2} \pm 3.92 \times 10^{-2}$	$5.09 \times 10^{-2} \pm 4.40 \times 10^{-2}$ (68 000)	$4.76 \times 10^{-1} \pm 5.40 \times 10^{-2}$ (292 000)
SGHN(ours)	$\mathbf{3.18 \times 10^{-7} \pm 4.13 \times 10^{-7}}$	$\mathbf{1.28 \times 10^{-6} \pm 1.95 \times 10^{-6}}$ (6390)	$\mathbf{2.72 \times 10^{-6} \pm 1.19 \times 10^{-5}}$ (13 110)

type SympNets [25]. The MSE of the predicted state variables is defined as

$$\mathrm{MSE}_{\mathrm{state}} = \sum_{\alpha\in\mathbb{Z}^{\mathrm{d}}} \left((\mathrm{q}_{\alpha,\mathrm{t}} - \mathrm{q}_{\theta,\alpha,\mathrm{t}})^2 + (\mathrm{p}_{\alpha,\mathrm{t}} - \mathrm{p}_{\theta,\alpha,\mathrm{t}})^2 \right).$$

5. Experimental supplement

Figure 4 shows a snapshot of the comparison between the predicted position and ground position at $t = 40s$, where the initial values are randomly given. The red circle represents the predicted particle position, while the black circle represents the ground truth particle position. It can be seen that there is almost no difference between the predicted position and the ground truth position of the SGHN model. The dynamic diagram can be seen in [31]. For the sake of rigor, we have increased the depth and width of the baseline models, and their network settings are shown in Table V. We recorded the MSE of the predicted energy in Table VI, and it can be seen that different network parameter settings did not improve the performance of the baseline model.

Observing Table VI, we found that for the 2D FK system, increasing the HNN width from 200 to 600 significantly improved the results of MSE. We continue to explore the impact of increasing the width of HNN on the results. We take three layers of 1200, 2400, and 3600 hidden units, and their predicted energy MSEs are $1.29 \times 10^{-2} \pm 7.82 \times 10^{-3}$, $9.31 \times 10^{-3} \pm 6.80 \times 10^{-3}$, and $1.90 \times 10^{-2} \pm 1.15 \times 10^{-2}$, respectively. It can be seen that the HNN model has reached its limit under certain parameters. Continuing to increase the model width will not further improve the performance of the model.

[1] D. Feng and G. Jin, *Introduction to Condensed Matter Physics* (World Scientific Publishing Company, Singapore, 2005), Vol. 1.

[2] I. Ichinose and T. Matsui, Lattice gauge theory for condensed matter physics: Ferromagnetic superconductivity as its example, Mod. Phys. Lett. B **28**, 1430012 (2014).

[3] V. G. Filev and D. O'Connor, The BFSS model on the lattice, J. High Energy Phys. 05 (2016) 167.

[4] Y. Kuno, Phase structure of the interacting Su-Schrieffer-Heeger model and the relationship with the Gross-Neveu model on lattice, Phys. Rev. B **99**, 064105 (2019).

[5] A. Caravano, E. Komatsu, K. D. Lozanov, and J. Weller, Lattice simulations of inflation, J. Cosmol. Astropart. Phys. 12 (2021) 010.

[6] M. Sakellariadou, Numerical experiments on string cosmology, Nucl. Phys. B **468**, 319 (1996).

[7] A. Spagnoli, A micromechanical lattice model to describe the fracture behaviour of engineered cementitious composites, Comput. Mater. Sci. **46**, 7 (2009).

[8] Y. Starosvetsky, K. Jayaprakash, M. A. Hasan, and A. F. Vakakis, *Topics on the Nonlinear Dynamics and Acoustics of Ordered Granular Media* (World Scientific, Singapore, 2017).

[9] C. Chong and P. G. Kevrekidis, *Coherent Structures in Granular Crystals: From Experiment and Modelling to Computation and Mathematical Analysis* (Springer, Cham, Switzerland, 2018).

[10] J. Dudowicz, K. F. Freed, and J. F. Douglas, Lattice model of equilibrium polymerization. IV. Influence of activation, chemical initiation, chain scission and fusion, and chain stiffness on polymerization and phase separation, J. Chem. Phys. **119**, 12645 (2003).

[11] M. Peyrard and A. R. Bishop, Statistical mechanics of a nonlinear model for DNA denaturation, Phys. Rev. Lett. **62**, 2755 (1989).

[12] G. Gaeta, C. Reiss, M. Peyrard, and T. Dauxois, Simple models of non-linear DNA dynamics, Riv. Nuovo Cim. **17**, 1 (1994).

[13] M. Peyrard, Nonlinear dynamics and statistical physics of DNA, Nonlinearity **17**, R1 (2004).

[14] S. Yomosa, Soliton excitations in deoxyribonucleic acid (DNA) double helices, Phys. Rev. A **27**, 2120 (1983).

[15] L. Yang, Finite heat conduction in a 2d disorder lattice, Phys. Rev. Lett. **88**, 094301 (2002).

[16] P. G. Kevrekidis, J. Cuevas-Maraver, and A. Saxena, *Emerging Frontiers in Nonlinear Science* (Springer, Cham, Switzerland, 2020).

[17] R. Wang and R. Yu, Physics-guided deep learning for dynamical systems: A survey, arXiv:2107.01272.

[18] G. Carleo, I. Cirac, K. Cranmer, and Da, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[19] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, Phys. Rep. **810**, 1 (2019).

[20] J. Li, J. Chen, and B. Li, Gradient-optimized physics-informed neural networks (GOPINNs): a deep learning method for solving the complex modified KdV equation, Nonlinear Dynamics **107**, 781 (2022).

[21] D. Zvyagintseva, H. Sigurdsson, V. K. Kozin, I. Iorsh, I. A. Shelykh, V. Ulyantsev, and O. Kyriienko, Machine learning of phase transitions in nonlinear polariton lattices, Commun. Phys. **5**, 8 (2022).

[22] S. Saqlain, W. Zhu, E. G. Charalampidis, and P. G. Kevrekidis, Discovering governing equations in discrete systems using PINNs, Commun. Nonlinear Sci. Numer. Simul. **126**, 107498 (2023).

[23] W. Zhu, W. Khademi, E. G. Charalampidis, and P. G. Kevrekidis, Neural networks enforcing physical symmetries in nonlinear dynamical lattices: The case example of the Ablowitz–Ladik model, Physica D **434**, 133264 (2022).

[24] P. Jin, Z. Zhang, I. G. Kevrekidis, and G. E. Karniadakis, Learning Poisson systems and trajectories of autonomous systems via Poisson neural networks, IEEE Trans. Neural Netw. Learn. Syst. **34**, 8271 (2023).

[25] P. Jin, Z. Zhang, A. Zhu, Y. Tang, and G. E. Karniadakis, SympNets: Intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems, Neural Networks **132**, 166 (2020).

[26] M. Yip, J. Mohle, and J. E. Bolander, Automated modeling of three-dimensional structural components using irregular lattices, Comput-Aided Civ Inf. **20**, 393 (2010).

[27] N. Laskin and G. Zaslavsky, Nonlinear fractional dynamics on a lattice with long range interactions, Physica A **368**, 38 (2006).

[28] N. Go and H. Taketomi, Respective roles of short-and long-range interactions in protein folding, Proc. Natl. Acad. Sci. **75**, 559 (1978).

[29] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, Neural ordinary differential equations, Adv. Neural Inf. Proc. Sys. **31**, 6572 (2018).

[30] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks, Adv. Neural Inf. Proc. Sys. **32**, 15379 (2019).

[31] https://github.com/GengRu93/DemoSGH.

[32] M. Johansson, Discrete nonlinear Schrödinger approximation of a mixed Klein–Gordon/Fermi–Pasta–Ulam chain: Modulational instability and a statistical condition for creation of thermodynamic breathers, Physica D **216**, 62 (2006).

[33] S. Paleari and T. Penati, Long time stability of small-amplitude Breathers in a mixed FPU-KG model, Z. Angew. Math. Phys. **67**, 148 (2016).

[34] T. Penati, V. Koukouloyannis, M. Sansottera, P. Kevrekidis, and S. Paleari, On the nonexistence of degenerate phase-shift multi-breathers in Klein–Gordon models with interactions beyond nearest neighbors, Physica D **398**, 92 (2019).

[35] V. Koukouloyannis, P. G. Kevrekidis, J. Cuevas, and V. Rothos, Multibreathers in Klein–Gordon chains with interactions beyond nearest neighbors, Physica D **242**, 16 (2013).

[36] O. M. Braun and Y. S. Kivshar, *The Frenkel-Kontorova Model: Concepts, Methods, and Applications* (Springer, Berlin, Heidelberg, 2004).

[37] D. Srolovitz and P. Lomdahl, Dislocation dynamics in the 2-D Frenkel-Kontorova model, Physica D **23**, 402 (1986).

[38] A. M. Popov, I. V. Lebedeva, A. A. Knizhnik, Y. E. Lozovik, and B. V. Potapkin, Commensurate-incommensurate phase transition in bilayer graphene, Phys. Rev. B **84**, 045404 (2011).

[39] N. Nikola, D. Hexner, and D. Levine, Entropic commensurate-incommensurate transition, Phys. Rev. Lett. **110**, 125701 (2013).

[40] J. Sokoloff, Sliding charge-density waves in periodic and disordered lattices, Phys. Rev. B **16**, 3367 (1977).

[41] C. Woods, L. Britnell, A. Eckmann, R. Ma, J. Lu, H. Guo, X. Lin, G. Yu, Y. Cao, R. V. Gorbachev *et al.*, Commensurate–incommensurate transition in graphene on hexagonal boron nitride, Nat. Phys. **10**, 451 (2014).

[42] E. Granato and S. C. Ying, Dynamical transitions and sliding friction in the two-dimensional Frenkel-Kontorova model, Phys. Rev. B **59**, 5154 (1999).

[43] G. Montavon, G. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade* (Springer, Berlin, Heidelberg, 2012), Vol. 7700.

[44] G. Raskutti, M. J. Wainwright, and B. Yu, Early stopping and non-parametric regression: An optimal data-dependent stopping rule, J. Mach. Learn. Res. **15**, 335 (2014).