MULTI-RESOLUTION ENERGY DIFFUSION GRAPH NEURAL NETWORK DISCOVERS SYSTEM CONSERVATION VARIABLES

Anonymous authors
Paper under double-blind review

ABSTRACT

Discovering conservation laws and assessing integrability are fundamental to understanding the evolution and long-term behavior of Hamiltonian dynamical systems. The neural deflation method provides a principled framework for iteratively learning conservation laws, but its performance depends on the representational capacity of the underlying neural network. Standard multilayer perceptrons lack structural priors and struggle to scale to high-dimensional lattice systems, leading to degraded accuracy and limited interpretability. We propose incorporating a multi-resolution diffusion graph neural network (MR-DGNN) as the approximator for conserved quantities within neural deflation. MR-DGNN uses multi-resolution graph diffusion operators to adaptively capture long-range and local particle interactions. This architecture aligns with the discrete geometric structure of lattice systems and achieves higher accuracy with significantly fewer parameters than baselines. Moreover, the learned diffusion weights reveal the spatial interaction range of each conserved quantity, supporting physical interpretation. We evaluate the method on several benchmark systems. Results show that MR-DGNN consistently outperforms baselines in test error, with advantages becoming more pronounced as the number of degrees of freedom increases. The interpretable weight patterns further indicate that different conserved quantities exhibit distinct interaction ranges. These findings suggest that the method can accurately discover conservation laws while providing physical insights into complex high-dimensional Hamiltonian systems.

1 INTRODUCTION

Discovering conservation laws and assessing the integrability of Hamiltonian systems are central to understanding their dynamical evolution and long-term behavior (Mouchet, 2016; Noether, 1983). Conserved quantities provide fundamental constraints for modeling complex phenomena across numerous disciplines, including optics, atomic physics, fluid dynamics, and plasma physics (Cicogna, 2013). Traditionally, physicists have relied on problem-specific pencil-and-paper derivations, making it difficult to develop general-purpose approaches. In recent years, the powerful pattern recognition capabilities of neural networks have driven the emergence of a data-driven scientific discovery paradigm—automatically extracting knowledge from experimental observations (Lu et al., 2023; Wetzel et al., 2020). Although a substantial body of research has made progress in learning conservation laws, existing methods still suffer from notable limitations: some approaches focus on the manifold associated with a single trajectory and are difficult to scale to high-dimensional systems (Liu & Tegmark, 2021); others require prior knowledge of the governing equations or the functional form of basis functions (Liu et al., 2022; Mebratie et al., 2025; Liu et al., 2024); some can only learn isolated conserved quantities and struggle with systems possessing multiple conservation laws (Ha & Jeong, 2021; Zhang et al., 2023); others fail to identify the complete set of conserved quantities (Chen et al., 2025). These shortcomings highlight the urgent need for novel approaches to conservation law discovery, particularly for high-dimensional complex lattice systems.

Zhu et al. (Zhu et al., 2023) introduced the neural deflation method, which iteratively discovers conservation laws and has demonstrated effectiveness on several benchmark systems. However, the practical performance of this method heavily depends on the representational capacity of the under-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

lying neural network. For low-dimensional problems, standard multi-layer perceptrons (MLPs) are often sufficient; but as dimensionality increases, MLPs not only suffer from a dramatic growth in parameter count but also fail to exploit the spatial locality inherent in lattice systems, leading to degraded approximation accuracy, insufficient generalization, and a lack of clear interpretability in the learned quantities—since MLPs do not explicitly encode interaction ranges, subsequent symbolic regression also becomes challenging.

To overcome these limitations, we propose embedding a multi-resolution diffusion graph neural network (MR-DGNN) into the neural deflation framework as an approximator for conserved quantities. The MR-DGNN adaptively captures long-range and local multi-scale dependencies among nodes through multi-resolution diffusion operators on graphs. This architecture naturally aligns with the discrete geometric structure of lattice systems, achieving superior approximation accuracy with far fewer parameters than MLPs. More importantly, the learned diffusion weights directly reveal the spatial interaction range corresponding to each conserved quantity, significantly enhancing physical interpretability. Moreover, the accurate approximation of conserved quantities provided by the MR-DGNN facilitates subsequent symbolic regression toward closed-form analytical expressions.

The main contributions of this paper are threefold:

1. We integrate the MR-DGNN into the neural deflation framework, enabling efficient and scalable discovery of complete sets of conservation laws in high-dimensional lattice systems;

2. Experiments on both integrable (Toda (Toda, 2012), Calogero-Moser (Calogero, 1971; Moser, 1976)) and non-integrable (Fermi-Pasta-Ulam-Tsingou(FPUT) (Gallavotti, 2007; Fermi et al., 1955), sine-Gordon (Williams, 2014)) systems demonstrate that our method significantly outperforms baselines in terms of test accuracy and parameter efficiency;

3. By visualizing the aggregated diffusion weights, we reveal distinct interaction ranges for different conserved quantities—providing not only quantitative performance gains but also qualitative physical insights that guide symbolic regression and deepen our understanding of the underlying system structure.

2 Hamiltonian Systems and Conservation Laws

The Hamiltonian system with d degrees of freedom described by the canonical coordinates $\mathbf{x} = (\mathbf{q}, \mathbf{p}) \in \mathbb{R}^{2d}$, where $\mathbf{q}$ and $\mathbf{p}$ denote the generalized positions and conjugate momenta, respectively. The dynamics are governed by Hamilton's equations:

$$\dot{\mathbf{q}} = \nabla_{\mathbf{p}} H(\mathbf{q}, \mathbf{p}), \qquad \dot{\mathbf{p}} = -\nabla_{\mathbf{q}} H(\mathbf{q}, \mathbf{p}),$$

where $H : \mathbb{R}^{2d} \to \mathbb{R}$ is the Hamiltonian function, representing the total energy of the system. More generally, the system can be written in the compact form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) = J(\mathbf{x}) \nabla H(\mathbf{x}),$$

with $J(\mathbf{x})$ being a skew-symmetric matrix, often the standard symplectic matrix in canonical coordinates.

A smooth function $I : \mathbb{R}^{2d} \to \mathbb{R}$ is called a **conservation law** (or a first integral) of the system if it remains constant along any trajectory $\mathbf{x}(t)$ satisfying the Hamiltonian dynamics, i.e.,

$$I(\mathbf{x}(t)) \equiv I(\mathbf{x}(0)), \quad \forall t.$$

This condition is equivalent to the vanishing of the Poisson bracket between I and the Hamiltonian H:

$$\{I, H\}(\mathbf{x}) = \nabla I(\mathbf{x})^\top J(\mathbf{x}) \nabla H(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in D,$$

where $D \subset \mathbb{R}^{2d}$ is the phase space domain. For two smooth functions F and G, their Poisson bracket is defined as

$$\{F, G\}(\mathbf{x}) = \nabla F(\mathbf{x})^\top J(\mathbf{x}) \nabla G(\mathbf{x}).$$

A set of K conservation laws $\{I_k\}_{k=1}^K$ is said to be **functionally independent** if their gradients $\{\nabla I_k(\mathbf{x})\}_{k=1}^K$ are linearly independent for almost every $\mathbf{x} \in D$. They are said to be **in involution**, or Poisson-commuting, if

$$\{I_j, I_k\} = 0, \quad \forall j \neq k.$$

For a Hamiltonian system with d degrees of freedom, there can exist at most d functionally independent conservation laws that are pairwise in involution. When such a maximal set of d conserved quantities exists, the system is said to be **completely integrable in the Liouville sense**. In such cases, the motion is confined to invariant tori, and the system admits a canonical transformation to action-angle variables.

3 Neural Deflation Method

The neural deflation method enables the systematic discovery of a maximal set of functionally independent and involutive conservation laws from data. This iterative procedure extends the classical idea of deflation — used in numerical bifurcation analysis to find multiple solutions of nonlinear equations — to the context of learning conserved quantities.

Each candidate conservation law $I_k(\mathbf{x})$ is parameterized by using a feedforward neural network $I_k(\mathbf{x}; \boldsymbol{\theta}_k)$, where $\boldsymbol{\theta}_k$ denotes the trainable parameters. Given a training set $\mathcal{T} \subset \mathbb{R}^{2d}$ sampled from the phase space, the first conserved quantity I_1 is learned by minimizing the conservation loss:

$$\mathcal{L}_1(\boldsymbol{\theta}_1; \mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \left| \hat{\mathbf{f}}(\mathbf{x}) \cdot \widehat{\nabla I_1}(\mathbf{x}; \boldsymbol{\theta}_1) \right|^2,$$

where $\hat{\mathbf{f}}$ and $\widehat{\nabla I_1}$ denote the l^2-normalized versions of the vector field and gradient, respectively. This normalization prevents the trivial solution $I_1 \equiv \text{const}$.

Once the first $K-1$ conservation laws have been learned, the K-th network is trained using a deflated loss function:

$$\mathcal{L}_K(\boldsymbol{\theta}_K; \mathcal{T}) = \frac{\ell_{\text{conserv}}[I_K] + \sum_{k=1}^{K-1} \ell_{\text{inv}}[\boldsymbol{\theta}_k^*, \boldsymbol{\theta}_K]}{\frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \left\| \text{Proj}^{\perp}_{\text{span}\{\nabla I_k(\mathbf{x}; \boldsymbol{\theta}_k^*)\}_{k=1}^{K-1}} \nabla I_K(\mathbf{x}; \boldsymbol{\theta}_K) \right\|^{2\alpha}},$$

where: $\ell_{\text{conserv}}[I_K]$ enforces that I_K is a conservation law (i.e., $\{I_K, H\} = 0$), ℓ_{inv} penalizes non-vanishing Poisson brackets with previously learned quantities, ensuring involution:

$$\ell_{\text{inv}}[\boldsymbol{\theta}_k^*, \boldsymbol{\theta}_K] = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} |\{I_k, I_K\}(\mathbf{x})|^2,$$

the denominator enforces functional independence by projecting the gradient of I_K onto the orthogonal complement of the span of the previously learned gradients, with a tunable deflation strength $\alpha > 0$.

This formulation ensures that the loss vanishes if and only if a new independent and commuting conserved quantity is found, thereby providing a consistent and interpretable stopping criterion. In practice, we monitor the validation loss $\mathcal{L}_K(\boldsymbol{\theta}_K^*; \mathcal{V})$ on a separate validation set $\mathcal{V}$. A sharp increase in this loss indicates that no further independent conservation laws exist, and the algorithm terminates.

4 METHOD: MULTI-RESOLUTION DIFFUSION GRAPH NEURAL NETWORK

Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ be a graph, where $\mathbf{V}$ denotes the set of nodes with $|\mathbf{V}| = \mathbf{N}$, and $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ is the set of edges. We denote by $\mathbf{A}$ the adjacency matrix of $\mathbf{G}$.

Definition 1 *The adjacency matrix $\mathbf{A}$ of a simple graph $\mathbf{G}$ is a symmetric matrix whose rows and columns are indexed by the same ordering of $\mathbf{V}$, with entries defined as*

$$\mathbf{A}[u, v] = \begin{cases} 1, & \textit{if } u \textit{ and } v \textit{ are adjacent}, \\ 0, & \textit{otherwise}. \end{cases}$$

Let $\mathbf{W}$ be the weight matrix of $\mathbf{G}$, also referred to as the random walk matrix or the Markov matrix. The normalized Laplacian is then defined as $\mathbf{L} = \mathbf{I}_N - \mathbf{W}$, where $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. For any signal $\mathbf{x}$ defined on the graph, the Laplacian acts as

$$\mathbf{L}\mathbf{x}(v_i) = \sum_{v_j \in \mathcal{N}_{v_i}} \mathbf{W}_{i,j} \big(\mathbf{x}(v_i) - \mathbf{x}(v_j)\big),$$

where $\mathcal{N}_{v_i}$ denotes the 1-hop neighborhood of node v_i. We further denote by $\mathcal{N}_{v_i}^k$ the set of k-hop neighbors of v_i.

In the context of energy diffusion, $\mathbf{L}$ plays the role of the graph Laplacian, and each node acts as a diffusion source. Let $\Psi = (\psi_1, \ldots, \psi_N)^\top$ represent the intensity distribution of the diffusion sources over the nodes, where ψ_i is the energy intensity at node v_i. The energy transfer between adjacent nodes follows the diffusion equation. With a diffusion coefficient $\kappa > 0$, the diffusion process on the graph $\mathbf{G}$ is governed by

$$\frac{d\Psi}{dt} = -\kappa(\mathbf{I}_N - \mathbf{W})\Psi = -\kappa\mathbf{L}\Psi.$$

Given an initial condition $\Psi(0)$, this linear system admits the unique solution

$$\Psi(t) = e^{-t\kappa\mathbf{L}}\Psi(0).$$

In the following, we fix a small constant $\kappa < 1$. Since we work with discrete-time dynamics on the graph, the continuous diffusion equation is discretized as

$$\Psi_{s+1} - \Psi_s = -\kappa(\mathbf{I}_N - \mathbf{W})\Psi_s,$$

which yields

$$\Psi_{s+1} = (1 - \kappa)\Psi_s + \kappa\mathbf{W}\Psi_s = W_\kappa \Psi_s,$$

where s denotes the discrete time step and

$$W_\kappa := (1 - \kappa)\mathbf{I}_N + \kappa\mathbf{W} = \mathbf{I}_N - \kappa\mathbf{L}. \tag{1}$$

Iterating this recurrence gives the solution after s steps:

$$\Psi_s = W_\kappa^s \, \Psi_0. \tag{2}$$

If the diffusion coefficient κ is allowed to vary at each step, the operator W_κ^s generalizes to the product form

$$W_{\boldsymbol{\kappa}}^s = \prod_{i=1}^{s} \big(\mathbf{I}_N - \kappa_i \mathbf{L}\big),$$

where $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_s)$ is a sequence of step-dependent coefficients.

This motivates the following definition of a multi-resolution energy diffusion operator.

Definition 2 *The multi-resolution energy diffusion operator is defined as*

$$W_{\boldsymbol{\kappa}}^s = \prod_{i=1}^{s} \big(\mathbf{I}_N - \kappa_i \mathbf{L}\big),$$

where s is the number of diffusion steps, $\boldsymbol{\kappa}$ is a learnable parameter. In this work, we set all graph edge weights to unity, that is, $\mathbf{L} = \mathbf{I}_N - \mathbf{A}$.

Based on this operator, we introduce a multi-resolution diffusion graph neural network. For an input node feature matrix $\mathbf{x} = (\mathbf{q}, \mathbf{p}) \in \mathbb{R}^{N\times d}$, the network computes

$$\mathbf{h}_1 = \bigoplus_{i=1}^{s} W^i_{\kappa_i}\,\mathbf{x} \in \mathbb{R}^{N\times s\cdot d},$$

where $\bigoplus$ denotes the concatenation operation along the feature dimension.

We then pass the multi-resolution node embeddings $\mathbf{h}_1$ through an encoder network, which maps the high-dimensional concatenated features to a compact per-node scalar representation:

$$\mathbf{h}_2 = \mathrm{Encoder}(\mathbf{h}_1) \in \mathbb{R}^{N\times 1}.$$

This encoder can be realized, for instance, as a MLP that aggregates information across diffusion scales for each node. Finally, a permutation-invariant aggregator—such as global sum, mean, or attention-based pooling—is applied over all nodes to fuse the node-wise features into a single scalar value, yielding the desired conserved quantity:

$$I(\mathbf{x};\boldsymbol{\theta}) = \mathrm{Aggregator}(\mathbf{h}_2) \in \mathbb{R}.$$

Here, $\boldsymbol{\theta}$ collectively denotes all trainable parameters of the diffusion operators, the encoder, and the aggregator. This parametric form of I is then used within the neural deflation framework to enforce the conservation and independence conditions.

4.1 Experimental Results

To evaluate the effectiveness of our proposed MR-DGNN in learning conserved quantities, we tested it on several benchmark lattice systems, including the integrable Toda lattice, the Calogero–Moser system, and the non-integrable FPUT and discrete sine-Gordon systems, whose specific forms are provided in the Appendix. The baselines include MLP, Convolutional Neural Network (CNN), Graph Convolutional Neural Network (GCN), and DeepSets (Zaheer et al., 2017). The network settings are as follows: the MLP uses 4 layers with 400 hidden units each; the CNN has 3 convolutional layers, 128 hidden channels, and a kernel size of 3, followed by sum pooling; the GCN has 3 convolutional layers with 200 hidden units; DeepSets consists of a ϕ network implemented as a 4-layer MLP and a ρ network implemented as a 3-layer MLP, both with 200 hidden units, followed by sum pooling; and MR-DGNN's encoder and aggregator are both implemented as 3-layer MLPs with 200 hidden units per layer. All activation functions are SiLU.

Accuracy and parameter efficiency. Fig. 1 shows the test error for learning conserved quantities with $N = 6$ particles, compared with the MLP. It can be seen that MR-DGNN achieves performance comparable to the MLP with fewer model parameters. In particular, for the Toda model, the six conserved quantities fitted by MR-DGNN are closer to the ground truth (i.e., have smaller errors) than those from the MLP. In Fig. 2, we further increase the number of particles to $N = 16$ and include comparisons with other models. As shown in the figure, CNN, GCN, and DeepSets can only learn one conserved quantity. For FPUT, MR-DGNN learns two conserved quantities with errors below 10^{-6}, whereas the MLP learns only one. For Toda, although the MLP learns 16 conserved quantities with errors around 10^{-4}, the 16 conserved quantities learned by MR-DGNN have errors around 10^{-6}, which is more accurate than the MLP, while using fewer parameters than the other baseline models. This indicates that the graph-based architecture effectively exploits the spatial locality of lattice systems, achieving superior generalization with a more compact model. As N grows, this improvement becomes increasingly pronounced, highlighting the scalability of our method to systems with large degrees of freedom.

Interpretability through diffusion weights. A key advantage of our MR-DGNN is that the learned multi-resolution diffusion operators $W^i_{\kappa_i}$ naturally reveal the spatial range of interactions encoded in each conserved quantity. By analyzing the aggregated diffusion weights $\sum_{i=1}^{s} W^i_{\kappa_i}$, we can identify which neighboring particles contribute to a given conservation law. Fig. 3, 4 and 5 display these aggregated weights for the Toda, Calogero-Moser systems, FPUT and sine-Gordon with $N = 6$, respectively.

For the Toda lattice (Fig. 3), we observe distinct interaction patterns for the six learned conserved quantities: H_1 (the Hamiltonian) mainly involves interactions up to the third nearest neighbor, consistent with the exponential coupling in the Toda potential. H_2 is dominated by second-order neighbors. H_3 is almost entirely localized to the particle itself, with negligible off-diagonal

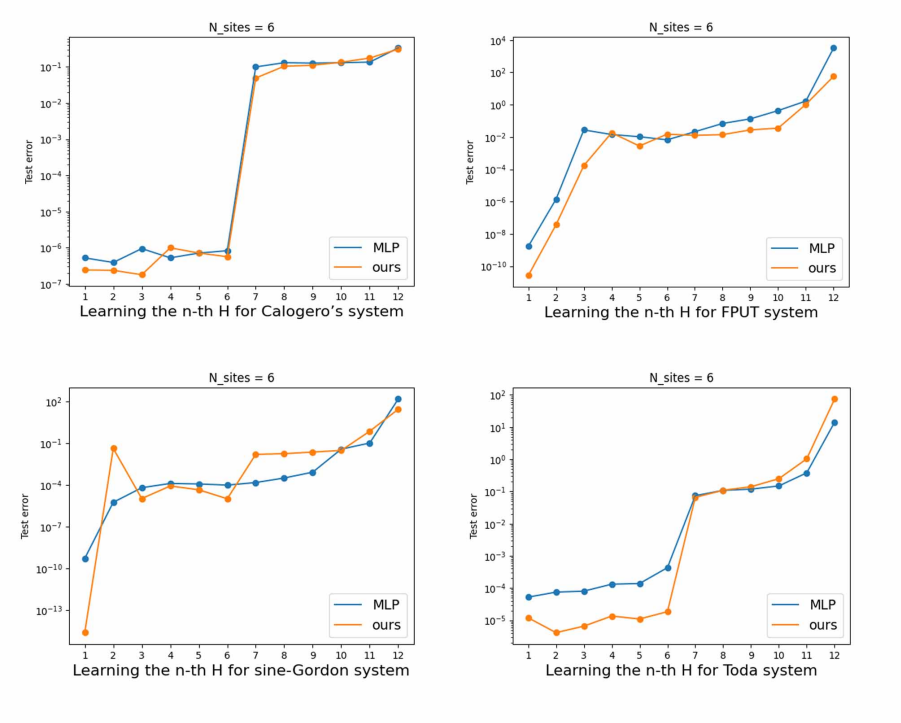

Figure 1: Comparison of test errors for learning conservation variables using MLP and MR-DGNN with $N = 6$. The MR-DGNN uses 84,817 parameters, whereas MLP uses 334,401.

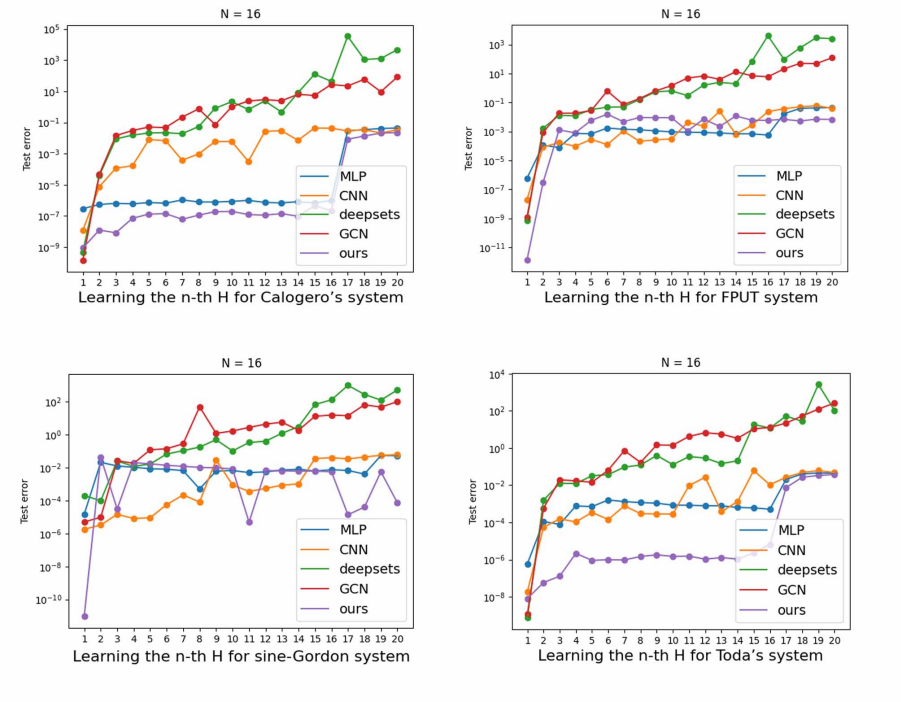

Figure 2: Comparison of test errors for learning conservation variables using MLP and MR-DGNN with $N = 16$. The MR-DGNN uses 90,922 parameters, while MLP uses 334,401, CNN uses 116,097, DeepSets uses 121,401, GCN uses 121,401.

weights—strongly suggesting that H_3 corresponds to the kinetic energy (or a function thereof). H_4 primarily involves first-order neighbors, indicating a short-range component. H_5 and H_6 extend to third- and fourth-order neighbors, respectively, capturing higher-order collective invariants.

Such fine-grained resolution is not achievable with a standard MLP, which treats all input features without spatial structure. By interpreting these weight distributions, we can infer the physical meaning of the discovered conservation laws.

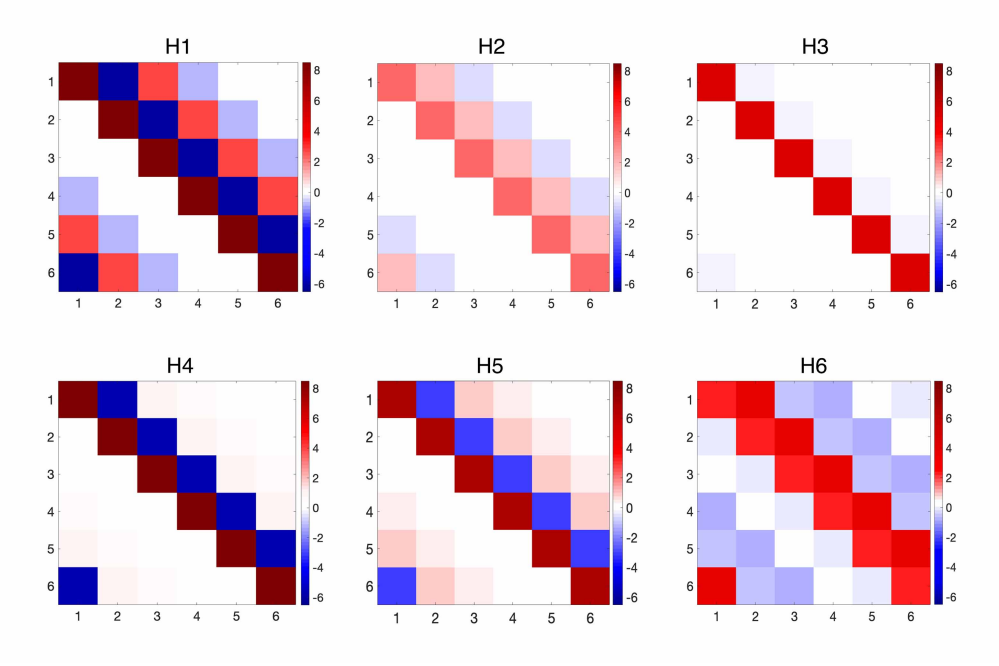

Figure 3: Aggregated diffusion weights $\sum_{i=1}^{5} W_{\kappa_i}^{i}$ for the Toda lattice with $N = 6$. The patterns indicate that H_1 depends on up to 3rd-neighbor interactions, H_2 on 2nd-neighbor, H_3 is self-localized (presumably kinetic energy), H_4 on 1st-neighbor, H_5 on 3rd-neighbor, and H_6 on 4th-neighbor.

Scaling to larger systems. We further evaluate our method on the Toda lattice with $N = 32$ particles. As shown in Fig. 6, the MR-DGNN maintains consistently low test errors across all six conserved quantities, whereas the MLP baseline suffers from severe performance degradation due to the substantial increase in parameter count and optimization difficulty. This result not only demonstrates the scalability of MR-DGNN in large-scale Hamiltonian lattices, but also highlights its ability to provide high-fidelity approximations of conserved quantities—an essential prerequisite for subsequent symbolic regression, as more accurate numerical approximations significantly enhance the success rate and reliability of recovering closed-form analytical expressions.

Overall, these results confirm that our multi-resolution diffusion graph neural network not only improves accuracy and parameter efficiency but also provides valuable physical insight into the interaction ranges of the learned conservation laws. This interpretability is particularly beneficial for guiding symbolic regression toward compact analytic expressions, thereby enabling a deeper understanding of the underlying integrable structures.

5 Conclusion and Future Work

In this paper, we have addressed the challenges of scalability and interpretability in discovering conservation laws for high-dimensional Hamiltonian lattice systems. We proposed a neural deflation method based on a MR-DGNN. This network leverages multi-scale diffusion operators on graphs to effectively capture the spatial locality and multi-scale interactions inherent in lattice systems, achieving superior approximation accuracy with a much more compact parameter count than baselines.

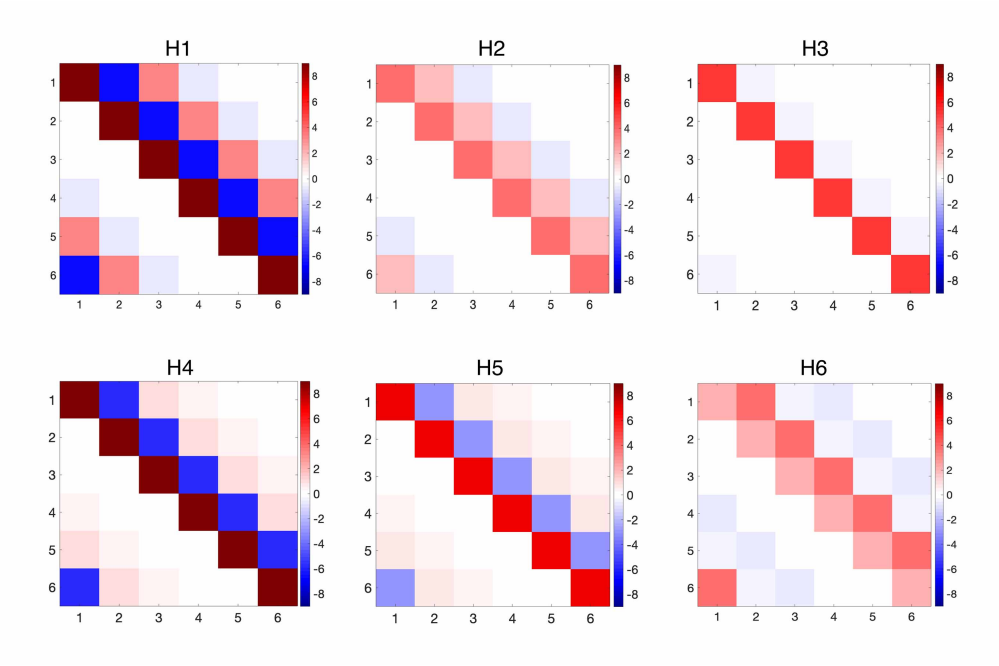

Figure 4: Aggregated diffusion weights for the Calogero-Moser system with $N = 6$, showing similar localization patterns that help identify the range of interactions for each conserved quantity.

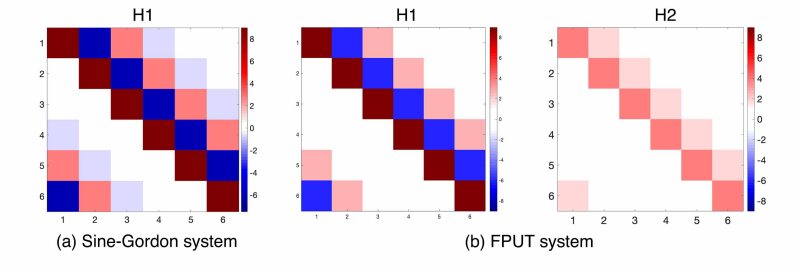

Figure 5: Aggregated diffusion weights for the sine-Gordon and FPUT systems with $N = 6$. The broad interaction ranges confirm the absence of additional local conservation laws.

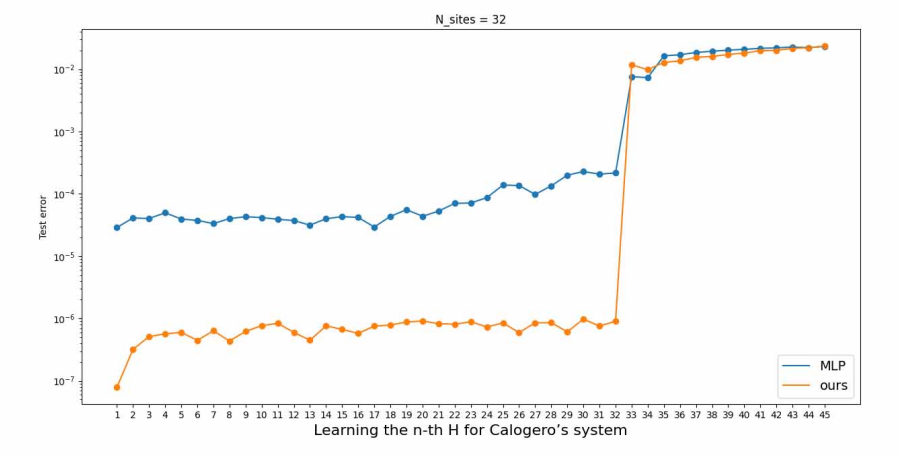

Figure 6: Test errors for learning conserved quantities in the Toda lattice with $N = 32$. The MR-DGNN continues to produce accurate approximations, whereas MLP deteriorates significantly.

We validated our method on a series of benchmark systems, including the integrable Toda lattice, Calogero-Moser system, and the non-integrable FPUT and discrete sine-Gordon systems. The experimental results demonstrate that:

Baseline models (CNN, GCN, DeepSets) can only find one conserved quantity, whereas the MLP can find more conserved quantities; however, MR-DGNN achieves lower test errors than the MLP across all tested systems, and this advantage becomes increasingly pronounced as the number of degrees of freedom grows..

By analyzing the aggregated diffusion weights, we can clearly identify the spatial interaction range corresponding to each conserved quantity. For example, in the Toda lattice, H_3 is identified as a self-localized term, likely associated with kinetic energy, while H_1 involves interactions up to third-nearest neighbors.

The method accurately captures the limited set of conserved quantities in non-integrable systems, confirming its general applicability.

Future work includes extending the method to continuum systems and more general high-dimensional dynamical systems, exploring adaptive diffusion step sizes and multi-scale designs, and integrating symbolic regression end-to-end with graph network training to enable automatic analytic derivation of conserved quantities. We believe that this direction offers a powerful new tool for data-driven studies of integrable systems.

AI USE STATEMENT

In this work, we used generative AI tools for manuscript polishing and baseline code generation. The LLM-generated code was verified and tested for correctness by the authors. We have not used generative AI tools for the proposed method or its code. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Francesco Calogero. Solution of the one-dimensional n-body problems with quadratic and/or inversely quadratic pair potentials. *Journal of Mathematical Physics*, 12(3):419–436, 1971.

Shaoxuan Chen, Panayotis G Kevrekidis, Hong-Kun Zhang, and Wei Zhu. Data-driven discovery of conservation laws from trajectories via neural deflation. *Communications in Nonlinear Science and Numerical Simulation*, 143:108563, 2025.

Giampaolo Cicogna. On the connections between symmetries and conservation rules of dynamical systems. *Mathematical Methods in the Applied Sciences*, 36(2):208–215, 2013.

Enrico Fermi, John R Pasta, Stanislaw M Ulam, and Mary Tsingou. Studies of nonlinear problems. 1955.

Giovanni Gallavotti. *The Fermi-Pasta-Ulam problem: a status report*, volume 728. Springer Science & Business Media, 2007.

Seungwoong Ha and Hawoong Jeong. Discovering invariants via machine learning. *Physical Review Research*, 3(4):L042035, 2021.

Ziming Liu and Max Tegmark. Machine learning conservation laws from trajectories. *Physical Review Letters*, 126(18):180604, 2021.

Ziming Liu, Varun Madhavan, and Max Tegmark. Machine learning conservation laws from differential equations. *Physical Review E*, 106(4):045307, 2022.

Ziming Liu, Patrick Obin Sturm, Saketh Bharadwaj, Sam J Silva, and Max Tegmark. Interpretable conservation laws as sparse invariants. *Physical Review E*, 109(2):L023301, 2024.

Peter Y Lu, Rumen Dangovski, and Marin Soljačić. Discovering conservation laws using optimal transport and manifold learning. *Nature Communications*, 14(1):4744, 2023.

Meskerem Abebaw Mebratie, Rüdiger Nather, Guido Falk von Rudorff, and Werner M Seiler. Machine learning conservation laws of dynamical systems. *Physical Review E*, 111(2):025305, 2025.

Jürgern Moser. Three integrable hamiltonian systems connected with isospectral deformations. In *Surveys in applied mathematics*, pp. 235–258. Elsevier, 1976.

Amaury Mouchet. Applications of noether conservation theorem to hamiltonian systems. *Annals of Physics*, 372:260–282, 2016.

Emmy Noether. Invariante variationsprobleme. In *Gesammelte abhandlungen-collected papers*, pp. 231–239. Springer, 1983.

Morikazu Toda. *Theory of nonlinear lattices*. Springer Science & Business Media, 2012.

Sebastian J Wetzel, Roger G Melko, Joseph Scott, Maysum Panju, and Vijay Ganesh. Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks. *Physical Review Research*, 2(3):033499, 2020.

Floyd Williams. *Sine-gordon Model and Its Applications-from Pendula and Josephson Junctio*. Springer International Publish, 2014.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

Wang Zhang, Tsui-Wei Weng, Subhro Das, Alexandre Megretski, Luca Daniel, and Lam M Nguyen. Concernet: a contrastive learning based framework for automated conservation law discovery and trustworthy dynamical system prediction. In *International Conference on Machine Learning*, pp. 41694–41714. PMLR, 2023.

Wei Zhu, Hong-Kun Zhang, and PG Kevrekidis. Machine learning of independent conservation laws through neural deflation. *Physical Review E*, 108(2):L022301, 2023.

A Appendix

In this appendix, we provide detailed descriptions of the four lattice systems used in our numerical experiments. All systems are Hamiltonian lattices with N particles and periodic boundary conditions, i.e., $q_{N+1} = q_1$ and $p_{N+1} = p_1$. The Hamiltonian takes the general form

$$H(\mathbf{q}, \mathbf{p}) = T(\mathbf{p}) + V(\mathbf{q}), \qquad T(\mathbf{p}) = \sum_{n=1}^{N} \frac{p_n^2}{2},$$

with different choices of the potential $V(\mathbf{q})$ giving rise to integrable or non-integrable dynamics.

A.1 Discrete Sine-Gordon System

The discrete sine-Gordon system models a one-dimensional chain of particles coupled by nonlinear springs, with an onsite periodic potential. Its Hamiltonian is given by

$$H = \sum_{n=1}^{N} \left[\frac{p_n^2}{2} + \frac{C}{2}(q_{n+1} - q_n)^2 + 1 - \cos q_n \right],$$

where $C > 0$ is the coupling strength. This system is non-integrable for generic parameters; it possesses only one independent conserved quantity, namely the total energy (the Hamiltonian itself). No additional local conservation laws exist.

A.2 CALOGERO-MOSER SYSTEM

The Calogero-Moser system describes particles interacting via a pairwise inverse-square potential. The potential energy is

$$V(\mathbf{q}) = \sum_{1 \le l < k \le N} \frac{1}{(q_k - q_l)^2}.$$

This model is completely integrable in the Liouville sense, possessing N functionally independent and Poisson-commuting conserved quantities. These invariants are related to the underlying symmetric structure of the system and can be explicitly constructed via Lax pair formulations.

A.3 TODA LATTICE

The Toda lattice is a classic integrable model of a nonlinear one-dimensional crystal, with exponential nearest-neighbor interactions. Its potential is

$$V(\mathbf{q}) = \sum_{n=1}^{N} \left[e^{q_n - q_{n-1}} + (q_{n+1} - q_n) - 1 \right],$$

where we adopt the convention $q_0 = q_N$ under periodic boundaries. The Toda lattice is fully integrable and admits exactly N independent conserved quantities in involution.

A.4 FERMI–PASTA–ULAM–TSINGOU SYSTEM

The FPUT system is historically significant as one of the first numerical experiments exploring nonlinear energy equipartition. Its potential is given by

$$V(\mathbf{q}) = \sum_{n=1}^{N} \left[\frac{1}{2}(q_n - q_{n+1})^2 - \frac{1}{6}(q_{n+1} - q_n)^3 \right].$$

Despite its close resemblance to the Toda lattice (both have nearest-neighbor interactions), the FPUT system is non-integrable. It conserves only two independent quantities: the total energy H and the total momentum $P = \sum_{n=1}^{N} p_n$. No other local invariants exist, and the system exhibits chaotic behavior for generic initial conditions.

The system properties are summarized in Table 1.

Table 1: Summary of properties of the four benchmark lattice systems.

System	**Integrability**	**Number of conserved quantities**
Discrete sine-Gordon	Non-integrable	1 (energy)
Calogero–Moser	Integrable	N
Toda lattice	Integrable	N
FPUT	Non-integrable	2 (energy and momentum)