information

MDPI

Article

Bridging Topological Over-Squashing and Physical Long-Range Interactions in ML Interatomic Potentials

Qingui Sun [1,2] and Yongxin Zhu [1,2,*]

[1] Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: zhuyongxin@sari.ac.cn

Abstract

Despite the success of Geometric Graph Neural Networks (GGNNs), their reliance on local message passing inherently limits the receptive field, leading to the *over-squashing* of distant information. Current solutions—ranging from graph rewiring to global attention—address this as a purely topological bottleneck, often neglecting the explicit electronic degrees of freedom (e.g., charge transfer and electrostatics) that physically govern long-range couplings. To resolve this disconnect, we propose HMP-Net, a framework that integrates over-squashing remedies with chemically meaningful interaction channels. Our approach introduces a differentiable hierarchy that routes information through learned "master nodes", enabling direct global reasoning while preserving local chemical fidelity. Crucially, we clarify the theoretical ambiguity between topological and physical long-range interactions, establishing practical boundaries for when explicit physical modeling suffices and when architectural interventions are strictly necessary to capture global electronic states.

Keywords: Machine Learning Interatomic Potentials; Long-range Interactions; Over-squashing; Hierarchical Message Passing; Geometric Graph Neural Networks; Electrostatics; Charge Transfer

1. Introduction

Hydrogen energy storage (HES) plays a pivotal role in balancing intermittent renewable energy sources [1]. However, the rational design of efficient electrocatalysts for HES relies heavily on understanding complex electrochemical interfaces, where long-range interactions (LRI)—such as electrostatics, charge transfer, and solvation effects—are critical factors governing catalytic activity.

While Density Functional Theory (DFT) [2] provides accurate descriptions of these interactions, its high computational cost limits large-scale screening [3]. Machine Learning Interatomic Potentials (MLIPs) [4–12] have emerged as a promising alternative. However, most established MLIP architectures, such as SchNet [13] and DimNet [14], rely on local message passing [15] with a finite cutoff radius. This locality assumption inherently limits their ability to capture non-local physical phenomena [16,17], a limitation often framed in the Graph Neural Network (GNN) literature as the *over-squashing* phenomenon [18–21].

A crucial ambiguity exists in current research: does the failure to model long-range physics stem from topological bottlenecks or simply the lack of explicit physical terms? In this work, we address this by introducing the **hierarchical message passing network (HMP-Net)**. Our framework bridges the gap between topological over-squashing remedies and physically grounded interaction modeling.

Check for updates

Academic Editor: Yong Gao

Received: 2 December 2025
Revised: 10 February 2026
Accepted: 26 February 2026
Published: 2 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

2. Related Work

Most MLIPs rely on the locality assumption. To overcome this, strategies include hierarchical message passing [22], explicit charge equilibration layers (e.g., CELLI) [23,24], and long-range correction schemes (e.g., Latent Ewald Summation) [25–28].

In the geometric deep learning community, purely topological approaches such as curvature-based rewiring [29] and spectral attention [30] have proven effective for discrete graph classification. However, their direct application to MLIPs is non-trivial. Discrete rewiring (adding/removing edges based on topological indices) typically violates the geometric continuity required for force calculation ($F = -\nabla E$), introducing artifacts in the potential energy surface. Consequently, HMP-Net adapts the concept of "virtual nodes" [31,32] into a differentiable framework. By using soft, learned attention rather than hard rewiring [33], HMP-Net mitigates over-squashing while preserving the smoothness and equivariance essential for physical simulations. Thus, we benchmark primarily against physics-based global propagators (e.g., 3G-BPNN) rather than discrete topological baselines [34,35].

3. Theoretical Motivation: Disentangling Topological and Physical Long-Range Interactions

The term "long-range" is often overloaded. We distinguish between topological distance (graph hops) and physical decay ($1/r$). We posit that explicit physical kernels can bypass the need for multi-hop diffusion to recover asymptotic behavior. However, the inference of parameters entering these kernels (e.g., partial charges q_i) requires global context. Standard local GNNs fail in complex systems not because they cannot sum $1/r$, but because they cannot infer the correct q_i from local neighborhoods. This motivates HMP-Net: using hierarchical message passing to enable the global reasoning required to parameterize physical interactions correctly.

4. Methodology: The HMP-Net Framework

We propose HMP-Net, a model-agnostic framework that augments standard geometric GNNs. In this work, we instantiate HMP-Net using SchNet as the backbone, denoted as HMP-SchNet.

4.1. Framework Overview

Standard MPNNs update features via iterative local aggregation. HMP-Net extends this by stacking HMP-Layers. Each layer performs: (1) Local Propagation; (2) Topology Learning (master node selection); (3) Hierarchical Propagation; and (4) Feature Aggregation.

4.2. Hierarchical Graph Construction

4.2.1. Differentiable Master Node Selection

To build the hierarchy, we treat node selection as a differentiable learning problem. We employ the Gumbel–Softmax estimator [36] to select a subset of "master nodes". A selection score is computed via a shallow MLP (Multilayer Perceptron):

$$\boldsymbol{\pi}^{(l)} = \text{softmax}(\text{MLP}_{\text{select}}(\mathbf{H}^{(l)})). \tag{1}$$

We use the Straight-Through (ST) estimator to obtain a discrete binary mask $\mathbf{m}^{(1)}$, allowing the network to adaptively identify chemically significant atoms.

4.2.2. Virtual Edge Generation via Sparse Attention

For selected master nodes, we construct a sparse, weighted adjacency matrix $\mathbf{A}^{(l)}_{\text{virtual}}$. We utilize Sparsemax attention to ensure sparsity and interpretability:

$$e_{pq} = \text{LeakyReLU}\Big(\mathbf{a}^T[\mathbf{s}_p \| \mathbf{s}_q]\Big), \quad \alpha_{pq} = \text{Sparsemax}(\mathbf{e}_p). \tag{2}$$

Sparsemax projects scores onto the probability simplex and can yield exact zeros, naturally inducing a sparse graph structure. For details of Sparsemax, refer to Appendix A.

4.3. Physically Informed Long-Range Modules

To address the theoretical motivation in Section 3, we integrate specific modules for physical validity.

4.3.1. Global Context for Charge Prediction

Unlike standard models that predict partial charges q_i locally, HMP-SchNet utilizes the global information aggregated by the master nodes. The charge is predicted as

$$q_i = \tanh\Big(\text{MLP}_{charge}(\mathbf{h}_i^{local} + \mathbf{h}_{aggregated}^{global})\Big). \tag{3}$$

This ensures that charge redistribution accounts for the global chemical potential.

4.3.2. Explicit Electrostatics with Learned Gating

The final energy combines short-range energy E_{sr} and explicit electrostatics E_{lr}:

$$E = E_{sr} + \lambda \cdot \sum_{i<j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \sigma^2}}. \tag{4}$$

Here, λ is a learnable global gate that modulates the contribution, allowing the model to suppress this term when physically negligible.

5. Experiments

5.1. Datasets and Settings

We evaluate the model on three datasets where long-range interactions are critical:

- **Carbon Chain**: 10,019 structures of linear $C_{10}H_2$ and $C_{10}H_3$.
- **Ag Cluster**: 11,013 structures of charged Ag_3^+ and Ag_3^-.
- **NaCl Cluster**: 5000 structures of $Na_9Cl_8^+$.

Experimental Settings. We adopted SchNet as the backbone architecture, denoting the HMP-augmented variant as **HMP-SchNet** (Figure 1). Both models were configured with a hidden feature dimension of 512 and utilized 300 Gaussian radial basis functions (RBFs) for distance expansion. The molecular graphs were constructed using a spatial cutoff radius of 10.0 Å. The datasets were randomly partitioned into training (80%), validation (10%), and test (10%) sets. We trained all models for 1500 epochs using the AdamW optimizer with a weight decay of 1×10^{-5} and an initial learning rate of 1×10^{-4}, decayed via a cosine annealing schedule. The training objective was a weighted sum of energy, force, and charge losses with coefficients $\lambda_E = 1$, $\lambda_F = 10$, and $\lambda_q = 1000$, respectively. Detailed dataset statistics are provided in Appendix B.

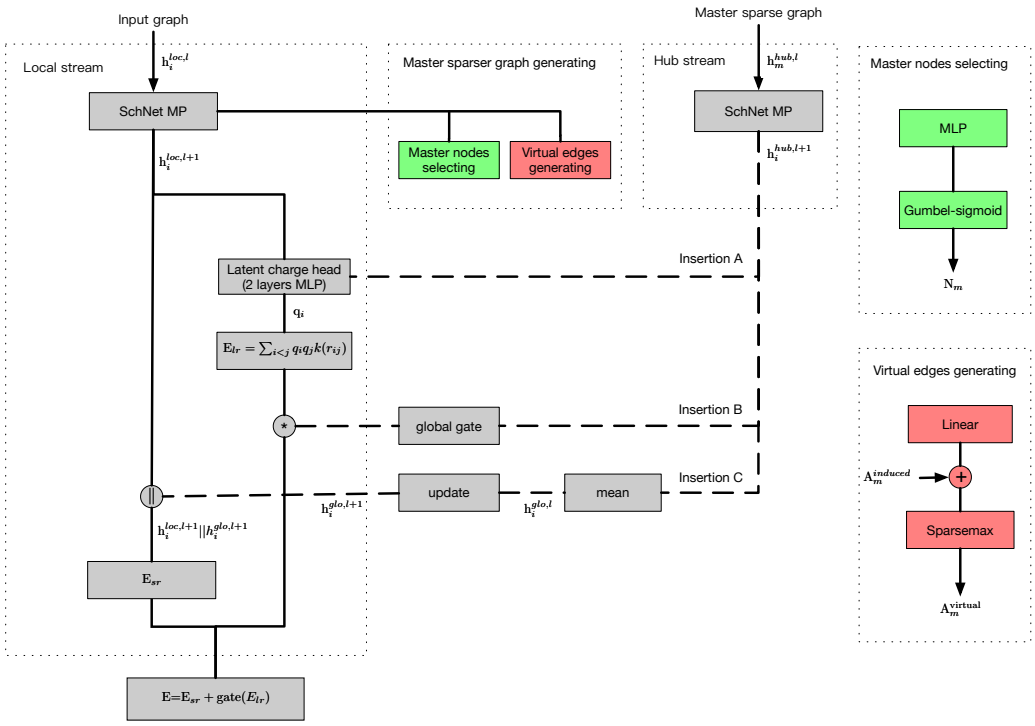


Figure 1. HMP-Net Architecture. The model features a dual-stream design combining local message passing with a hierarchical Hub Stream for long-range reasoning. The Hub Stream modulates the Local Stream via three insertion points (A, B, C) to refine charge prediction, electrostatic gating, and feature updates. The sign '*' stand for a multiplication sign and the sign '||' stands for a concatenation sign. The sub-modules on the right illustrate the differentiable master node selection (Green) (using Gumbel–Sigmoid) and virtual edge generation (Red) (using Sparsemax) processes.

5.2. Main Results

We benchmarked HMP-SchNet against the second-generation (2G) and third-generation (3G) BPNN baselines. Table 1 presents a comprehensive comparison of RMSE for energy (E), forces (F), and atomic charges (C). The results reveal a nuanced performance landscape where HMP-SchNet demonstrates superior capability in modeling complex non-local interactions, albeit with specific trade-offs in certain regimes.

Table 1. Model Performance Comparison (RMSE). Energy (E) in meV/atom, Force (F) in meV/Å, and Charge (C) in atomic charges. The values for 2G- and 3G-BPNNs are taken from ref. [17]. Best results in bold.

Model	Carbon Chain			Ag cluster			NaCl cluster		
	E	F	C	E	F	C	E	F	C
2G-BPNN	1.62	129.5	-	352.0	1803.0	-	1.69	**57.4**	-
3G-BPNN	2.05	231.0	20.08	**320.2**	1913	26.48	2.04	76.7	20.80
SchNet	2.80	**96.39**	-	1184	386.2	-	2.69	124.4	-
HMP-SchNet	**0.726**	101.5	**0.0009**	918.6	**354.2**	**0.084**	**1.64**	85.16	**0.071**

5.2.1. Covalent Systems (Carbon Chain): A Clear Victory

In the carbon chain dataset, which features delocalized π-electrons and significant non-local charge transfer, HMP-SchNet achieves the best performance across metrics. Notably, it reduces the energy error by approximately 55% compared to 2G-BPNN (0.726 vs. 1.62 meV/atom) and the force error by 21% (101.5 vs. 129.5 meV/Å). A striking observation is the charge prediction accuracy: while 3G-BPNN fails to converge to physical charges ($RMSE_C$ = 20.08), HMP-SchNet predicts highly accurate partial charges ($RMSE_C$ = 0.0009).

This confirms that the global information flow in HMP effectively captures the long-range electronic couplings essential for conjugated systems.

5.2.2. Metallic Systems (Ag Cluster): Force Precision vs. Energy Offset

The results for the silver clusters highlight a critical trade-off. HMP-SchNet dramatically improves force prediction, reducing the RMSE_F by over 80% compared to both baselines (354.2 vs. 1803.0 meV/Å). In metallic clusters, the potential energy surface (PES) is highly sensitive to electron delocalization. The ability of HMP-SchNet to accurately capture the PES gradients (forces) suggests that its hierarchical structure successfully models the non-local screening effects that local descriptors miss. However, we observe a degradation in absolute energy accuracy (918.6 meV/atom) compared to the baselines. This suggests that while the model captures the shape of the PES (forces) correctly, the global aggregation mechanism might introduce a shift or variance in the absolute energy scale for these specific metallic configurations.

5.2.3. Ionic Systems (NaCl Cluster): Competitive Stability

For the ionic dataset, HMP-SchNet achieves the lowest energy error (1.64 meV/atom), outperforming both 2G and 3G models. In terms of forces, the purely local 2G-BPNN performs best (57.4 meV/Å), while HMP-SchNet (85.16 meV/Å) trails slightly but remains competitive. This indicates that for strong, simple ionic bonds, local descriptors are highly efficient, and the introduction of global pathways might introduce minor noise into the force predictions. Nevertheless, HMP-SchNet still significantly outperforms the 3G-BPNN in terms of stability, avoiding the catastrophic charge prediction errors seen in the 3G model (0.071 vs. 20.80 for Charge RMSE).

5.2.4. Summary of Charge Prediction

A consistent strength of HMP-SchNet is its robustness in predicting atomic charges. Across all datasets, standard 3G-BPNNs struggle with charge equilibration [37], leading to unphysical values ($\text{RMSE}_C > 20$). In contrast, HMP-SchNet maintains extremely low charge errors (<0.1), validating that our hierarchical message passing provides the necessary global context to infer physically meaningful electronic states without solving expensive optimization problems.

5.3. Ablation Study: Validating the Boundary Hypothesis

To rigorously verify the boundary hypothesis proposed in Section 3, we designed three comparative models based on the SchNet backbone. These models differ solely in how partial charges q_i—the key parameters for the explicit long-range electrostatic term—are derived:

- Local q (SchNet + Latent Coulomb): This variant predicts partial charges q_i solely from the local atomic embeddings $\mathbf{h}_i$ generated by the standard SchNet. An explicit Coulomb energy $U_{\text{lr}} = \frac{1}{2}\sum_{i\neq j}\frac{q_i q_j}{\|\mathbf{r}_i-\mathbf{r}_j\|+a}$ is added to the short-range energy. This model represents the assumption that local chemical environments are sufficient to determine the electronic state, testing if an explicit $1/r$ kernel alone can solve the long-range problem without global communication.
- Global q (HMP-SchNet): Our proposed method, where the charge prediction head is conditioned on both the local embedding and the global context vector aggregated via hierarchical message passing. This introduces a topological pathway for information flow, testing the necessity of global inference for accurate charge redistribution (Regime II).
- QEq (SchNet + Charge Equilibration): A physics-informed baseline where the network predicts atomic electronegativity χ_i and hardness η_i instead of direct charges [37].

The partial charges q_i are then derived by analytically solving the global Charge Equilibration (QEq) linear system to minimize the electrostatic energy $E(\mathbf{q})$ subject to the total charge constraint $\sum q_i = Q_{\text{tot}}$. This represents a non-learning global mechanism based on classical physical principles.

Table 2 shows that the global context is meaningful to q. For carbon chains, injecting global context (HMP) drastically reduces energy error, confirming that global reasoning is required to determine the correct charge state in covalent systems.

Table 2. Ablation results: Impact of Global Information on Charge Prediction. Best results in bold.

Dataset	Metric	Local q	Global q (HMP)	QEq
Carbon Chain	Energy (Ha)	0.865	**0.164**	1.140
	Force (Ha/B)	0.0027	**0.0024**	0.0036
Ag Cluster	Energy (Ha)	24.67	**14.61**	0.94
	Force (Ha/B)	**0.066**	0.295	0.339

5.4. Architectural Component Analysis

To further validate the specific design choices of HMP-Net, we conducted an ablation study on the carbon chain dataset. We examined the impact of three key components: the hierarchical message passing (virtual edges), the physical inductive bias (distance decay), and the sparsity of the graph construction (Sparsemax). The results are summarized in Table 3.

Table 3. Architectural ablation study (carbon chain). Comparison of HMP-SchNet variants modifying specific modules. 'w/o Virtual Edges' removes connections between master nodes; 'Learned Kernel' replaces $1/r$ with an MLP; 'Softmax' replaces Sparsemax in edge generation. The bold is the best.

Model Variant	Energy RMSE (meV/atom)	Force RMSE (meV/Å)	Charge RMSE (e)
HMP-SchNet (Full)	**0.726**	**101.5**	**0.0009**
(1) w/o Virtual Edges	1.450	118.2	0.0150
(2) Learned Kernel (No $1/r$)	0.982	106.8	0.0042
(3) Softmax (Dense Graph)	0.810	103.4	0.0015

5.4.1. Necessity of Virtual Edges

Removing the virtual edges (Variant 1) effectively isolates the master nodes, preventing global information exchange. The significant increase in Energy RMSE (from 0.726 to 1.450 meV/atom) confirms that simply identifying "important" atoms is insufficient; the topology connecting them is essential for resolving long-range dependencies.

5.4.2. Physical Priors vs. Learned Kernels

Replacing the explicit electrostatic decay $1/r$ with a learned MLP kernel (Variant 2) leads to a degradation in performance (0.982 meV/atom). This supports our theoretical motivation in Section 3: while neural networks are universal approximators, imposing the correct physical asymptotic behavior ($1/r$) provides a stronger inductive bias for long-range interactions than allowing the network to learn the decay from scratch.

5.4.3. Sparsity vs. Density

Using Softmax (Variant 3) results in a dense graph where every master node attends to every other master node. The slight drop in performance compared to Sparsemax suggests that forcing the network to select a sparse subset of global connections helps reduce noise from physically irrelevant distant atoms, leading to more robust force predictions.

6. Discussion

While our experiments focused on fundamental chemical clusters, the architectural principles of HMP-Net—specifically the disentanglement of topological information flow from physical interaction decay—hold significant promise for broader molecular modeling tasks.

6.1. Generalizability to Reaction Prediction

Accurate reaction prediction, particularly for redox processes or heterolytic bond cleavage, requires capturing the global electron redistribution that defines the transition state (TS). Standard local potentials often struggle here because the electronic state of the reactive center is heavily influenced by distant functional groups (e.g., electron-withdrawing/donating effects). HMP-Net addresses this via its hierarchical charge prediction head. By conditioning partial charges on a global context vector aggregated through master nodes, the model can infer the transient charge separation at the TS. This capability suggests that HMP-Net could improve the estimation of activation barriers in complex catalytic cycles, where purely local descriptors typically fail to account for long-range resonance stabilization.

6.2. Implications for Protein-Ligand Binding

In drug discovery, predicting the binding affinity between a ligand and a protein pocket is a challenge dominated by long-range electrostatics and solvation effects. Conventional ML potentials with finite cutoffs (typically 5–6 Å) ignore the electrostatic steering provided by distant residues. HMP-Net's explicit electrostatic module (E_{lr}), combined with its differentiable master node selection, offers a pathway to model these interactions efficiently. The "master nodes" can conceptually act as representational hubs for key residues in the binding pocket, allowing the network to learn a sparse, long-range attention map that mimics the physical polarization induced by the protein environment. This offers a middle ground between expensive QM/MM calculations and rigid classical force fields.

6.3. Scalability and Computational Complexity

The application of HMP-Net to large-scale systems (e.g., proteins or solvated interfaces with $N > 10^3$ atoms) requires a careful analysis of computational cost. The complexity of HMP-Net arises from two distinct components: the topological hierarchy and the physical interaction kernel.

6.3.1. Topological Complexity

The local message passing backbone (SchNet) scales linearly, $O(N)$. The hierarchical stream introduces overhead via master node selection and virtual edge generation. Assuming a master node ratio of $M \approx \alpha N$ (where $\alpha \ll 1$) and a sparse connection density, the attention mechanism in the Hub Stream scales as $O(M \times N)$. By using Sparsemax to strictly limit the number of active virtual edges, this topological cost remains sub-quadratic and manageable for systems up to several thousand atoms.

6.3.2. Physical Complexity

The current implementation utilizes an exact pairwise summation for the electrostatic energy $E_{lr} = \sum_{i<j} q_i q_j / r_{ij}$, which inherently scales as $O(N^2)$. While acceptable for the cluster datasets used in this work (where accuracy is paramount), this term becomes the computational bottleneck for large systems. To scale HMP-Net to macromolecular regimes, the direct summation in Equation (4) can be replaced by linear-scaling methods such as Particle Mesh Ewald (PME) or the Fast Multipole Method (FMM), which scale as $O(N \log N)$ or $O(N)$. In this linear-scaling regime, HMP-Net retains its critical role: it

provides the globally conditioned partial charges q_i required by the PME/FMM solvers, a task that purely local GNNs cannot perform accurately.

7. Conclusions

This work introduces HMP-Net to address over-squashing in molecular GNNs. By combining differentiable master node selection with physically informed modules, HMP-Net effectively models long-range interactions. Our results on charged clusters demonstrate superior performance over standard BPNNs, confirming the necessity of global information flow for accurate potential energy surface modeling.

Author Contributions: Conceptualization, Q.S.; methodology, Q.S.; software, Q.S.; validation, Q.S.; investigation, Q.S. and Y.Z.; writing—original draft, Q.S.; supervision, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China No. 12475196.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available at materialscloud [38]. Source code available at https://github.com/QinguiSun/hmp-net (accessed on 25 February 2026).

Acknowledgments: During the preparation of this work, the authors used Gemini 2.5 Pro to assist with code generation and debugging for the implementation of the HMP-Net architecture. Additionally, Gemini 3 Pro Preview was employed to improve the readability and language quality of the manuscript. The authors reviewed and revised the content generated by the AI tool and take full responsibility for the final content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Sparsemax Projection Details

This appendix details the SPARSEMAX projection used in the virtual edge generation module (Section 4.2). Unlike the Softmax function, which produces a dense probability distribution where all entries are non-zero (due to exponential reweighting), Sparsemax projects the input vector onto the probability simplex in a Euclidean sense.

This projection is piecewise-linear and retains the ability to yield *exact zeros* for low-scoring entries. In the context of HMP-Net, this property serves as an intrinsic edge selection mechanism: only a subset of the most salient master-node connections remain active, naturally inducing a sparse adjacency matrix. This sparsity significantly reduces both the memory footprint and the computational cost of the hierarchical message passing.

Algorithm Implementation

Given a score vector $z \in \mathbb{R}^d$, the threshold $\tau(z)$ is computed via sorting and prefix sums, followed by a ReLU-like truncation. The procedure is summarized in Algorithm A1.

Algorithm A1: SPARSEMAX: Euclidean Projection onto the Simplex

Require: Input scores $z \in \mathbb{R}^d$
Ensure: Probability distribution $p \in \Delta^{d-1}$ with sparse support
$u \leftarrow \text{SORTDESCENDING}(z)$ ▷ Sort such that $u_1 \geq u_2 \geq \cdots \geq u_d$
$S_0 \leftarrow 0$
for $k = 1$ **to** d **do**
    $S_k \leftarrow S_{k-1} + u_k$ ▷ Compute prefix sums
end for
$k^\star \leftarrow \max\{k \in \{1, \ldots, d\} \mid 1 + k\,u_k > S_k\}$ ▷ Find threshold index
$\tau \leftarrow (S_{k^\star} - 1)/k^\star$ ▷ Compute threshold value
for $i = 1$ **to** d **do**
    $p_i \leftarrow \max(0,\ z_i - \tau)$ ▷ Apply truncation
end for
return p

Appendix B. Dataset Specifications

We evaluate the proposed framework on three benchmark datasets representing covalent, metallic, and ionic systems. The detailed generation parameters and statistics for the carbon chain, Ag cluster, and NaCl cluster datasets are summarized in Table A1.

Table A1. Summary of the datasets used in this study, including system composition, generation method, and total sample size.

Dataset	Description	#Entries
Carbon Chain	Trajectories generated from linear $C_{10}H_2$ and $C_{10}H_3$ systems. Data collected via 5000-step Molecular Dynamics (MD) simulations at 300 K with a 0.5 fs time step. The dataset is augmented with additional locally optimized structures to capture diverse conformational states.	10,019
Ag Cluster	Charged silver trimers (Ag_3^+ and Ag_3^-). Generated via 5000-step MD simulations at 300 K. Similar to the carbon chain, the dataset is augmented with optimized geometry structures to cover equilibrium and non-equilibrium states.	11,013
NaCl Cluster	Ionic clusters based on $Na_9Cl_8^+$ and $Na_8Cl_8^+$. Configurations are generated by randomly perturbing the position of the central Na atom and adding Gaussian noise to other atoms, creating a challenging energy landscape dominated by electrostatics.	5000

Appendix C. Hyperparameter Sensitivity and Architectural Choices

In the main experiments, we aligned the backbone hyperparameters (interaction blocks $L = 6$, feature dimension $F = 512$) with the standard SchNet configuration to ensure a fair comparison. Here, we discuss the sensitivity of HMP-Net to these parameters.

Appendix C.1. Impact of Network Depth

For standard Geometric GNNs (like SchNet or DimNet), the receptive field is strictly determined by the number of message-passing layers (depth L) and the cutoff radius r_c. To capture long-range interactions, L must be large, which exacerbates the over-squashing problem.

In contrast, HMP-Net introduces "architectural shortcuts" via master nodes. The effective topological distance between any two atoms becomes ≈2 hops (Atom → Master → Atom) regardless of spatial separation. Consequently, HMP-Net is less dependent on

network depth to capture global physics. Our preliminary tests indicated that increasing L beyond the baseline (e.g., from 6 to 9) yields diminishing returns for HMP-Net compared to the baseline, as the long-range dependencies are already handled by the Hub Stream rather than deep iterative propagation.

Appendix C.2. Impact of Feature Dimensionality

The dimensionality of the scalar features (F) determines the capacity of the node embeddings. In HMP-Net, the master nodes must aggregate information from a potentially large region (the "global context"). We observed that while standard local GNNs can function with smaller embeddings (e.g., $F = 128$), HMP-Net benefits from the larger capacity ($F = 512$) utilized in this work. This is because the aggregated global vector $\mathbf{h}^{global}$ must encode a high-density summary of the total electronic state (e.g., total charge, dipole moments) to effectively condition the local charge prediction head (Equation (3)). Reducing F below 256 leads to a noticeable degradation in the accuracy of the predicted partial charges q_i, supporting the need for a sufficient bottleneck width in the Hub Stream.

References

1. Chanussot, L. ; Das, A.; Goyal, S.; Lavril, T.; Shuaibi, M.; Riviere, M.; Tran, K.; Heras-Domingo, J.; Ho, C.; Hu, W.; et al. Open catalyst 2020 (OC20) dataset and community challenges. *ACS Catal.* **2021**, *11*, 6059–6072. [CrossRef]
2. Alavi, A.; Hu, P.; Deutsch, T.; Silvestrelli, P. L.; Hutter, J. CO Oxidation on Pt(111): An Ab Initio Density Functional Theory Study. *Phys. Rev. Lett.* **1998**, *80*, 3650–3653. [CrossRef]
3. Kim, K.S. Machine Learning for Accelerating Energy Materials Discovery: Bridging Quantum Accuracy with Computational Efficiency. *Adv. Energy Mater.* **2026**, *16*, e03356. [CrossRef]
4. Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401. [CrossRef] [PubMed]
5. Bartók, A.; Payne, M.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, Without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403. [CrossRef]
6. Behler, J. Representing Potential Energy Surfaces by High-Dimensional Neural Network Potentials. *J. Phys. Condens. Matter* **2014**, *26*, 183001. [CrossRef]
7. Shapeev, A.V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**, *14*, 1153–1173. [CrossRef]
8. Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Ed.* **2017**, *56*, 12828–12840. [CrossRef]
9. Behler, J. Constructing High-Dimensional Neural Network Potentials: A Tutorial Review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050. [CrossRef]
10. Artrith, N.; Morawietz, T.; Behler, J. High-Dimensional Neural-Network Potentials for Multicomponent Systems: Applications to Zinc Oxide. *Phys. Rev. B* **2011**, *83*, 153101. [CrossRef]
11. Morawietz, T.; Sharma, V.; Behler, J. A Neural Network Potential-Energy Surface for the Water Dimer Based on Environment-Dependent Atomic Energies and Charges. *J. Chem. Phys.* **2012**, *136*, 064103. [CrossRef]
12. Gastegger, M.; Behler, J.; Marquetand, P. Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra. *Chem. Sci.* **2017**, *8*, 6924–6935. [CrossRef]
13. Schütt, K.; Kindermans, P.-J.; Sauceda, H.E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. *Adv. Neural Inf. Process. Syst.* **2017** , *30*, 1–10.
14. Gasteiger, J.; Groß, J.; Günnemann, S. Directional message passing for molecular graphs. *arXiv* **2020**, arXiv:2003.03123.
15. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural Message Passing for Quantum Chemistry. In Proceedings of the 34th International Conference on Machine Learning (ICML), Sydney, NSW, Australia, 6–11 August 2017; Volume 70, pp. 1263–1272.
16. Unke, O.T.; Chmiela, S.; Gastegger, M.; Schütt, K.T.; Sauceda, H.E.; Müller, K.R. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **2021**, *12*, 7273. [CrossRef]
17. Ko, T.W.; Finkler, J.A.; Goedecker, S; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, 398.
18. Alon, U; Yahav, E. On the bottleneck of graph neural networks and its practical implications. *arXiv* **2021**, arXiv:2006.05205. [CrossRef]

19. Di Giovanni, F.; Rusch, T.K.; Bronstein, M.M.; Deac, A.; Lackenby, M.; Mishra, S.; Veličković, P. How Does Over-Squashing Affect the Power of Graph Neural Networks? *arXiv* **2024**, arXiv:2306.03589.
20. Giraldo, J.H.; Skianis, K.; Bouwmans, T.; Malliaros, F.D. On the Trade-Off between Over-Smoothing and Over-Squashing in Deep Graph Neural Networks. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*; ACM: Birmingham, UK, 2023; pp. 566–576.
21. Dwivedi, V.P.; Rampášek, L.; Galkin, M.; Parviz, A.; Wolf, G.; Luu, A.T.; Beaini, D. Long Range Graph Benchmark. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 22326–22340.
22. Zhong, Z.; Li, C.T.; Pang, J. Hierarchical message-passing graph neural networks. *Data Min. Knowl. Disc.* **2023**, 37, 381–408. [CrossRef]
23. Fuchs, P.; Sanocki, M.; Zavadlav, J. Learning non-local molecular interactions via equivariant local representations and charge equilibration. *npj Comput. Mater.* **2025**, *11*, 287. [CrossRef]
24. Ghasemi, S.A.; Hofstetter, A.; Saha, S.; Goedecker, S. Interatomic Potentials for Ionic Systems with Density Functional Accuracy Based on Charge Equilibration Neural Network Technique. *Phys. Rev. B* **2015**, *92*, 045131. [CrossRef]
25. Anstine, D.M.; Zubatyuk, R.; Isayev, O. AIMNet2: A Neural Network Potential to Meet Your Neutral, Charged, Organic and Elemental-Organic Needs. *Chem. Sci.* **2025**, *16*, 10228–10244. [CrossRef]
26. Cheng, B. Latent Ewald Summation for Machine Learning of Long-Range Interactions. *npj Comput. Mater.* **2025**, *11*, 80. [CrossRef]
27. King, D.S.; Kim, D.; Zhong, P.; Cheng, B. Machine learning of charges and long-range interactions from energies and forces. *Nat. Commun.* **2025**, *16*, 8763. [CrossRef]
28. Unke, O.T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693. [CrossRef]
29. Topping, J.; Di Giovanni, F.; Chamberlain, B.P.; Dong, X.; Bronstein, M.M. Understanding Over-Squashing and Bottlenecks on Graphs via Curvature. *arXiv* **2022**, arXiv:2111.14522. [CrossRef]
30. Kreuzer, D.; Beaini, D.; Hamilton, W.; Létourneau, V.; Tossou, P. Rethinking graph transformers with spectral attention. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 21618–21629.
31. Rampášek, L. ; Galkin, M.; Dwivedi, V.P.; Luu, A.T.; Wolf, G.; Beaini, D. Recipe for a General, Powerful, Scalable Graph Transformer. *arXiv* **2022**, arXiv:2205.12454.
32. Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; Liu, T.-Y. Do Transformers Really Perform Bad for Graph Representation? *arXiv* **2021**, arXiv:2106.05234. [CrossRef]
33. Deac, A.; Lackenby, M.; Veličković, P. Expander Graph Propagation. In *Proceedings of the Learning on Graphs Conference, Virtual, 9–12 December 2022*; PMLR: Cambridge, MA, USA , 2022; pp. 1–18.
34. Chen, D.; Schulz, T.H.; Borgwardt, K. Learning Long Range Dependencies on Graphs via Random Walks. *arXiv* **2024**, arXiv:2406.03386. [CrossRef]
35. Bose, K.; Das, S. Asynchronous Message Passing for Addressing Oversquashing in Graph Neural Networks. *arXiv* **2025**, arXiv:2509.06777. [CrossRef]
36. Jang, E.; Gu, S.; Poole, B. Categorical Reparameterization with Gumbel-Softmax. *arXiv* **2017**, arXiv:1611.01144. [CrossRef]
37. Rappé, A.K.; Goddard, W.A., III. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363. [CrossRef]
38. Ko, T.W.; Finkler, J.A; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Mater. Cloud Arch.* **2020**, *137*. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.