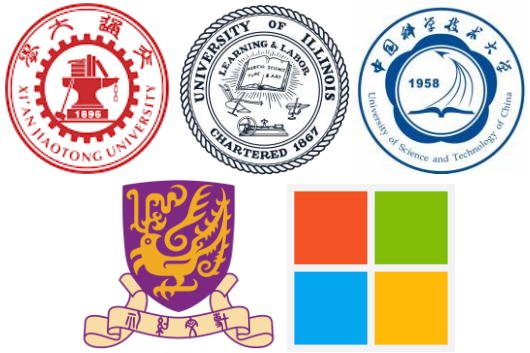
Neural P^3M: A Long-Range Interaction Modeling Enhancer for Geometric GNNs

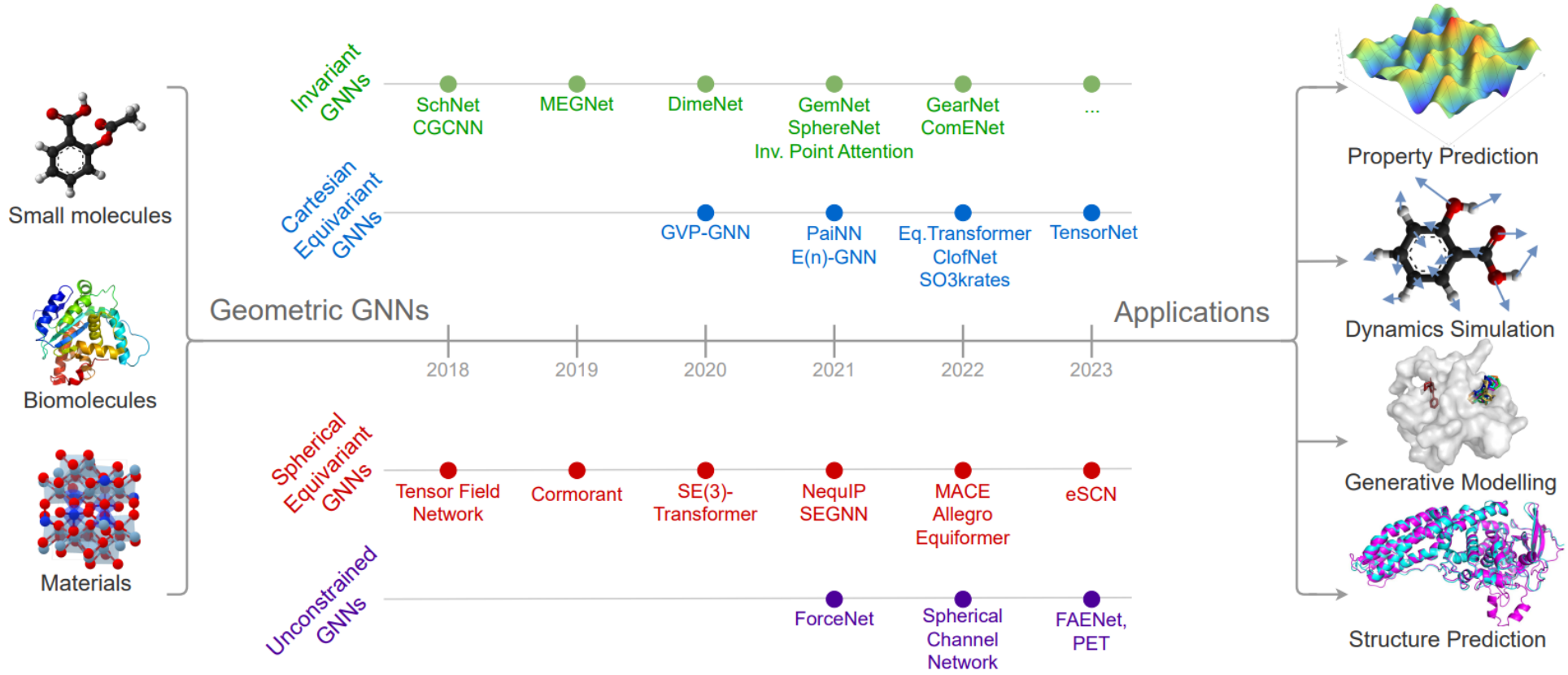
Yusong Wang[1,#], Chaoran Cheng[2,#], Shaoning Li[3,#], Yuxuan Ren[4], Bin Shao[5], Ge Liu[2], Pheng-Ann Heng[3], Nanning Zheng[1,*]

[1] Xi'an Jiaotong University
[2] University of Illinois Urbana-Champaign
[3] The Chinese University of Hong Kong
[4] University of Science and Technology of China
[5] Microsoft Research AI4Science
[#] Equal Contribution [*] Corresponding Author



Project URL: https://github.com/OnlyLoveKFC/Neural_P3M

Background – Introduction to Geometric GNNs



- Geometric graph neural networks have emerged as powerful tools for modeling molecular geometry.
- However, they encounter **limitations in effectively capturing long-range interactions** in large molecular systems due to **the localization assumption of GNN**.

Preliminary – Ewald Summation

Consider the pair-wise electrostatic potential as $\psi(r_{ij}) = 1/||r_{ij}||_2$ .The total electrostatic potential energy E can be evaluated as the infinite summation over pairs under the periodic boundary condition (PBC) as:

$$E = \frac{1}{2}\sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}{}' \iint \rho_i(\mathbf{r})\rho_j(\mathbf{r}')\psi(\|\mathbf{r}-\mathbf{r}'+\mathbf{n}\cdot\mathbf{c}\|_2)d^3\mathbf{r}d^3\mathbf{r}' = \frac{1}{2}\sum_{i=1}^{N}\int \rho_i(\mathbf{r})\ \phi_{[i]}(\mathbf{r})d^3\mathbf{r}$$

- $\rho_i(\mathbf{r})$ is charge density.
- $\mathbf{c}$ is the cell vector.
- N is the number of atoms in a cell.
- The $'$ summation is introduced to exclude the term $j = i$, if and only if $\mathbf{n} = 0$. $\phi_{[i]}(\mathbf{r})$ represents the potential generated by all particles excluding the particle i.

A continuous partition function that delays rapidly with respect to the distance is used to separate the short-range and long-range terms:

$$\psi^{\mathrm{sr}}(\mathbf{r}) = \frac{1-\mathrm{erf}(\beta\|\mathbf{r}\|_2)}{\|\mathbf{r}\|_2}, \psi^{\mathrm{lr}}(\mathbf{r}) = \frac{\mathrm{erf}(\beta\|\mathbf{r}\|_2)}{\|\mathbf{r}\|_2}$$

Preliminary – Ewald Summation

With the rapid delay of the partition function, it is safe to assume convergence by only considering the interaction pairs within a specific cutoff distance as:

$$E^{\mathrm{sr}} = \frac{1}{2}\sum_{i=1}^{N}\int \rho_i(\mathbf{r})\phi_{[i]}^{\mathrm{sr}}(\mathbf{r})d^3\mathbf{r} = \frac{1}{2}\sum_{(i,j)\in\mathcal{E}} q_i q_j \psi^{\mathrm{sr}}(\mathbf{r}_{ij})$$

By the Parseval's theorem, the corresponding long-range term can be expressed as the summation in the Fourier domain as:

$$E^{\mathrm{lr}} = \frac{1}{2}\sum_{i=1}^{N}\int \rho_i(\mathbf{r})\phi_{[i]}^{\mathrm{lr}}(\mathbf{r})d^3\mathbf{r} = \frac{1}{2V}\sum_{\mathbf{m}\neq 0} \tilde{g}(\mathbf{m})\tilde{\gamma}(\mathbf{m})\|\tilde{\rho}(\mathbf{m})\|_2^2$$

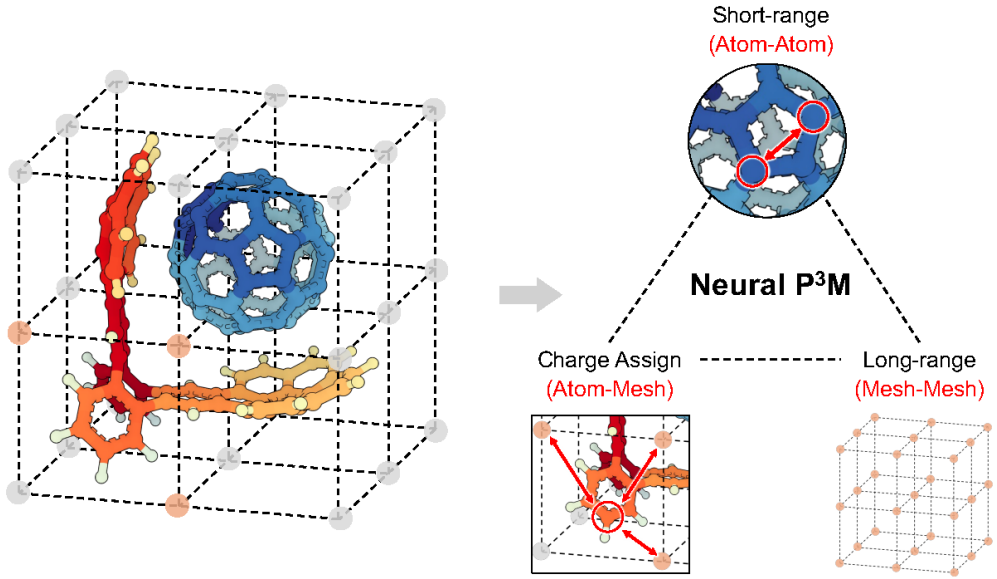
Using the convolution theory: $E^{\mathrm{lr}} = \frac{1}{2}\sum_{j=1}^{N} q_j[g \otimes \gamma \otimes \rho](\mathbf{r}_j) = \frac{1}{2}\sum_{j=1}^{N} q_j[G \otimes \rho](\mathbf{r}_j)$

As the long-range term introduces the self-interaction energy, a correction term is also applied to the final potential energy as:

$$E^{\mathrm{self}} = -\frac{1}{2}\sum_{i=1}^{N}\rho_i(\mathbf{r})\phi_i^{\mathrm{lr}}(\mathbf{r})d^3\mathbf{r} = -\frac{\beta}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2$$

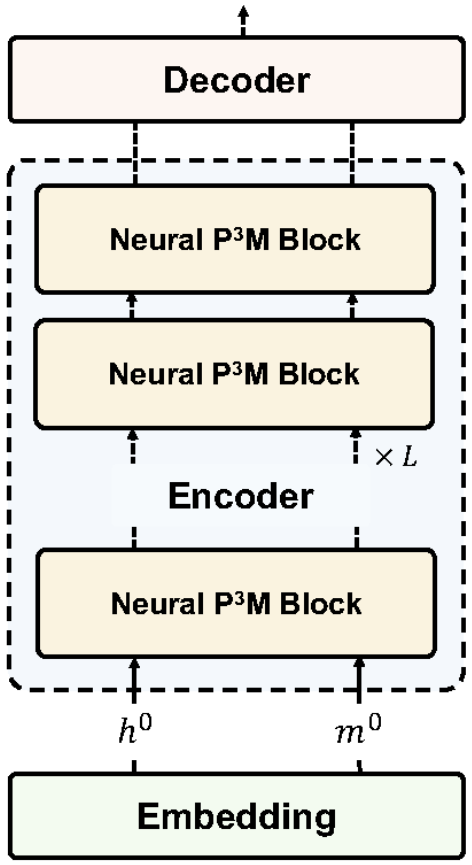
Methods – Meshing up the Ewald Summation with the Trainable Version

- Particles with their continuous coordinates, must be scattered onto grid-based densities (meshes) [1].
- The discrete approximation for E^{lr} can be expressed as [2]
- Reimage the traditional mathematical operations in mesh-based methods in a trainable manner, laying the foundation of Neural P^{3}M framework.

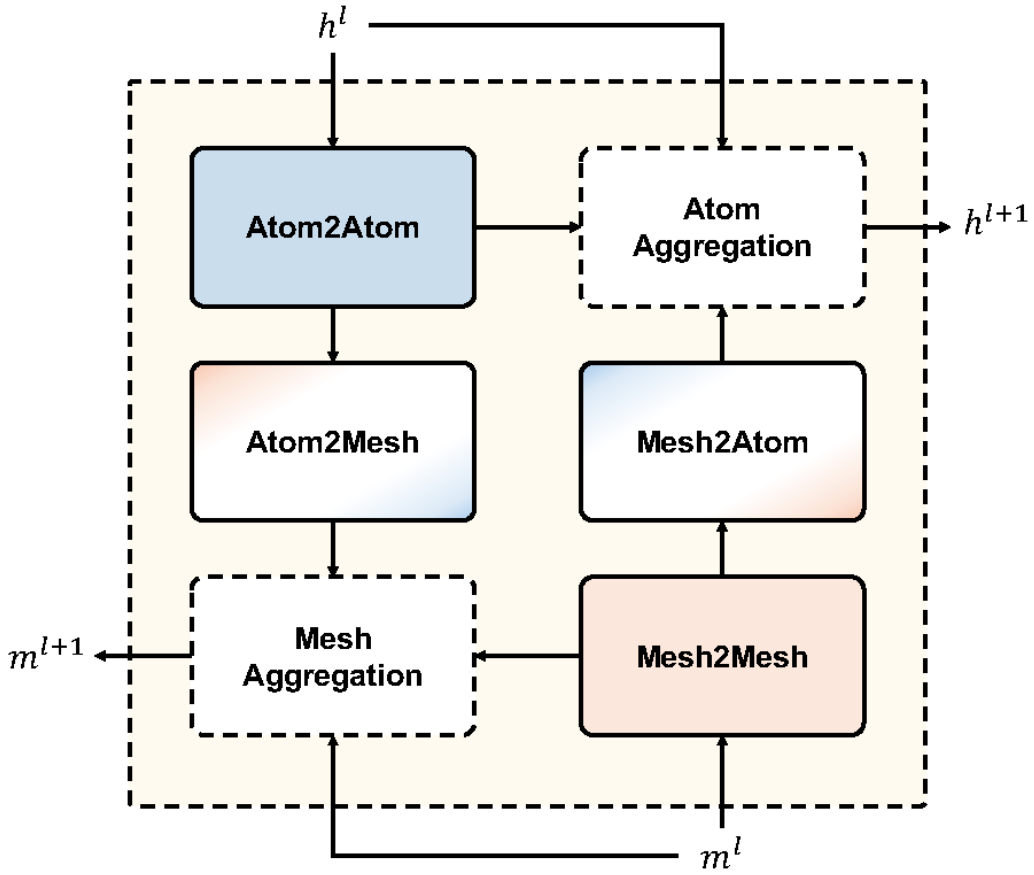


	Meshing up Methods	Neural Network
Short-range Block	$\frac{1}{2}\sum_{(i,j)\in\mathcal{E}} q_i q_j \psi^{\mathrm{sr}}(\mathbf{r}_{ij})$	Geometric GNNs
Charge/Representation Assignment	$\rho_M(\mathbf{r}_p) = \frac{1}{V_{\mathrm{grid}}}\sum_{i=1}^{N} q_i W(\mathbf{r}_p - \mathbf{r}_i)$ [1]	$(m \leftarrow a)_i^l = \mathrm{MLP}\left(\tilde{h}_j^l \cdot W_{m\leftarrow a}^l f_{ij}^{\mathrm{assign}}\right)$
Long-range Block	$E^{\mathrm{lr}} \approx \frac{1}{2}\sum_{\mathbf{r}_p\in\mathcal{V}} V_{\mathrm{grid}}\,\rho_M(\mathbf{r}_p)[G \otimes \rho_M](\mathbf{r}_p)$ [2]	$\tilde{m}^l \leftarrow \sigma\left(W^{\mathrm{long}} m^l + \left(\mathcal{F}^{-1}(\tilde{G}\cdot\mathcal{F})\right)(m^l)\right)$

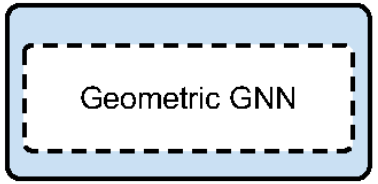
Methods – Overall Neural P³M framework architecture



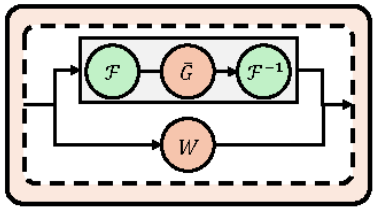
(a) Model Architecture



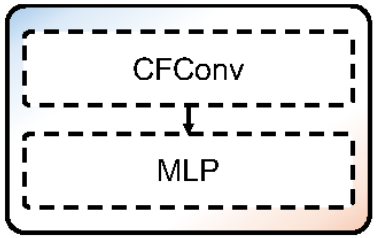
(b) Neural P³M Block



(c) Short-range Block



(d) Long-range Block



(e) Repr. Assignment

Experiments – MD22

Mean absolute errors (MAE) of energy and forces on 7 large molecules in MD22 datasets compared with state-of-the-art algorithms

Molecule	Diameter (Å)		sGDML	SO3KRATES	Allegro	Equiformer	MACE	ViSNet			
								Baseline	Ewald	LSRM	Neural P^3M
Ac-Ala3-NHMe	10.75	energy	0.3902	0.337	0.1019	0.0828	**0.0620**	0.0796	0.0775	0.0654	0.0719
		forces	0.7968	0.244	0.1068	0.0804	0.0876	0.0972	0.0814	0.0902	**0.0788**
DHA	14.58	energy	1.3117	0.379	0.1153	0.1788	0.1317	0.1526	0.0932	0.0873	**0.0712**
		forces	0.7474	0.242	0.0732	**0.0506**	0.0646	0.0668	0.0664	0.0598	0.0679
Stachyose	13.87	energy	4.0497	0.442	0.2485	0.1404	0.1244	0.1283	0.1089	0.1055	**0.0856**
		forces	0.6744	0.435	0.0971	**0.0635**	0.0876	0.0869	0.0976	0.0767	0.0940
AT-AT	17.63	energy	0.7235	0.178	0.1428	0.1309	0.1093	0.1688	0.1487	0.0772	**0.0714**
		forces	0.6911	0.216	0.0952	0.0960	0.0992	0.1070	0.0885	0.0781	**0.0740**
AT-AT-CG-CG	21.29	energy	1.3885	0.345	0.3933	0.1510	0.1578	0.1995	0.1571	0.1135	**0.1124**
		forces	0.7028	0.332	0.1280	0.1252	0.1153	0.1563	0.1115	0.1063	**0.0993**
Buckyball catcher	15.89	energy	1.1962	0.381	0.5258	0.3978	0.4812	0.4421	0.3575	0.4220	**0.3543**
		forces	0.6820	0.237	0.0887	0.1114	0.0853	0.1335	0.0989	0.1026	**0.0846**
Double-walled nanotube	32.39	energy	4.0122	0.993	2.2097	1.1945	1.6553	1.0339	0.7909	1.8230	**0.7751**
		forces	0.5231	0.727	0.3428	0.2747	0.2767	0.3959	0.2875	0.3391	**0.2561**

- **Flexibility.** Neural P3M is well-suited for a wide range of molecular systems without constraints, whereas LSRM relies on fragmentation algorithms like BRICS.

Experiments – OE62

Model	Variant	OE62-val		OE62-test		Forward Pass		Forward & Backward Pass	
		MAE meV ↓	Rel. % ↑	MAE meV ↓	Rel. % ↑	Runtime ms/struct. ↓	Rel. % ↓	Runtime ms/struct. ↓	Rel. % ↓
SchNet	Baseline	133.5	-	131.3	-	**0.13**	-	**0.28**	-
	Embeddings	144.7	-8.4	136.7	-4.1	0.14	15.2	0.33	17.8
	Cutoff	257.4	-92.8	254.8	-94.1	0.14	13.6	0.31	11.6
	SchNet-LR	86.6	35.1	89.2	32.1	0.32	156.0	0.75	171.7
	Ewald	79.2	40.7	81.1	38.2	0.70	461.6	1.03	271.4
	Neural P^{3}M	**70.2**	**47.4**	**69.1**	**47.4**	0.37	184.6	0.57	103.6
PaiNN	Baseline	61.4	-	63.3	-	**1.52**	-	**3.16**	-
	Embeddings	63.5	-3.4	63.1	-0.2	1.54	1.4	3.28	3.8
	Cutoff	65.1	-6.0	64.4	-2.2	1.84	20.9	3.91	23.6
	SchNet-LR	58.3	5.1	58.2	7.7	1.84	20.7	4.21	33.1
	Ewald	57.9	5.7	59.7	5.7	2.29	50.5	4.57	44.4
	Neural P^{3}M	**54.1**	**11.9**	**52.9**	**16.4**	2.17	42.8	4.19	32.6
DimeNet++	Baseline	51.2	-	53.8	-	**1.99**	-	**4.26**	-
	Embeddings	50.4	1.6	53.4	0.7	2.25	12.9	4.93	15.8
	Cutoff	48.3	5.7	48.1	10.6	2.68	34.7	6.10	43.4
	SchNet-LR	51.4	-0.5	54.4	-1.1	2.37	19.0	4.73	11.2
	Ewald	46.5	9.2	48.1	10.6	2.70	35.5	5.93	39.5
	Neural P^{3}M	**40.9**	**20.1**	**41.5**	**22.9**	3.11	56.3	5.62	31.9
GemNet-T	Baseline	51.5	-	53.1	-	**3.07**	-	**6.96**	-
	Embeddings	52.7	-2.3	53.9	-1.5	3.11	1.5	6.98	0.4
	Cutoff	47.8	7.2	47.7	10.2	4.02	31.2	8.88	27.7
	SchNet-LR	51.2	0.6	52.8	0.5	3.32	8.3	7.73	11.1
	Ewald	47.4	8.0	47.5	10.5	4.05	32.0	8.86	27.4
	Neural P^{3}M	**47.2**	**8.3**	**47.4**	**10.7**	3.93	28.0	7.71	10.8

Energy MAEs and computation times per input structure on the OE62 dataset when integrating various GNNs into Neural P^{3}M

- **Enhancement and Versatility.** Combined with various models, Neural P^{3}M shows a consistent improvement.
- **Efficiency.** Thanks to fast Fourier transformation, Neural P^{3}M is faster than Ewald MP in most cases.

Thanks

NEURAL INFORMATION
PROCESSING SYSTEMS