

Algorithm 1 Message Passing with Φ -Module

Require: Mini-batch $\mathcal{B} = \{G_i = (V_i, E_i)\}_{i=1}^B$, # message-passing layers T , modes k

Ensure: Total loss \mathcal{L} for back-propagation

```

1:  $L \leftarrow \text{BLOCKDIAG}(\{\text{GRAPHLAPLACIAN}(G_i)\}_{i=1}^B)$ 
2:  $(U, \Lambda) \leftarrow \text{LOBPCG}(L, k)$  ▷ batched eigendecomposition
3:  $\{h_v^0\}_{v \in V} \leftarrow \text{EMBEDDING}(\mathcal{B})$ 
4:  $\phi^0 \leftarrow \mathbf{0}; \quad \rho^0 \leftarrow \mathbf{0}$ 
5: for  $t = 0$  to  $T - 1$  do
6:    $\{h_v^{t+1}\} \leftarrow \text{MESSAGEPASSING}(\{h_v^t\}, \mathcal{B})$ 
7:   if  $t == 0$  then
8:      $\alpha_\phi^0, \alpha_\rho^0 \leftarrow \text{ALPHANET}(\{h_v^{t+1}\})$ 
9:      $\phi^1 \leftarrow U \alpha_\phi^0; \quad \rho^1 \leftarrow U \Lambda \alpha_\rho^0$ 
10:  else
11:     $\alpha_\phi^t, \alpha_\rho^t \leftarrow \text{ALPHANET}(\{h_v^{t+1}\})$ 
12:     $\phi^{t+1} \leftarrow \phi^t + U \alpha_\phi^t$ 
13:     $\rho^{t+1} \leftarrow \rho^t + U \Lambda \alpha_\rho^t$ 
14:  end if
15: end for
16:  $\mathbf{E}^{\text{model}} \leftarrow \text{READOUTENERGY}(\{h_v^T\}, \mathcal{B})$ 
17:  $\mathbf{E}^{\text{ES}} \leftarrow \frac{1}{2} \sum_{v \in V_i} (\phi_v \rho_v)$  ▷ electrostatic energy term
18:  $\mathbf{r} \leftarrow L \phi^T - \rho^T$  ▷ PDE residual
19:  $\mathcal{L} \leftarrow \underbrace{\ell(\mathbf{E}^{\text{model}} + \mathbf{E}^{\text{ES}}, \mathbf{E}^{\text{target}})}_{\mathcal{L}_{\text{model}}} + \underbrace{\beta \|\mathbf{r}\|_2}_{\mathcal{L}_{\text{PDE}}} + \underbrace{\gamma \left| \sum_{v \in V_i} \rho_v^T \right|}_{\mathcal{L}_{\text{net}}}$ 
20: return  $\mathcal{L}$ 

```

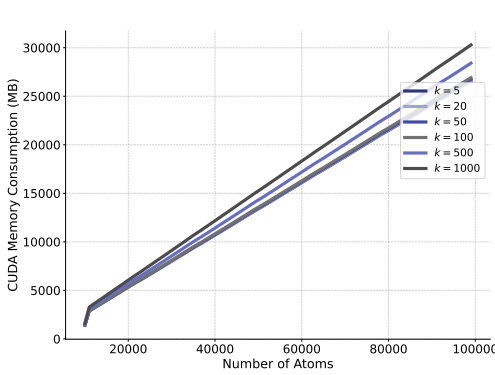
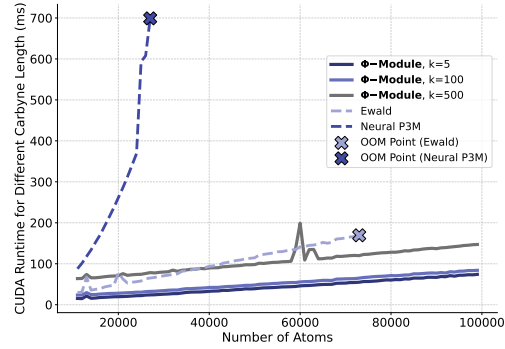
(a) Memory scaling of Φ -Module with varying k .(b) GPU runtime comparison of Φ -Module with Ewald and Neural P3M.

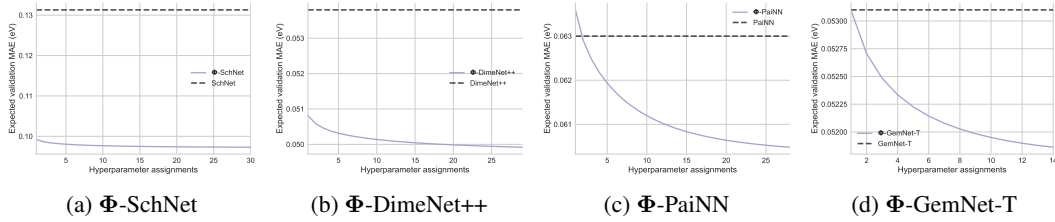
Figure 9: Additional memory and runtime scaling experiments.

A CODE AVAILABILITY

We provide the source code to reproduce the experiments in the supplementary material to the submission as a file archive. The code will be released to public upon acceptance.

B PSEUDOCODE FOR Φ -MODULE

Complete and detailed pseudocode for Φ -Module can be examined in Algorithm 1.

Figure 10: Energy-variance plots for Φ -variants.Table 2: Φ -Module hyperparameters for the reported models on OE62.

Model	Hyperparameters		
	k	β	γ
Φ -SchNet	9	10^{-4}	10^{-4}
Φ -DimeNet++	9	10^{-2}	10^{-4}
Φ -PaiNN	10	10^{-3}	10^{-1}
Φ -GemNet-T	3	$5 * 10^{-1}$	10^{-3}
Φ -E ₂ GNN	5	10^{-1}	0

C MEMORY AND RUNTIME SCALING WITH INCREASING k .

In this experiment, we follow the design of the task involving the linear carbyne chain in Section 4, but test the memory and runtime trends of Φ -Module with respect to the increasing number of estimated eigenvalues. In Figure 9a, you can see that memory consumption increases at a very slow rate which allows efficient processing of large systems. Moreover, Φ -Module scales favorably in terms of GPU runtime for large systems (starting from 10^4 atoms) in comparison to Ewald and Neural P3M based on Figure 9b.

D HYPERPARAMETERS

Hyperparameter Search. We run a hyperparameter search with random uniform sampling for the Φ -Module with the following configuration for each model in this study: k : $\{3, 5, 7, 9, 10, 15\}$, β : $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 5 * 10^{-1}\}$, γ : $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 5 * 10^{-1}\}$. All experiments were conducted on one NVIDIA 80G H100 GPU.

OE62. We follow the same hyperparameters for the baselines as in Kosmala et al. (2023) in most cases. The main difference are the usage of Adam (Kingma & Ba, 2014) optimizer and cosine learning rate schedule (Loshchilov & Hutter, 2016) without warm restarts as well as gradient clipping of 10^3 . For hyperparameters related to Φ -Module, refer to Table 2. For E₂GNN we use 256 hidden channels, 6 layers, 128 Gaussian RBFs, cutoff of 6.0 Å with a maximum 50 neighbors. Batch size is set to 64 and the training of E₂GNN was performed for 400 epochs with the same optimizer and scheduler as for other models without gradient clipping.

MD22. For the baseline ViSNet, we employ the same hyperparameters as in (Wang et al., 2022). Optimizer and scheduler choice is the same as for OE62. The hyperparameters for Φ -ViSNet are the following: k - 9, β - 10^{-3} , γ - 10^{-4} .

E OE62 RESULTS

In Table 3 the exact numerical comparison for the Section 4 OE62 experiment is shown.

Table 3: Energy MAEs and computation time of baselines and their alternatives with Ewald module and Φ -Module on OE62. Lowest errors and fastest runtimes compared to baselines are highlighted in bold. See Section 4.

MODEL	VERSION	OE62-VAL		OE62-TEST		MEAN EPOCH TIME	
		MAE, MEV \downarrow	REL, % \downarrow	MAE, MEV \downarrow	REL, % \downarrow	RUNTIME, S \downarrow	REL, % \downarrow
SCHNET	BASILINE	133.5	-	131.3	-	16.91	-
	EWALD	79.2	40.7	81.1	38.2	68.4	304.5
	Φ -SCHNET	92.2	30.9	92.0	29.9	18.65	10.3
DIME.NET++	BASILINE	51.2	-	53.8	-	90.66	-
	EWALD	46.5	9.2	48.1	10.6	212.1	134.0
	Φ -DIME.NET++	47.1	8.0	48.8	9.3	94.36	4.1
PAIINN	BASILINE	61.4	-	63.0	-	56.70	-
	EWALD	57.9	5.7	59.7	5.2	193.2	240.9
	Φ -PAIINN	57.7	6.0	58.8	6.7	62.24	9.8
GEMNET-T	BASILINE	51.2	-	53.1	-	179.01	-
	EWALD	46.5	9.2	47.5	10.5	501.0	179.9
	Φ -GEMNET-T	47.3	7.6	48.2	9.2	187.40	4.7
E ₂ GNN	BASILINE	60.9	-	61.6	-	130.8	-
	EWALD	60.3	1.0	61.0	1.0	185.1	41.5
	Φ -E ₂ GNN	59.2	2.8	60.7	1.5	162.0	23.9

The results show that the Φ -Module provides consistent performance improvements across all baselines, with gains of at least 5% in most cases and around 3% for E₂GNN. In addition, it outperforms the Ewald block in 2 out of 5 settings while requiring noticeably less computational overhead.

F ARCHITECTURAL DETAILS OF α -NET

Permutational Invariance. The convolutions in the α -Net are applied over the node dimension. The permutational invariance is lost only given the kernel size is not 1. Convolutions with 1x1 filter can also serve as a competitive option. Below (see Table 4) are results comparing SchNet and DimeNet++ with regular α -Net and the one with 1x1 convolutions preserving invariance. Both options show similar performance.

Separate Eigenbasis Coefficients α_ϕ and α_ρ . In this paragraph, we discuss the idea of learning separated Laplacian eigenbasis coefficients for potential and charges. Theorem F.1 describes the symmetric nature of residual gradient w.r.t α_ϕ and α_ρ , given the parametrization of $\rho = U\Lambda\alpha_\rho$ aligned with the eigenbasis of $L\phi = U\Lambda U^\top U\alpha_\phi = U\Lambda\alpha_\phi$. We expect this parametrization to benefit training dynamics as such symmetries guarantee equal update rate for both α_ϕ and α_ρ . On the other hand, plain $\rho = U\alpha_\rho$ results in the dependence on λ_i making an optimization process dominated by specific modes and neglecting others.

Proposition F.1 (Symmetric vs. asymmetric gradients for the Poisson residual). *Preserving the notations from Theorem 3.1, let the potential be $\phi = U\alpha_\phi$ and define the Poisson residual loss*

$$\mathcal{L}(\alpha_\phi, \alpha_\rho) = \beta \|L\phi - \rho\|_2^2, \quad \beta > 0.$$

Then:

Case (A): ($\rho = U\Lambda\alpha_\rho$). *Writing $r = L\phi - \rho = U\Lambda(\alpha_\phi - \alpha_\rho)$, the gradients are*

$$\nabla_{\alpha_\phi} \mathcal{L} = 2\beta \Lambda^2(\alpha_\phi - \alpha_\rho), \quad \nabla_{\alpha_\rho} \mathcal{L} = -2\beta \Lambda^2(\alpha_\phi - \alpha_\rho).$$

Per mode i :

$$\frac{\partial \mathcal{L}}{\partial (\alpha_\phi)_i} = 2\beta \lambda_i^2 ((\alpha_\phi)_i - (\alpha_\rho)_i), \quad \frac{\partial \mathcal{L}}{\partial (\alpha_\rho)_i} = -2\beta \lambda_i^2 ((\alpha_\phi)_i - (\alpha_\rho)_i).$$

Hence the updates are equal in magnitude and opposite in sign, with identical per-mode scaling λ_i^2 resulting in mode-wise symmetry.

Case (B): ($\rho = U\alpha_\rho$). Writing $r = L\phi - \rho = U(\Lambda\alpha_\phi - \alpha_\rho)$, the gradients are

$$\nabla_{\alpha_\phi} \mathcal{L} = 2\beta \Lambda(\Lambda\alpha_\phi - \alpha_\rho), \quad \nabla_{\alpha_\rho} \mathcal{L} = -2\beta(\Lambda\alpha_\phi - \alpha_\rho).$$

Per mode i :

$$\frac{\partial \mathcal{L}}{\partial (\alpha_\phi)_i} = 2\beta \lambda_i (\lambda_i (\alpha_\phi)_i - (\alpha_\rho)_i), \quad \frac{\partial \mathcal{L}}{\partial (\alpha_\rho)_i} = -2\beta (\lambda_i (\alpha_\phi)_i - (\alpha_\rho)_i).$$

Thus the two updates differ by a factor λ_i resulting in mode-wise asymmetry.

Proof. Let $r = L\phi - \rho$. Then, $\nabla_\phi \|r\|_2^2 = 2L^\top r = 2Lr$ and $\nabla_\rho \|r\|_2^2 = -2r$. Mapping to $\phi = U\alpha_\phi$ and applying the chain rule we acquire $\nabla_{\alpha_\phi} \|r\|_2^2 = U^\top (2Lr)$.

Case (A). If $\rho = U\Lambda\alpha_\rho$, then $r = U\Lambda(\alpha_\phi - \alpha_\rho)$ and

$$\nabla_{\alpha_\phi} \mathcal{L}_{\text{PDE}} = \beta (U\Lambda)^\top (2r) = 2\beta \Lambda U^\top U \Lambda (\alpha_\phi - \alpha_\rho) = 2\beta \Lambda^2 (\alpha_\phi - \alpha_\rho),$$

$$\nabla_{\alpha_\rho} \mathcal{L}_{\text{PDE}} = \beta (-U\Lambda)^\top (2r) = -2\beta \Lambda^2 (\alpha_\phi - \alpha_\rho).$$

Case (B). If $\rho = U\alpha_\rho$, then $r = U(\Lambda\alpha_\phi - \alpha_\rho)$ and

$$\nabla_{\alpha_\phi} \mathcal{L}_{\text{PDE}} = \beta (U\Lambda)^\top (2r) = 2\beta \Lambda(\Lambda\alpha_\phi - \alpha_\rho), \quad \nabla_{\alpha_\rho} \mathcal{L}_{\text{PDE}} = \beta (-U)^\top (2r) = -2\beta (\Lambda\alpha_\phi - \alpha_\rho).$$

□

Nature of the Laplacian. For a molecular graph $G = (V, E)$ with node set V and positive symmetric edge weights $w_{ij} > 0$, we employ the *symmetric normalized Laplacian*

$$L = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}},$$

where $W \in \mathbb{R}^{|V| \times |V|}$ is the weighted adjacency matrix with entries $W_{ij} = w_{ij}$ if $(i, j) \in E$ and 0 otherwise, and $D = \text{diag}(d_1, \dots, d_{|V|})$ is the diagonal degree matrix with $d_i = \sum_j W_{ij}$. By construction L is real, symmetric, and positive semidefinite. We use the edge weights $w_{ij} = d_{ij}$ given by the interatomic distances between atoms i and j , which preserves symmetry and ensures that L encodes geometric information about molecular conformations.

G ADDITIONAL BACKGROUND.

Microcanonical Ensemble. In classical molecular dynamics, the microcanonical ensemble (NVE) models an isolated system with constant particle number (N), volume (V), and total energy (E). The dynamics follow Newton’s equations of motion:

$$m_i \ddot{\mathbf{r}}_i(t) = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \dots, \mathbf{r}_N), \quad (5)$$

where $\mathbf{r}_i(t)$ denotes the position of particle i , m_i its mass, and U the potential energy function. In the absence of thermostats or external driving, this guarantees conservation of total energy.

To integrate trajectories numerically, Verlet-type schemes are widely adopted due to their symplecticity and time reversibility. The basic Verlet update is given by

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \mathbf{F}_i(t), \quad (6)$$

with forces $\mathbf{F}_i(t) = -\nabla_{\mathbf{r}_i} U$. A more practical variant is the velocity Verlet integrator:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{1}{2} \Delta t^2 \mathbf{a}_i(t), \quad (7)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{1}{2} \Delta t (\mathbf{a}_i(t) + \mathbf{a}_i(t + \Delta t)), \quad (8)$$

where $\mathbf{a}_i(t) = \mathbf{F}_i(t)/m_i$. These integrators achieve $\mathcal{O}(\Delta t^2)$ accuracy while requiring only a single force evaluation per timestep. Crucially, their symplectic structure ensures stable long-time energy behavior, making NVE with Verlet the de facto baseline.

Table 4: Comparison of α -net variants for SchNet and DimeNet++.

MODEL	DEFAULT α -NET	1×1 CONV α -NET
SCHNET	92.0	93.8
DIMENET++	48.8	49.5

Expected Validation Performance. Expected Validation Performance (EVP) Dodge et al. (2019) curve represents how on average performance changes as the number of hyperparameter assignments increases during the random search. The X-axis represents the number of hyperparameter trials. The Y-axis represents the expected best performance for a given number of hyperparameter trials.

The expected best performance is computed as

$$\mathbb{E}[V_n^*|n] = \sum_v v \cdot (P(V_i \leq v)^n - P(V_i < v)^n),$$

where $V_n^* = \max_{i \in \{1, \dots, n\}} V_i$ is the maximum for model performance evaluations V_i given a series of n i.i.d. hyperparameter configurations, which are acquired empirically from the random hyperparameter search process and $P(V_n^*|n)$ is the probability mass function for the max-random variable.

The EVP curves for Φ -SchNet, Φ -DimeNet++, Φ -PaiNN and Φ -GemNet-T can be seen in ??.

Φ -Module demonstrates hyperparameter stability for all of the baseline models.

you removed all text completely. listen again, make it as close as possible to the initial version below

here are proofs for theorems, check briefly if they are correct

H PROOFS

In this part we restate Theorem 3.1 and Theorem 3.2 from the main body and proof them accordingly. Note that we prove theorems for the surrogate L2 objective for the tractability, and it is interchangeable.

Theorem H.1 (Exact inner minimizer over ρ). *Define $a = E - E_{\text{model}}$. Fix $\phi \in \text{span}(U_k)$. The unique minimizer of $\rho \mapsto \mathcal{L}(\phi, \rho)$ over $\text{span}(U_k)$ is*

$$\rho^*(\phi) = L\phi - t^*(\phi)\phi, \quad t^*(\phi) = \frac{a + \frac{1}{2}\phi^\top L\phi}{2\beta + \frac{1}{2}\|\phi\|^2}.$$

Proof. Let $e(\phi, \rho) := a + \frac{1}{2}\phi^\top \rho$. Using $\nabla_\rho \|L\phi - \rho\|^2 = -2(L\phi - \rho)$ and $\nabla_\rho e(\phi, \rho)^2 = 2e(\phi, \rho) \cdot \frac{1}{2}\phi = e(\phi, \rho)\phi$, the first-order condition is

$$\nabla_\rho \mathcal{L}(\phi, \rho) = 2\beta(\rho - L\phi) + e(\phi, \rho)\phi = 0.$$

Hence $\rho - L\phi$ is colinear with ϕ , then $\rho = L\phi - t\phi$ for some $t \in \mathbb{R}$. Substituting back gives

$$-2\beta t\phi + \left(a + \frac{1}{2}(\phi^\top L\phi - t\|\phi\|^2)\right)\phi = 0,$$

and we obtain $-2\beta t + a + \frac{1}{2}\phi^\top L\phi - \frac{1}{2}t\|\phi\|^2 = 0$, i.e.

$$t^*(\phi) = \frac{a + \frac{1}{2}\phi^\top L\phi}{2\beta + \frac{1}{2}\|\phi\|^2}.$$

Uniqueness follows because the Hessian w.r.t. ρ is $2\beta I + \frac{1}{2}\phi\phi^\top \succ 0$ for $\beta > 0$. \square

Theorem H.2 (Monotone objective decrease in optimization towards ρ^*). *Define $A(\phi) := a + \frac{1}{2}\phi^\top L\phi$. Then substituting $\rho^*(\phi)$ from Theorem 3.1 yields*

$$\tilde{\mathcal{L}}(\phi) := \mathcal{L}(\phi, \rho^*(\phi)) = A(\phi)^2 \frac{4\beta}{4\beta + \|\phi\|^2} \leq A(\phi)^2,$$

with equality if and only if $A(\phi) = 0$ or $\phi = 0$.

Proof. Along the affine line $\rho(t) = L\phi - t\phi$ we have

$$\mathcal{L}(\phi, \rho(t)) = \beta \|t\phi\|^2 + \left(A(\phi) - \frac{1}{2}t \|\phi\|^2 \right)^2 = \underbrace{\beta \|\phi\|^2}_p t^2 + \underbrace{\left(A(\phi) - \frac{1}{2}\|\phi\|^2 t \right)^2}_q.$$

This is a strictly convex quadratic in t (for $\|\phi\|^2 > 0$) with minimizer $t^* = \frac{A(\phi)q}{p+q^2}$ and minimum value

$$\mathcal{L}(\phi, \rho^*(\phi)) = A(\phi)^2 \frac{p}{p+q^2} = A(\phi)^2 \frac{\beta \|\phi\|^2}{\beta \|\phi\|^2 + \frac{1}{4}\|\phi\|^4} = A(\phi)^2 \frac{4\beta}{4\beta + \|\phi\|^2}.$$

Since $\frac{4\beta}{4\beta + \|\phi\|^2} \in (0, 1]$, the inequality follows; equality holds exactly when $A(\phi) = 0$ or $s(\phi) = 0$. \square