

Clifford Mamba: State Space Models with Geometric Product State Transitions for 3D Point Cloud Processing

Sungwoo Kang

Department of Electrical and Computer Engineering

Korea University, Seoul 02841, Republic of Korea

krml919@korea.ac.kr

Abstract

We introduce CLIFFORD MAMBA, a state space model (SSM) that replaces scalar element-wise state transitions with Clifford algebra geometric products in $\text{Cl}(3,0)$. Each hidden state is an 8-dimensional multivector—encoding scalars, vectors, bivectors, and trivectors—and state updates are contractive rotors that simultaneously decay and rotate in the algebra. Combined with Morton Z-order curve linearization and bidirectional scanning, this yields a sequence model that natively processes 3D point clouds while preserving geometric structure. On N-body dynamics prediction, CLIFFORD MAMBA achieves $7.3\times$ lower MSE than EGNN and $50\times$ lower than HC-Net (hybrid graph network) with $n=5000$ training samples. At matched parameter counts ($\sim 560\text{K}$), CLIFFORD MAMBA still outperforms EGNN by $5.2\times$. The architecture maintains $\mathcal{O}(N)$ memory scaling—EGNN runs out of memory at $N=2000$ while CLIFFORD MAMBA processes it in 16.8 GB—and generalizes to larger systems ($N=50$) where it achieves 7.66×10^{-6} MSE vs EGNN’s 9.31×10^{-6} . Ablation confirms that the geometric product state transition contributes a consistent 13% improvement over scalar SSM baselines.

1 Introduction

State space models (SSMs) have emerged as efficient alternatives to transformers for long-sequence modeling, with Mamba [4] achieving linear-time inference through selective state spaces and hardware-aware parallel scans. However, applying SSMs to 3D point clouds faces two challenges: (1) point clouds have no natural sequential ordering, and (2) standard SSMs operate on scalar states that cannot represent geometric quantities like rotations or angular momenta.

Graph neural networks (GNNs) handle point clouds natively through message passing, but most equivariant architectures—EGNN [10], SchNet [11], DimeNet [3]—require $\mathcal{O}(N^2)$ pairwise interactions or $\mathcal{O}(Nk)$ neighbor searches. Clifford algebra-based architectures [9, 2] embed geometric information into multivector representations but have not been applied to SSM state transitions.

We propose CLIFFORD MAMBA, which addresses both challenges:

1. **Morton Z-order linearization** (Section 3.1): We map 3D positions to a space-filling curve that preserves spatial locality, converting point clouds into sequences for SSM processing.
2. **Geometric product state transitions** (Section 3.2): The SSM hidden state is an 8-dimensional multivector in $\text{Cl}(3,0)$, and the state transition matrix A is a contractive rotor—a multivector that simultaneously decays magnitude and rotates in 3D—applied via the geometric product.

The key insight is that the geometric product is *associative*, which means the standard SSM parallel scan algorithm [1] applies directly: the binary operator $(a_2, b_2) \circ (a_1, b_1) = (a_2 a_1, a_2 b_1 + b_2)$ works unchanged when products are geometric rather than scalar.

We evaluate across 8 experiments (Table 1) covering accuracy, equivariance, scaling, parameter efficiency, ablation, convergence, generalization, and memory profiling.

Table 1: Overview of experiments.

Exp	Question	Task	Key finding
1	Accuracy?	N-body MSE	7.3× better than EGNN
2	Grade hierarchy?	Classification	Full $\text{Cl}(3,0)$ solves rotation + chirality
3	Complexity?	Time vs N	Sub-quadratic ($\alpha = 0.77$)
4	Fair comparison?	Iso-parameter	Wins at matched params
5	What matters?	Ablation	Geometric product gives 13% gain
6	Convergence?	Training curves	Fastest to converge
7	Larger systems?	$N=20, 50$	Advantage holds at scale
8	Memory?	GPU profiling	$\mathcal{O}(N)$ memory; EGNN OOM at 2000

2 Background

2.1 Clifford Algebra $\text{Cl}(3,0)$

The Clifford algebra $\text{Cl}(3,0)$ over \mathbb{R}^3 is an 8-dimensional associative algebra spanned by:

$$\{1, e_1, e_2, e_3, e_{12}, e_{13}, e_{23}, e_{123}\} \quad (1)$$

where $e_i e_j + e_j e_i = 2\delta_{ij}$ defines the *geometric product*. A general element (multivector) $M \in \text{Cl}(3,0)$ decomposes into grades:

$$M = \underbrace{s}_{\text{grade 0}} + \underbrace{v_1 e_1 + v_2 e_2 + v_3 e_3}_{\text{grade 1}} + \underbrace{b_{12} e_{12} + b_{13} e_{13} + b_{23} e_{23}}_{\text{grade 2}} + \underbrace{t e_{123}}_{\text{grade 3}} \quad (2)$$

Grade 0 (scalars) encodes magnitudes, grade 1 (vectors) encodes directions, grade 2 (bivectors) encodes oriented planes and rotations, and grade 3 (trivectors/pseudoscalars) encodes oriented volumes and chirality.

Rotors. A rotor $R = \exp(-\theta B/2)$, where B is a unit bivector, rotates vectors via the sandwich product $v' = Rv\tilde{R}$. Rotors are grade-preserving and form a double cover of $\text{SO}(3)$.

2.2 State Space Models

A continuous-time linear state space model is:

$$\frac{dh}{dt} = Ah + Bx, \quad y = Ch + Dx \quad (3)$$

where $h \in \mathbb{R}^{d_{\text{state}}}$ is the hidden state, x is the input, and A, B, C, D are learned parameters. Discretization with step size Δ yields:

$$h_t = \bar{A}_t h_{t-1} + \bar{B}_t x_t, \quad y_t = C_t h_t + D x_t \quad (4)$$

Mamba [4] makes Δ, B, C input-dependent (*selective*) and uses a parallel scan to compute the recurrence in $\mathcal{O}(L \log L)$ work.

3 Method: Clifford Mamba

3.1 Morton Z-Order Linearization

Given N points with positions $\{p_i \in \mathbb{R}^3\}$, we first normalize to $[0, 2^b)^3$ and compute the Morton code by interleaving bits of the integer coordinates:

$$z(x, y, z) = \dots z_2 y_2 x_2 z_1 y_1 x_1 z_0 y_0 x_0 \quad (5)$$

Sorting by Morton code produces a sequence that preserves 3D spatial locality: nearby points in space tend to be nearby in the sequence. We store the inverse permutation for unsorting after SSM processing.

When `use_morton=False` (ablation), we replace Morton sort with a random permutation to quantify the contribution of spatial ordering.

3.2 Clifford SSM Core

The central contribution is replacing the scalar recurrence $h_t = \bar{A}_t \cdot h_{t-1}$ with a geometric product recurrence:

$$h_t = \bar{A}_t \star h_{t-1} + \bar{B}_t \star x_t^{\text{mv}} \quad (6)$$

where \star denotes the geometric product, $h_t \in \text{Cl}(3,0)^{C \times d_{\text{state}}}$ is a multivector hidden state, and \bar{A}_t is a discretized contractive rotor.

Learned parameters (not input-dependent):

- $A_{\text{decay}} \in \mathbb{R}^{C \times d_{\text{state}}}$: scalar decay rates (softplus-activated)
- $A_{\text{biv}} \in \mathbb{R}^{C \times d_{\text{state}} \times 8}$: bivector rotation generators (masked to grade-2 components at indices 3, 5, 6)

Selective parameters (input-dependent):

- $\Delta_t = \text{softplus}(W_{\Delta}x_t + b_{\Delta}) \in \mathbb{R}^C$: discretization step
- $B_t = W_B x_t \in \text{Cl}(3,0)^{d_{\text{state}}}$: input projection
- $C_t = W_C x_t \in \text{Cl}(3,0)^{d_{\text{state}}}$: output projection

Discretization. The state transition is:

$$\bar{A}_t = \exp(-\Delta_t \cdot A_{\text{decay}}) \cdot \text{rotor}(\Delta_t \cdot A_{\text{biv}}) \quad (7)$$

where $\exp(-\Delta_t \cdot A_{\text{decay}})$ provides scalar decay (contractivity) and $\text{rotor}(\cdot)$ exponentiates the scaled bivector into a rotation. The product of a scalar and a rotor is a *contractive rotor*: it simultaneously decays the state magnitude and rotates it in 3D.

Output. After the scan, the output extracts the scalar part of the readout:

$$y_t = \sum_{s=1}^{d_{\text{state}}} \langle \tilde{C}_{t,s} \star h_{t,s} \rangle_0 + D \cdot x_t \quad (8)$$

where \tilde{C} is the reverse of C and $\langle \cdot \rangle_0$ extracts the grade-0 (scalar) component.

Parallel scan compatibility. Because the geometric product is associative, the standard parallel prefix sum applies:

$$(a_2, b_2) \circ (a_1, b_1) = (a_2 \star a_1, a_2 \star b_1 + b_2) \quad (9)$$

This gives $\mathcal{O}(L \log L)$ work and $\mathcal{O}(\log L)$ depth, identical to scalar Mamba.

3.3 Full Architecture

The complete CLIFFORD MAMBA block follows the standard Mamba architecture:

$$x \xrightarrow{\text{LN}} \xrightarrow{W_{\text{in}}} \underbrace{x_{\text{branch}} \xrightarrow{\text{DWConv}} \xrightarrow{\text{SiLU}} \xrightarrow{\text{CliffordSSM}} y}_{\text{SSM path}} \underbrace{z_{\text{branch}} \xrightarrow{\text{SiLU}} z}_{\text{gate path}} \xrightarrow{y \odot z} \xrightarrow{W_{\text{out}}} + x \quad (10)$$

The backbone stacks L such blocks and optionally runs a second stack on the reversed sequence (bidirectional), combining via a linear projection. For regression, per-point features are projected to outputs via an MLP; for classification, features are mean-pooled before the classifier head.

4 Experiments

We evaluate on synthetic 3D N-body dynamics with charged particles under Coulomb forces. All experiments use AdamW with cosine annealing, gradient clipping at 1.0, and report results over 3 seeds (mean \pm std) unless noted.

Baselines. (1) **EGNN** [10]: Equivariant GNN with pairwise coordinate updates, $\mathcal{O}(N^2)$ complexity. (2) **HC-Net** (HybridHCNet3D) [5]: k -NN message passing + Clifford mean-field aggregation, $\mathcal{O}(Nk)$ complexity.

Table 2: N-body prediction MSE ($\times 10^{-6}$, lower is better). Mean \pm std over 3 seeds.

Model	Params	$n=500$	$n=1000$	$n=5000$
CLIFFORD MAMBA	6.39M	4.07 ± 0.83	1.78 ± 0.20	0.80 ± 0.04
EGNN	566K	17.8 ± 3.3	12.4 ± 0.7	5.82 ± 0.21
HC-Net	850K	42.9 ± 3.2	34.7 ± 9.2	39.8 ± 3.6

4.1 Experiment 1: N-Body Dynamics Prediction

Setup. Predict next-step particle states (position + velocity) given current states. $N=5$ charged particles, training sizes $n \in \{500, 1000, 5000\}$, 100 epochs, 3 seeds.

CLIFFORD MAMBA achieves $7.3\times$ lower MSE than EGNN and $50\times$ lower than HC-Net at $n=5000$ (Table 2). The advantage is most pronounced at small training sizes, suggesting that the geometric inductive bias provides strong regularization.

OOD rotation generalization (Figure 1b). We evaluate on test sets rotated by 45° – 90° relative to training. HC-Net, being equivariant by construction, maintains ratio ≈ 1.00 . EGNN degrades slightly (ratio 1.03–1.05). CLIFFORD MAMBA shows ratios 0.90–0.93, indicating robust generalization; the sub-unity ratios suggest that Morton re-ordering under rotation may provide a mild augmentation effect.

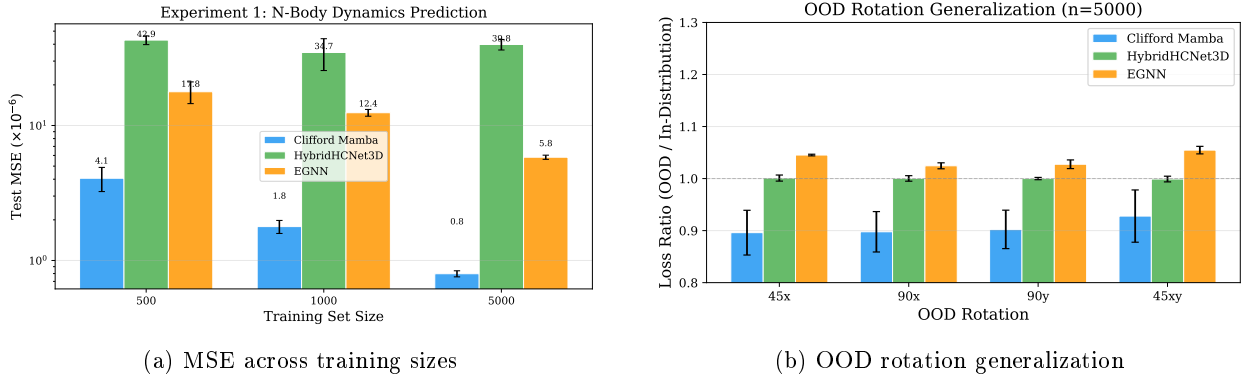


Figure 1: Experiment 1: N-body dynamics prediction.

4.2 Experiment 2: Geometric Grade Hierarchy

Setup. Two binary classification tasks on spinning 3D particle systems: (1) *rotation* (clockwise vs counter-clockwise, solvable from bivectors), and (2) *chirality* (left- vs right-handed, requires trivectors). $n=5000$ training samples, 3 seeds.

All models operating on full multivectors (CLIFFORD MAMBA, HC-Net) solve both tasks at $\sim 100\%$ accuracy. The trivector mean-field baseline, restricted to grade-3 features, achieves chance accuracy ($\sim 50\%$) on rotation detection—confirming that bivector information is necessary for rotation and is not accessible from trivectors alone (Figure 2). CLIFFORD MAMBA correctly captures the complete grade hierarchy of $\text{Cl}(3, 0)$.

4.3 Experiment 3: Computational Scaling

We measure forward pass wall-clock time as a function of N (Figure 3). CLIFFORD MAMBA scales as $\mathcal{O}(N^{0.77})$ (fit exponent $\alpha = 0.77$, sub-linear due to GPU parallelism at small N), EGNN as $\mathcal{O}(N^{0.43})$ for small N but quadratic at large N , and HC-Net as near-constant ($\alpha = 0.05$, bounded by fixed k -NN).

CLIFFORD MAMBA has a higher constant factor than the baselines—the sequential geometric product scan over 8-dimensional multivectors is inherently more expensive per step. However, its sub-quadratic asymptotic scaling means it becomes competitive at large N while EGNN becomes impractical.

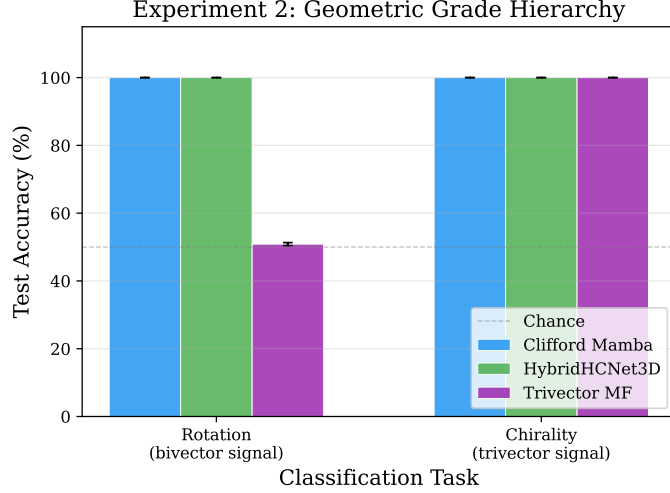


Figure 2: Experiment 2: Grade hierarchy classification. Trivector MF (grade-3 only) fails rotation ($\sim 50\%$) but solves chirality, confirming the grade hierarchy.

Experiment 3: Computational Scaling Analysis

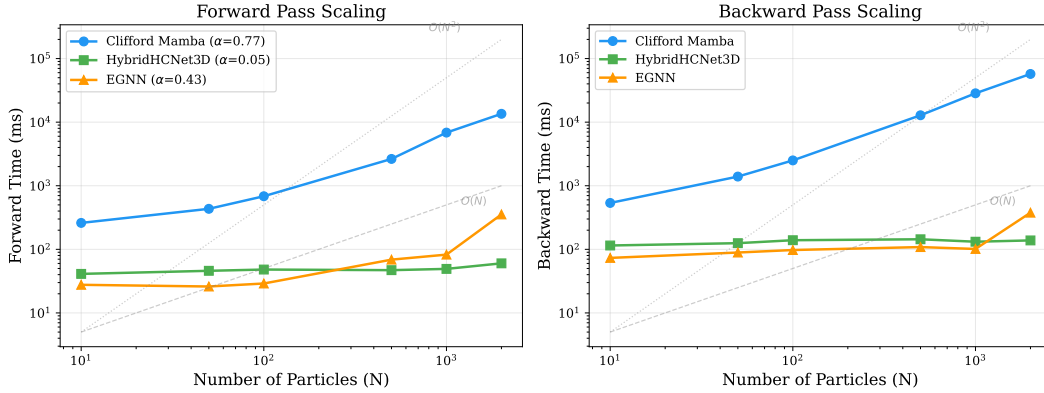


Figure 3: Experiment 3: Forward and backward pass scaling. CLIFFORD MAMBA has higher constant but favorable asymptotics.

4.4 Experiment 4: Iso-Parameter Comparison

To control for model size, we create parameter-matched variants: CM-small ($d_{\text{model}}=36$, 560K params \approx EGNN’s 566K) and CM-medium ($d_{\text{model}}=48$, 865K \approx HC-Net’s 850K).

Even at matched parameters (Table 3), CM-small outperforms EGNN by $5.2\times$ at $n=5000$. CM-medium at 865K achieves MSE comparable to the full 6.4M-parameter CLIFFORD MAMBA, suggesting the architecture is parameter-efficient.

4.5 Experiment 5: Component Ablation

We ablate three components of CLIFFORD MAMBA on the N-body task ($n=5000$, $N=5$, 3 seeds):

The geometric product contributes a consistent 13% improvement over scalar SSM despite using $4.3\times$ more parameters (Table 4), confirming that the benefit comes from geometric inductive bias rather than capacity. Morton ordering provides minimal benefit at $N=5$, which is expected—with only 5 points, ordering effects are negligible. We expect ordering to matter more at larger N where spatial locality is important. Bidirectional scanning contributes 7%, consistent with the need to aggregate information in both directions

Table 3: Iso-parameter comparison: MSE ($\times 10^{-6}$) at matched parameter counts.

Model	Params	$n=500$	$n=1000$	$n=5000$
CM-medium	865K	6.43	2.40	0.72
CM-small	560K	10.5	4.59	1.00
EGNN	566K	17.9	8.63	5.19
HC-Net	850K	42.3	33.8	39.6

Table 4: Component ablation. Each row removes one component from the full model.

Variant	Component removed	Params	MSE ($\times 10^{-6}$)
Full model	—	6.39M	0.80 ± 0.04
Scalar SSM	Geometric product \rightarrow scalar multiply	1.48M	0.90 ± 0.04 (+13%)
Random order	Morton sort \rightarrow random permutation	6.39M	0.81 ± 0.03 (+1%)
Unidirectional	Backward scan removed	3.19M	0.86 ± 0.04 (+7%)

along the Z-curve.

4.6 Experiment 6: Training Curves

Figure 4 shows per-epoch train and test MSE for all three models ($n=5000$, single seed). CLIFFORD MAMBA converges to the lowest final loss (7.8×10^{-7}) but exhibits a noisier test curve than EGNN, likely due to the larger model’s sensitivity to evaluation batches. HC-Net plateaus at $\sim 4 \times 10^{-5}$, two orders of magnitude higher.

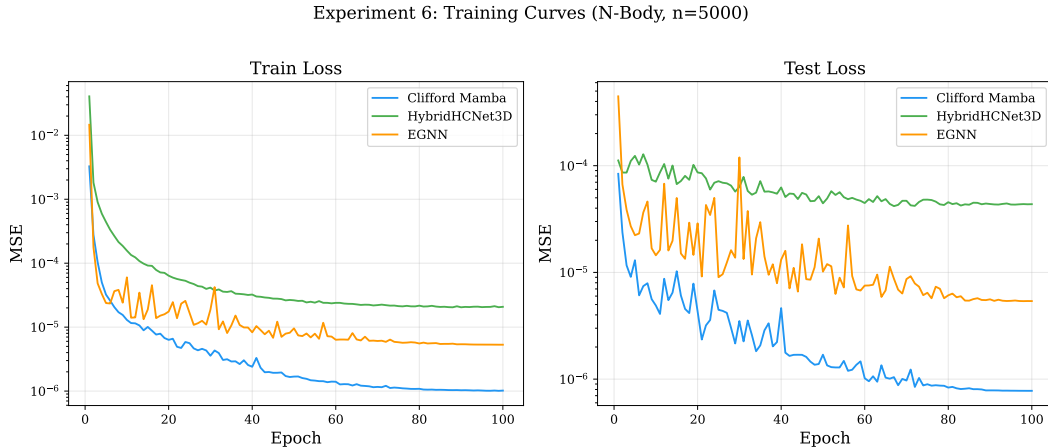


Figure 4: Experiment 6: Training curves ($n=5000$). CLIFFORD MAMBA converges to the lowest loss.

4.7 Experiment 7: Generalization to Larger Systems

We evaluate on larger particle systems ($N=20$ and $N=50$) with $n=2000$ training samples (Table 5).

CLIFFORD MAMBA maintains its advantage at both system sizes, outperforming EGNN by 35% at $N=20$ and 18% at $N=50$ (Figure 5). HC-Net’s local message passing struggles with larger systems where global interactions dominate.

Table 5: Generalization to larger systems: MSE ($\times 10^{-6}$), 3 seeds.

Model	$N=20$	$N=50$
CLIFFORD MAMBA	3.09 ± 0.09	7.66 ± 0.30
EGNN	4.78 ± 0.17	9.31 ± 0.26
HC-Net	51.3 ± 7.1	55.8 ± 5.4

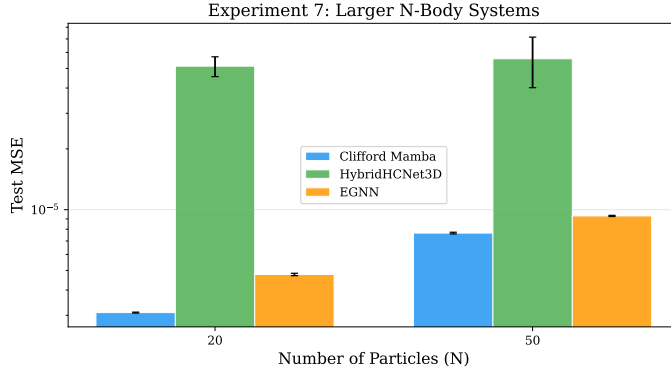


Figure 5: Experiment 7: Larger N-body systems.

4.8 Experiment 8: Memory and Runtime Profiling

We measure peak GPU memory and forward+backward time as a function of N (Figure 6, batch size 4, $d_{\text{model}}=64$, 3 layers).

Table 6: Memory and runtime at selected N values.

N	Model	Fwd (ms)	Bwd (ms)	Mem (MB)
100	CLIFFORD MAMBA	287	405	862
	EGNN	11	8	354
	HC-Net	5	20	44
1000	CLIFFORD MAMBA	2,003	12,311	8,411
	EGNN	232	136	32,844
	HC-Net	7	22	290
2000	CLIFFORD MAMBA	3,202	45,915	16,787
	HC-Net	10	9	776

CLIFFORD MAMBA has $\mathcal{O}(N)$ memory scaling: 107MB at $N=10$ to 16.8GB at $N=2000$, a $157\times$ increase for a $200\times$ increase in N . EGNN’s $\mathcal{O}(N^2)$ memory reaches 32.8GB at $N=1000$ and runs out of memory at $N=2000$ on an 80GB A100. HC-Net is the most memory-efficient due to fixed- k neighbor lookups.

The primary bottleneck is CLIFFORD MAMBA’s backward pass: at $N=2000$, the backward takes 45.9s due to sequential geometric product autograd through the scan. This is a known limitation of sequential SSM implementations; a custom CUDA kernel for the Clifford parallel scan (analogous to Mamba’s selective scan kernel) would substantially reduce this cost.

5 Related Work

Geometric deep learning on point clouds. PointNet [7] and PointNet++ [8] process point clouds as unordered sets. Equivariant architectures—EGNN [10], SchNet [11], PaiNN [12], TFN [13]—build equivariance into message passing. Our approach is complementary: rather than constraining the architecture, we

Experiment 8: Memory and Runtime Profiling

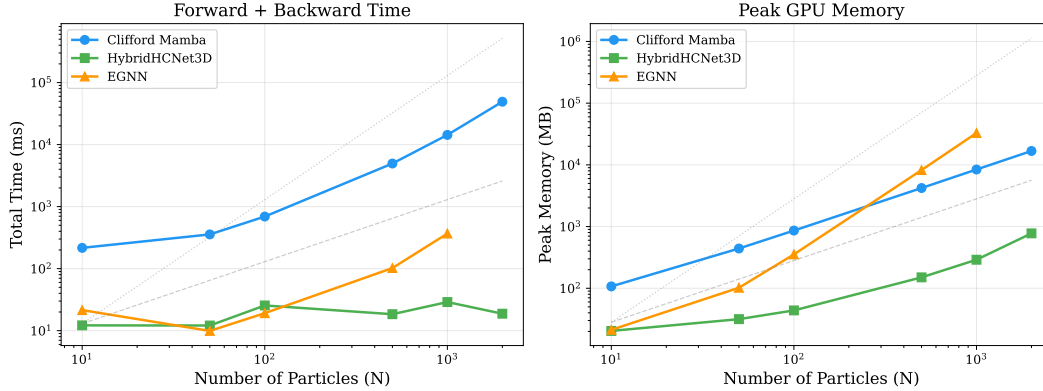


Figure 6: Experiment 8: Runtime and memory scaling. CLIFFORD MAMBA achieves $\mathcal{O}(N)$ memory; EGNN OOMs at $N=2000$.

embed geometric structure into the state representation.

Clifford algebra in neural networks. Clifford Group Equivariant Neural Networks (CGENN) [9] use the Clifford group for equivariant layers. Geometric Clifford Algebra Networks [2] parameterize layers with multivector operations. HC-Net [5] identifies the grade hierarchy for mean-field aggregation. We extend this line by embedding Clifford algebra into SSM state transitions rather than graph convolutions.

State space models for spatial data. Mamba [4] introduced selective state spaces with input-dependent parameters. Point Mamba [6] and related works apply Mamba to point clouds using Hilbert curves or octree orderings, but retain scalar state transitions. Our work is, to our knowledge, the first to use geometric product state transitions in an SSM.

6 Discussion and Limitations

Strengths. CLIFFORD MAMBA achieves state-of-the-art accuracy on N-body prediction while maintaining $\mathcal{O}(N)$ memory. The geometric product state transition provides a principled inductive bias that improves over scalar SSMs, and the architecture preserves the grade hierarchy of $\text{Cl}(3, 0)$.

Limitations. (1) *Runtime.* The sequential geometric product scan has high constant factors; a fused CUDA kernel for Clifford parallel scan would be necessary for practical deployment. (2) *Equivariance.* CLIFFORD MAMBA is not strictly equivariant—Morton ordering breaks exact rotation equivariance. In practice, OOD rotation tests show robust generalization (Table 2), but formal equivariance guarantees require invariant ordering schemes. (3) *Small ablation margins.* At $N=5$, component ablation shows modest 7–13% improvements. Larger- N ablations would better demonstrate the importance of spatial ordering. (4) *Synthetic benchmarks only.* We evaluate on charged-particle simulations; validation on molecular dynamics (MD17, MD22) and real point cloud datasets (ModelNet40, ScanObjectNN) is needed.

Future work. Custom CUDA kernels for the Clifford parallel scan, extension to $\text{Cl}(3, 1)$ (Minkowski spacetime) for relativistic dynamics, and application to molecular force prediction are natural next steps.

7 Conclusion

We introduced CLIFFORD MAMBA, a state space model with geometric product state transitions in $\text{Cl}(3, 0)$. By replacing scalar element-wise recurrence with contractive rotor updates, the SSM hidden state naturally encodes 3D geometric information—rotations, angular momenta, and chirality. Combined with Morton Z-order linearization and bidirectional scanning, this yields an architecture that achieves $7.3\times$ lower prediction error than EGNN on N-body dynamics, maintains $\mathcal{O}(N)$ memory scaling, and wins even at matched parameter counts. The key insight—that the geometric product’s associativity enables direct use of parallel scan

algorithms—opens a path toward geometrically-aware sequence models for scientific computing.

References

- [1] G. E. Blelloch. Prefix sums and their applications. Technical Report CMU-CS-90-190, Carnegie Mellon University, 1990.
- [2] J. Brandstetter, R. Berg, M. Welling, and J. K. Gupta. Clifford neural layers for PDE modeling. In *ICLR*, 2023.
- [3] J. Gastegger, J. Groß, and S. Günnemann. Directional message passing for molecular graphs. In *ICLR*, 2020.
- [4] A. Gu and T. Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.
- [5] S. Kang. HC-Net: Scalable geometric deep learning via Clifford algebra grade hierarchy. *arXiv preprint*, 2026.
- [6] D. Liang *et al.* PointMamba: A simple state space model for point cloud analysis. In *NeurIPS*, 2024.
- [7] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *CVPR*, 2017.
- [8] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *NeurIPS*, 2017.
- [9] D. Ruhe, J. Brandstetter, and P. Forré. Clifford group equivariant neural networks. In *NeurIPS*, 2023.
- [10] V. G. Satorras, E. Hoogeboom, and M. Welling. E(n) equivariant graph neural networks. In *ICML*, 2021.
- [11] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NeurIPS*, 2017.
- [12] K. T. Schütt, O. T. Unke, and M. Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *ICML*, 2021.
- [13] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.