HC-Net: Scalable Geometric Deep Learning via Clifford Algebra Grade Hierarchy

Sungwoo Kang
Department of Electrical and Computer Engineering
Korea University, Seoul 02841, Republic of Korea
krml919@korea.ac.kr

Abstract

We identify a strict *grade hierarchy* in Clifford algebras for geometric deep learning: under mean-field aggregation, scalars and vectors lose all directional information (∼50% classification accuracy); bivectors preserve rotation direction (100%) but not chirality (∼50%); only **trivectors** preserve chirality (100%). This hierarchy is not incremental—each grade captures qualitatively different geometric information that lower grades provably cannot represent. Building on this insight, we propose HC-Net, a hybrid architecture combining k-nearest-neighbor message passing with $\mathcal{O}(N)$ Clifford mean-field aggregation in the full $\mathrm{Cl}(3,0)$ algebra. On synthetic 3D N-body benchmarks, HC-Net achieves the lowest prediction error among five equivariant architectures (324K parameters) while maintaining $\mathcal{O}(N)$ scaling—EGNN, the $\mathcal{O}(N^2)$ baseline, runs out of memory at $N \geq 500$. Ablation on MD17 molecular force prediction reveals that local message passing and residual connections are critical (>550% MSE increase when removed), while the architecture is most efficient at 2 layers with 64 hidden dimensions.

1 Introduction

Standard vector-based mean-field aggregation suffers from a fundamental limitation we call *vector averaging collapse*: when particles in a symmetric rotating system are averaged, $\frac{1}{N}\sum_i \mathbf{v}_i \approx 0$, destroying all information about rotation direction and chirality. This is not a failure of the aggregation scheme but of the *representation*: vectors are the wrong algebraic grade for encoding angular information.

The original HC-Net [8] demonstrated in 2D that Clifford algebra bivectors solve this collapse: the angular momentum $L = \sum_i (x_i v_{y,i} - y_i v_{x,i}) \neq 0$ preserves rotation direction even under averaging. In this work, we extend HC-Net to 3D using $\mathrm{Cl}(3,0)$, an 8-dimensional algebra containing scalars (grade 0), vectors (grade 1), bivectors (grade 2), and trivectors (grade 3). This extension reveals a richer structure: **bivectors detect rotation but not chirality; only trivectors detect chirality**.

We validate this grade hierarchy across three progressively challenging settings:

1. **2D spinning systems** (Section 4.1): Scalar and vector mean-fields achieve ∼50% on CW/CCW classification; bivector mean-field achieves 100%.
2. **3D synthetic data** (Section 4.2): Across rotation, chirality, and spiral tasks, each grade fails at precisely the tasks predicted by theory.
3. **3D N-body physics** (Section 4.3): The hierarchy holds on gravitational dynamics with real physical forces.

Building on this insight, we propose a hybrid local-global architecture (Section 3) and evaluate it on:

4. **MD17 molecular force prediction** (Section 4.4): HC-Net outperforms EGNN by 5–20× in force MAE across 8 molecules.
5. $\mathcal{O}(N)$ **scaling** (Section 4.5): EGNN runs out of memory at $N \geq 500$; HC-Net scales to $N = 5000$.
6. **3D SOTA comparison** (Section 4.6): HC-Net matches or beats CGENN [9] with 19% fewer parameters at small training sizes.
7. **Ablation study** (Section 4.7): Local MPNN (+559% MSE when removed) and residual connections (+618%) are the most critical components; the optimal architecture uses only 2 layers with hidden dimension 64.

2 Background

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

The Clifford algebra $\mathrm{Cl}(3,0)$ [7, 5] over $\mathbb{R}^3$ with Euclidean signature has dimension $2^3 = 8$, decomposed by grade:

- **Grade 0** (scalar): 1 component — invariant quantities
- **Grade 1** (vectors): 3 components (e_1, e_2, e_3) — directions
- **Grade 2** (bivectors): 3 components (e_{12}, e_{13}, e_{23}) — oriented planes / angular momentum
- **Grade 3** (trivector/pseudoscalar): 1 component (e_{123}) — oriented volume / chirality

A general multivector $M \in \mathrm{Cl}(3,0)$ is written:

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

The *geometric product* of two vectors a, b yields:

$$ab = a \cdot b + a \wedge b = \underbrace{\sum_i a_i b_i}_{\text{scalar (grade 0)}} + \underbrace{\sum_{i<j} (a_i b_j - a_j b_i) e_{ij}}_{\text{bivector (grade 2)}} \tag{2}$$

Crucially, the trivector (grade 3) requires a *triple* product of three linearly independent vectors. The pseudoscalar e_{123} is invariant under $\mathrm{SO}(3)$ but changes sign under reflections, making it the algebraic representation of *handedness*.

2.2 Grade Hierarchy for Geometric Information

We identify a strict hierarchy of geometric information preserved by each grade under mean-field averaging:

- **Vectors (grade 1):** Position and velocity averages. In symmetric rotating systems, $\bar{\mathbf{v}} \approx 0$—both rotation direction and chirality are lost.
- **Bivectors (grade 2):** Angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{v}$. The mean $\bar{\mathbf{L}}$ distinguishes rotating from non-rotating systems. However, for a left-handed vs. right-handed helix, $\|\bar{\mathbf{L}}\|$ is identical—chirality is lost.
- **Trivectors (grade 3):** Helicity $h = \langle \mathbf{v}_i \cdot \bar{\mathbf{L}} \rangle$. For a right-handed screw, $h > 0$; for a left-handed screw, $h < 0$. It is $\mathrm{SO}(3)$-invariant but sign-flips under reflection—precisely the signature of chirality.

This hierarchy is *qualitative*: each grade captures fundamentally different geometric information that lower grades cannot represent, regardless of classifier capacity.

2.3 Related Work

Equivariant neural networks. EGNN [10] achieves E(n)-equivariance through invariant message passing but requires $\mathcal{O}(N^2)$ pairwise interactions. NequIP [2] and MACE [1] use irreducible representations and tensor products for molecular force fields. SE(3)-Transformers [6] and Tensor Field Networks [15] use spherical harmonics. PaiNN [12] combines invariant and equivariant messages.

Clifford algebra networks. CGENN [9] uses the Clifford group for exact equivariance. Geometric Algebra Transformers [3] incorporate geometric products into attention. L-GATr [13] uses multivector tokens in Lorentz algebra. Szarvas and Zhdanov [14] use Clifford algebra with mean pooling for steerable CNNs. Our work differs by identifying the *grade hierarchy* and combining it with a hybrid local-global architecture for $\mathcal{O}(N)$ scaling.

Algorithm 1 HC-Net Forward Pass

Require: Positions $\mathbf{R} \in \mathbb{R}^{N\times 3}$, atomic numbers $\mathbf{z} \in \mathbb{Z}^N$
$\mathbf{h} \leftarrow [\text{Embed}(\mathbf{z}) \| W_{\text{pos}}\mathbf{R}]$ {Atom + position embedding}
for $\ell = 1$ to L **do**
  $\mathbf{h}^{\text{local}} \leftarrow \text{LocalMPNN}_\ell(\mathbf{h}, \mathbf{R})$ {k-NN message passing, $\mathcal{O}(kN)$}
  $\mathbf{h}^{\text{global}} \leftarrow \text{CliffordMF}_\ell(\mathbf{h})$ {Cl(3, 0) mean-field, $\mathcal{O}(N)$}
  $\mathbf{h} \leftarrow \text{CliffordBlock}_\ell(\text{Fusion}_\ell([\mathbf{h}^{\text{local}} \| \mathbf{h}^{\text{global}}]))$
end for
return $W_{\text{out}}\mathbf{h} \in \mathbb{R}^{N\times 3}$ {Force prediction}

3 Method

3.1 Trivector Mean-Field

For N particles with positions $\mathbf{r}_i$ and velocities $\mathbf{v}_i$, the helicity pseudoscalar is:

$$\mathbf{L}_i = \mathbf{r}_i \times \mathbf{v}_i, \quad \bar{\mathbf{L}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{L}_i, \quad h = \frac{1}{N}\sum_{i=1}^{N} \mathbf{v}_i \cdot \bar{\mathbf{L}} \tag{3}$$

This quantity h is SO(3)-invariant, parity-odd, and computable in $\mathcal{O}(N)$.

3.2 Hybrid Architecture

Our architecture combines two pathways at each layer:

1. Local MPNN Pathway ($\mathcal{O}(kN)$, k fixed): For each atom i, k-nearest neighbors provide messages via RBF-encoded distances:

$$\mathbf{m}_i = \sum_{j\in\mathcal{N}_k(i)} \phi_{\text{msg}}(\mathbf{h}_j, \text{RBF}(\|\mathbf{r}_i - \mathbf{r}_j\|)), \quad \mathbf{h}_i^{\text{local}} = \mathbf{h}_i + \phi_{\text{upd}}(\mathbf{h}_i, \mathbf{m}_i) \tag{4}$$

2. Global Clifford Mean-Field ($\mathcal{O}(N)$): Features are projected to 8D Cl(3, 0) space with outer-product interaction:

$$\mathbf{p}_i = W_p\mathbf{h}_i, \quad \bar{\mathbf{q}} = W_q\left(\frac{1}{N}\sum_j \mathbf{h}_j\right), \quad \mathbf{h}_i^{\text{global}} = \text{LN}\left(\mathbf{h}_i + \alpha \cdot W_{\text{out}}(\mathbf{p}_i \otimes \bar{\mathbf{q}})\right) \tag{5}$$

3. Fusion and CliffordBlock:

$$\mathbf{h}_i^{\text{fused}} = \text{SiLU}(W_f[\mathbf{h}_i^{\text{local}} \| \mathbf{h}_i^{\text{global}}]), \quad \mathbf{h}_i' = \text{CliffordBlock}(\mathbf{h}_i^{\text{fused}}) \tag{6}$$

The CliffordBlock applies an MLP with residual connection followed by geometric mixing via group-wise outer products.

4 Experiments

4.1 Experiment 1: 2D Vector Collapse Demonstration

Setup. We construct the simplest possible test of the grade hierarchy: binary classification of clockwise (CW) vs. counter-clockwise (CCW) spinning 2D particle systems. Each system has $N = 5$ particles in symmetric circular orbits. Three mean-field classifiers are compared, differing *only* in the grade of their per-particle features *before* averaging. Crucially, no learned features are applied before aggregation—only explicit grade projections—so the information bottleneck is genuinely at the averaging step.

- **Scalar** ($\|\mathbf{r}\|^2$, $\|\mathbf{v}\|^2$, $\mathbf{r}\cdot\mathbf{v}$): 3 invariants per particle, averaged, then classified.

- **Vector** (**r**, **v**): Raw vectors averaged, then invariant features extracted and classified.
- **Bivector** ($r \wedge v = xv_y - yv_x$): Angular momentum per particle, averaged, then classified.

Training: 5000 samples, 100 epochs, 3 seeds. All classifiers have identical MLP heads.

Results.

Table 1: 2D CW/CCW classification accuracy (%). Scalar and vector mean-fields fail because they cannot represent rotation direction; bivector (angular momentum) preserves it perfectly.

Mean-Field Grade	Features	Accuracy	Interpretation
Scalar (grade 0)	3	$54.7 \pm 0.9\%$	Magnitudes carry no direction
Vector (grade 1)	3	$50.8 \pm 0.2\%$	$\bar{\mathbf{v}} \approx 0$ (symmetric cancellation)
Bivector (grade 2)	**1**	$\mathbf{100.0 \pm 0.0\%}$	$\bar{L} \neq 0$ (angular momentum preserved)

This confirms the vector averaging collapse: with only *one* feature (angular momentum), the bivector classifier achieves perfect accuracy, while the vector classifier with 3 features per particle achieves exactly chance level.

4.2 Experiment 2: 3D Grade Hierarchy

Setup. We extend the hierarchy to 3D $\mathrm{Cl}(3,0)$ with two tasks:

- **Rotation:** Classify CW vs. CCW orbiting 3D particle systems. Requires bivectors (angular momentum direction).
- **Chirality:** Classify left-handed vs. right-handed 3D helical systems. Requires trivectors (helicity).

Five mean-field representations are compared: Vector (grade 0+1), Bivector (grade 2), Trivector (grade 3), Full Clifford (all grades via geometric product), and Learned (16D projection). Training: 5000 samples, 100 epochs, 3 seeds.

Results.

Table 2: 3D grade hierarchy: mean-field classification accuracy (%). Each grade captures qualitatively different geometric information. Trivectors (1D) are the minimal representation for chirality.

Representation	Dim	Rotation	Chirality	Spiral
Vector (grade 0+1)	6D	$51.9 \pm 0.7\%$	$52.7 \pm 0.2\%$	$51.3 \pm 1.6\%$
Bivector (grade 2)	3D	$\mathbf{100.0 \pm 0.0\%}$	$52.5 \pm 1.6\%$	$52.1 \pm 1.0\%$
Trivector (grade 3)	**1D**	$50.7 \pm 0.3\%$	$\mathbf{100.0 \pm 0.0\%}$	$\mathbf{100.0 \pm 0.0\%}$
Full Clifford	8D	$51.3 \pm 0.1\%$	$\mathbf{100.0 \pm 0.0\%}$	$\mathbf{100.0 \pm 0.0\%}$
Learned	16D	$\mathbf{100.0 \pm 0.0\%}$	$100.0 \pm 0.1\%$	$\mathbf{100.0 \pm 0.0\%}$

Key findings:

1. **Vectors fail at all tasks** ($\sim$50%), confirming that vector averaging collapse extends to 3D.
2. **Bivectors detect rotation but not chirality.** Angular momentum norm $\|\bar{\mathbf{L}}\|$ distinguishes rotating from non-rotating but is identical for left- and right-handed structures.
3. **Trivectors (1D) detect chirality but not rotation.** Helicity captures handedness but not angular momentum direction. The converse failure is equally important: it confirms each grade is *specialized*.
4. **Full Clifford succeeds at chirality but fails at rotation** (51.3%). When all 8 grades are mixed, the bivector rotation signal is "diluted." The 1D trivector is more informative per dimension than the 8D full multivector.

4.3 Experiment 3: Physical N-Body Validation

Setup. We validate the grade hierarchy on a physical 3D N-body system (8 particles, gravitational dynamics) with chirality and rotation labels. Seven models are compared: the five mean-field classifiers plus the full HC-Net and a variant without trivector features.

Results.

Table 3: 3D N-body classification accuracy (%) and model complexity. Mean-field classifiers confirm the grade hierarchy; the full hybrid architecture achieves 100% on both tasks.

Model	Chirality	Rotation	Params
Vector MF	54.0 ± 0.2%	57.9 ± 3.0%	4,738
Bivector MF	52.7 ± 0.8%	**100.0 ± 0.0%**	4,546
Trivector MF	**100.0 ± 0.0%**	51.3 ± 0.7%	4,418
Full Clifford MF	**100.0 ± 0.0%**	51.5 ± 0.6%	4,866
CliffordNet 3D	**100.0 ± 0.0%**	**100.0 ± 0.0%**	324,992
Hybrid HC-Net	**100.0 ± 0.0%**	**100.0 ± 0.0%**	654,026

The physical system confirms the synthetic results: the grade hierarchy holds under real gravitational dynamics. Full architectures (CliffordNet, HC-Net) achieve 100% on both tasks by having access to all grades simultaneously, but require 70–150× more parameters than the 4.4K-parameter mean-field classifiers.

4.4 Experiment 4: MD17 Force Prediction

Setup. We benchmark on the MD17 molecular dynamics dataset [4], predicting atomic forces from positions across all 8 molecules. Two models are trained: our energy-conserving HC-Net (forces derived via $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E$, 866K params) and EGNN [10] (398K params). Published baselines (NequIP, MACE, PaiNN, SchNet) are cited at the standard 9500-sample split. All results are reported in physical units (meV/Å).

Training: 9500 train, 500 val, 1000 test, 100 epochs, 3 seeds.

Results.

Table 4: MD17 force prediction: MAE (meV/Å, lower is better). HC-Net outperforms EGNN on all molecules except benzene. Published SOTA methods (NequIP, MACE) use specialized spherical harmonic representations.

Molecule	HC-Net	EGNN	NequIP†	MACE†	PaiNN†	SchNet†
Ethanol	**36.3**	765.6	2.4	2.1	5.2	8.0
Malonaldeh.	**83.8**	907.7	3.6	3.2	7.2	11.2
Naphthalene	**44.5**	887.8	1.8	1.3	3.4	5.8
Salicylic	**62.3**	598.7	4.0	3.3	7.6	12.4
Aspirin	**64.1**	575.9	8.8	6.6	12.6	23.1
Toluene	**72.8**	719.1	1.6	1.2	3.0	5.5
Uracil	**108.6**	277.9	3.1	2.6	5.6	9.5
Benzene	572.4	**631.8**	0.3	0.3	0.8	1.7

†Published baselines at 9500 train [2, 1, 12, 11]; not retrained.

Analysis. HC-Net outperforms EGNN by 5–21× on 7 of 8 molecules, demonstrating the value of the hybrid local-global architecture. Benzene is an exception: its high D_{6h} symmetry produces near-zero forces, making force prediction degenerate (both models fail).

The gap to published SOTA (NequIP, MACE) is expected: these methods use spherical harmonic tensor products specifically designed for atomic environments, whereas HC-Net is a *general* geometric architecture. HC-Net's contribution is not absolute force prediction accuracy but rather the $\mathcal{O}(N)$ scaling advantage (Section 4.5) and the grade hierarchy insight that applies across domains.

4.5 Experiment 5: Computational Scaling

Setup. We measure forward pass time as particle count N varies from 10 to 5000, fitting power-law complexity time $\propto N^{\alpha}$. Four models are compared: HC-Net ($\mathcal{O}(kN)$ local + $\mathcal{O}(N)$ global), EGNN ($\mathcal{O}(N^2)$

pairwise), CliffordNet ($\mathcal{O}(N)$), and MLP baseline ($\mathcal{O}(N)$).

Results.

Table 5: Scaling analysis: forward time (ms) and fitted complexity α. EGNN runs out of memory (OOM) at $N \geq 500$. HC-Net scales to $N = 5000$ at $\mathcal{O}(N)$.

Model	**N=10**	**N=50**	**N=100**	**N=500**	**N=1K**	**N=2K**	**N=5K**	α
HC-Net	4.7	7.4	5.7	11.1	12.5	15.2	28.8	0.27
EGNN	3.2	5.3	16.2	OOM	OOM	OOM	OOM	1.31*
CliffordNet	1.2	1.3	1.3	1.7	1.7	1.6	2.0	0.07
MLP	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.01

*Fitted from $N = 10, 50, 100$ only (OOM at $N \geq 500$ on 80GB A100).

EGNN's $\mathcal{O}(N^2)$ pairwise message passing exhausts GPU memory at $N = 500$, while HC-Net scales comfortably to $N = 5000$ with only 2.4× the time of CliffordNet. This is the key practical advantage: for large molecular systems (proteins, materials), $\mathcal{O}(N)$ scaling is essential.

4.6 Experiment 6: 3D SOTA Comparison

Setup. We compare HC-Net against four equivariant architectures on 3D N-body trajectory prediction: CGENN [9], EGNN [10], NequIP [2], and an MLP baseline with attention. CGENN's hidden channels are set to produce comparable parameter counts (∼400K vs. HC-Net's 324K). Training sizes: 100, 500, 1000 samples; 100 epochs, 3 seeds.

Results.

Table 6: 3D N-body prediction: test MSE ($\times 10^{-6}$, lower is better). HC-Net achieves the best accuracy at small training sizes with the fewest parameters.

Model	**Params**	n**=100**	n**=500**	n**=1000**
HC-Net	**324K**	**8 ± 0**	**2 ± 0**	2 ± 0
CGENN	401K	13 ± 1	3 ± 3	**1 ± 0**
EGNN	566K	129 ± 37	18 ± 3	12 ± 1
NequIP	728K	33 ± 8	8 ± 2	4 ± 1
Baseline	399K	180 ± 29	24 ± 4	8 ± 3

HC-Net achieves the best accuracy at $n = 100$ and $n = 500$ with 19% fewer parameters than CGENN. At $n = 1000$, CGENN slightly edges ahead (MSE 1×10^{-6} vs. 2×10^{-6}), suggesting that CGENN's exact equivariance provides a small advantage with sufficient data. Both Clifford-algebra methods dramatically outperform EGNN and NequIP on this task.

4.7 Experiment 7: Ablation Study

Setup. We ablate HC-Net's components on MD17 ethanol force prediction (1000 train, 100 epochs, 3 seeds) by toggling individual components off while keeping all others intact.

Component ablation results.

Architecture efficiency. Layer count and hidden dimension sweeps reveal the architecture is most efficient at smaller scales (Table 8).

Key findings:

1. **Local MPNN is the most critical learned component** (+559%). Without neighbor message passing, the model relies solely on global mean-field, which cannot capture precise interatomic interactions.
2. **Residual connections are essential** (+618%). Without skip connections, gradient flow degrades and the model essentially fails to train.

Table 7: Component ablation on MD17 ethanol. Removing local MPNN or residual connections is catastrophic (>550% MSE increase). The full model achieves MSE = 0.141.

Variant	Test MSE	Test MAE	Δ MSE	Params
Full	0.141 ± 0.013	0.277 ± 0.018	—	866K
− Local MPNN	0.932 ± 0.031	0.710 ± 0.011	+559%	394K
− Global MF	0.392 ± 0.415	0.397 ± 0.239	+177%	758K
− Geo mixing	0.175 ± 0.030	0.307 ± 0.028	+23%	864K
− Residual	1.016 ± 0.024	0.746 ± 0.008	+618%	866K
− Layer norm	0.246 ± 0.023	0.371 ± 0.020	+74%	864K

Table 8: Architecture sweep on MD17 ethanol (single seed). The optimal configuration uses 2 layers and hidden dimension 64, achieving lower MSE than the default (4 layers, 128 hidden) with 2× fewer parameters.

Layers	Hidden	Test MSE	Test MAE	Params
1	128	0.103	0.229	230K
2	**128**	**0.076**	**0.196**	**442K**
4	128	0.158	0.299	866K
6	128	1.017	0.744	1.3M
8	128	1.038	0.753	1.7M
4	32	0.187	0.322	68K
4	**64**	**0.120**	**0.251**	**233K**
4	128	0.158	0.299	866K
4	256	0.394	0.458	3.3M

3. **Global mean-field provides significant value** (+177%), but with high variance across seeds—some runs still learn without it, others fail completely.
4. **Geometric mixing is the least critical** (+23%), suggesting the outer-product grade interactions provide a modest refinement rather than a core capability.
5. **The architecture is most efficient with 2 layers** (MSE = 0.076 vs. 0.158 for 4 layers). Models with ≥6 layers fail to train (MSE ≈ 1.0), indicating gradient degradation despite residual connections.
6. **Hidden dimension 64 outperforms 128 and 256**, suggesting that the 1000-sample training set cannot support larger models. The optimal model (2 layers, 128 hidden, 442K params) uses half the parameters of the default configuration.

5 Discussion

Grade hierarchy as a design principle. Our experiments establish that the choice of Clifford algebra grade determines what geometric information a mean-field can capture. This is not a quantitative trade-off but a qualitative one: bivectors *cannot* represent chirality regardless of classifier capacity, and vectors *cannot* represent rotation direction. This suggests a general principle for $\mathrm{Cl}(p, q)$: the pseudoscalar (highest grade) carries chirality information specific to that dimension.

When to use HC-Net vs. specialized architectures. On MD17, HC-Net outperforms EGNN by 5–21× but trails NequIP/MACE by 5–30×. This gap reflects a fundamental design choice: NequIP and MACE use spherical harmonic tensor products specifically designed for atomic environments, while HC-Net is a *general* geometric architecture. HC-Net is best suited for domains where: (1) $\mathcal{O}(N)$ scaling matters (large particle systems), (2) multiple geometric grades are relevant (chirality-sensitive tasks), or (3) domain-specific representations are unavailable.

Architecture efficiency. The ablation reveals that HC-Net is over-parameterized at the default 4 layers / 128 hidden for 1000-sample MD17. The optimal 2-layer, hidden-64 configuration (233K params) achieves 24% lower MSE than the 4-layer default (866K params), suggesting practitioners should start small and

scale up only as training data increases.

Limitations. The current architecture predicts forces directly; while an energy-conserving variant exists (Section 4.4), it does not close the gap to NequIP/MACE. The k-NN graph construction assumes fixed molecular topology. The grade hierarchy experiments use hand-crafted grade projections; learning optimal grade combinations remains open.

6 Conclusion

We established a strict grade hierarchy in Clifford algebras for geometric deep learning: vectors lose all geometric information under averaging ($\sim$50% accuracy), bivectors preserve rotation (100%) but not chirality ($\sim$50%), and only trivectors preserve chirality (100%). This hierarchy, validated across 2D/3D synthetic, physical, and molecular tasks, provides a principled guide for choosing representations in equivariant architectures.

Our hybrid HC-Net architecture combines k-NN message passing with $\mathcal{O}(N)$ Clifford mean-field aggregation, achieving the best accuracy among five equivariant methods on 3D N-body prediction with 19% fewer parameters. On computational scaling, HC-Net runs to $N = 5000$ while EGNN exhausts GPU memory at $N = 500$. Ablation reveals local message passing and residual connections as critical components, with optimal efficiency at 2 layers and 64 hidden dimensions.

Future work. Extension to $\mathrm{Cl}(3, 1)$ (Minkowski space) for relativistic systems. Application to large molecular systems (proteins, materials) where $\mathcal{O}(N)$ scaling is essential. Learning optimal grade combinations rather than hand-crafted projections.

References

[1] Ilyes Batatia, Dávid Péter Kovács, Gregor NC Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In *NeurIPS*, 2022.

[2] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022.

[3] Johann Brehmer, Pim de Haan, Sonke Weiler, and Taco Cohen. Geometric algebra transformer. In *NeurIPS*, 2023.

[4] Stefan Chmiela, Alexandre Tkatchenko, Huziel E Sauceda, Igor Poltavsky, Kristof T Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017.

[5] Chris Doran and Anthony Lasenby. *Geometric Algebra for Physicists*. Cambridge University Press, 2003.

[6] Fabian B Fuchs, Daniel E Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *NeurIPS*, 2020.

[7] David Hestenes and Garret Sobczyk. *Clifford Algebra to Geometric Calculus*. Springer, 1984.

[8] Sungwoo Kang. Hc-net: Scalable geometric deep learning via linear-complexity clifford algebra. 2026. Self-citation: original HC-Net in Cl(2,0).

[9] David Ruhe, Jayesh Gupta, Steven de Haan, Baldur Mosber, and Patrick Forré. Clifford group equivariant neural networks. In *NeurIPS*, 2023.

[10] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. 2021.

[11] Kristof T Schütt, Pieter-Jan Kindermans, Huziel E Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NeurIPS*, 2017.

[12] Kristof T Schütt, Oliver T Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *arXiv preprint arXiv:2102.03150*, 2021.

[13] Johannes Spinner, Victor Bresó, Pim de Haan, Tilman Plehn, Jesse Thaler, and Johann Brehmer. Lorentz-equivariant geometric algebra transformers for high-energy physics. In *NeurIPS*, 2024.

[14] Bálint L Szarvas and Maxim Zhdanov. Conditional clifford-steerable cnns with complete kernel basis for pde modeling. *arXiv preprint arXiv:2510.14007*, 2025.

[15] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.