CliffordIP: Clifford Algebra Equivariant Interatomic Potentials for Heterogeneous Catalysis

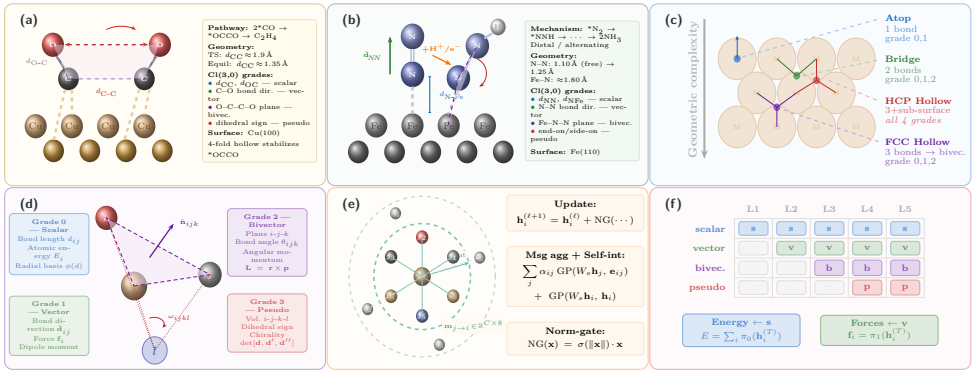
Can Polat[1], Erchin Serpedin[1], Mustafa Kurban[2,3*], Hasan Kurban[4*]

[1]Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA.
[2]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar.
[3]Department of Prosthetics and Orthotics, Ankara University, Ankara, Turkey.
[4]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar.

*Corresponding author(s). E-mail(s): kurbanm@ankara.edu.tr; hkurban@hbku.edu.qa;
Contributing authors: can.polat@tamu.edu; eserpedin@tamu.edu;

Abstract

Decarbonizing the global economy requires efficient catalysts for key electrochemical transformations, including CO_2 reduction, nitrogen reduction, and C–C coupling. Yet systematic catalyst discovery is limited by the cost of density functional theory calculations needed to evaluate adsorbate–surface energetics across chemically diverse surfaces. Machine-learning interatomic potentials (MLIPs) can accelerate this workflow, but many leading equivariant architectures rely on spherical harmonics and Clebsch–Gordan tensor products, increasing computational complexity and implementation overhead. Here we introduce CliffordIP, a message-passing interatomic potential built on the Clifford algebra $\mathbf{Cl(3,0)}$, representing atomic environments as 8-dimensional multivectors—scalars, vectors, bivectors, and pseudoscalars—coupled through the geometric product. This naturally captures bond directions, reaction planes, and local chirality relevant to catalytic intermediates, while enforcing full $\mathbf{O(3)}$ equivariance, including reflections, via the $\mathbf{Pin(3)}$ group. On catalysis-focused adsorbate subsets, CliffordIP reduces energy MAE by 20–24% relative to the next-best baseline on CO_2RR, N_2RR, and C_2 formation, and achieves the strongest in-distribution



energy performance among the compared baselines on OC20 and OC22 under a unified, compute-constrained protocol, while remaining competitive in force magnitude. These results establish Clifford algebra as a promising foundation for energy-focused MLIPs in computational catalysis.

Keywords: machine learning interatomic potentials, Clifford algebra, geometric algebra, equivariant neural networks, heterogeneous catalysis, Open Catalyst

Introduction

Electrocatalyst design is governed by challenges that emerge across multiple length and computational scales. At the atomic level, reaction selectivity can depend sensitively on the geometry of adsorbed intermediates, including fragment orientation, adsorption configuration, and the symmetry of the catalytic site that stabilizes them [1]. At the scale required for catalyst discovery, however, resolving these structural effects across large chemical spaces with first-principles methods remains computationally prohibitive [2]. Addressing the tradeoff between geometric fidelity and high-throughput screening forms the central motivation of this work.

Heterogeneous electrocatalysis proceeds through reaction intermediates whose binding energies and transition-state geometries at the catalyst surface collectively determine selectivity and activity [3, 4]. In CO_2 electroreduction (CO_2RR), the branching between CO, formate, methanol, and multi-carbon products is governed by the relative stability of *CO, *COOH, and *CHO intermediates, whose binding geometries range from linear end-on adsorption to bent bidentate configurations [5, 6]. Electrochemical N_2 reduction (N_2RR) proceeds through a hydrogenation sequence of $^*N_2 \rightarrow {}^*NNH \rightarrow {}^*NHNH$ in which each protonation step alters the molecular orientation at the surface, creating a cascade of geometrically distinct binding modes that control both ammonia yield and selectivity over competing hydrogen evolution [7, 8]. C–C coupling (C_2), which unlocks multi-carbon products such as ethylene and ethanol from CO_2RR, involves the dimerization of *CO intermediates and the stabilization of multidentate species such as $^*CCH_2$ and *CHCO that span multiple surface adsorption

sites with out-of-plane dihedral character [9, 10]. Across these three reaction families, the same geometric theme recurs: the bent, planar, and dihedral configurations of key intermediates encode the orbital interactions of d-band hybridization, back-donation, and adsorbate rehybridization [11, 12] that ultimately set selectivity and activity.

Rational catalyst design for these reactions requires systematically mapping adsorption energies across the space of surface compositions, facets, and coverages [13]. Density functional theory (DFT) provides accurate energetics but at a cost that renders exhaustive screening prohibitive: even the Open Catalyst 2020 (OC20) dataset [14] covers only a small fraction of the viable catalyst space. The Open Catalyst 2022 (OC22) dataset [15] extends the scope to oxide electrocatalyst surfaces, adding the chemical diversity essential for oxygen evolution and reduction reactions. Machine learning interatomic potentials (MLIPs) address this bottleneck by learning to approximate the potential energy surface directly from DFT data, achieving speedups of 10^3–$10^6\times$ over explicit electronic-structure calculations [16, 17]. In the context of computational catalysis, MLIPs enable structural relaxations, adsorption energy predictions, and reaction pathway searches at scales previously inaccessible to DFT.

The pursuit of accurate and generalizable MLIPs has driven a rapid evolution of neural network architectures for atomic systems, progressing through three generations of increasing geometric sophistication. The first generation of invariant models exemplified by crystal graph convolutional neural networks [18], SchNet [19], which introduced continuous-filter convolutions on interatomic distances. DimeNet [20], which incorporated angular information via triplet interactions operated on scalar (rotation-invariant) features and were therefore blind to the directional character of interatomic interactions. The second generation introduced vector-equivariant representations: PaiNN [21] augmented scalar features with equivariant vector channels that transform under SO(3); E(n)-equivariant graph neural networks [22] extended equivariance to the full Euclidean group while avoiding higher-order tensor representations; and TorchMD-Net [23] combined equivariant message passing with attention mechanisms. The third and current generation employs higher-order equivariant representations based on irreducible representations (irreps) of SO(3) or O(3): Cormorant [24] and NequIP [25] use Clebsch–Gordan (CG) tensor products with learnable radial functions; MACE [26] and Allegro [27] introduce many-body message passing through iterated tensor products; SEGNNs [28] demonstrated steerable equivariant message passing with physical quantities; SE(3)-transformers [29] first combined equivariant features with dot-product attention, a concept extended by Equiformer [30] to equivariant graph attention; and GotenNet [31] proposes a geometric O(3)-equivariant tensor network with efficient tensor contractions that avoids CG coefficients entirely.

Beyond this generational arc, two further developments shape the current MLIP landscape. The atomic cluster expansion [32] provides a polynomial, systematically improvable many-body basis that is equivariant by construction, underpinning the parameterisation strategy of several leading models [33]. Alongside it, universal interatomic potentials trained on broad materials databases demonstrate transferability across wide chemical spaces, from graph-network approaches covering the periodic

table [34] to large-scale foundation models trained on millions of DFT structures spanning diverse chemistries [35, 36]. A consistent empirical finding across these generations is that richer geometric representations yield improved accuracy on catalysis benchmarks, but at the cost of increasing architectural complexity and computational overhead [37].

The technical cost of those richer representations stems from a specific algebraic choice. The total energy of a system of atoms is invariant under translations, rotations, and reflections (the Euclidean group E(3)), while forces must transform equivariantly [38]. The dominant framework for enforcing these symmetries represents atomic features as irreps of SO(3), labeled by angular momentum number l and interacting via CG tensor products [39], whose computational cost scales as $O(l_{\text{max}}^6)$ [40]. Higher l captures finer angular detail (with l=2 encoding d-orbital-like environments and l=3 encoding f-orbital-like features), but incurs steep computational penalties. EquiformerV2 [41] reduces this bottleneck via SO(2) convolutions, yet still requires nontrivial engineering for representations up to l=6. More broadly, Wigner D-matrices and CG coefficients are difficult to optimize with modern automatic differentiation tools, imposing significant implementation complexity [42]. This raises a natural question: is there a mathematically principled alternative that natively encodes bond planes, dihedral angles, and chirality without the algebraic machinery of spherical harmonics?

Clifford algebra — also known as geometric algebra — provides precisely such an alternative [43, 44]. The Clifford algebra $\text{Cl}(3,0)$ encodes scalars, vectors, bivectors, and pseudoscalars in a compact 8-dimensional multivector, with the geometric product serving as a single algebraic operation that unifies inner products, cross products, and rotations by replacing the suite of CG tensor products required by spherical harmonics [45]. Critically for catalysis applications, each grade carries direct physical meaning: grade-0 scalars encode energy-like invariant quantities; grade-1 vectors encode bond directions and forces; grade-2 bivectors encode oriented rotation planes corresponding to bond angles and dihedral environments; and the grade-3 pseudoscalar encodes chirality of the local coordination environment [46]. The bent geometry of adsorbed CO_2 and *COOH, the planar dihedral of *NHNH, and the bridging geometry of C_2 intermediates are all naturally represented by combinations of bivector components that the geometric product mixes in a single closed algebraic operation [47] (Fig. 1). Full O(3) equivariance, including reflections, important for distinguishing enantiomeric surface adsorption sites, is guaranteed by the Pin(3) group structure without Wigner D-matrices or CG coefficients [48]. Clifford algebra has been applied to machine learning on physical systems in n-body dynamics [49] and molecular trajectories [50], but not to interatomic potential energy surfaces for heterogeneous catalysis.

In this work, we introduce CliffordIP: to our knowledge, the first Clifford-algebra-based message-passing neural network designed as an interatomic potential for heterogeneous catalysis. CliffordIP represents each atom's features as $\text{Cl}(3,0)$ multivectors and constructs equivariant messages via the geometric product, achieving full O(3) equivariance without spherical harmonics. The architecture incorporates several innovations tailored to interatomic potential prediction including progressive

grade activation, grade-sparse geometric product dispatch, L=2 feature augmentation, MACE-style multi-body interactions [26], and equivariant attention [41].

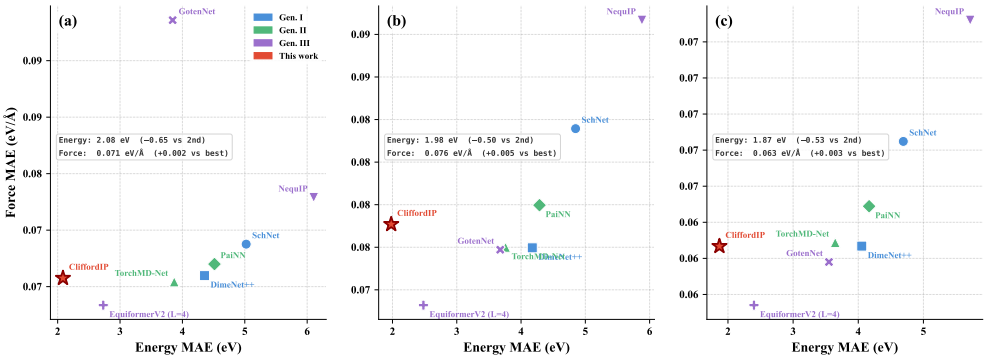
To place CliffordIP within the broader landscape of modern MLIPs, we compare it against twelve state-of-the-art architectures spanning all three generations of geometric learning methods, including invariant graph networks, vector-equivariant models, and higher-order equivariant transformer and tensor-network approaches. The benchmark suite further includes systematic angular-order sweeps over MACE ($L\in\{1,2\}$), ICTP ($L\in\{1,2\}$), and EquiformerV2 ($L\in\{2,4\}$) in order to examine the effect of increasing equivariant complexity. All thirteen architectures are evaluated on the Open Catalyst benchmarks across four tasks: OC20-S2EF, consisting of 1.3 million adsorbate–surface structures with 82 adsorbates on metallic catalysts; OC20-IS2RE, which focuses on direct relaxed-energy prediction; OC22-S2EF, which extends the benchmark to oxide electrocatalyst surfaces; and OC22-IS2RE, the corresponding oxide relaxed-energy prediction task. Cross-domain generalization is further assessed on the rMD17 [51, 52] molecular dynamics benchmark for small organic molecules using an expanded comparison suite to eighteen by incorporating additional higher-order variants of MACE, ICTP, and EquiformerV2, together with FAENet and ViSNet. Our main contributions are:

- We introduce CliffordIP, to our knowledge the first interatomic potential for heterogeneous catalysis built on the Clifford algebra $\mathrm{Cl}(3,0)$, achieving full $\mathrm{O}(3)$ equivariance through the geometric product without spherical harmonics or Clebsch–Gordan tensor products.
- We develop a Clifford-algebraic MLIP architecture that combines progressive grade activation, grade-sparse geometric-product dispatch, $L = 2$ feature augmentation, multi-body interactions, and equivariant attention.
- On catalysis-focused adsorbate subsets of OC20 evaluated without task-specific fine-tuning, CliffordIP achieves the lowest energy MAE among the compared baselines, improving over the next-best model by 24% on CO_2RR, 20% on N_2RR, and 22% on C_2 formation.
- Under this unified, compute-constrained protocol, CliffordIP achieves the strongest in-distribution energy performance among the compared models on both OC20 and OC22 S2EF benchmarks, while ablation analysis identifies attention and multi-scale readout as the dominant contributors to accuracy. Cross-domain evaluation on rMD17 shows the same energy advantage transfers beyond catalysis.

The remainder of this paper is organized as follows. The Results section reports and analyzes the benchmark results. The Discussion section discusses the implications of our findings, directions for future work, and concludes the paper. The Methods section describes the CliffordIP architecture and the experimental protocol.

Results

All models share a single training setup on OC20 (200K split) and OC22 (full set); infrastructure, schedules, and per-model configurations are detailed in the Supplementary Information (Benchmark Settings). Evaluation proceeds in three stages: a



zero-shot analysis on catalysis-focused OC20 adsorbate subsets for CO_2RR, N_2RR, and C_2 formation (eight baselines, the Catalysis-Specific Transfer Performance subsection); the full S2EF and IS2RE benchmarks on OC20 and OC22 (twelve baselines, the S2EF, IS2RE, and OOD-Generalisation subsections); and a cross-domain extension to rMD17 with five additional architectures (the Cross-Domain Robustness on rMD17 subsection). Numerical results across all splits are tabulated in the Supplementary Information (Detailed Experiment Analysis). Figures report mean ± standard deviation over three seeds (four for rMD17).

Catalysis-Specific Transfer Performance

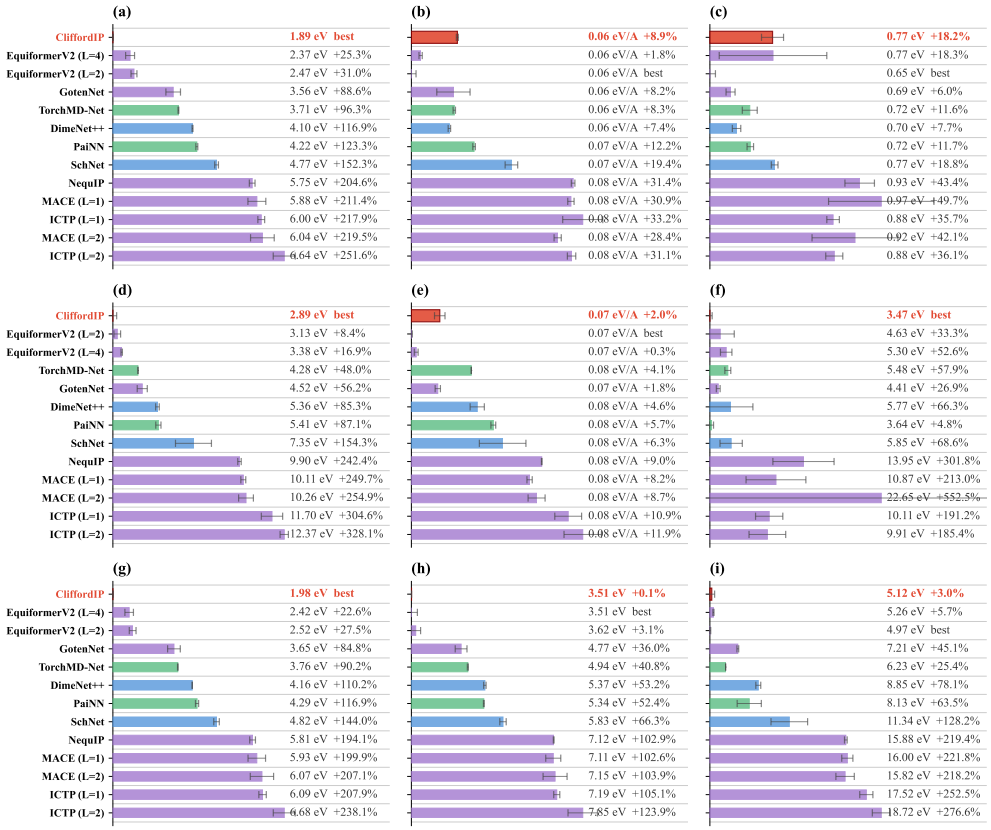
Fig. 2 reports energy and force MAE on the three catalysis subsets, with subset compositions defined in the Methods (Benchmark Settings).

On the 13-adsorbate CO_2RR subset (Fig. 2a), CliffordIP reaches an energy MAE of 2.08 eV against EquiformerV2's 2.73 eV — a 24% reduction. The bent and multi-coordinate species that dominate this set (*COOH, *HCOO, *CHO) are exactly the geometries the bivector channels of $\mathrm{Cl}(3,0)$ are designed to capture, and force MAE remains within the best-baseline cluster.

The pattern repeats on the 13-intermediate N_2RR cascade from $*N_2$ through $*ONNH_2$ (Fig. 2b), where CliffordIP attains 1.98 eV against 2.48 eV for EquiformerV2 — a 20% reduction. The end-on/side-on binding distinction of $*N_2$ and the planar dihedral of *NHNH demand orientational structure beyond pairwise distances, and the geometric product captures both in a single closed operation.

On the ten C–C coupling intermediates of the C_2 subset (Fig. 2c), CliffordIP reaches 1.87 eV versus 2.40 eV — a 22% reduction. Bridge-bonded and multidentate species in this set span multiple surface sites with out-of-plane coordination, the regime in which higher-grade features carry the most information.

Across the three subsets, CliffordIP delivers the lowest energy MAE without any task-specific fine-tuning while remaining within the best force-MAE cluster. The consistency suggests that the geometric product is a useful inductive bias for adsorbate–surface environments characterised by bond planes, multidentate coordination, and local chirality. For high-throughput catalyst screening [53, 54], where energy ranking matters more than directional force fidelity, this profile is well-suited.



Structure-to-Energy-and-Forces (S2EF)

Under the unified, compute-constrained protocol, CliffordIP leads S2EF energy on both datasets among the compared baselines (Fig. 3a, d). On OC20 it reaches 1.889 eV, cutting EquiformerV2's MAE by 20% at L=4 (2.366 eV) and 24% at L=2 (2.474 eV); GotenNet (3.562 eV) trails by 88%, and the MACE and ICTP families by more than 200%. The same ordering holds on OC22, where CliffordIP attains 2.890 eV against 3.133 eV for EquiformerV2 (L=2, +8%), 3.378 eV at L=4 (+17%), and 9.897 eV for NequIP (+242%). The energy advantage thus survives the shift from metallic to oxide surface chemistry.

Force MAE tells a different story (Fig. 3b, e). On OC20 the EquiformerV2 family leads at 0.059–0.060 eV/Å, with CliffordIP within 0.005 eV/Å; on OC22 the top three models are essentially indistinguishable on force magnitude. Force cosine similarity is sharper still — CliffordIP trails EquiformerV2 there (Supplementary Information, Detailed Experiment Analysis), so competitive magnitudes do not translate into equally strong directional fidelity. The picture is that of an energy specialist: top accuracy on energy, competitive but not leading on forces.

Initial-Structure-to-Relaxed-Energy (IS2RE)

IS2RE results split by dataset (Fig. 3c, f). On OC20, EquiformerV2 (L=2) leads at 0.647 eV, with GotenNet (0.686 eV, +6.0%) and DimeNet++ (0.698 eV, +7.7%) close behind; CliffordIP ranks sixth at 0.765 eV (+18.2%) within a tight eight-model cluster spanning ∼0.12 eV. On OC22 the order inverts — CliffordIP ranks first at 3.471 eV, 5% below PaiNN (3.639 eV) and 21% below GotenNet (4.405 eV), with NequIP at 13.946 eV (+302%) and MACE (L=2) at 22.650 eV (+553%). The reversal likely reflects OC22's greater structural diversity in surface–adsorbate configurations, where the richer geometric encoding of $\mathrm{Cl}(3,0)$ provides a stronger prior for predicting relaxed geometries from unrelaxed inputs.

Out-of-Distribution Generalisation and Generational Analysis

OOD splits reproduce the in-distribution picture (Fig. 3g–i). CliffordIP leads OC20 OOD-Ads at 1.977 eV — 18% below EquiformerV2 (L=4, 2.424 eV) and 22% below L=2 (2.520 eV) — with all Generation I–II baselines above 3.7 eV and the MACE/ICTP family above 5.9 eV. On OC22 OOD, EquiformerV2 (L=2, 4.971 eV) edges CliffordIP (5.120 eV, +3.0%) and EquiformerV2 (L=4, 5.256 eV, +5.7%), while NequIP reaches 15.880 eV (+219%). On OC20 OOD-Both, EquiformerV2 (L=4, 3.507 eV) and CliffordIP (3.510 eV) are tied to within 0.1%.

A generational stratification is visible across all nine panels: Generation I (SchNet, DimeNet++, PaiNN) sits in the high-error regime; Generation II (TorchMD-Net, NequIP) occupies an intermediate band; Generation III spans a wide range, with EquiformerV2 leading on forces and the MACE/ICTP family lagging on energy. CliffordIP leads the energy frontier despite being limited to native l=1 plus an L=2 scalar augmentation — a strictly lower native angular order than several spherical-harmonic baselines — suggesting that Clifford algebra is a complementary prior to the attention and higher-order harmonics that characterise the rest of Generation III. Training is stable throughout, with the lowest validation energy MAE at convergence (Supplementary Fig. S1).

Cross-Domain Robustness on rMD17

To verify that the inductive biases of $\mathrm{Cl}(3,0)$ generalise beyond catalysis, we evaluate CliffordIP on the rMD17 small-molecule benchmark across all ten molecules under a 900 / 100 / 1000 train / val / test split with four seeds per molecule. The smaller scale of rMD17 also lets us include FAENet, ViSNet, and GotenNet, which were not feasible to train at OC20/OC22 scale within our compute budget. Aggregate test MAEs are reported in Table 1; per-molecule values appear in Supplementary Table S7.

Three patterns emerge. CliffordIP ranks second in aggregate energy MAE (0.034 eV), behind only FAENet (0.024 eV) and ahead of every remaining baseline by more than an order of magnitude — the third-place TorchMD-Net (0.711 eV) trails by 20×. On aggregate force MAE, CliffordIP and FAENet tie at 0.021 eV/Å, both an order of magnitude behind the EquiformerV2 family at ∼0.004 eV/Å across $L_{\max}\in\{2,3,4\}$ — the energy/force split observed on OC20/OC22 thus reproduces in another chemistry.

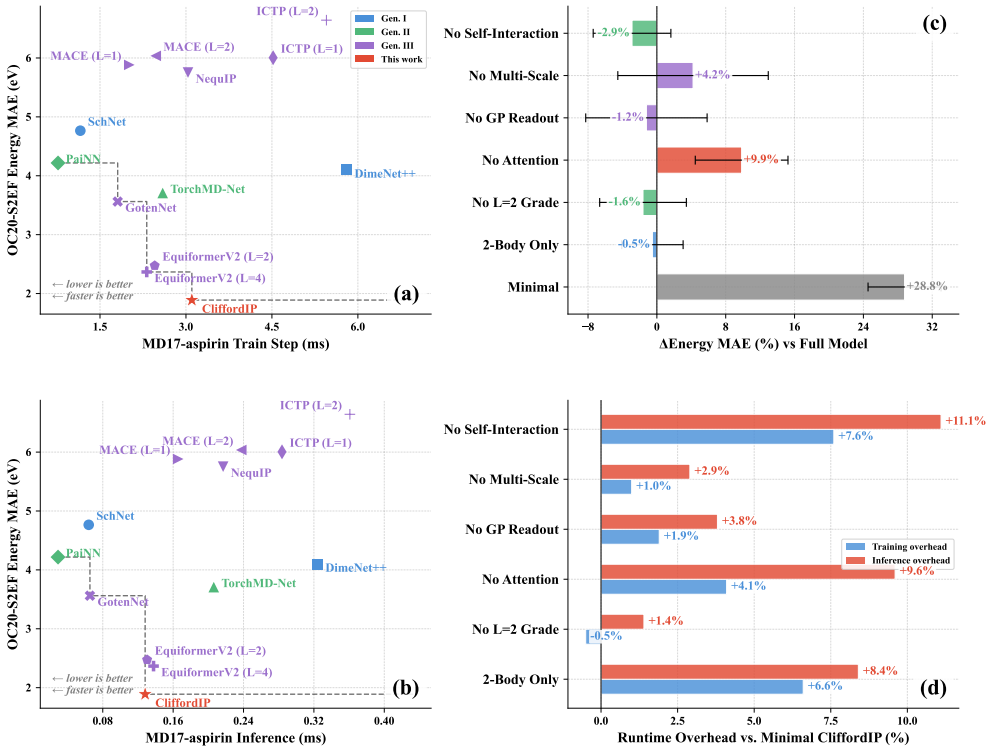
The EquiformerV2 family's rMD17 energy MAE (247.4–247.7 eV) sits orders of magnitude above every other energy-capable architecture and is uniform across all ten molecules; this behaviour is consistent with an energy-head calibration failure under strongly force-weighted direct-force training on small datasets, where the energy head does not calibrate against the reference total-energy scale. The advantage of CliffordIP therefore extends beyond catalysis to small-molecule chemistry, with FAENet's frame-averaging the only stronger small-molecule energy predictor in this comparison.

Table 1 rMD17 cross-domain benchmark. Aggregate test MAE across all ten rMD17 molecules. Mean ± std over 4 seeds × molecules. Best in bold, runner-up underlined.

Model	E-MAE (eV) ↓	F-MAE (eV/Å) ↓
FAENet	$\mathbf{0.024}_{\pm 0.019}$	$0.021_{\pm 0.001}$
CliffordIP	$\underline{0.034}_{\pm 0.017}$	$0.021_{\pm 0.000}$
TorchMD-Net	$0.711_{\pm 0.390}$	$0.318_{\pm 0.038}$
SchNet	$1.132_{\pm 0.568}$	$0.719_{\pm 0.041}$
NequIP	$1.229_{\pm 0.068}$	$2.640_{\pm 0.021}$
MACE (L=3)	$1.283_{\pm 0.321}$	$3.645_{\pm 0.760}$
DimeNet++	$1.326_{\pm 0.788}$	$0.765_{\pm 0.086}$
PaiNN	$1.686_{\pm 1.031}$	$0.849_{\pm 0.147}$
MACE (L=2)	$1.801_{\pm 0.469}$	$4.664_{\pm 0.877}$
ICTP (L=2)	$1.946_{\pm 0.595}$	$5.008_{\pm 1.195}$
ICTP (L=3)	$2.030_{\pm 0.329}$	$5.256_{\pm 1.033}$
ICTP (L=1)	$2.168_{\pm 0.608}$	$5.117_{\pm 1.240}$
MACE (L=1)	$2.192_{\pm 0.364}$	$5.631_{\pm 0.842}$
ViSNet	$2.383_{\pm 1.669}$	$0.935_{\pm 0.260}$
GotenNet	$27.757_{\pm 33.206}$	$8.152_{\pm 8.166}$
EquiformerV2 (L=2)	$247.431_{\pm 0.040}$	$0.004_{\pm 0.000}$
EquiformerV2 (L=3)	$247.610_{\pm 0.033}$	$\underline{0.004}_{\pm 0.000}$
EquiformerV2 (L=4)	$247.721_{\pm 0.023}$	$\mathbf{0.004}_{\pm 0.000}$

Computational Efficiency

Panels (a) and (b) of Fig. 4 pair MD17-aspirin (21 atoms, batch 32, 10 warmup + 20 timed steps, two-seed average) per-step timing with OC20-S2EF in-distribution energy MAE for the thirteen models that have both measurements. At training time, CliffordIP sits at the accuracy-optimal end of the Pareto front (3.10 ms/step, 1.889 eV); the chain runs PaiNN (0.77 ms, 4.22 eV) → GotenNet (1.81 ms, 3.56 eV) → EquiformerV2 L=4 (2.32 ms, 2.37 eV) → CliffordIP, with each step yielding a meaningful accuracy gain. At inference, CliffordIP drops to 0.128 ms/step — faster than EquiformerV2 L=4 (0.137 ms), on par with L=2 (0.130 ms) — so the model dominates both spherical-harmonic baselines at equal or lower latency. The MACE family at $L \in \{1, 2\}$ scales monotonically in cost (2.01 → 2.47 ms) without overtaking CliffordIP



on energy. The full 18-architecture wall-clock and memory benchmark, including ViSNet, FAENet, and the L=3 spherical-harmonic variants, appears in Supplementary Table S10.

Ablation Study

Panels (c) and (d) of Fig. 4 decompose component contributions on OC22-S2EF. From a full-model baseline of 3.241 eV mean energy MAE, removing attention degrades performance most severely (+9.9%), followed by multi-scale aggregation (+4.2%) (panel c). Self-interaction (−2.9%), L=2 grade features (−1.6%), GP readout (−1.2%), and 2-body restriction (−0.5%) yield marginal or slightly negative changes, consistent with modest regularisation rather than primary representational capacity. The minimal variant, retaining only the message-passing backbone, degrades by +28.8%, confirming that the full architecture is necessary for competitive performance.

Runtime overhead (panel d) shows attention and self-interaction dominating inference cost (+9.6% and +11.1%), while multi-scale aggregation contributes only +2.9% inference for its +4.2% accuracy gain — a favourable accuracy-to-cost ratio. GP readout adds +3.8% inference for −1.2% accuracy, marking it prunable in latency-constrained settings. Removing L=2 grade features cuts training time by −0.5% with

negligible accuracy impact, consistent with grade sparsity being most beneficial in early layers where higher-grade features have not yet been activated.

We additionally ablate the underlying Clifford algebra on rMD17, comparing $\mathrm{Cl}(3,1)$, $\mathrm{Cl}(2,0)$, and $\mathrm{Cl}(3,0)$ at matched architecture and training budget. Aggregate energy MAEs are 0.086 eV for $\mathrm{Cl}(3,1)$, 0.101 eV for $\mathrm{Cl}(2,0)$, and 0.108 eV for $\mathrm{Cl}(3,0)$; aggregate force MAEs are essentially tied at 0.021 eV/Å across the three algebras. $\mathrm{Cl}(3,1)$ wins on rMD17, where finite organic molecules activate only a subset of the available grades. We retain $\mathrm{Cl}(3,0)$ as the headline choice because adsorbate–surface chemistry simultaneously activates all four grades of $\mathrm{Cl}(3,0)$ (scalars for distances and coordination energies, vectors for bond and adsorption directions, bivectors for surface and reaction planes, pseudoscalars for face selectivity and chirality), which is the geometric setting the architecture was designed for.

Discussion

The central finding of this benchmark is that Clifford algebra provides a viable and in several settings superior alternative to spherical harmonics for constructing equivariant interatomic potentials. The $\mathrm{Cl}(3,0)$ algebra encodes E(3)-equivariant features in 8-dimensional multivectors via the geometric product, whereas spherical-harmonic approaches require $\sum_{l=0}^{L}(2l+1)$ channels to reach angular order L (49 dimensions for L=6). Despite this much lower-dimensional feature space, CliffordIP matches or exceeds the energy accuracy of the strongest spherical-harmonic baseline tested, EquiformerV2, under the present unified, compute-constrained comparison that also includes MACE and ICTP. The geometric product, which fuses scalar, vector, bivector, and pseudoscalar interactions in a single algebraically closed operation, evidently captures chemical environment information with greater parameter efficiency than the Clebsch–Gordan tensor products required by spherical-harmonic representations. This is consistent with theoretical analyses showing that the Clifford group action yields multiple non-equivalent subrepresentations whose multiplicative structure bypasses explicit CG decomposition, a property recently exploited for higher-order graph networks [55] and simplicial message passing [56].

This representational efficiency carries over to chemically distinct domains. CliffordIP leads in-distribution energy on both OC20 and OC22 S2EF and on OC22-IS2RE (see Results), with its top-two ranking alongside the EquiformerV2 family stable across metallic and oxide chemistries. Under distribution shift the picture is consistent: CliffordIP leads OC20 OOD-Ads (1.977 eV, 18% below EquiformerV2 at L=4, 2.424 eV), finishes a close second on OC22-OOD (5.120 eV vs. 4.971 eV, +3.0%), and is essentially tied with EquiformerV2 (L=4) on the hardest OC20 OOD-Both split (3.510 vs. 3.507 eV), evidence that the energy lead is preserved or amplified out of distribution rather than reflecting in-distribution overfitting. Given recent evidence that data composition can matter more than architectural sophistication for surface-property prediction [57], the persistence of this lead under distribution shift makes the case for representational priors (such as the multivector grading of $\mathrm{Cl}(3,0)$) that generalise across chemical domains.

The practical consequence is most visible on catalysis-specific transfer. On the CO_2RR, N_2RR, and C_2 formation adsorbate subsets (evaluated zero-shot), CliffordIP reduces energy MAE by 20–24% relative to EquiformerV2. The mechanism is straightforward: bivectors encode the bond planes that distinguish bent and dihedral intermediates, and the pseudoscalar resolves the chirality of multidentate adsorption sites. In a screening workflow this translates to more reliable energy rankings across the intermediate space, and to more confident selection of candidate surfaces for downstream DFT validation [58]. Recent benchmarking of universal MLIPs has identified adsorption-energy accuracy as the critical bottleneck for high-throughput screening reliability [59, 60], making the energy-specialist profile of CliffordIP well-suited to this stage of the discovery pipeline; force fidelity becomes critical only at the subsequent structural-relaxation stage.

However, CliffordIP shows a clear limitation in force prediction. On OC20 S2EF its force MAE of 0.064 eV Å^{-1} trails the best EquiformerV2 (L=2, 0.059 eV Å^{-1}) by 8%, with EquiformerV2 (L=4) at 0.060 eV Å^{-1} and the remaining baselines clustered between 0.063 and 0.078 eV Å^{-1}. On OC22 the spread narrows further (0.073 vs. 0.072 eV Å^{-1}, +1%), and on OC20-IS2RE CliffordIP ranks sixth at 0.765 eV (+18% above EquiformerV2 at L=2). The structural origin is that the 8-dimensional $\mathrm{Cl}(3,0)$ basis natively encodes angular information only up to l=1; the L=2 scalar augmentation injects higher-order directional features but does not close the gap to models maintaining l=6 spherical harmonics throughout. Because forces are gradients of the energy, angular-resolution limits in the learned potential surface are amplified in the derivative. Fu et al. [61] have argued that energy conservation and surface smoothness, rather than raw force MAE on static test sets, are the properties most predictive of downstream molecular-dynamics performance, suggesting that the small absolute force gaps observed here may overstate the practical impact for energy-based screening. Importantly, the absolute force-MAE comparison above is performed in the direct-equivariant-force-head setting used uniformly across all baselines for protocol fairness; CliffordIP additionally supports conservative-force prediction in autograd mode, where forces are obtained as the energy gradient and the resulting potential is exactly energy-conserving by construction. On rMD17 we find that switching to this conservative mode improves test energy MAE by 25% with marginal change in test force MAE, at a $\sim$3$\times$ wall-clock cost per training step from the second backward pass (Supplementary Information, Force Consistency). For applications requiring high-fidelity force fields, the angular-resolution gap motivates extensions to richer Clifford algebras.

The ablation study (see the Ablation Study subsection) confirms that attention-weighted aggregation and multi-scale readout are the dominant accuracy contributors. Removing them degrades OC22-S2EF energy MAE by 9.9% and 4.2% respectively against the full-model baseline of 3.241 eV. Self-interaction, L=2 grade features, GP readout, and 2-body restriction each show marginal isolated effects (-2.9% to -0.5%), pointing to a regularisation rather than primary-capacity role. The gap between the minimal-variant degradation (+28.8%) and the sum of individual component effects indicates strong synergistic interactions that emerge only when the full architecture is co-trained.

On the timing side, CliffordIP sits at the accuracy-optimal end of the OC20-S2EF training Pareto front (3.10 ms/step) and drops to 0.128 ms/step at inference, faster than EquiformerV2 at L=4 (0.137 ms) and on par with L=2 (0.130 ms), reflecting the efficiency of the grade-sparse geometric-product dispatch at forward pass. Among ablated components, attention and self-interaction carry the highest inference overhead (+9.6% and +11.1%), while multi-scale aggregation adds only +2.9% for its +4.2% accuracy gain. Recent parameter-light Clifford architectures [62] and efficient multivector neuron designs [63] indicate that the training-time overhead can be substantially reduced without sacrificing equivariance; hybrid invariant–equivariant strategies in the spirit of HIENet [64] offer a complementary route by reserving Clifford layers for the equivariant pathway and delegating bulk computation to cheaper invariant layers. Training convergence curves for all models are provided in the Supplementary Information (Training Convergence).

The success of CliffordIP on catalysis benchmarks adds to a growing body of evidence that Clifford algebra provides a powerful inductive bias for physical systems, spanning n-body dynamics, signed-volume regression, top-tagging in particle physics, and molecular generation via diffusion models [65]. CliffordIP is, to our knowledge, the first application of this framework to interatomic potential energy surfaces, demonstrating that the benefits of geometric algebra extend from small-molecule dynamics to the large-scale, chemically diverse setting of heterogeneous catalysis.

Several directions follow naturally from these limitations. The angular-resolution gap could be addressed by extending to $\mathrm{Cl}(3,1)$ or $\mathrm{Cl}(4,1)$, or by combining Clifford-algebraic energy prediction with a spherical-harmonic force-correction head, analogous to recent hybrid strategies for balancing invariant efficiency with equivariant expressiveness. Our experiments are conducted at $\sim$1 M parameters, and scaling to 10–100 M parameters may shift relative rankings, since larger models can better exploit high-order representations [66]. The uniform training protocol prioritises comparability over absolute performance; model-specific tuning, longer training, and denoising strategies such as DeNS [67] would likely benefit all architectures. A separate axis of improvement concerns the reference data: the OC20 and OC22 labels are computed at the GGA (RPBE / RPBE+U) level, which is known to suffer from self-interaction error in charge-transfer chemistry and transition-metal-oxide regimes that overlap with parts of these benchmarks; retraining MLIPs on self-interaction-corrected datasets [68, 69] is an orthogonal route to higher absolute accuracy, complementary to the architectural axis explored here. Beyond static benchmarks, evaluating CliffordIP on structural relaxation, transition-state searches, and molecular-dynamics stability remains important for validating its practical utility in catalyst-discovery workflows [70].

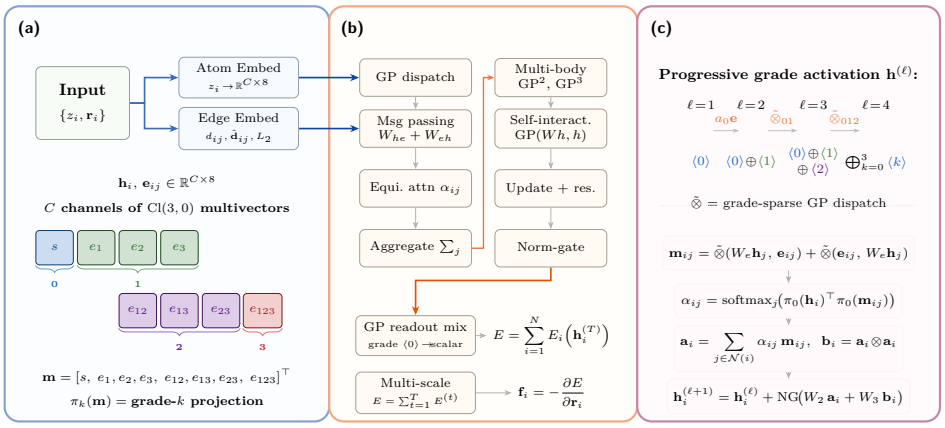
A scope caveat about absolute scales is in order, and it differs by task. The S2EF energy MAEs reported throughout this paper are produced under a setting chosen to isolate architectural effects under matched data, parameter count, and compute. They are therefore elevated by approximately 8$\times$ relative to the OCP S2EF leaderboard's headline entries, which are trained at orders-of-magnitude larger scale (e.g., EquiformerV2-153M on S2EF-All+MD, $\sim$172 M training frames, $\sim$1,571 V100-GPU-days; in-distribution energy MAE $\sim$0.23 eV). The IS2RE energy values, by contrast,

land on the same absolute scale as the comparable-compute baselines reported in the OC20 reports test-ID IS2RE energy MAEs of 0.61 eV (CGCNN), 0.64 eV (SchNet), and 0.56 eV (DimeNet++) when trained on the All (∼460K) split, with relaxation-based DimeNet++ variants reaching as low as 0.50 eV [14]; the lowest IS2RE MAE in our benchmark is 0.65 eV (EquiformerV2 at L=2), within this range. Our 1 M→10 M parameter-scaling experiment (Supplementary Table S9) shows that scaling parameters at fixed 200K data closes part of the S2EF gap for EquiformerV2 (1.7×) and MACE (1.2×), while CliffordIP is essentially flat (1.0×), indicating that its capacity is already saturated by 200K frames at this parameter count, with the corresponding compute envelope mapped in Supplementary Table S10. Two follow-ups suggest themselves. The OC20-IS2RE leaderboard is the more accessible near-term target, given that our absolute IS2RE values already sit within the published-baseline range; a scaled-up CliffordIP at the 10–30 M parameter range trained on the official IS2RE split is a realistic next step. Closing the S2EF gap requires the joint params-and-data scaling regime occupied by the leaderboard. Training CliffordIP at the 10–150 M parameter range on the S2EF-2M or S2EF-All+MD splits at leaderboard-scale compute is identified as longer-term future work. Whether the in-distribution energy advantage of $\mathrm{Cl}(3,0)$ persists at full leaderboard scale on either task remains an open empirical question.

In summary, CliffordIP brings Clifford-algebra equivariance to interatomic potentials for heterogeneous catalysis, replacing Clebsch–Gordan tensor products with the geometric product while delivering the strongest in-distribution energy performance among the compared baselines on OC20 and OC22 S2EF and matching that lead out-of-distribution. The energy advantage transfers to small-molecule chemistry on rMD17 as a cross-domain robustness check: although finite organic molecules activate only a subset of the four $\mathrm{Cl}(3,0)$ grades, CliffordIP is rank 1 or 2 on every molecule and runner-up on aggregate energy MAE (0.034 vs. 0.024 eV for FAENet), with the third-place baseline trailing by ∼21×. This is the graceful-degradation pattern predicted when the architecture's geometric advantages are largest in geometrically rich systems and retained, rather than lost, in less geometrically rich ones. Force fidelity remains a step behind the strongest spherical-harmonic models in the direct-force-head setting, although CliffordIP natively supports conservative-force prediction in autograd mode for applications where energy conservation is critical. The energy accuracy combined with near-parity inference cost relative to EquiformerV2 makes CliffordIP a promising foundation for energy-focused MLIPs in computational catalysis and electrocatalyst screening. Whether this advantage persists at the joint params-and-data scaling of the OCP S2EF leaderboard remains the principal open question; on IS2RE, where our absolute values are already in the published-baseline range, a scaled-up submission is a realistic near-term direction.

Methods

This section presents the CliffordIP architecture, progressing from the algebraic foundations of the Clifford algebra $\mathrm{Cl}(3,0)$ through the feature representations,



message-passing mechanism, and output readout. An architectural overview is given in Fig. 5.

Clifford algebra Cl(3, 0) and equivariance

The Clifford algebra $\mathrm{Cl}(3,0)$ over $\mathbb{R}^3$ is the $2^3 = 8$-dimensional associative algebra generated by three orthonormal basis vectors $\{e_1, e_2, e_3\}$ satisfying $e_i\, e_j + e_j\, e_i = 2\,\delta_{ij}$ [43, 44]. The resulting eight basis elements are organised into four grades: scalars $\{1\}$ (grade 0, 1D), vectors $\{e_1, e_2, e_3\}$ (grade 1, 3D), bivectors $\{e_{12}, e_{13}, e_{23}\}$ (grade 2, 3D), and the pseudoscalar $\{e_{123}\}$ (grade 3, 1D), where $e_{ij} \equiv e_i\, e_j$. A general element $M \in \mathrm{Cl}(3,0)$—called a multivector—admits a unique grade decomposition

$$M = \langle M\rangle_0 + \langle M\rangle_1 + \langle M\rangle_2 + \langle M\rangle_3\,. \tag{1}$$

The central operation of the algebra is the geometric product (GP), which subsumes the inner and outer (wedge) products in a single algebraic operation. For two vectors $\mathbf{a}, \mathbf{b} \in \langle\cdot\rangle_1$ it decomposes as

$$\mathbf{a}\,\mathbf{b} = \underbrace{\mathbf{a}\cdot\mathbf{b}}_{\text{grade 0}} + \underbrace{\mathbf{a}\wedge\mathbf{b}}_{\text{grade 2}}\,, \tag{2}$$

while for arbitrary multivectors the product is determined by the Cayley multiplication table, a fixed $8\times 8 \to 8$ tensor of structure constants $c_{ijk} \in \{-1, 0, +1\}$:

$$(A \odot B)_k = \sum_{i,j} c_{ijk}\, A_i\, B_j\,. \tag{3}$$

Each evaluation requires exactly 64 multiply-add operations, independent of angular-momentum order. This contrasts with the spherical harmonics (SH) framework employed in tensor field networks [38, 39] and equivariant transformers [30, 41], where

Clebsch–Gordan tensor products scale as $O(l_{\max}^6)$ and require Wigner D-matrices for equivariant coupling [40, 48, 49].

For physically meaningful predictions an MLIP must yield energy invariance and force equivariance under the full orthogonal group O(3). In $\mathrm{Cl}(3,0)$ the reversion operator $\widetilde{M} = \langle M\rangle_0 + \langle M\rangle_1 - \langle M\rangle_2 - \langle M\rangle_3$ reverses the order of all basis-vector products. The sandwich product $\rho(R)\colon M \mapsto R\,M\,\widetilde{R}$ then defines a representation of Pin(3)—the double cover of O(3)—on multivectors, where a rotor $R = \cos(\theta/2) + \sin(\theta/2)\,\hat{B}$ encodes rotation by θ about the plane of the unit bivector $\hat{B}$.

By the associativity of Clifford multiplication and $R\,\widetilde{R} = 1$, the GP respects this action:

$$\rho(R)(A \odot B) = \rho(R)(A) \odot \rho(R)(B)\,. \tag{4}$$

Because the sandwich product preserves grades, grade-specific operations (projection, grade-wise norms) commute with $\rho(R)$; consequently, grade-0 outputs are O(3)-invariant and grade-1 outputs are O(3)-equivariant [48, 50]. Pin(3) equivariance includes improper rotations (reflections), which is strictly stronger than SO(3) equivariance alone [48].

Each grade carries a direct physical meaning: grade-0 scalars encode invariant quantities such as energy; grade-1 vectors encode positions, forces, and momenta; grade-2 bivectors encode oriented rotation planes and angular momentum; and the grade-3 pseudoscalar encodes orientation (chirality). $\mathrm{Cl}(3,0)$ thus provides a compact, fixed 8-dimensional algebra that naturally encompasses l=0 (grade 0) and l=1 (grade 1) representations together with the bivector (grade 2, which transforms as l=1 under rotations but with opposite parity) and the pseudoscalar (grade 3) [49].

$\mathrm{Cl}(3,0)$ is the minimal Clifford algebra that is also complete for 3D Euclidean geometry, which makes it the natural choice for non-relativistic atomic systems. Lower-dimensional alternatives such as $\mathrm{Cl}(2,0)$ omit the grade-2 (planes) and grade-3 (pseudoscalar) content needed to distinguish features such as FCC vs HCP hollow sites, dihedral signs, and out-of-plane reaction geometries, and therefore cannot represent the catalytic and chiral systems that motivate this work. Higher-dimensional alternatives such as $\mathrm{Cl}(3,1)$ introduce a fourth (timelike) basis vector whose additional grades have clear meaning in relativistic or spacetime settings but are not required in this regime; their inclusion enlarges the multivector from 8 to 16 dimensions, doubling per-message compute without a corresponding gain in expressiveness for non-relativistic atomic simulations. We empirically validate this choice in the Supplementary Information (Algebra Ablation), where $\mathrm{Cl}(2,0)$, $\mathrm{Cl}(3,0)$, and $\mathrm{Cl}(3,1)$ produce statistically indistinguishable force MAE on rMD17 at matched compute; $\mathrm{Cl}(3,0)$ is therefore the principled default at no measurable cost.

Feature representations

Each atom's state is represented as C multivectors, yielding a feature tensor $\mathbf{h}_i \in \mathbb{R}^{C\times 8}$.

The atomic number z_i is mapped through a learnable embedding table to a C-dimensional vector placed in the grade-0 (scalar) slots; all higher-grade components

are initialised to zero:

$$\mathbf{h}_i^{(0)}[\ldots, 0] = \mathrm{Emb}(z_i)\,, \qquad \mathbf{h}_i^{(0)}[\ldots, 1{:}8] = \mathbf{0}\,. \tag{5}$$

For each edge (i, j) the interatomic distance d_{ij} and unit direction $\hat{\mathbf{d}}_{ij}$ are encoded as a grade-$(0, 1)$ multivector $\mathbf{e}_{ij} \in \mathbb{R}^{C\times 8}$:

$$\mathbf{e}_{ij}[\ldots, 0] = \mathbf{g}_0\,, \qquad \mathbf{e}_{ij}[\ldots, 1{:}4] = \mathbf{g}_1 \otimes \hat{\mathbf{d}}_{ij}\,, \tag{6}$$

where $\mathbf{g}_0$ and $\mathbf{g}_1$ are produced by two-layer MLPs with SiLU activations (σ_s and σ_v, respectively) that act on the concatenation of Gaussian radial basis functions (n_rbf=50), l=2 spherical features (described below), and a cosine cutoff envelope at $r_\mathrm{cut} = 6.0$ Å.

Because the $\mathrm{Cl}(3, 0)$ basis natively spans angular information up to l=1, we augment the edge input with the five independent components of the symmetric traceless outer product $\hat{\mathbf{d}} \otimes \hat{\mathbf{d}}$ to capture d-orbital-like angular dependence:

$$L_2(\hat{\mathbf{d}}) = \left[\, d_x^2 - \tfrac{1}{3},\ d_x d_y,\ d_x d_z,\ d_y^2 - \tfrac{1}{3},\ d_y d_z \,\right], \tag{7}$$

corresponding to the five real l=2 spherical harmonics (the sixth component $d_z^2 - \frac{1}{3}$ is redundant by tracelessness). Because these features enter the edge MLPs as invariant scalar inputs rather than as separate multivector components, overall equivariance is preserved.

Naïvely initialising all eight multivector components and applying the full GP from the first layer leads to grade explosion: random high-grade inputs generate large magnitudes in every output grade, destabilising early training. CliffordIP addresses this with a progressive schedule in which the active grade subspace of node features expands layer by layer according to the GP grade table $T(g_a, g_b)$:

$$\mathbf{h}^{(L)} \in \begin{cases} \langle 0\rangle \oplus \langle 1\rangle & L = 1 \\ \bigoplus_{k=0}^{2} \langle k\rangle & L = 2 \\ \bigoplus_{k=0}^{3} \langle k\rangle & L = 3 \\ \mathrm{Cl}(3, 0) = \bigoplus_{k=0}^{3} \langle k\rangle & L \geq 4 \end{cases} \tag{8}$$

(see Fig. 5c). The schedule is determined at model construction time by iteratively computing the union of the current node grades and the output grades produced by the GP between node and edge features; since edges carry grades $(0, 1)$ and atoms are initialised in grade 0, the first GP already produces grades $(0, 1)$, reaching the full algebra at layer 3. With the default T=5 interaction layers, this ensures stable convergence without auxiliary losses or gradient clipping. To our knowledge, this strategy is novel among Clifford-algebraic neural networks [48–50, 55, 63].

Message passing and interaction blocks

CliffordIP follows the message-passing neural network framework [21, 28] and stacks T=5 interaction blocks, each described below.

For each directed edge (i, j) a multivector message is constructed as

$$\mathbf{m}_{ij} = \alpha_{ij} \cdot \left[W_{he} \cdot \mathrm{GP}(\mathbf{h}_j, \tilde{\mathbf{e}}_{ij}) + W_{eh} \cdot \mathrm{GP}(\tilde{\mathbf{e}}_{ij}, \mathbf{h}_j) + \mathrm{skip}(\mathbf{h}_j) \right], \tag{9}$$

where $\tilde{\mathbf{e}}_{ij} = W_e\, \mathbf{e}_{ij}$ is a grade-preserving linear pre-projection of the edge multivector, W_{he} and W_{eh} are grade-preserving linear projections, and skip$(\cdot)$ extracts the grade-0 features through a standard linear layer, providing a scalar residual path. Both GP orderings are used because the product is non-commutative: $\mathbf{b}\,\mathbf{a}$ flips the sign of the wedge product relative to $\mathbf{a}\,\mathbf{b}$ (Eq. 2), so the two orderings allow the network to independently weight symmetric (inner product) and antisymmetric (wedge product) contributions.

Because not all 64 multiply-adds of the full GP are needed in every layer, a dispatch function selects the minimal GP variant consistent with the active grades: layer 1, where nodes carry only grade 0, uses scalar×multivector (8 ops); layer 2, where nodes carry grades $(0, 1)$ and edges carry grades $(0, 1)$, uses the grades-$(0, 1)$×$(0, 1)$ kernel (24 ops); layer 3, where nodes have expanded to grades $(0, 1, 2)$, uses the grades-$(0, 1, 2)$×$(0, 1)$ kernel against the grade-$(0, 1)$ edges (also 24 ops); and layers 4–5, where all grades are active, use the full GP (64 ops). This yields a 2.7–8× reduction in message computation cost for early layers with no approximation.

Attention weights α_{ij} are computed via multi-head scaled dot-product attention restricted to grade-0 (invariant) features, ensuring that the resulting scalars preserve equivariance:

$$\mathbf{q}_i = W_Q\, \mathbf{h}_i[\ldots, 0], \quad \mathbf{k}_j = W_K\, \mathbf{h}_j[\ldots, 0], \quad \mathbf{r}_{ij} = W_R \cdot \mathrm{RBF}(d_{ij}), \tag{10}$$

$$\alpha_{ij} = \frac{1}{n_h} \sum_{h=1}^{n_h} \mathrm{softmax}_j^{(h)}\!\left(\frac{\mathbf{q}_i^{(h)} \odot \mathbf{k}_j^{(h)} \odot \mathbf{r}_{ij}^{(h)}}{\sqrt{d_{\mathrm{head}}}} \right), \tag{11}$$

where $\odot$ is element-wise multiplication, $d_{\mathrm{head}} = C/n_h$ (n_h=4 heads), the softmax is applied independently per head over the neighbourhood of each receiver atom i, and the resulting per-head weights are averaged to produce a single scalar attention coefficient per edge.

Standard message passing captures pairwise (2-body) interactions. Following the MACE paradigm [26], CliffordIP extends to higher body orders via iterated geometric products. After aggregating messages $\mathbf{A}_i = \sum_j \mathbf{m}_{ij}$:

$$\mathbf{A}_i^{(2)} = \mathbf{A}_i, \qquad \mathbf{A}_i^{(3)} = \mathrm{GP}(\mathbf{A}_i, \mathbf{A}_i), \qquad \mathbf{A}_i^{(4)} = \mathrm{GP}(\mathbf{A}_i^{(3)}, \mathbf{A}_i), \tag{12}$$

where $\mathbf{A}_i^{(3)}$ encodes 3-body angular correlations between neighbours and $\mathbf{A}_i^{(4)}$ encodes 4-body correlations (optional; default body order is 3). The combined multi-body

feature is $\mathbf{M}_i = W_2\,\mathbf{A}_i^{(2)} + W_3\,\mathbf{A}_i^{(3)} + W_4\,\mathbf{A}_i^{(4)}$ with grade-preserving linear maps W_2, W_3, W_4.

Each node additionally undergoes a per-atom GP self-interaction

$$\mathbf{h}_i^{\text{self}} = \text{GP}(W_{\text{proj}} \cdot \mathbf{h}_i,\, \mathbf{h}_i)\,, \tag{13}$$

which mixes information across channels and grades in a nonlinear fashion that grade-preserving linear maps alone cannot achieve. The node features are then updated as

$$\mathbf{h}_i' = \text{Norm}\big(W_{\text{out}} \cdot \sigma\big(W_{\text{in}} \cdot [\mathbf{h}_i \,\|\, \mathbf{M}_i \,\|\, \mathbf{h}_i^{\text{self}}]\big) + \mathbf{h}_i\big)\,, \tag{14}$$

where $\|$ denotes channel-wise concatenation and σ is a gated activation: SiLU is applied directly to scalar grades $(0, 3)$, while non-scalar grades $(1, 2)$ are norm-gated—a learned gate network maps the per-channel norm through a sigmoid to produce a scalar scale factor applied element-wise, preserving direction while modulating magnitude. Grade-wise normalisation applies standard-deviation scaling to scalar grades and root-mean-square scaling to vector and bivector grades. The final addition is a residual connection.

Output readout

Energy and force predictions are extracted from the final multivector features via a GP readout mixing step followed by grade-specific projections.

A final geometric product mixes information across all grades:

$$\mathbf{h}_i^{\text{mix}} = \text{GP}(W_{\text{pre}} \cdot \mathbf{h}_i,\, \mathbf{h}_i)\,, \tag{15}$$

routing information from every grade into grades 0 and 1, which enriches the readout representation.

The total energy is obtained by summing per-atom contributions predicted from grade-0 (scalar) features, which are Pin(3)-invariant by construction (Eq. 4):

$$E = \sum_i \text{MLP}\big([\,\mathbf{h}_i[\ldots, 0] \,\|\, \mathbf{h}_i^{\text{mix}}[\ldots, 0]\,]\big)\,, \tag{16}$$

where a three-layer network with SiLU activations maps $2C$ scalar features to a single per-atom energy. Following MACE, multi-scale readout collects energy contributions from every interaction layer: $E_{\text{total}} = E_{\text{out}} + \sum_{t=1}^{T} E_t$, where $E_t = \sum_i \text{MLP}_t(\mathbf{h}_i^{(t)}[\ldots, 0])$ operates on the grade-0 features after layer t, allowing both local (early-layer) and global (late-layer) chemical environments to contribute directly to the prediction [33].

Forces are predicted from grade-1 (vector) features, whose Pin(3)-equivariance guarantees the correct transformation behaviour under rotations and reflections:

$$\mathbf{f}_i = W_f\,\big[h_i[\ldots, 1{:}4] \,\|\, h_i^{\text{mix}}[\ldots, 1{:}4]\big]^\top\,, \tag{17}$$

where W_f is a linear projection over the channel dimension.

Because the geometric product is smooth and differentiable, $\mathrm{Cl}(3,0)$ message passing is end-to-end differentiable through the full network, and conservative forces can alternatively be obtained as

$$\hat{\mathbf{f}}_i = -\frac{\partial E}{\partial \mathbf{r}_i} \tag{18}$$

from the predicted energy by a single configuration flag. The autograd variant is force-consistent by construction. An optional denoising auxiliary loss [67] provides additional regularisation during training in either mode.

Datasets and evaluation splits

Table 2 summarises the training and validation splits. OC20 uses the standard 200K S2EF training split; OC22 uses the complete training set for both tasks; all validation subsets are evaluated in full. Graph connectivity uses radius graphs at $r_{\mathrm{cut}} = 6.0\,\text{Å}$ with at most 50 neighbours per atom. Force training and evaluation are restricted to free atoms (adsorbate and surface-layer); fixed subsurface atoms are excluded from all force objectives and metrics.

Table 2 Dataset configurations. N_{train} denotes the number of training structures; validation sets were evaluated in full.

Dataset	Task	N_{train}	Validation splits
OC20	S2EF	200,000	ID, OOD-Ads., OOD-Cat., OOD-Both
OC20	IS2RE	200,000	ID, OOD-Ads., OOD-Cat., OOD-Both
OC22	S2EF	full	ID, OOD
OC22	IS2RE	full	ID, OOD

The CO_2 reduction reaction, C–C coupling, and nitrogen reduction reaction results in the main text were obtained by evaluating the OC20-trained S2EF model on adsorbate-specific subsets of the OC20 S2EF validation set; no additional fine-tuning was performed. Structures were selected by filtering on adsorbate identity using the OC20 metadata mapping. The CO2RR subset includes adsorbates $^*CO$, $^*COOH$, $^*CHO$, $^*CO_2$, $^*HCOO$, $^*COH$, $^*CH_2O$, $^*H$, $^*OH$, $^*O$, $^*CH$, $^*CH_2$, and $^*C$; the C2 subset includes $^*C^*C$, $^*CCH$, $^*CCH_2$, $^*CCH_3$, $^*CH^*CH$, $^*CHCH_2$, $^*CH_2CH_3$, $^*CCO$, $^*CHCO$, and CH_2^*CO; and the N2RR subset includes $^*N$, $^*N_2$, $^*NH$, $^*NH_3$, $^*NHNH$, $^*N^*NH$, $^*N^*NO$, $^*NO$, $^*NO_2$, $^*NO_3$, $^*ONH$, $^*NONH$, and $^*ONNH_2$.

Training protocol and losses

All models train with the Adam optimiser [71] at a fixed learning rate of 1×10^{-4} and no weight decay, with mini-batches of 16 structures. Training stops at the epoch cap or after 30 (50 for OC22-IS2RE) epochs without improvement on the primary validation metric, whichever comes first (Table 3). Each configuration is run under three random seeds; reported results are mean ± standard deviation over seeds.

Table 3 Training schedules were adjusted according to the benchmark task. Because the OC22 IS2RE task has a relatively small sample size, a longer training schedule was used to achieve adequate learning.

Dataset	Task	Max. epochs	Early-stopping patience
OC20	S2EF	50	30
OC20	IS2RE	50	30
OC22	S2EF	50	30
OC22	IS2RE	200	50

The S2EF loss is a weighted sum of a per-atom energy MAE and a per-atom force-vector norm loss, consistent with the FairChem evaluation framework:

$$\mathcal{L}_{\mathrm{S2EF}} = \frac{w_E}{|\mathcal{B}|} \sum_{s \in \mathcal{B}} \frac{|\hat{E}_s - E_s|}{N_s} + \frac{w_F}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} \|\hat{\mathbf{f}}_i - \mathbf{f}_i\|_2, \tag{19}$$

where $\hat{E}_s$ and E_s are the predicted and reference total energies of structure s containing N_s atoms, $\hat{\mathbf{f}}_i$ and $\mathbf{f}_i$ are the predicted and reference force vectors for free atom i, $\mathcal{B}$ is the set of structures in the mini-batch, $\mathcal{F}$ is the set of free atoms across the batch, and $w_E = 1.0$, $w_F = 3.0$ are fixed scalar weights.

IS2RE training minimises the mean absolute error on total relaxed energies:

$$\mathcal{L}_{\mathrm{IS2RE}} = \frac{1}{|\mathcal{B}|} \sum_{s \in \mathcal{B}} |\hat{E}_s - E_s|. \tag{20}$$

S2EF performance is characterised by four metrics: (i) energy MAE (mean absolute error over structures, eV); (ii) force MAE (mean absolute error over all free-atom force components, eV Å^{-1}); (iii) force cosine similarity (mean cosine similarity between predicted and reference force vectors per free atom); and (iv) energy-and-force within threshold (EFwT), the fraction of structures whose absolute energy error is below 0.02 eV and whose maximum per-atom force error is below 0.03 eV Å^{-1}. IS2RE performance is assessed by energy MAE and energy within threshold (EwT), the fraction of structures with absolute energy error at most 0.02 eV.

Baseline model configurations

Table 4 reports the key architectural hyperparameters across baselines.

Computational setup and compute budget

All models were trained and evaluated on single NVIDIA RTX 5090 GPUs, with each independent experimental run assigned exclusively to one device. No distributed or multi-GPU training, gradient accumulation, or mixed-precision arithmetic was employed.

Table 4 Architectural configurations used across all benchmark experiments. L = number of interaction layers (or blocks); d = primary feature dimension; RBF = radial basis functions. All models use $r_{\text{cut}} = 6.0\,\text{Å}$ and at most 50 neighbours per atom.

Model	L	d	Radial basis	Key design choices
SchNet† [19]	6	192	50 Gaussian filters	Continuous-filter convolution on interatomic distances
DimeNet++† [72]	4	128	6 radial + 7 spherical Bessel	Angular triplet interactions; int. emb. 64
PaiNN [21]	4	144	64 sinc RBF	Equivariant scalar–vector message passing
TorchMD-Net [23]	6	104	56 expnorm RBF	Equivariant transformer; 8 attention heads
NequIP [25]	4	44	8 Bessel RBF	E(3)-equivariant; L_{max}=1; radial MLP depth 2
EquiformerV2 (L=2) [41]	2	64	—	SO(3) equivariant transformer; L_{max}=2, m_{max}=2; 4 heads
EquiformerV2 (L=4) [41]	2	30	—	SO(3) equivariant transformer; L_{max}=4, m_{max}=2; 4 heads
GotenNet [31]	5	96	—	Geometric tensor network; cosine cutoff
MACE (L=1) [26]	2	128	8 Bessel RBF	E(3)-equivariant body-ordered; correlation order 3; L_{max}=1
MACE (L=2) [26]	2	96	8 Bessel RBF	E(3)-equivariant body-ordered; correlation order 3; L_{max}=2
ICTP (L=1) [73]	4	96	8 Bessel RBF	Irreducible Cartesian Tensor Potential; L_{max}=1
ICTP (L=2) [73]	4	80	8 Bessel RBF	Irreducible Cartesian Tensor Potential; L_{max}=2
CliffordIP	5	52×8	50 RBF	Cl(3, 0)-equivariant; progressive grade activation; 100 atom types

†Baseline implementation via the PyTorch Geometric library [74].

The full benchmark comprises 13 models × 4 dataset–task combinations × 3 random seeds = 156 independent training runs, each trained for up to 50 (200 for OC22-IS2RE) epochs, totalling ∼1,755 GPU-hours. Per-model and per-dataset hours from the training logs appear in Table 5; per-step timings for the full 18-architecture benchmark are in Supplementary Table S10.

Per-model: CliffordIP is the most expensive at 216.6 hours (about 18.1 hours per run), reflecting unoptimised 8-dimensional multivector operations. The ICTP family (200–202 hours), DimeNet++ (187.1), and GotenNet (185.9) form the next-most-expensive tier — driven by triplet interactions, tensor contractions, and Cartesian-tensor products respectively. MACE at L=2 (147.0) and EquiformerV2 (105.6 at L=2, 117.7 at L=4) are notably more efficient at moderate angular orders, with EquiformerV2 benefiting from SO(2) convolution optimisations [40]. NequIP and

TorchMD-Net occupy a middle tier (∼104–105 hours); MACE (L=1, 86.3 hours), PaiNN (55.2), and SchNet (42.3) are the cheapest, roughly 4–5× below CliffordIP.

Per-dataset: the two S2EF tasks dominate (OC20-S2EF 648.7 hours, OC22-S2EF 614.1 hours), since each step computes both energy and force losses. OC20-IS2RE drops to 385.3 hours — force head disabled, but the OC20 dataset is large — and OC22-IS2RE is the cheapest task at 107.5 hours, combining the smaller OC22 training set with the absence of force computation.

Table 5 Wall-clock training hours by model and dataset for the full 13-baseline benchmark (13 models × 4 datasets × 3 seeds = 156 runs). Per-step timings are reported in Supplementary Table S10.

Model	Hrs	Runs	Dataset	Hrs	Runs
CliffordIP	216.6	12	OC20-S2EF	648.7	39
ICTP (L=2)	202.2	12	OC22-S2EF	614.1	39
ICTP (L=1)	200.3	12	OC20-IS2RE	385.3	39
DimeNet++	187.1	12	OC22-IS2RE	107.5	39
GotenNet	185.9	12			
MACE (L=2)	147.0	12			
EquiformerV2 (L=4)	117.7	12			
EquiformerV2 (L=2)	105.6	12			
NequIP	105.2	12			
TorchMD-Net	104.1	12			
MACE (L=1)	86.3	12			
PaiNN	55.2	12			
SchNet	42.3	12			
Total	1,755.5	156	Total	1,755.6	156

Data Availability

The OC20 and OC22 datasets used in this study are publicly available through the Open Catalyst Project (https://opencatalystproject.org). The rMD17 dataset is publicly available on figshare at https://doi.org/10.6084/m9.figshare.12672038.

Code Availability

The source code, trained CliffordIP model checkpoints, training logs, and scripts to reproduce all benchmark results are available at https://github.com/KurbanIntelligenceLab/CliffordIP. The package can be installed from PyPI as `cliffordip` (`pip install cliffordip`).

Acknowledgements

No funding was received for this research.

Author Contributions

C.P. conceived and implemented the CliffordIP architecture, conducted all computational simulations and benchmark experiments, performed the data analysis, and wrote the original manuscript. E.S., M.K., and H.K. supervised the research, contributed to the interpretation of results, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Competing Interests

The authors declare no competing financial or non-financial interests.

References

[1] Cho, J. *et al.* Importance of broken geometric symmetry of single-atom pt sites for efficient electrocatalysis. *Nature Communications* **14**, 3233 (2023).

[2] Ning, M. *et al.* Dynamic active sites in electrocatalysis. *Angewandte Chemie* **136**, e202415794 (2024).

[3] Nørskov, J. K., Studt, F., Abild-Pedersen, F. & Bligaard, T. *Fundamental concepts in heterogeneous catalysis* (John Wiley & Sons, 2014).

[4] Seh, Z. W. *et al.* Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **355**, eaad4998 (2017).

[5] Nitopi, S. *et al.* Progress and perspectives of electrochemical co2 reduction on copper in aqueous electrolyte. *Chemical Reviews* **119**, 7610–7672 (2019).

[6] Kuhl, K. P., Cave, E. R., Abram, D. N. & Jaramillo, T. F. New insights into the electrochemical reduction of carbon dioxide on metallic copper surfaces. *Energy & Environmental Science* **5**, 7050–7059 (2012).

[7] Skulason, E. *et al.* A theoretical evaluation of possible transition metal electrocatalysts for n 2 reduction. *Physical Chemistry Chemical Physics* **14**, 1235–1245 (2012).

[8] Singh, A. R. *et al.* Electrochemical ammonia synthesis—the selectivity challenge (2017).

[9] Garza, A. J., Bell, A. T. & Head-Gordon, M. Mechanism of co2 reduction at copper surfaces: pathways to c2 products. *Acs Catalysis* **8**, 1490–1499 (2018).

[10] Montoya, J. H., Shi, C., Chan, K. & Nørskov, J. K. Theoretical insights into a co dimerization mechanism in co2 electroreduction. *The Journal of Physical Chemistry Letters* **6**, 2032–2037 (2015).

[11] Hammer, B. & Nørskov, J. K. Electronic factors determining the reactivity of metal surfaces. *Surface Science* **343**, 211–220 (1995).

[12] Hammer, B. & Nørskov, J. K. Theoretical surface science and catalysis: calculations and concepts. *Advances in Catalysis* **45**, 71–129 (2000).

[13] Nørskov, J. K., Bligaard, T., Rossmeisl, J. & Christensen, C. H. Towards the computational design of solid catalysts. *Nature Chemistry* **1**, 37–46 (2009).

[14] Chanussot, L. *et al.* Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis* **11**, 6059–6072 (2021).

[15] Tran, R. *et al.* The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *Acs Catalysis* **13**, 3066–3084 (2023).

[16] Unke, O. T. *et al.* Machine learning force fields. *Chemical Reviews* **121**, 10142–10186 (2021).

[17] Friederich, P., Häse, F., Proppe, J. & Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nature Materials* **20**, 750–761 (2021).

[18] Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters* **120**, 145301 (2018).

[19] Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **148** (2018).

[20] Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123* (2020).

[21] Schütt, K., Unke, O. & Gastegger, M. Meila, M. & Zhang, T. (eds) *Equivariant message passing for the prediction of tensorial properties and molecular spectra.* (eds Meila, M. & Zhang, T.) *Proceedings of the 38th International Conference on Machine Learning*, Vol. 139 of *Proceedings of Machine Learning Research*, 9377–9388 (PMLR, 2021).

[22] Satorras, V. G., Hoogeboom, E. & Welling, M. Meila, M. & Zhang, T. (eds) *E(n) equivariant graph neural networks.* (eds Meila, M. & Zhang, T.) *Proceedings of the 38th International Conference on Machine Learning*, Vol. 139 of *Proceedings of Machine Learning Research*, 9323–9332 (PMLR, 2021).

[23] Thölke, P. & De Fabritiis, G. Torchmd-net: equivariant transformers for neural network based molecular potentials. *arXiv preprint arXiv:2202.02541* (2022).

[24] Anderson, B., Hy, T. S. & Kondor, R. Cormorant: Covariant molecular neural networks. *Advances in Neural Information Processing Systems* **32** (2019).

[25] Batzner, S. *et al.* E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications* **13**, 2453 (2022).

[26] Batatia, I., Kovacs, D. P., Simm, G., Ortner, C. & Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems* **35**, 11423–11436 (2022).

[27] Musaelian, A. *et al.* Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications* **14**, 579 (2023).

[28] Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E. J. & Welling, M. Geometric and physical quantities improve e (3) equivariant message passing. *arXiv preprint arXiv:2110.02905* (2021).

[29] Fuchs, F., Worrall, D., Fischer, V. & Welling, M. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems* **33**, 1970–1981 (2020).

[30] Liao, Y.-L. & Smidt, T. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990* (2022).

[31] Aykent, S. & Xia, T. GotenNet: Rethinking efficient 3d equivariant graph neural networks (2025). URL https://openreview.net/forum?id=5wxCQDtbMo.

[32] Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B* **99**, 014104 (2019).

[33] Batatia, I. *et al.* The design space of e (3)-equivariant atom-centred interatomic potentials. *Nature Machine Intelligence* **7**, 56–67 (2025).

[34] Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* **2**, 718–728 (2022).

[35] Yuan, E. C.-Y. *et al.* Foundation models for atomistic simulation of chemistry and materials. *Nature Reviews Chemistry* 1–19 (2026).

[36] Yang, H. *et al.* Mattersim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967* (2024).

[37] Duval, A. *et al.* A hitchhiker's guide to geometric gnns for 3d atomic systems. *arXiv preprint arXiv:2312.07511* (2023).

[38] Thomas, N. *et al.* Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219* (2018).

[39] Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453* (2022).

[40] Passaro, S. & Zitnick, C. L. Krause, A. *et al.* (eds) *Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs.* (eds Krause, A. *et al.*) *Proceedings of the 40th International Conference on Machine Learning*, Vol. 202 of *Proceedings of Machine Learning Research*, 27420–27438 (PMLR, 2023).

[41] Liao, Y.-L., Wood, B., Das, A. & Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059* (2023).

[42] Luo, S., Chen, T. & Krishnapriyan, A. S. Enabling efficient equivariant operations in the fourier basis via gaunt tensor products. *arXiv preprint arXiv:2401.10216* (2024).

[43] Hestenes, D. & Sobczyk, G. *Clifford algebra to geometric calculus: a unified language for mathematics and physics* (Springer Science & Business Media, 2012).

[44] Doran, C. & Lasenby, A. *Geometric algebra for physicists* (Cambridge University Press, 2003).

[45] Dorst, L., Fontijne, D. & Mann, S. *Geometric algebra for computer science (revised edition): An object-oriented approach to geometry* (Morgan Kaufmann, 2009).

[46] Gopalan, V. Wedge reversion antisymmetry and 41 types of physical quantities in arbitrary dimensions. *Foundations of Crystallography* **76**, 318–327 (2020).

[47] Calle-Vallejo, F. & Koper, M. Theoretical considerations on the electroreduction of co to c 2 species on cu (100) electrodes. *Angewandte Chemie* **125** (2013).

[48] Ruhe, D., Brandstetter, J. & Forré, P. Clifford group equivariant neural networks. *Advances in Neural Information Processing Systems* **36**, 62922–62990 (2023).

[49] Ruhe, D., Gupta, J. K., De Keninck, S., Welling, M. & Brandstetter, J. Krause, A. *et al.* (eds) *Geometric Clifford algebra networks.* (eds Krause, A. *et al.*) *Proceedings of the 40th International Conference on Machine Learning*, Vol. 202 of *Proceedings of Machine Learning Research*, 29306–29337 (PMLR, 2023).

[50] Brehmer, J., De Haan, P., Behrends, S. & Cohen, T. S. Geometric algebra transformer. *Advances in Neural Information Processing Systems* **36**, 35472–35496 (2023).

[51] Chmiela, S. *et al.* Machine learning of accurate energy-conserving molecular force fields. *Science advances* **3**, e1603015 (2017).

[52] Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature communications* **9**, 3887 (2018).

[53] Lan, J. *et al.* Adsorbml: a leap in efficiency for adsorption energy calculations using generalizable machine learning potentials. *npj Computational Materials* **9**, 172 (2023).

[54] Kolluru, A. *et al.* Open challenges in developing generalizable large-scale machine-learning models for catalyst discovery. *Acs Catalysis* **12**, 8572–8581 (2022).

[55] Tran, V.-H., Vo, T. N., Huu, T. T. & Nguyen, T. M. A clifford algebraic approach to e (n)-equivariant high-order graph neural networks. *arXiv preprint arXiv:2410.04692* (2024).

[56] Liu, C., Ruhe, D., Eijkelboom, F. & Forré, P. Clifford group equivariant simplicial message passing networks. *arXiv preprint arXiv:2402.10011* (2024).

[57] Mehdizadeh, A. & Schindler, P. Surface stability modeling with universal machine learning interatomic potentials: a comprehensive cleavage energy benchmarking study. *AI for Science* **1**, 025002 (2025).

[58] Mok, D. H. *et al.* Data-driven discovery of electrocatalysts for co2 reduction using active motifs-based machine learning. *Nature Communications* **14**, 7303 (2023).

[59] Tang, D., Ketkaew, R. & Luber, S. Machine learning interatomic potentials for heterogeneous catalysis. *Chemistry–A European Journal* **30**, e202401148 (2024).

[60] Omranpour, A., Elsner, J., Lausch, K. N. & Behler, J. Machine learning potentials for heterogeneous catalysis. *Acs Catalysis* **15**, 1616–1634 (2025).

[61] Fu, X. *et al.* Learning smooth and expressive interatomic potentials for physical property prediction. *arXiv preprint arXiv:2502.12147* (2025).

[62] Filimoshina, E. & Shirokov, D. Glgenn: a novel parameter-light equivariant neural networks architecture based on clifford geometric algebras. *arXiv preprint arXiv:2506.09625* (2025).

[63] Liu, C., Ruhe, D. & Forré, P. Multivector neurons: Better and faster o (n)-equivariant clifford graph neural networks. *arXiv preprint arXiv:2406.04052* (2024).

[64] Yan, K. *et al.* A materials foundation model via hybrid invariant-equivariant architectures. *arXiv preprint arXiv:2503.05771* (2025).

[65] Liu, C., Vadgama, S., Ruhe, D., Bekkers, E. & Forré, P. Clifford group equivariant diffusion models for 3d molecular generation. *arXiv preprint arXiv:2504.15773* (2025).

[66] Barroso-Luque, L. *et al.* Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771* (2024).

[67] Liao, Y.-L., Smidt, T., Shuaibi, M. & Das, A. Generalizing denoising to non-equilibrium structures improves equivariant force fields. *arXiv preprint arXiv:2403.09549* (2024).

[68] Shinde, R., Yamijala, S. S. & Wong, B. M. Improved band gaps and structural properties from wannier–fermi–löwdin self-interaction corrections for periodic systems. *Journal of Physics: Condensed Matter* **33**, 115501 (2021).

[69] Janesko, B. G. Replacing hybrid density functional theory: motivation and recent advances. *Chemical Society Reviews* **50**, 8470–8495 (2021).

[70] Olajide, G., Baral, K., Ezendu, S., Soyemi, A. & Szilvasi, T. Application of machine learning interatomic potentials in heterogeneous catalysis. *Journal of Catalysis* **448**, 116202 (2025).

[71] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[72] Gasteiger, J., Giri, S., Margraf, J. T. & Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115* (2020).

[73] Zaverkin, V. *et al.* Higher-rank irreducible cartesian tensors for equivariant message passing. *Advances in Neural Information Processing Systems* **37**, 124025–124068 (2024).

[74] Fey, M. & Lenssen, J. E. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428* (2019).

Figure Legends

Fig. 1 Geometric algebra representations for catalytic surface chemistry and equivariant graph neural networks. (a) CO_2 electroreduction via C–C coupling of two adsorbed *CO intermediates on a Cu surface, illustrating the four grades of $Cl(3,0)$ needed to fully describe the reaction geometry: interatomic distances (scalar), bond directions (vector), the O–C–C–O reaction plane (bivector), and the dihedral chirality (pseudoscalar). (b) Electrochemical N_2 reduction on Fe, showing end-on $*N_2$ adsorption and its protonated *NNH intermediate; the binding mode and protonation geometry require vector and bivector representations that scalar-only models cannot capture. (c) Top-view schematic of surface adsorption sites on a close-packed metal surface, ranging from atop (one bond, grades 0–1) to bridge (two bonds, grades 0–2) and hollow sites (three or more bonds, up to all four grades); distinguishing FCC from HCP hollow sites requires bivector and pseudoscalar information encoding the plane orientation and above/below-surface parity. (d) The four grades of the Clifford algebra $Cl(3,0)$ mapped onto a local atomic motif: grade 0 (scalar) captures bond lengths and energies, grade 1 (vector) encodes bond directions and forces, grade 2 (bivector) represents the plane spanned by three atoms and angular momentum, and grade 3 (pseudoscalar) describes oriented volumes and dihedral chirality. (e) Message-passing scheme on a cutoff graph, where atom i aggregates geometric product messages from neighbours j within a cutoff radius; three equivariant operations — message aggregation with attention, self-interaction, and norm-gating — preserve the full multivector structure across layers. (f) Grade activation schedule and readout: scalar features are present from the first layer, with vector, bivector, and pseudoscalar channels introduced progressively in later layers to stabilise training; the final scalar and vector channels are projected to per-atom energies and equivariant forces, respectively.

Fig. 2 Energy–force accuracy trade-off on catalysis-focused zero-shot subset evaluation for three electrocatalytic reaction families: (a) CO_2 reduction, (b) N_2 reduction, and (c) C_2 formation. Each point shows mean energy MAE (x-axis) and force MAE (y-axis) over three random seeds on adsorbate-filtered subsets of the OC20 val-id split; no task-specific fine-tuning was performed. CliffordIP achieves the lowest energy MAE on all three subsets, with margins of 0.50–0.65 eV over the next-best baseline while remaining within 0.005 eV/Å of the best force MAE.

Fig. 3 Benchmark comparison on OC20 and OC22. Horizontal bars: mean MAE ± std over three seeds for S2EF energy (a, d), S2EF force (b, e), IS2RE relaxed energy (c, f), and out-of-distribution splits (g–i); percentage offsets are relative to the best model in each panel. CliffordIP ranks first on in-distribution and OOD-Ads S2EF energy across both datasets and on OC22 IS2RE; the EquiformerV2 family leads on force MAE.

Fig. 4 Efficiency and ablation analysis. (a, b) Per-step timing on MD17 aspirin plotted against OC20-S2EF in-distribution energy MAE for the thirteen baselines with both measurements; CliffordIP lies on the inference Pareto front and is the most accurate model overall. (c) Ablation on OC22-S2EF reporting ΔEnergy MAE relative to the full model: removing attention (+9.9%) or multi-scale aggregation (+4.2%) causes the largest degradation. (d) Per-component runtime overhead relative to the minimal CliffordIP variant on OC20-S2EF, separated into training and inference. The full 18-architecture wall-clock and memory benchmark, including ViSNet, FAENet, and the L=3 variants of MACE, ICTP, and EquiformerV2, appears in Supplementary Table S10.

Fig. 5 CliffordIP architecture. (a) Input and embedding: atomic numbers z_i and Cartesian positions $\mathbf{r}_i$ are mapped to grade-$\langle 0\rangle$ atom features and grade-$(\langle 0\rangle, \langle 1\rangle)$ edge multivectors; the $Cl(3,0)$ basis layout $\mathbf{m}=[s, e_1, e_2, e_3, e_{12}, e_{13}, e_{23}, e_{123}]^\top$ with grade projections π_k is shown at the bottom. (b) Interaction block ($\times T$=5): grade-sparse geometric-product (GP) messages, equivariant scalar attention α_{ij}, MACE-style multi-body GP terms (GP^2, GP^3), self-interaction, and a residual norm-gated node update; the output sub-block performs GP readout mixing and multi-scale energy summation, while forces are predicted from grade-1 vector features. Conservative forces can alternatively be obtained as energy gradients when required for downstream applications. (c) Grade schedule: progressive activation of multivector grade subspaces $\mathbf{h}^{(L)}$ across interaction layers (top), and the four key model equations for symmetric GP message construction, scalar attention, neighbourhood aggregation, and norm-gated node update (bottom); $\tilde{\otimes}$ denotes the grade-sparse GP dispatch.