WEYL ALGEBRA EQUIVARIANT NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Encoding physical symmetries into neural network architectures is key to improving generalization and physical consistency in domains such as molecular dynamics and particle physics. Existing equivariant networks rely primarily on Clifford algebras, which realize orthogonal-group equivariance via the sandwich action, or on partial differential operator (PDO) networks, which implicitly exploit Weyl algebra structure without an explicit unifying framework. Both approaches are limited: Clifford algebras cannot capture symplectic or differential-operator symmetries, while PDO networks lack a common algebraic language across groups. We show that the Weyl algebra—generated by position and momentum operators under the canonical commutation relation—unifies both lines of work: its subalgebra structure recovers orthogonal-group equivariance, while its differential-operator generators extend naturally to the symplectic group. Building on this, we introduce a Weyl-equivariant neural network supporting both symmetries within a single architecture, along with degree-truncation approximations for computational tractability. Empirical results on tasks involving symplectic and orthogonal symmetries demonstrate that our method matches or outperforms existing equivariant networks while offering substantially broader symmetry coverage.

1 INTRODUCTION

Deep learning has achieved remarkable success across vision, language, and scientific modeling (LeCun et al., 2015), yet many real-world data domains carry intrinsic geometric structure and physical symmetry: molecular energies are invariant to absolute atomic coordinates (Batzner et al., 2022; Schütt et al., 2021), fields in particle physics transform covariantly under Lorentz boosts (Bogatskiy et al., 2020), and point clouds are equivariant to rotation and translation (Thomas et al., 2018; Fuchs et al., 2020). Geometric deep learning (Bronstein et al., 2021; Cohen & Welling, 2016) encodes such symmetries directly into network architecture so that the model is equivariant or invariant to a prescribed transformation group, improving sample efficiency and physical interpretability. This program has recently been placed on a unified representation-theoretic footing: Kondor (2025) argues that essentially all equivariant architectures used in physics and chemistry can be understood by expressing network features in the generalized Fourier space associated with the symmetry group and by realizing equivariant nonlinearities as Clebsch–Gordan products of group representations, while Weber (2025) surveys the wider landscape of geometric machine learning on non-Euclidean data—graphs, matrix spaces, and symmetric spaces—of which the phase-space, Weyl-algebra construction we study here is a further instance.

Clifford algebras and their limits. A major line of work builds equivariant networks on Clifford (geometric) algebras (Ruhe et al., 2023; Brandstetter et al., 2023; Brehmer et al., 2023), in which orthogonal, Lorentz, projective, and conformal transformations are all realized as a single sandwich (conjugation) action $v \mapsto MvM^{-1}$ for a suitable invertible algebra element M. This unifies many finite-dimensional linear isometry groups under one algebraic recipe, but it is intrinsically a *configuration-space* geometric algebra: it cannot express (i) equivariance of differential operators acting on function spaces, as required by many field equations, nor (ii) the symplectic symmetry $\mathrm{Sp}(2n)$ of classical and quantum phase space, where position and momentum operators satisfy $[\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij}$.

Weyl algebras: from geometric algebra to operator algebra. These limitations point to a more fundamental algebraic object, the Weyl algebra (algebra of differential operators) $\mathcal{A}_n$, defined by $2n$ generators $\{x_1, \ldots, x_n, \partial_1, \ldots, \partial_n\}$ subject to $[\partial_i, x_j] = \delta_{ij}$, $[x_i, x_j] = [\partial_i, \partial_j] = 0$. When the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

commutator is replaced by an anticommutator, $\mathcal{A}_n$ degenerates to a Clifford algebra; representation-theoretically, both are quotients of universal enveloping algebras of finite-dimensional Lie algebras ($\mathfrak{sp}$ and $\mathfrak{so}$, respectively). The Weyl algebra therefore does not replace Clifford algebras so much as extend their equivariance framework from the orthogonal group to the symplectic group and to differential-operator groups. Crucially, it also suggests a new equivariance mechanism: network layers act on feature *functions* through parameterized differential operators rather than through matrix multiplication on feature vectors, which is naturally suited to field equations, density estimation, and kernel methods.

Related work. Group-equivariant networks trace back to G-CNNs (Cohen & Welling, 2016) and were extended to spherical/gauge convolutions (Cohen et al., 2018), tensor-field and SE(3)-equivariant networks for point clouds and molecules (Thomas et al., 2018; Fuchs et al., 2020; Batzner et al., 2022; Schütt et al., 2021), and Lorentz-equivariant networks for particle physics (Bogatskiy et al., 2020). Clifford-algebra equivariant networks (Brandstetter et al., 2023; Brehmer et al., 2023; Ruhe et al., 2023; de Haan et al., 2024) have since given a unified account of $O(n)/SO(n)$ (CGENN, Ruhe et al., 2023; Clifford Neural Layers, Brandstetter et al., 2023), affine/projective (GATr, Brehmer et al., 2023), conformal (de Haan et al., 2024), and Lorentz/Poincaré (Ruhe et al., 2023; Bogatskiy et al., 2020) equivariance, but remain confined to finite-dimensional linear transformation groups; the symplectic group $\mathrm{Sp}(2n)$ and differential-operator symmetries lie outside all of them. A second, largely disjoint, line of work on steerable PDOs (Jenner & Weiler, 2022) and PDO-equivariant convolutions (Shen et al., 2020) implicitly parameterizes elements of finite degree in $\mathcal{A}_n$ as convolution kernels, but does not identify or exploit the Weyl-algebra structure explicitly, so different symmetry groups are treated in an ad hoc, non-unified manner. A third, representation-theoretic line of work studies equivariant maps on homogeneous spaces G/K directly: building on the linear steerable-CNN theory of Cohen et al. (2018), Nyholm et al. (2026) derive generalized steerability constraints that any—possibly nonlinear—G-equivariant layer on a homogeneous space must satisfy, and prove a universality result for constructions meeting them. As we discuss in Section 3.1, the symplectic group $\mathrm{Sp}(2n,\mathbb{R})$ acting on the Weyl algebra is itself a homogeneous-space construction once one fixes the maximal compact subgroup $U(n)$, and the Moyal-product layer of Section 4 can be read as a concrete, Clebsch–Gordan-structured instance of such a nonlinear steerable map on the symmetric space $\mathrm{Sp}(2n,\mathbb{R})/U(n)$.

Contributions. We (i) build a unified theoretical framework in which the symplectic group $\mathrm{Sp}(2n)$ acts on the Weyl algebra $\mathcal{W}(V,\omega)$ by algebra automorphisms, and prove that every Wigner–Weyl degree projection, and every polynomial in Weyl-algebra elements, is $\mathrm{Sp}(2n)$-equivariant (Section 3); (ii) show that the orthogonal group $O(n)$ embeds canonically and diagonally into $\mathrm{Sp}(2n,\mathbb{R})$, so that $O(n)$-equivariance—the setting covered by Clifford networks—arises as a special case; (iii) construct equivariant linear, Moyal-product, normalization, and gated-activation layers from this theory (Section 4); (iv) resolve two engineering obstacles absent in the Clifford setting—the infinite-dimensionality of $\mathcal{W}(V,\omega)$ and the non-existence of a positive-definite $\mathrm{Sp}(2n)$-invariant inner product on the non-compact symplectic group—via degree truncation and four alternative normalization schemes (Section 5); and (v) validate the resulting WeylNN on symplectic-capacity regression, quantum Wigner-function classification, and Hamiltonian dynamics prediction, including a real-data counterpart on PETRA III accelerator beam dynamics (Section 6).

2 Preliminaries: The Weyl Algebra and Phase-Space Quantization

Definition 1 (Symplectic vector space)**.** *A symplectic vector space (V,ω) is a $2n$-dimensional real vector space equipped with a non-degenerate antisymmetric bilinear form $\omega : V \times V \to \mathbb{R}$. A Darboux basis $q_1,\dots,q_n,p_1,\dots,p_n$ satisfies $\omega(q_i,p_j)=\delta_{ij}$, $\omega(q_i,q_j)=\omega(p_i,p_j)=0$.*

Definition 2 (Weyl algebra)**.** *Fix $\hbar > 0$ (a learnable scalar in our networks). The Weyl algebra $\mathcal{W}(V,\omega)$ is the unital associative $\mathbb{R}$-algebra generated by V subject to $v_1v_2 - v_2v_1 = \hbar\,\omega(v_1,v_2)\cdot 1$ for all $v_1,v_2 \in V$; equivalently $\mathcal{W}(V,\omega) = T(V)/\mathcal{I}(\omega)$ for the tensor algebra $T(V)$ and the two-sided ideal $\mathcal{I}(\omega)$ generated by these relations. In a Darboux basis, $[q_i,p_j] = \hbar\delta_{ij}$, $[q_i,q_j] = [p_i,p_j] = 0$; setting $\hbar = 1$ and $p_i \leftrightarrow \partial_i$ recovers the standard differential-operator algebra $[\partial_i, x_j] = \delta_{ij}$.*

Wigner–Weyl quantization and degree decomposition. The symmetric algebra $\mathrm{Sym}(V)$ (real polynomials on V) admits a canonical vector-space isomorphism $\sigma_W : \mathrm{Sym}(V) \to \mathcal{W}(V,\omega)$, the Wigner–Weyl map, sending a monomial to its symmetrized (Weyl-ordered) product, $\sigma_W(v_1 \cdots v_m) = \frac{1}{m!}\sum_{\tau\in\mathfrak{S}_m} v_{\tau(1)} \cdots v_{\tau(m)}$. Although the natural filtration of $\mathcal{W}(V,\omega)$ by tensor degree is *not* a direct-sum grading (the commutation relation lowers degree by two), the PBW theorem guarantees that σ_W is a bijection between graded pieces $\mathrm{Sym}^m(V)$ and canonical subspaces $\mathcal{W}^{(m)}(V,\omega) := \sigma_W(\mathrm{Sym}^m(V))$, giving a vector-space (not algebra) decomposition $\mathcal{W}(V,\omega) = \bigoplus_{m\geq 0}\mathcal{W}^{(m)}(V,\omega)$ with $\dim \mathcal{W}^{(m)}(V,\omega) = \binom{2n+m-1}{m}$. We call the linear map $(\cdot)^{(m)} : \mathcal{W} \to \mathcal{W}^{(m)}$ the *degree-m projection*. Transporting the algebra product through σ_W gives the *Moyal star product* on $\mathrm{Sym}(V)$,

$$(f \star_\hbar g)(z) = \sum_{k=0}^{\infty} \frac{1}{k!}\Big(\frac{\hbar}{2}\Big)^k \omega^{a_1 b_1}\cdots\omega^{a_k b_k}(\partial_{a_1}\cdots\partial_{a_k} f)(z)(\partial_{b_1}\cdots\partial_{b_k} g)(z), \tag{1}$$

a finite sum for polynomial f, g, reducing to ordinary multiplication as $\hbar \to 0$ and to the Poisson bracket at first order, $f \star_\hbar g - fg = \frac{\hbar}{2}\{f,g\}_\omega + O(\hbar^2)$.

3 Symplectic Equivariance of the Weyl Algebra

Definition 3 (Symplectic group and its action)**.** $\mathrm{Sp}(V,\omega) = \{S \in GL(V) : \omega(Sv_1, Sv_2) = \omega(v_1,v_2)\ \forall v_1, v_2\}$. *For $S \in \mathrm{Sp}(V,\omega)$, let $\rho(S) : \mathcal{W}(V,\omega) \to \mathcal{W}(V,\omega)$ be the unique algebra automorphism extending $\rho(S)(v) = Sv$ on generators; equivalently, on the symbol side, $\rho(S)(f)(z) = f(S^{-1}z)$.*

Theorem 1 (Properties of the symplectic action)**.** *$\rho(S)$ is well defined for every $S \in \mathrm{Sp}(V,\omega)$ and satisfies: (1) linearity; (2) multiplicativity, $\rho(S)(AB) = \rho(S)(A)\rho(S)(B)$; (3) $\rho(S^{-1}) = \rho(S)^{-1}$; (4) $\rho(S_2)\rho(S_1) = \rho(S_2S_1)$, i.e. $\rho : \mathrm{Sp}(V,\omega) \to \mathrm{Aut}(\mathcal{W}(V,\omega))$ is a group homomorphism; (5) restricted to a compact subgroup $O(n) \hookrightarrow \mathrm{Sp}(V,\omega)$, $\rho^{(m)}(R)$ is orthogonal with respect to a Haar-averaged $O(n)$-invariant inner product on $\mathcal{W}^{(m)}(V,\omega)$.*

Theorem 2 (Degree projections are equivariant)**.** *For all $S \in \mathrm{Sp}(V,\omega)$, $A \in \mathcal{W}(V,\omega)$, $m \in \mathbb{N}_0$: $\rho(S)(A^{(m)}) = (\rho(S)(A))^{(m)}$. Consequently every $\mathcal{W}^{(m)}(V,\omega)$ is $\mathrm{Sp}(V,\omega)$-invariant, and the degree-m projection is an equivariant linear map.*

Corollary 1 (Polynomials in Weyl-algebra elements are equivariant)**.** *For any real polynomial $F \in \mathbb{R}[T_1,\ldots,T_\ell]$ and $A_1,\ldots,A_\ell \in \mathcal{W}(V,\omega)$, $\rho(S)\big(F(A_1,\ldots,A_\ell)\big) = F\big(\rho(S)(A_1),\ldots,\rho(S)(A_\ell)\big)$, and the same holds after applying any fixed degree projection to both sides.*

Proofs are given in Appendix A; the key facts used throughout are that $\rho(S)$ preserves the ideal defining $\mathcal{W}(V,\omega)$ because S preserves ω (Definition 2), and that on the symbol side $\rho(S)$ acts as a linear pullback, which manifestly preserves polynomial degree.

Proposition 1 (Canonical embedding of $O(n)$)**.** *Let W be an n-dimensional Euclidean space and $V = W \oplus W^*$ its phase space with the canonical symplectic form $\omega((q_1,p_1),(q_2,p_2)) = \langle p_1, q_2\rangle - \langle p_2, q_1\rangle$. The diagonal embedding $\iota(R)(q,p) = (Rq, Rp)$, i.e. $\iota(R) = \mathrm{diag}(R,R)$, satisfies $\iota(R) \in \mathrm{Sp}(V,\omega)$ for every $R \in O(n)$, because $R^\top R = I_n$. Hence $\rho \circ \iota : O(n) \to \mathrm{Aut}(\mathcal{W}(V,\omega))$ equips the Weyl algebra with an $O(n)$-equivariant action, and all of Theorem 1–Corollary 1 apply with $S = \iota(R)$.*

Proposition 1 is what allows WeylNN to reproduce $O(n)/SO(n)$ equivariance—the setting covered by Clifford-algebra networks—as the restriction of the more general $\mathrm{Sp}(2n)$-equivariant theory.

Remark 1 (Physical interpretation of the diagonal embedding)**.** *Under $\iota(R)$, a rotation $R \in O(n)$ acts on phase space as $(q,p) \mapsto (Rq, Rp)$: position and momentum coordinates are rotated simultaneously (identifying $W^* \cong W$ via $\langle\cdot,\cdot\rangle$), and on a Weyl-ordered monomial, $\rho(\iota(R))\,\sigma_W(q^\alpha p^\beta) = \sigma_W\big((Rq)^\alpha (Rp)^\beta\big)$. Scalar (degree-0) features are invariant; vector (degree-1) features rotate as ordinary vectors; and every higher-degree component transforms under the corresponding tensor representation of $O(n)$, exactly as multivectors of matching grade transform under $\mathrm{Spin}(n)$ in Clifford-algebra networks. This correspondence is what lets a WeylNN linear layer (Section 4) reduce to an ordinary $O(n)$-equivariant tensor layer when restricted to $\iota(O(n)) \subset \mathrm{Sp}(2n,\mathbb{R})$.*

Extending the embedding to pseudo-orthogonal groups $O(p,q)$ (e.g. the Lorentz group) requires replacing ω by an η-twisted symplectic form, under which $\iota(R) = \mathrm{diag}(R,R)$ need no longer be symplectic for the standard form J (since $R^\top R \neq I_n$ in general for $R \in O(p,q)$); establishing the corresponding equivariant structure on the η-twisted Weyl algebra $\mathcal{W}(V,\omega_\eta)$ is left to future work.

3.1 Representation-Theoretic Structure: Clebsch–Gordan Decomposition under $U(n)$

The results above show that $\mathcal{W}(V,\omega)$ carries an exact, global $\mathrm{Sp}(2n,\mathbb{R})$ action. We now refine this picture along the lines of Kondor (2025): rather than treating each degree-m block $\mathcal{W}^{(m)}(V,\omega)$ as an opaque vector space, we decompose it into irreducible pieces of the maximal compact subgroup and identify the resulting Clebsch–Gordan structure of the Moyal product. This is what ultimately licenses the $O(D^3)$ parameter count of the Moyal-product layer in Section 4 and connects WeylNN to the homogeneous-space equivariance framework of Nyholm et al. (2026).

Definition 4 (Compatible complex structure and $U(n)$)**.** *An endomorphism $J \in GL(V)$ is a* compatible complex structure *for (V,ω) if $J^2 = -\mathrm{id}_V$, $\omega(Ju,Jv) = \omega(u,v)$ for all $u,v \in V$, and $g_J(u,v) := \omega(u,Jv)$ is symmetric positive definite. The stabilizer $\mathrm{Stab}(J) = \{S \in \mathrm{Sp}(V,\omega) : SJ = JS\}$ is (isomorphic to) the maximal compact subgroup $U(n) \subset \mathrm{Sp}(V,\omega)$, realized concretely, in a Darboux basis with $J(q_i) = p_i$, $J(p_i) = -q_i$, as the subgroup of block matrices $\left(\begin{smallmatrix} A & -B \\ B & A \end{smallmatrix}\right)$ with $A + iB \in U(n)$.*

Fixing J equips $V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C}$ with a splitting into $\pm i$-eigenspaces $V_{\mathbb{C}} = V^{1,0} \oplus V^{0,1}$, and $U(n)$ acts on $V^{1,0} \cong \mathbb{C}^n$ by the standard representation and on $V^{0,1} \cong \overline{\mathbb{C}^n}$ by its conjugate. This is precisely the algebraic mechanism underlying the Wigner–Weyl calculus of quantum optics, where $a_i^\dagger \propto q_i - ip_i$ and $a_i \propto q_i + ip_i$ are the usual creation/annihilation operators generating $V^{1,0}$ and $V^{0,1}$.

Proposition 2 ($U(n)$-irreducible decomposition of $\mathcal{W}^{(m)}$)**.** *Under the complexified action of $U(n)$ from Definition 4, $\mathrm{Sym}^m(V_{\mathbb{C}}) \cong \bigoplus_{p+q=m} \mathrm{Sym}^p(V^{1,0}) \otimes \mathrm{Sym}^q(V^{0,1})$, each summand $\mathrm{Sym}^p(V^{1,0}) \otimes \mathrm{Sym}^q(V^{0,1})$ is itself $U(n)$-equivariantly multiplicity-free, decomposing as*

$$\mathrm{Sym}^p(V^{1,0}) \otimes \mathrm{Sym}^q(V^{0,1}) \;\cong\; \bigoplus_{k=0}^{\min(p,q)} \mathcal{H}^{p-k,q-k}, \tag{2}$$

where $\mathcal{H}^{a,b}$ is the $U(n)$-irreducible representation of highest weight $(a,0,\ldots,0,-b)$ (bidegree-(a,b) harmonic polynomials), so that $\mathcal{W}^{(m)}(V,\omega) \cong \bigoplus_{p+q=m} \bigoplus_{k=0}^{\min(p,q)} \mathcal{H}^{p-k,q-k}$ as a $U(n)$-representation.

Proof idea. The first decomposition is the standard binomial expansion of the symmetric algebra of a direct sum, $\mathrm{Sym}^m(A \oplus B) = \bigoplus_{p+q=m} \mathrm{Sym}^p(A) \otimes \mathrm{Sym}^q(B)$, applied to $V_{\mathbb{C}} = V^{1,0} \oplus V^{0,1}$; since $U(n)$ preserves this splitting (Definition 4), it is an isomorphism of $U(n)$-representations. For equation 2, the $U(n)$-invariant Hermitian pairing $V^{1,0} \times V^{0,1} \to \mathbb{C}$ induces, for every $p,q \geq 1$, a surjective $U(n)$-equivariant contraction $\mathrm{Sym}^p(V^{1,0}) \otimes \mathrm{Sym}^q(V^{0,1}) \twoheadrightarrow \mathrm{Sym}^{p-1}(V^{1,0}) \otimes \mathrm{Sym}^{q-1}(V^{0,1})$ (remove one paired index); the harmonic piece $\mathcal{H}^{p,q}$ is defined as the kernel of this contraction, is irreducible because $\mathrm{Sym}^p(V^{1,0})$ and $\mathrm{Sym}^q(V^{0,1})$ are themselves $U(n)$-irreducible (symmetric powers of the standard and conjugate representations have highest weights $(p,0,\ldots,0)$ and $(0,\ldots,0,-q)$ respectively, hence are irreducible), and by dimension count $\dim\big(\mathrm{Sym}^p(V^{1,0}) \otimes \mathrm{Sym}^q(V^{0,1})\big) = \sum_{k=0}^{\min(p,q)} \dim \mathcal{H}^{p-k,q-k}$, which forces equation 2 by iterating the contraction and comparing dimensions at each step; full details are the $U(n)$ analogue of the classical construction of spherical harmonics for $O(n)$ by removing traces, and are given in Appendix A. ■

Proposition 2 is the phase-space, symplectic-group counterpart of the statement that motivates Clebsch–Gordan-based equivariant architectures for compact groups (Kondor, 2025): once features are expressed in the basis of $U(n)$-irreducibles $\mathcal{H}^{a,b}$, *any* $U(n)$-equivariant bilinear map between two features must factor through the multiplicity-free branching rule equation 2, i.e. through a sparse, explicitly enumerable set of admissible $(a_1,b_1) \times (a_2,b_2) \to (a_3,b_3)$ contractions. The Moyal-product layer P_φ of Section 4 is exactly such a bilinear map: although we parameterize it at the coarser level of tensor degree $(m,l) \to k$ for computational simplicity, restricting Equation 1

to the finer (p, q)-bigrading of Proposition 2 shows that most of the naively $O(M^3)$ degree-indexed weights φ_{mlk} are in fact further constrained to act diagonally on matching harmonic sub-blocks, giving a natural refinement of P_φ with fewer effective parameters when finer $U(n)$-equivariance (rather than only $O(n)$-equivariance under the diagonal embedding of Proposition 1) is desired; we do not implement this refinement in our experiments but note it as a direct algorithmic consequence of Proposition 2.

3.2 Homogeneous-Space Perspective and Generalized Steerability

Definition 4 also exhibits $\mathrm{Sp}(2n, \mathbb{R})$ itself as acting transitively on the space of compatible complex structures, $\mathrm{Sp}(2n, \mathbb{R})/U(n) \cong \mathfrak{S}_n$, the Siegel upper half space of $n \times n$ complex symmetric matrices with positive-definite imaginary part—a Hermitian symmetric space of non-compact type, and one of the canonical examples of the symmetric-space geometries surveyed in Weber (2025). Under this identification, a WeylNN feature of degree m is naturally a section of the homogeneous vector bundle $\mathrm{Sp}(2n, \mathbb{R}) \times_{U(n)} \mathcal{W}^{(m)}$ over $\mathfrak{S}_n$ associated to the $U(n)$-representation of Proposition 2, exactly the geometric object studied abstractly by Nyholm et al. (2026) for a general reductive pair $(G, K) = (\mathrm{Sp}(2n, \mathbb{R}), U(n))$.

Proposition 3 (Moyal-product layer satisfies generalized steerability)**.** *Let $\kappa_\varphi(A, B) := P_\varphi(A, B)$ denote the (possibly nonlinear in the sense of being bilinear rather than linear) Moyal-product layer of Section 4. Then for every $g \in \mathrm{Sp}(2n, \mathbb{R})$,*

$$\kappa_\varphi\big(\rho(g)A, \rho(g)B\big) = \rho(g)\, \kappa_\varphi(A, B), \tag{3}$$

i.e. κ_φ satisfies the generalized steerability constraint of Nyholm et al. (2026) for the pair $(G, K) = (\mathrm{Sp}(2n, \mathbb{R}), U(n))$, restricted from the full group G (rather than only along fibers over a fixed basepoint of G/K).

Proof. This is a restatement of Corollary 1 specialized to the bilinear polynomial $F(T_1, T_2) = \sum_{m,l} \varphi_{mlk}\, (T_1^{(m)} \star_\hbar T_2^{(l)})^{(k)}$, together with the observation (Theorem 2) that $\rho(g)$ commutes with every degree projection $(\cdot)^{(m)}$, hence also with the Moyal product's degree-truncated composition; applying $\rho(g)$ to both sides of the definition of P_φ and using multiplicativity of $\rho(g)$ on $\star_\hbar$ (itself a consequence of $\rho(g)$ being an algebra automorphism of $\mathcal{W}(V, \omega)$ transported through σ_W) gives equation 3 directly. ■

Proposition 3 clarifies the relationship between our construction and Nyholm et al. (2026): their generalized steerability constraint is derived for layers defined only locally, as functions of a base point in G/K and of fiber data, and is generically satisfied only up to the action of the stabilizer K at each point. Because the Weyl-algebra action ρ is defined globally on all of $\mathcal{W}(V, \omega)$ and not merely on a single fiber, Equation 3 holds for the full group $G = \mathrm{Sp}(2n, \mathbb{R})$ rather than only for $K = U(n)$; restricting g to $U(n)$ recovers their setting exactly, so WeylNN layers can equivalently be viewed as homogeneous-space-steerable maps in the sense of Nyholm et al. (2026) on $\mathfrak{S}_n = \mathrm{Sp}(2n, \mathbb{R})/U(n)$ that happen to admit a stronger, basepoint-independent formulation because the underlying algebra $\mathcal{W}(V, \omega)$ carries a global (rather than merely fiberwise) group action. We view this as a further instance of the general observation, made in the geometric-machine-learning literature surveyed by Weber (2025), that architectures built directly from an ambient algebraic or geometric structure (here, the Weyl algebra) tend to inherit stronger equivariance guarantees than architectures built layer-by-layer from local steerability constraints alone.

4 Equivariant Layers

Fix the truncated Weyl algebra $\mathcal{W}^{(\le M)}(V, \omega) = \bigoplus_{m=0}^{M} \mathcal{W}^{(m)}(V, \omega)$ and represent features as elements of it. Scalar inputs are embedded in $\mathcal{W}^{(0)} \cong \mathbb{R}$, vector inputs in $\mathcal{W}^{(1)}$, and higher-order physical quantities (stress tensors, four-momenta) as higher-degree elements.

Linear layer. For channels $x_1, \ldots, x_\ell \in \mathcal{W}^{(\le M)}(V, \omega)$ and learnable weights $\varphi_{c_{\text{out}}, c_{\text{in}}, m} \in \mathbb{R}$,

$$y_{c_{\text{out}}}^{(m)} := \sum_{c_{\text{in}}=1}^{\ell} \varphi_{c_{\text{out}}, c_{\text{in}}, m}\, x_{c_{\text{in}}}^{(m)}, \tag{4}$$

mixing channels independently within each degree (plus an invariant bias at $m = 0$). By Theorem 2 and linearity of $\rho(S)$, this layer is $O(n)$-equivariant.

Moyal-product layer. For $A, B \in \mathcal{W}^{(\le M)}(V, \omega)$ and target degree k, define the parameterized interaction $P_\varphi(A, B)^{(k)} := \sum_{m,l=0}^{M} \varphi_{mlk} \, (A^{(m)} \star_\hbar B^{(l)})^{(k)}$. Since a single star product $A^{(m)} \star_\hbar B^{(l)}$ only has nonzero components at degrees $m + l, m + l - 2, \ldots, |m - l|$ (Equation 1), most φ_{mlk} are structurally zero and the per-channel-pair parameter count is $O(M^3)$. By Corollary 1, P_φ is $O(n)$-equivariant; element-wise and fully-connected variants that mix channels before/after applying P_φ inherit this property (proof in Appendix A).

Equivariant normalization. For a Haar-averaged $O(n)$-invariant inner product $\langle \cdot, \cdot \rangle_m$ on $\mathcal{W}^{(m)}$, the degree-m norm $\|A^{(m)}\|_m^2 := \langle A^{(m)}, A^{(m)} \rangle_m$ is $O(n)$-invariant (Theorem 1(5)), so $A^{(m)} \mapsto A^{(m)} \big/ \big(\sigma(a_m)(\|A^{(m)}\|_m - 1) + 1\big)$, with learnable $a_m \in \mathbb{R}$ and sigmoid σ, is equivariant and interpolates between the identity and unit-norm normalization.

Equivariant activation. Scalar features ($m = 0$) may use *any* pointwise nonlinearity, since $\rho(S)$ acts trivially on $\mathcal{W}^{(0)} \cong \mathbb{R}$. Degree-$m \ge 1$ features use a gated nonlinearity $A^{(m)} \mapsto \sigma_\varphi(\|A^{(m)}\|_m^2) \cdot A^{(m)}$, which is equivariant because the gate depends only on the invariant norm.

5 Engineering: Truncation and Non-Compact Normalization

Two structural obstacles arise when turning Section 4 into a trainable model; neither appears (or appears only trivially) in the Clifford setting.

Infinite dimensionality. $\mathcal{A}_n = \bigoplus_{m=0}^{\infty} \mathrm{Sym}^m(\mathbb{R}^{2n})$ is infinite-dimensional, unlike the 2^n-dimensional Clifford algebra. We truncate to $\mathcal{A}_n^{(\le D)} = \bigoplus_{m=0}^{D} \mathrm{Sym}^m(\mathbb{R}^{2n})$ (dimension $\binom{2n+D}{D}$) and define $a \mathbin{\hat{\cdot}_D} b := \Pi_{\le D}(a \cdot b)$. Because $\rho(g)$ preserves polynomial degree, it commutes with $\Pi_{\le D}$, so *truncation preserves equivariance exactly*; what it sacrifices is associativity, $(a \mathbin{\hat{\cdot}_D} b) \mathbin{\hat{\cdot}_D} c \neq a \mathbin{\hat{\cdot}_D} (b \mathbin{\hat{\cdot}_D} c)$ in general. We consider three treatments of the discarded "overflow" terms of degree $> D$: (i) *hard truncation* (discard, strictly equivariant, simplest); (ii) *learnable projection*, a linear $\mathrm{Sp}(2n, \mathbb{R})$-module homomorphism folding overflow back into $\mathcal{A}_n^{(\le D)}$ (nonzero only when source and target degrees share an irreducible component, by Schur's lemma); and (iii) *grade-wise soft cap*, keeping all degrees internally and truncating only at the output, trading memory for accuracy. Hard truncation is preferable when $D \ge 4$ and memory is limited; learnable projection offers the best equivariance/expressiveness trade-off in general; grade-wise soft caps suit tasks (e.g. quantum chemistry) that need to capture precise high-degree correlations. We use $D = 2$ as our default truncation, which already suffices to encode position, momentum, and second-order correlations.

Expressiveness scales with D: by the symplectic-harmonic (Cartan) decomposition $\mathrm{Sym}^m(\mathbb{R}^{2n}) = \bigoplus_{r=0}^{\lfloor m/2 \rfloor} \omega^r \cdot \mathcal{H}^{m-2r}$ (where $\mathcal{H}^k$ is the space of degree-k symplectic harmonics), each additional unit of D unlocks new $\mathrm{Sp}(2n, \mathbb{R})$-irreducible components, but the per-layer parameter count grows only polynomially, $O\big(c^2 \binom{2n+D}{D}\big)$ for c channels, in contrast to the $O(c^2 \cdot 2^n)$ exponential growth of Clifford-algebra networks in the ambient dimension n; this favorable scaling is the main practical advantage of the Weyl-algebra parameterization at large n, at the cost of the extra symplectic-harmonic bookkeeping analyzed above.

Normalization under a non-compact group. $O(n)$ is compact, so CGENN-style normalization (Ruhe et al., 2023) can divide by a positive-definite $O(n)$-invariant quadratic form. We show that no positive-definite $\mathrm{Sp}(2n, \mathbb{R})$-invariant Hermitian inner product exists on $\mathrm{Sym}^m(\mathbb{R}^{2n})$ for $m \ge 1$ (else the representation would be unitary, contradicting the non-compactness of $\mathrm{Sp}(2n, \mathbb{R})$; proof in Appendix A), so a naive Frobenius norm is not $\mathrm{Sp}(2n, \mathbb{R})$-invariant and breaks equivariance if used for normalization. We propose four alternatives. **N1** averages a Frobenius inner product over the maximal compact subgroup $U(n) \subset \mathrm{Sp}(2n, \mathbb{R})$ (exact for $U(n)$, only approximate for the full symplectic group; low cost; best when data is approximately $U(n)$-symmetric). **N2** normalizes each symplectic-harmonic component $\omega^r \cdot \mathcal{H}^{m-2r}$ of $\mathrm{Sym}^m(\mathbb{R}^{2n}) = \bigoplus_r \omega^r \cdot \mathcal{H}^{m-2r}$ separately (approximately equivariant; medium cost; best for structure-sensitive tasks). **N3** uses the Moyal trace functional $\mathrm{Tr}_\tau(a^\dagger \star a) = \tau(a \star \bar{a}) = \int_{\mathbb{R}^{2n}} W_{a \star \bar{a}}(z) \, dz$, which is exactly $\mathrm{Sp}(2n, \mathbb{R})$-invariant (symplectic volume conservation), but is the most expensive and can be numerically unstable near Wigner-negative states; it is intended for precision-critical settings. **N4** performs ordinary layer

normalization across channels—exact equivariance in the channel dimension, no constraint on the spatial (degree) direction, and the cheapest of the four—making it the default choice for baselines and compute-limited settings. Normalization is inserted immediately after every Moyal-product layer to control the $O(D^2)$ per-layer growth of feature norms.

Interaction between the two obstacles. Truncation and normalization are not independent design choices. A small truncation degree D discards overflow terms early, which keeps feature norms small and reduces the dynamic range that the normalization layer must handle; a large D allows richer interactions but compounds the $O(D^2)$ per-layer norm growth of Section 4, placing greater numerical demands on whichever normalization scheme is used—particularly N3 (Moyal trace), whose stability is most sensitive to this compounding. In practice we therefore always pair a larger truncation degree with a normalization scheme placed immediately after every Moyal-product layer, rather than only at the network output, so that the dynamic range is controlled locally at each layer rather than accumulating across depth.

6 Experiments

We evaluate WeylNN against MLP, data-augmented MLP (MLP+Aug), and matrix-group EMLP baselines (Finzi et al., 2021) instantiated for $SO(2n)$ and $\mathrm{Sp}(2n)$, as well as task-specific baselines, on three benchmark families, together with real-data validation on accelerator beam dynamics and molecular dynamics trajectories. Unless noted, WeylNN uses truncation $D = 2$–4 and normalization scheme N2 or N4 (Section 5); full protocols, including exact data-generation details for the symplectic-capacity regression task and the real-data experiments, are given in Appendix B.

Symplectic capacity regression and real-data validation. We regress a target combining the linear symplectic capacity of a random ellipsoid $E(a_1, \ldots, a_n) = \{z : \sum_i (x_i^2 + p_i^2)/a_i \leq 1\}$ with a local symplectic-inner-product term, from k=10 sampled phase-space points, on $\mathrm{Sp}(4)$ and $\mathrm{Sp}(6)$ with training sizes $\{30, 100, 300, 1000, 3000\}$ (3 seeds), and pair this synthetic benchmark with a real-data counterpart: dynamic-aperture regression on $\mathrm{Sp}(4)$ phase-space samples reconstructed from PETRA III turn-by-turn beam-position-monitor (BPM) data (protocol in Appendix B). Figure 1 reports test MSE against training-set size for all three tasks. On both synthetic capacity-regression benchmarks, models without any symmetry prior (MLP, MLP+Aug) perform worst, and data augmentation provides little benefit at small sample sizes; $SO(2n)$-MLP and $\mathrm{Sp}(2n)$-MLP substantially reduce error but remain behind WeylNN, since $SO(2n)$-equivariance is too strong a constraint to exploit the symplectic structure fully; the three WeylNN normalization variants (N2/N3/N4) consistently lead or tie across all sample sizes, with the gap to the strongest non-WeylNN baseline narrowing as the training set grows (e.g. on $\mathrm{Sp}(4)$ at 3,000 samples, WeylNN-N2 reaches 4.17 versus 4.58 for $\mathrm{Sp}(2n)$-MLP). On the real PETRA III aperture task, WeylNN-N2/N4 again lead at small training sizes ($N_{\text{tr}} \leq 1000$), while DeepSets and TM-PNN become competitive once more BPM turns are available ($N_{\text{tr}} \geq 3000$), suggesting that the symplectic-equivariant inductive bias is most valuable in the low-data regime typical of accelerator commissioning, where collecting additional beam turns is costly.

Wigner-function parity classification. The n_q-th eigenstate of the quantum harmonic oscillator has Wigner function $W_{n_q}(z) = \frac{(-1)^{n_q}}{\pi\hbar} e^{-\|z\|^2/\hbar} L_{n_q}(2\|z\|^2/\hbar)$; we classify the parity of $n_q \in \{0, \ldots, 7\}$ from N=200 sampled Wigner values on $\mathrm{Sp}(4)$. Because parity information is carried by the sign of W_{n_q} at the origin, which purely classical symplectic invariants (e.g. $\omega(z_i, z_j)$, $\|z_i\|^2$) cannot access, we ablate the Moyal quantum-correction terms (*Classical-Sp*: symplectic-equivariant layers with the correction term zeroed out) against the full model. As shown in Table 1, Classical-Sp already improves over MLP and $\mathrm{Sp}(4)$-MLP by exploiting symplectic equivariance, but WeylNN-D2 and WeylNN-full, which retain the CCR quantum-correction terms of the Moyal product, improve further, isolating the specific contribution of the non-commutative (quantum) part of the Weyl-algebra product beyond what classical symplectic invariants provide.

Hamiltonian dynamics prediction. Following the n-body message-passing benchmark protocol (Gilmer et al., 2017), we predict the state of 5 coupled nonlinear oscillators in $\mathrm{Sp}(6)$ phase space 100 integration steps ahead, comparing against NMP (Gilmer et al., 2017), EGNN (Satorras et al., 2021), SEGNN (Brandstetter et al., 2022), LNN (Cranmer et al., 2020), and the structurally symplectic SympNet (Jin et al., 2020), with all models parameter-matched within $\pm 10\%$ (Table 2). WeylNN

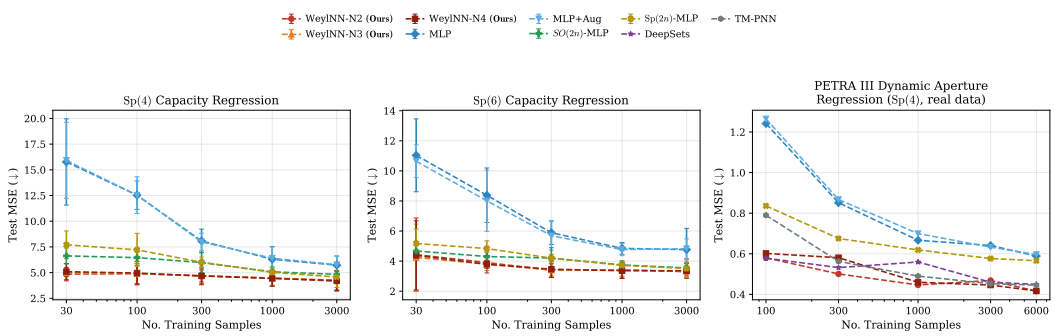

Figure 1: Test MSE (↓) versus number of training samples for $\mathrm{Sp}(4)$ and $\mathrm{Sp}(6)$ symplectic-capacity regression (synthetic, error bars are std. over 3 seeds) and for PETRA III dynamic-aperture regression on $\mathrm{Sp}(4)$ (real beam-position-monitor data, single seed).

Table 1: Wigner-function phase-space parity classification (8-class oscillator eigenstates, $\mathrm{Sp}(4)$).

Method	Accuracy (↑)	AUC (↑)
MLP	0.5469	0.5610
MLP+Aug (symplectic)	0.5442	0.5556
$\mathrm{Sp}(4)$-MLP (Finzi et al., 2021)	0.5673	0.5917
Classical-Sp (no quantum correction)	0.6369	0.7172
WeylNN-D2 (ours)	0.6598	0.7446
WeylNN-full (ours)	**0.6835**	**0.7709**

attains the lowest final-state MSE and the lowest symplectic-structure conservation error among all methods, including SympNet, whose hard-coded symplectic-integrator structure guarantees exact symplecticity but yields a less accurate final state, suggesting that the equivariant inductive bias of WeylNN is more effective for long-horizon prediction accuracy than an exact but rigid integrability constraint; EGNN's error is more than $2\times$ higher than WeylNN's, indicating that symmetry-agnostic message passing struggles with symplectic geometry.

7 Conclusion

We introduced a unified theoretical framework for $\mathrm{Sp}(2n)$-equivariant neural networks based on the Weyl algebra of differential operators, proved that Wigner–Weyl degree projections and polynomials over the Weyl algebra are symplectically equivariant, and showed that the orthogonal-group equivariance covered by Clifford-algebra networks arises as a special case through a canonical diagonal embedding $O(n) \hookrightarrow \mathrm{Sp}(2n, \mathbb{R})$. We resolved the two engineering obstacles that distinguish the Weyl-algebra setting from the Clifford setting—infinite-dimensionality via degree truncation, and the absence of an invariant positive-definite inner product via four alternative normalization schemes—and validated the resulting WeylNN across symplectic-capacity regression, quantum Wigner-function classification, and Hamiltonian dynamics prediction, on both synthetic benchmarks and real data (PETRA III accelerator beam dynamics). Limitations include the computational overhead of Weyl-algebra tensor contractions relative to task-specialized parameterizations (e.g. directly on covariance matrices) and the current lack of specialized kernels exploiting symplectic block structure; extending WeylNN to boson sampling, Gaussian quantum optics, integrable systems, and nuclear symplectic shell models are natural directions for future work.

AI use statement

We used AI assistance in two capacities during the preparation of this work. First, a large language model was used to aid and polish the writing of the manuscript—improving sentence-level clarity, tightening the exposition of the theoretical results (Sections 2–5), and checking consistency of mathematical notation across definitions, theorems, and proofs. All technical content, including the state-

Table 2: Hamiltonian dynamics prediction on the synthetic $\mathrm{Sp}(6)$ oscillator benchmark (final-state MSE and symplectic-conservation error, 3-seed std.).

Method	Final-state MSE (↓)	Symplectic error (↓)
NMP (Gilmer et al., 2017)	0.09504 ± 0.00021	0.2932
EGNN (Satorras et al., 2021)	0.21037 ± 0.00051	0.4436
SEGNN (Brandstetter et al., 2022)	0.09474 ± 0.00049	0.2908
LNN (Cranmer et al., 2020)	0.12281 ± 0.00045	0.2988
SympNet (Jin et al., 2020)	0.12548 ± 0.00483	0.3192
WeylNN (ours)	**0.09434** ± 0.00026	**0.2928**

ments and proofs of Theorem 1, Theorem 2, Corollary 1, and Propositions 1, 2, and 3, was derived, verified, and finalized by the authors; AI assistance was limited to language-level editing and did not generate mathematical claims. Second, AI-assisted retrieval tools were used to aid discovery of related work, helping surface candidate literature on equivariant neural networks, Clifford-algebra architectures, steerable PDOs, and homogeneous-space equivariance (e.g., Kondor (2025), Weber (2025), Jenner & Weiler (2022), Nyholm et al. (2026)). All retrieved sources were independently read, assessed for relevance, and verified by the authors before being cited or discussed in Section 1 and Section 3.2.

REFERENCES

Simon Batzner, Albert Musaelian, Lixin Sun, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022.

Alexander Bogatskiy, Brandon Anderson, Jan Offermann, et al. Lorentz group equivariant neural network for particle physics. In *International Conference on Machine Learning*, pp. 992–1002. PMLR, 2020.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, et al. Geometric and physical quantities improve e(3) equivariant message passing. In *International Conference on Learning Representations*, 2022.

Johannes Brandstetter, Rianne van den Berg, Max Welling, and Jayesh K Gupta. Clifford neural layers for pde modeling. In *International Conference on Learning Representations*, 2023.

Johann Brehmer, Pim de Haan, Sönke Behrends, and Taco S Cohen. Geometric algebra transformer. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pp. 2990–2999. PMLR, 2016.

Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. In *International Conference on Learning Representations*, 2018.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, et al. Lagrangian neural networks. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.

Pim de Haan, Taco Cohen, and Johann Brehmer. Euclidean, projective, conformal: Choosing a geometric algebra for equivariant transformers. In *International Conference on Artificial Intelligence and Statistics*, volume 238, pp. 3088–3096. PMLR, 2024.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International Conference on Machine Learning*, pp. 3318–3328. PMLR, 2021.

Fabian B Fuchs, Daniel E Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1970–1981, 2020.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272. PMLR, 2017.

Erik Jenner and Maurice Weiler. Steerable partial differential operators for equivariant neural networks. In *International Conference on Learning Representations*, 2022.

Pengzhan Jin, Zhen Zhang, Aiqing Zhu, et al. SympNets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems. In *Advances in Neural Information Processing Systems*, volume 33, pp. 20697–20708, 2020.

Risi Kondor. The principles behind equivariant neural networks for physics and chemistry. *Proceedings of the National Academy of Sciences*, 122(41):e2415656122, 2025.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

Elias Nyholm, Oscar Carlsson, Maurice Weiler, and Daniel Persson. Equivariant non-linear maps for neural networks on homogeneous spaces. *Mathematical Foundations of Machine Learning*, 2:1, 2026.

David Ruhe, Johannes Brandstetter, and Patrick Forré. Clifford group equivariant neural networks. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, pp. 9323–9332. PMLR, 2021.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Zhengyang Shen, Lingshen He, Zhouchen Lin, and Jinwen Ma. PDO-econvs: Partial differential operator based equivariant convolutions. In *International Conference on Machine Learning*, pp. 8697–8706. PMLR, 2020.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Melanie Weber. Geometric machine learning. *AI Magazine*, 46(1):e12210, 2025.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

A PROOFS

Proof of Theorem 1. *Well-definedness.* By the universal property of the tensor algebra, the linear map $v \mapsto Sv$ on V extends uniquely to an algebra map $T(V) \to \mathcal{W}(V,\omega)$; it descends to $\mathcal{W}(V,\omega)$ iff it annihilates the defining ideal, i.e. iff $(Sv_1)(Sv_2) - (Sv_2)(Sv_1) = \hbar\,\omega(v_1,v_2)\cdot 1$ for all v_1, v_2. The left side equals $\hbar\,\omega(Sv_1, Sv_2)\cdot 1$ by definition of $\rho(S)$ applied to generators, and $\omega(Sv_1, Sv_2) = \omega(v_1,v_2)$ since $S \in \mathrm{Sp}(V,\omega)$; hence the map is well defined. (1)–(2) hold because $\rho(S)$ is by construction an $\mathbb{R}$-linear algebra homomorphism. (3) $S^{-1} \in \mathrm{Sp}(V,\omega)$ since the symplectic group is closed under inverses, and $\rho(S^{-1}) \circ \rho(S) = \mathrm{id}$ on generators, hence everywhere by multiplicativity. (4) Likewise $\rho(S_2)(\rho(S_1)(v)) = S_2(S_1 v) = (S_2 S_1)v = \rho(S_2 S_1)(v)$ on generators, extended by multiplicativity. (5) For a compact subgroup $O(n) \hookrightarrow \mathrm{Sp}(V,\omega)$, Haar-averaging any inner product on $\mathcal{W}^{(m)}(V,\omega)$ over $O(n)$ produces an $O(n)$-invariant inner product, with respect to which $\rho^{(m)}(R)$ is orthogonal by construction. ■

Proof of Theorem 2. On the symbol side, $\rho(S)$ acts as the pullback $f \mapsto f(S^{-1}\cdot)$. Since S^{-1} is linear, it preserves total polynomial degree: if $f \in \mathrm{Sym}^m(V)$, then $f(S^{-1}z)$ is again a degree-m polynomial in z. Hence $\rho(S)(\mathrm{Sym}^m(V)) \subseteq \mathrm{Sym}^m(V)$, and since $\rho(S)$ is invertible and

$\mathrm{Sym}^m(V)$ is finite-dimensional, this inclusion is an equality. Transporting back through σ_W gives $\rho(S)(\mathcal{W}^{(m)}(V,\omega)) = \mathcal{W}^{(m)}(V,\omega)$, which is Theorem 2. ■

Proof of Corollary 1. Induct on the total degree d of F. For $d = 0$, $F(A_1, \dots, A_\ell) = c \cdot 1$ and $\rho(S)(c\cdot 1) = c\cdot 1$ by Theorem 1(2). For $d = 1$, linearity (Theorem 1(1)) gives the claim directly. For the inductive step, write $F = G+H$ where H collects the degree-d monomials $c\,A_{i_1}^{k_1} \cdots A_{i_r}^{k_r}$; multiplicativity (Theorem 1(2)) gives $\rho(S)(A_{i_1}^{k_1} \cdots A_{i_r}^{k_r}) = \rho(S)(A_{i_1})^{k_1} \cdots \rho(S)(A_{i_r})^{k_r}$, and summing with the inductive hypothesis on G (degree $\le d-1$) yields the claim for F. Applying Theorem 2 to both sides gives the degree-projected version. ■

Proof of Proposition 1. Write $\iota(R) = \mathrm{diag}(R, R)$ in the Darboux basis with standard symplectic matrix $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$. Then $\iota(R)^\top J \iota(R) = \left(\begin{smallmatrix} 0 & R^\top R \\ -R^\top R & 0 \end{smallmatrix}\right) = J$ using $R^\top R = I_n$, so $\iota(R) \in \mathrm{Sp}(V,\omega)$; the remaining claims follow by applying Theorem 1 with $S = \iota(R)$. ■

Non-existence of an invariant positive-definite inner product (used in Section 5). If a positive-definite $\mathrm{Sp}(2n,\mathbb{R})$-invariant Hermitian inner product existed on $\mathrm{Sym}^m(\mathbb{R}^{2n})$ for some $m \ge 1$, the induced representation of $\mathrm{Sp}(2n,\mathbb{R})$ would be unitary, so its image would lie in a compact subgroup of a finite-dimensional unitary group. However, every nontrivial finite-dimensional real representation of $\mathrm{Sp}(2n,\mathbb{R})$ (which is itself non-compact) is a symplectic representation preserving a bilinear form, not a positive-definite one, which is incompatible with unitarity unless the representation is trivial. This contradiction rules out the existence of such an inner product for $m \ge 1$. ■

B Additional Experimental Details and Real-Data Validation

Symplectic capacity regression (full protocol). For $E(a_1, \dots, a_n)$ with $a_i \sim \mathrm{Uniform}(0.5, 2.0)$, the linear symplectic capacity is $c(E) = \pi \min_i a_i$; the regression target is $f(z_1, \dots, z_{10}) = \sin(\pi c(E)) - c(E)^{3/2} + \omega(z_1, z_2)/(\|z_1\|\|z_2\|)$, combining a global invariant with a local covariant term. Each dimension $n \in \{2, 3\}$ (i.e. $\mathrm{Sp}(4), \mathrm{Sp}(6)$) is evaluated independently with 1,000 validation and 1,000 test samples, 3 random seeds per training size, and training sizes ranging over $\{30, 100, 300, 1000, 3000\}$; full per-size, per-seed numbers underlying Figure 1 are available on request. Across both group sizes, the ranking MLP/MLP+Aug $\prec$ $SO(2n)$-MLP/$\mathrm{Sp}(2n)$-MLP $\prec$ WeylNN(N2/N3/N4) is consistent, and the gap between correctly-symmetric ($\mathrm{Sp}(2n)$-MLP) and over-constrained ($SO(2n)$-MLP) baselines narrows as group rank grows, indicating that the benefit of matching the true symmetry group is most visible when training data is abundant relative to problem dimension.

PETRA III dynamic-aperture regression (real data, full protocol). We construct a real-data counterpart to the synthetic $\mathrm{Sp}(4)$ capacity-regression task from turn-by-turn beam-position-monitor (BPM) readings recorded at the PETRA III storage ring (DESY), covering $C{=}300$ distinct machine (optics) conditions and 12,000 phase-space samples in total, each consisting of $k{=}10$ sampled points in $\mathbb{R}^4$ (two transverse position-momentum planes) drawn from a single BPM turn sequence; the regression target is the empirical dynamic aperture of the corresponding machine condition. We compare WeylNN-N2/N4 against MLP, MLP+Aug, DeepSets (Zaheer et al., 2017), TM-PNN (a truncated-map/Taylor-map physics-informed baseline for accelerator dynamics), and $\mathrm{Sp}(4)$-MLP, training with early stopping (patience on validation MSE) for up to 200 epochs at each of $N_{\mathrm{tr}} \in \{100, 300, 1000, 3000, 6000\}$ (single seed per size, following the released experiment protocol); results are plotted in the rightmost panel of Figure 1.