DEGREE BOUNDS FOR SEPARATING POLYNOMIAL AND NEURAL INVARIANTS FOR POINT CLOUDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural networks that respect the symmetries of point clouds commonly compute high-order representations of the rotation group to improve their expressivity. Such models are known to be universal, and thus able to separate any pair of point clouds, when using representations of sufficiently high order, yet no precise bound on this order, as a function of the number of points, was previously known. We establish the first quadratic lower and upper degree bounds for polynomials that separate point clouds of n points up to rotations and permutations. By showing Tensor Field Networks and MACE can express all invariant polynomials of a given degree with sufficient representation order, we additionally derive the first explicit, quadratic bound on the representation order sufficient for these architectures to be complete, replacing previous existence results with concrete guarantees. We further show that, under local regularity assumptions, point cloud separation can be achieved with a message-passing-based model using invariants whose degree depends only on the size of a local neighborhood, rather than on the size of the full point cloud. Finally, we provide counterexamples showing MACE must compute high order irreducible representations to separate locally regular point clouds.

1 INTRODUCTION

Learning functions that are invariant, or equivariant, to a group of symmetries is a central problem in geometric deep learning Bronstein et al. (2021). One of the most important instances of this problem is the case of *point clouds*, or finite sets of points, that efficiently represent molecules, particle systems and 3D shapes, where the relevant symmetry group is generated by permutations of the points together with the rigid motions of the ambient space Guo et al. (2019); Han et al. (2022); Sanchez-Gonzalez et al. (2020). Here the goal is to design architectures whose internal mechanisms respect the symmetries of point clouds, that is point ordering and orientation in space, and whose predictions depend only on the underlying geometry rather than on ambiguous choices of point order or choice of frame of reference Schütt et al. (2017); Thomas et al. (2018).

A natural theoretical desideratum of invariant architectures is *completeness*: informally, a family of invariant functions is complete if it can separate every pair of inputs that are not related by a symmetry, that is, that do not lie in the same group orbit. This follows from completeness being a necessary and sufficient condition for the family to be able to approximate *every* continuous invariant function, and hence for the corresponding architecture to be a universal approximator of invariant functions Chen et al. (2019); Dym & Gortler (2025).

The completeness of invariant architectures for point clouds has consequently received a great deal of attention. For example, the DeepSets model family was proven to be complete on set-structured data Zaheer et al. (2017), while message-passing neural networks for graphs are not Morris et al. (2023). For point clouds, it is known that distance-based message-passing approaches are *not* complete Pozdnyakov & Ceriotti (2022), while high-order invariant graph neural networks can attain completeness Hordan et al. (2024b); Rose et al. (2023); Li et al. (2023). Yet, these approaches suffer from high computational cost, making them difficult to scale.

A popular methodology to increase the expressive power of geometric message passing neural networks while maintaining tractable runtime is to use higher order tensor representations of the rotation group. This approach is successfully used in many modern day models for point clouds such as MACE Batatia et al. (2022) and Equiformer Liao & Smidt (2023). In Dym & Maron (2021) it

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.

is shown that these higher order approaches are complete. The mechanism behind these results is the following: an arbitrarily high-order representation can express polynomials of an arbitrarily high degree, and the class of invariant polynomials of sufficiently high degree is complete. What has been missing, up to this work, is any quantitative control on *how* high the representation order needs to be for completeness to hold.

In this work we provide the first answer to this question, in the form of explicit degree bounds for separating polynomial invariants of the relevant group action, that is the minimal polynomial degree needed to separate all orbits (defined precisely in Section 2). For point clouds of n points in $\mathbb{R}^d$, we show that the separation degree is lower bounded by a bound proportional to n^2 (Theorem 1), and upper bounded by a bound proportional to dn^2 (Theorem 4). Thus, in the main case of interest where $d = 3$, this bound is tight, up to a constant. As a corollary, we deduce a quadratic lower bound and a cubic upper bound for the polynomial degree sufficient for separating weighted graphs (Proposition 5).

We then use our bounds to derive concrete completeness guarantees for existing invariant architectures: as we show in Section 3, it upgrades classical results in the geometric deep learning literature, which established completeness only for *some* sufficiently high, unspecified, representation order, into an explicit quadratic bound on that order for models such as Tensor Field Networks Thomas et al. (2018) and MACE Batatia et al. (2022). Moreover, we show that, under locality assumptions on the point cloud, our results can be leveraged to obtain completeness for MACE with a representation order that depends quadratically only on the size of local neighborhood rather than on the size of the entire point cloud. We then show, by a counterexample construction, that if the representation order and correlation degree aren't sufficiently high then MACE cannot separate locally regular configurations (Proposition 12).

In summary, the main results of this paper are:

1. Tight quadratic degree bounds for separation of 3D point clouds by polynomials.
2. First quantitative completeness results for MACE and TFN, with quadratic dependence on point cloud cardinality.
3. Under locality assumptions on point clouds, an improvement of this bound to quadratic in the neighborhood size, and a lower bound showing that, with bounded depth and correlation order, the irreducible order must grow at least linearly in the neighborhood size.

1.1 Related Work

Invariant Theory The origins of Invariant Theory, which is the study of invariant polynomials, go back to the 19^{th} century. Its goal is to understand invariant polynomials, including explicit description of the generators of the invariant ring, or bounds on the degree of these generators. A celebrated result by Hilbert (1890) states that the ring of invariant polynomials is generated by a finite number of polynomials. Classical results in invariant theory focus on explicit derivation of these generators for specific group actions, or on bounding the cardinality or maximal degree of these generators. More recently there has been growing interest in the notion of separating invariants, which for compact groups are identical to the notion of complete invariants defined here. Generators are also separating invariants, and as a result the degree needed for separation cannot be larger than the maximal degree of the generators.

To the best of our knowledge, there are no known results regarding the separation degree for point clouds. For weighted graphs, Thiéry proved in Thiéry (2000) an $O(n^4)$ upper bound on the degree of the generators. A bound with similar asymptotics can also be obtained from a general bound on permutation groups Göbel (1995). In this paper we discuss the separation degree of weighted graphs, and prove a quadratic lower bound and a cubic upper bound.

For more background on invariant theory and further related work see Appendix F and Appendix A.

1.2 Notation

We consider the setting of a group G acting on a set V. We denote the orbit of an element $v \in V$ by $[v] := \{gv | g \in G\}$ and say that $q : V \to \mathbb{R}$ is G invariant if $q(u) = q(v)$ for all $v \in V, u \in [v]$.

Let $\mathcal{Q}$ be a family of G invariant functions from V to $\mathbb{R}$. We say that $\mathcal{Q}$ is (G, V) *complete*, if for all $v, u \in V$ with $[v] \neq [u]$, there exists a function $q \in \mathcal{Q}$ such that $q(v) \neq q(u)$.

Now assume additionally that V is a finite dimensional real vector space, so that the notion of polynomials is defined. We denote the set of all polynomials $p : V \to \mathbb{R}$ with real coefficients by $\mathbb{R}[V]$, and the subset of polynomials of degree $\leq D$ by $\mathbb{R}[V]_{\leq D}$. The subset of G invariant polynomials (of degree $\leq D$) are denoted by $\mathbb{R}[V]^G$ (and $\mathbb{R}[V]^G_{\leq D}$). The *separation degree* of (G, V), denoted by $D(G, V)$, is the smallest nonnegative integer D such that $\mathbb{R}[V]^G_{\leq D}$ is (G, V) complete.

Our interest in this paper will be in completeness and separation degree for point clouds. A *point cloud* with n points in $\mathbb{R}^d$ is a tuple $X = (x_1, \ldots, x_n)$ residing in the vector space $V = \mathbb{R}^{d\times n}$. We are interested in the action of rigid motions and permutations on this space. Explicitly, for a vector $t \in \mathbb{R}^d$, an orthogonal matrix $R \in O(d)$, and a permutation $\sigma \in S_n$, this action is defined by

$$(t, R, \sigma) \cdot X := \big(R(x_{\sigma^{-1}(1)} - t),\ \ldots,\ R(x_{\sigma^{-1}(n)} - t)\big).$$

We are interested in four variants of this problem, which come up in different applications: we will consider groups of the form $G = H \times S_n$ where $H = E(d)$ (all rigid motions), $H = O(d)$ (the translation component t is set to zero), $H = SO(d)$ (no translations, only proper rotations), and $H = SE(d)$ (with translations, only proper rotations). For each of these groups, we will derive bounds for the separation degree $D(G, \mathbb{R}^{d\times n})$.

We will also be interested in the related notion of separation for *weighted graphs*. A weighted graph is a symmetric $n \times n$ matrix with real coefficients. We denote the set of such matrices by $\mathcal{H}_n$. The permutation group S_n acts on $\mathcal{H}_n$ by conjugation. Namely, if $\tau \in S_n$ and $A \in \mathcal{H}_n$ then $\tau \cdot A$ is the matrix defined by

$$(\tau \cdot A)_{ij} = A_{\tau^{-1}(i),\tau^{-1}(j)}.$$

If A is a (weighted) adjacency matrix of a graph with some fixed node ordering, then $\tau \cdot A$ is the adjacency matrix of the same graph, with its nodes reordered by τ. We use 1_n (or 0_n) to denote the n-dimensional vector whose entries are all 1 (or 0).

2 Degree Bounds for Separating Polynomials on Point Clouds

In this section we answer the following questions: what is a polynomial degree required for separation and what polynomial degree is sufficient for separation of point clouds? We derive a lower bound for the polynomial degree of separation for point clouds, which is proportional to n^2 (Subsection 2.1), and an upper bound proportional to dn^2 (Subsection 2.2). From this we derive a quadratic lower bound and a cubic upper bound on the separation degree for weighted graphs (Section 2.3).

2.1 Lower Bound for Point Clouds

We begin by stating our lower bound, that is a polynomial degree required for separating point clouds.

Theorem 1. *Let $d \geq 2, n \geq 3$. Consider the action of $G = H \times S_n$ on $V = \mathbb{R}^{d\times n}$, where $H \in \{E(d), SE(d), O(d), SO(d)\}$. Then the separation degree is lower bounded by*

$$D(G, V) \geq n^2/4 - 1.$$

The remainder of this subsection is devoted to proving this theorem. We begin with some preliminaries. Firstly, the following lemma establishes a simple relationship between the separation degree with and without translations:

Lemma 2. *Let $V = \mathbb{R}^{d\times n}$ and V_{cent} denote the subspace of point clouds $X \in \mathbb{R}^{d\times n}$ with $X1_n = 0_d$. Then*

$$D(E(d) \times S_n, V) = D(O(d) \times S_n, V_{\text{cent}}) \leq D(O(d) \times S_n, V) \tag{1}$$

$$D(SE(d) \times S_n, V) = D(SO(d) \times S_n, V_{\text{cent}}) \leq D(SO(d) \times S_n, V) \tag{2}$$

Proof idea. Equation 1 follows from the fact that two point clouds X, Y are in the same $E(d) \times S_n$ orbit, if and only if the centralized point clouds $X - \frac{1}{n}X1_n1_n^T, Y - \frac{1}{n}Y1_n1_n^T$ are in the same $O(d) \times S_n$ orbit. The proof for equation 2 is identical. □

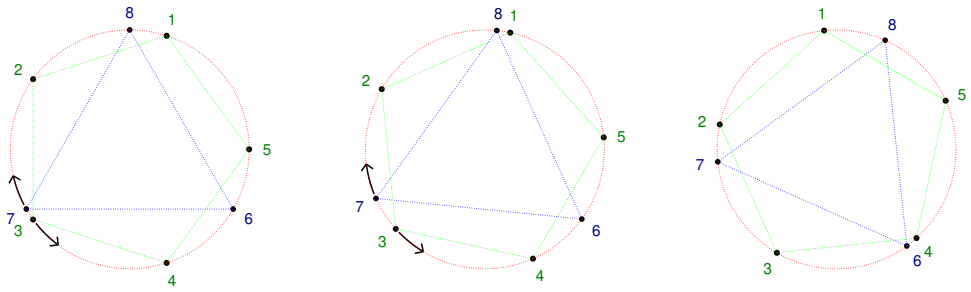


Figure 1: From left to right we have plotted the coordinates of $w(0)$, $w(0.1)$ and $w(\frac{2\pi}{15}) = w(0.4189)$ in the complex plane for $n = 8$. We have $p = 5$, $q = 3$ and $r = 4$. As t goes from 0 to 0.1 the pentagon moves counterclockwise and the triangle moves clockwise. The point clouds of $w(0)$ and $w(\frac{2\pi}{15})$ are isometric. The latter is obtained from the first by rotation over $\frac{8\pi}{15}$ and applying the permutation $(1\ 2\ 3\ 4\ 5)(8\ 7\ 6)$ to the columns of w.

The proof of the theorem is strongly based on the following proposition:

Proposition 3. *Let G be a compact group acting orthogonally on a finite dimensional real inner product space V. Let $k \geq 1$ be an integer. Suppose there exists a 2-dimensional subspace $W \subseteq V$ and an element $g \in G$ such that $g \cdot W \subseteq W$ and g acts by rotation over $2\pi/k$ on W. If the unit circle $S(W) \subset W$ contains more than 1 orbit, then we have $D(G, V) \geq k$.*

Proof. We choose an orthonormal basis of W and identify $\mathbb{R}[W] \cong \mathbb{R}[x_1, x_2]$. Restricting polynomials gives a surjective ring homomorphism $\mathbb{R}[V] \to \mathbb{R}[x_1, x_2]$. Let H be the subgroup of G generated by g. It is well-known that the invariant ring $\mathbb{R}[x_1, x_2]^H$ is generated by $x_1^2 + x_2^2$ and the polynomials $f_1, f_2 \in \mathbb{R}[x_1, x_2]^H$ of degree k given by $(x_1 + ix_2)^k = f_1 + if_2$ (see for example Yasinsky (2022)). In particular, every element of $\mathbb{R}[x_1, x_2]^H$ of degree $< k$ is a polynomial in $x_1^2 + x_2^2$ and is constant on $S(W)$. Since not all elements in $S(W)$ lie in the same G-orbit, there exists an invariant $f \in \mathbb{R}[V]^G$ of degree $\leq D(G, V)$ that is not constant on $S(W)$. If f' is the restriction of f to W, then $f' \in \mathbb{R}[W]^H = \mathbb{R}[x_1, x_2]^H$ is not constant on $S(W)$, so we have $D(G, V) \geq \deg(f) \geq \deg(f') \geq k$. □

We now return to the proof of Theorem 1. By Lemma 2 it is sufficient to prove a bound for all four groups in the lemma, by considering the two groups $G = O(d) \times S_n$ or $G = SO(d) \times S_n$, and the vector space V_{cent}- the space of centralized point clouds. Throughout, we will assume G is one of these two groups.

We first consider the case $d = 2$. We identify $\mathbb{R}^2$ with $\mathbb{C}$ and $\mathbb{R}^{2\times n}$ with $\mathbb{C}^n$. If $n = 2m$ is even, we define $p = m + 1$, $q = m - 1$, $u = 1$ and $r = m$. If $n = 2m + 1$ then we define $p = m + 1$, $q = m$, $u = 2$ and $r = 2m + 1$. We have $n = p + q$, $r = up - 1 = uq + 1$ and $pq \geq n^2/4 - 1$. For $t \in \mathbb{R}$ and all $j \in \mathbb{Z}$, define

$$a_j(t) = e^{i(2\pi j/p + t)} \quad \text{and} \quad b_j(t) = e^{i(\pi/2 - 2\pi j/q - t)}.$$

So $a_1(t), a_2(t), \ldots, a_p(t)$ are the counterclockwise consecutive vertices of a regular p-gon centered around the origin. The vertices $b_1(t), b_2(t), \ldots, b_q(t)$ are clockwise consecutive vertices of a regular q-gon. As t increases, the p-gon rotates counter-clockwise and the q-gon rotates clockwise. We easily verify that $a_j(t + \frac{\pi}{2}) = ia_j(t)$, $b_j(t + \frac{\pi}{2}) = -ib_j(t)$, $a_{j+p}(t) = a_j(t)$, $b_{j+q}(t) = b_j(t)$ and

$$\begin{aligned} a_j(s + t) &= e^{is} a_j(t) = (\cos(s) + i \sin(s)) a_j(t) = \cos(s) a_j(t) + \sin(s) a_j(t + \tfrac{\pi}{2}) \\ b_j(s + t) &= e^{-is} b_j(t) = (\cos(s) - i \sin(s)) b_j(t) = \cos(s) b_j(t) + \sin(s) b_j(t + \tfrac{\pi}{2}) \end{aligned}$$

Define

$$w(t) = (a_1(t) \quad a_2(t) \quad \cdots \quad a_p(t) \quad b_1(t) \quad b_2(t) \quad \cdots \quad b_q(t)).$$

The coordinates of $w(t)$ form a point cloud in the complex plane (see figure 1) and this point cloud is centralized. We have $w(t + s) = \cos(s) w(t) + \sin(s) w(t + \frac{\pi}{2})$. This shows that $w(t)$ lies in the real span of the two linearly independent matrices $w(0)$ and $w(\frac{\pi}{2})$. Now we take $s = \frac{2\pi}{pq}$. Because

$r = up - 1 = uq + 1$, we have

$$\begin{aligned} a_i\big(t + \tfrac{2\pi}{pq}\big) &= a_i\big(t + \tfrac{2\pi r}{pq} - \tfrac{2\pi u}{p}\big) = a_{i-u}\big(t + \tfrac{2\pi r}{pq}\big) = e^{2\pi i r/pq} a_{i-u}(t) \\ b_j\big(t + \tfrac{2\pi}{pq}\big) &= b_j\big(t - \tfrac{2\pi r}{pq} + \tfrac{2\pi u}{q}\big) = b_{j+u}\big(t - \tfrac{2\pi r}{pq}\big) = e^{2\pi i r/pq} b_{j+u}(t) \end{aligned}$$

and

$$w\big(t + \tfrac{2\pi}{pq}\big) = e^{2\pi i r/pq} \begin{pmatrix} a_{1-u}(t) & a_{2-u}(t) & \cdots & a_{p-u}(t) & b_{1+u}(t) & b_{2+u}(t) & \cdots & b_{q+u}(t) \end{pmatrix}$$

The action of $\mathrm{SO}(2)$ on $\mathbb{R}^{2\times n} \cong \mathbb{C}^n$ is just multiplication with unit scalars in $\mathbb{C}$. The permutation

$$\sigma = (1\ 2\ \ldots\ p)^u (p+1\ p+2\ \cdots\ p+q)^{-u} \in S_n$$

acts on $\mathbb{C}^n$ by permuting the entries (or permuting the columns in $\mathbb{R}^{2\times n}$). For $g = (e^{2\pi i r/pq}, \sigma) \in \mathrm{SO}(2) \times S_n \subseteq G$ we have $g \cdot w(t) = w(t + \frac{2\pi}{pq})$. Recall that W is the span of $w(0)$ and $w(\frac{\pi}{2})$. From this it is clear that $g \cdot W \subseteq W$. If we restrict the action of $g \in G$ to the subspace W, then it is just rotation over the angle $2\pi/pq$.

If we forget about the labeling of the vertices, then $w(t)$ is a point cloud consisting of the union of the vertices of a p-gon and the vertices of a q-gon. Now $w(t)$ is obtained from $w(0)$ by rotating the p-gon counterclockwise and the q-gon clockwise. From this it is clear that for small $t > 0$, $w(0)$ and $w(t)$ are not isometric point clouds. This proves that not all the elements of $S(W) \subset W$ lie in the same $\mathrm{O}(2) \times S_n$-orbit. By Proposition 3 we get $D(G, V) \geq pq \geq n^2/4 - 1$.

Now assume that $d > 2$. We view $\mathbb{R}^{2\times n}$ as the subspace of $\mathbb{R}^{d\times n}$ of $d \times n$ matrices whose rows $3, \ldots, d$ are all equal to 0, and let $W \subseteq \mathbb{R}^{2\times n}$ be the 2-dimensional subspace that was constructed in the $d = 2$ case. Now there is again an element $g \in \mathrm{SO}(2) \times S_n \subseteq G \subseteq \mathrm{O}(d) \times S_n$ that acts on W by rotation over the angle $2\pi/pq$. For similar reasons as before, not all elements of $S(W) \subset W$ lie in the same $\mathrm{O}(d) \times S_n$-orbit. So we again can apply Proposition 3 to conclude that $D(G, V) \geq pq \geq n^2/4 - 1$.

□

2.2 Upper bound for point clouds

Having established a lower bound on the degree required for separation of point clouds, we proceed to provide an upper bound. In the proof of Theorem 4, we introduce a polynomial construction with degree that depends linearly on the dimension d and quadratically on the number of points that separates d-dimensional point clouds, thereby establishing an upper bound on the degree sufficient for separation.

Theorem 4. *Let $n \geq 2$ and $1 \leq d \leq n$. Then*

$$\begin{aligned} D\big(H \times S_n, \mathbb{R}^{d\times n}\big) &\leq 6dn^2 - 4dn - d(d-1)(n-1), && \text{for } H \in \{\mathrm{O}(d), E(d)\} \\ D\big(H \times S_n, \mathbb{R}^{d\times n}\big) &\leq 6dn^2 - 4dn - d(d-1)(n-1) + d, && \text{for } H \in \{\mathrm{SO}(d), SE(d)\} \end{aligned}$$

Proof idea. By Lemma 2 it is sufficient to prove the claim for $O(d) \times S_n$ and $SO(d) \times S_n$. The goal is to find, for any given $X, Y \in \mathbb{R}^{d\times n}$ from different orbits, an invariant polynomial of the advertised degree that separates them. Since two clouds are $\mathrm{O}(d) \times S_n$-equivalent iff their Gram matrices $A = X^T X$, $B = Y^T Y$ are equal up to simultaneous row/column permutation, it suffices to separate A and B by an S_n-invariant polynomial Q (at the cost of doubling the degree). We choose a polynomial of the form

$$Q(C) = \sum_{\tau \in S_n} \chi(\tau \cdot C)$$

where χ is a polynomial which takes values in $\{0, 1\}$ on the S_n-orbits of A and B, with $\chi(A) = 1$ and $\chi(\tau \cdot B) = 0$ for all permutations τ. The task is thus reduced to realizing such an indicator χ by a low-degree polynomial. Clouds of different rank are separated by $\mathrm{tr}\big((X^T X)^k\big)$ for some $k \leq n$, so we may assume $\mathrm{rank}(A) = \mathrm{rank}(B) = r \leq d$. Then A can be relabeled so that its leading $r \times r$ block is nonsingular, and among matrices of rank r it is determined by its first r rows, so at most $d \cdot n$ entries must be checked. Each entry is checked by a Lagrange polynomial of degree $< n$ interpolating on the values of the corresponding row of A, which is valid thanks to a row indicator

of degree $O(n^2)$ (via power sums), which ensures the row of $\tau \cdot C$ is a rearrangement of that of A. The total degree is $O(dn^2)$. For $\mathrm{SO}(d)$, if the point clouds have rank d, we multiply each summand by an affine function of the determinant of the first d points, which detects orientation and adds d to the degree; if their rank is smaller than d, the $\mathrm{SO}(d) \times S_n$-orbit of X is equal to its $\mathrm{O}(d) \times S_n$-orbit, and the polynomial for $\mathrm{O}(d)$ separates them. □

Theorem 4 shows that it is possible to obtain separating invariants of degree quadratic in n for point clouds. It follows that neural networks that are able to express polynomials up to that degree bound can provably separate point clouds with n points. Many popular machine learning equivariant models are able to express high order polynomials using tensor representations, allowing us to show how this upper bound proves the separation capabilities of prominent such models in Section 3.

2.3 Degree bounds for Weighted graphs

Our degree bounds for point clouds automatically imply degree bounds for weighted graphs (see definition in Subsection 1.2).

Proposition 5. *For any natural n, we have that $D(O(n) \times S_n, \mathbb{R}^{n \times n}) = 2D(S_n, \mathcal{H}_n)$. In particular it follows from Theorem 1 and Theorem 4, that for $n \geq 3$,*

$$\frac{1}{2}(n^2/4 - 1) \leq D(S_n, \mathcal{H}_n) \leq \frac{1}{2}\left(5n^3 - 2n^2 - n\right)$$

Proof idea. The proof that $D(O(n) \times S_n, \mathbb{R}^{n \times n}) = 2D(S_n, \mathcal{H}_n)$ is based on the observation that two positive semi-definite weighted graphs A, B are in the same S_n orbit, if and only if there are $O(n) \times S_n$ equivalent point clouds X, Y satisfying $X^T X = A, Y^T Y = B$. The full proof is in the appendix. □

3 Quantitative approximation bounds for Invariant Models

Our upper bound in Theorem 4 shows universality and separation for $SE(d) \times S_n$ invariant polynomials of degree proportional to $d \cdot n^2$. These results can be extended to similar quantitative universality results for $SE(d) \times S_n$ invariant architectures which are based on message passing and high dimensional representations of the rotation group, since such architectures can construct invariant polynomials. We give two specific architectures for which such results can be obtained: Tensor Field Networks (TFN) and MACE.

In this section we will often specialize to the case $d = 3$, since this is the most relevant dimension in practice, and as a result many invariant models, including TFN and MACE, are defined for $d = 3$ only.

3.1 Tensor Field Networks

Tensor Field Networks (TFN) Thomas et al. (2018) pioneered the approach of using high-dimensional representations of the rotation group for $SE(3) \times S_n$ invariant models. We give a brief background on this approach.

A real representation of $SO(3)$ is a real vector space on which $SO(3)$ acts linearly. A standard example is the action of $SO(3)$ on $\mathbb{R}^3$ by matrix multiplication. More generally, for every $\ell = 0, 1, 2, \ldots$, there is a $2\ell + 1$ dimensional representation $W_\ell \cong \mathbb{R}^{2\ell+1}$ of $SO(3)$. This representation is obtained via spherical harmonics: For a fixed ℓ, the ℓ-th spherical harmonics are a collection of $2\ell + 1$ polynomials $Y_\ell^m : \mathbb{R}^3 \to \mathbb{R}, m = -\ell, \ldots, \ell$ which are (a) ℓ-homogeneous and (b) are harmonic and satisfy that the mapping $x \mapsto (Y_\ell^m(x), m = -\ell, \ldots, \ell)$ is a rotation equivariant function mapping the standard representation $\mathbb{R}^3 \cong W_1$ into a high dimensional representation W_ℓ.

To what extent does the employment of higher dimensional features increase the expressive power of the model? In Pozdnyakov & Ceriotti (2022) it was shown that methods which use geometric message passing on the distances only (corresponding to $\ell = 0$ features) are not complete. In contrast, in Dym & Maron (2021) it was shown that TFN can construct all invariant polynomials up to degree D, if the feature order and depth are proportional to D. Specifically, the following is a simple corollary of Theorem 2 in that paper:

Proposition 6. *Let $D \in \mathbb{N}$. Any $SE(3) \times S_n$ invariant polynomial $p : \mathbb{R}^{3\times n} \to \mathbb{R}$ can be expressed by a TFN model with $2D$ layers, and irreducible features of maximal order $\ell_{\max} = D$.*

In Dym & Maron (2021), this proposition was used to prove that one can always obtain a complete model if the maximal irreducible order is taken to be arbitrarily large, where this result is based on the observation that invariant polynomials can approximate all continuous invariant functions. With our result, Theorem 4, we can now for the first time obtain a bound on the order of $\ell_{\max}$ needed:

Corollary 7. *Let $n \in \mathbb{N}$ and set $D = 18n^2 - 18n + 9$. Then an invariant TFN model with $2D$ layers, and irreducible features of order $\leq D$, is $(SE(3) \times S_n, \mathbb{R}^{3\times n})$ complete.*

The derivation of Proposition 6 from the results in Dym & Maron (2021), and some more details of the TFN model used, are discussed in Appendix D.1.

3.2 MACE

MACE Batatia et al. (2022) is a very successful $SE(3) \times S_n$ invariant model. Its main innovation is the employment of tensor products of several high order features. Namely, if $A_1, \ldots, A_\nu$ are features in $W_{\ell_1}, \ldots, W_{\ell_\nu}$ respectively, then one can obtain a new feature $A_1 \otimes A_2 \otimes \ldots \otimes A_\nu$ in the tensor product representation $W_{\ell_1} \otimes W_{\ell_2} \otimes \ldots \otimes W_{\ell_\nu}$. Moreover, using the Clebsch-Gordan coefficients, this tensor product representation can be mapped equivariantly to a direct sum of W_ℓ representations. While previous works like TFN employed such tensor products between pairs of tensors ($\nu = 2$), MACE was the first model to use such decompositions for $\nu > 2$. Our next proposition gives an explanation for the effectiveness of this procedure, by showing that unlike TFN, for which the construction of Dym & Maron (2021) uses depth $2D$ to express invariant polynomials of degree D, MACE can do this with depth=1, as long as the maximal number of tensor products (known as the body order) $\nu_{\max}$, and the maximal representation order $\ell_{\max}$, satisfy $\nu_{\max} = \min\{D, n-1\}$ and $\ell_{\max} = \lfloor D/2 \rfloor$, respectively.

We modify the step of MACE that forms tensor products of the aggregated features so that it follows the atomic cluster expansion (ACE) Drautz (2019): each factor of the tensor product has its own channel mixing, as each factor of an ACE basis function has its own radial index. This modification, and the other differences from the original model, are detailed in Appendix D.2.

Proposition 8. *MACE with $\nu_{\max} = \min\{D, n-1\}$ and $\ell_{\max} = \lfloor D/2 \rfloor$ can express any $SE(3)\times S_n$ invariant polynomial of degree $\leq D$.*

As a corollary of this proposition and Theorem 4, we obtain the following completeness result for MACE

Corollary 9. *Let $n \in \mathbb{N}$ and set $D = 18n^2 - 18n + 9$. Then a MACE model with depth 1, $\nu_{\max} = n - 1$ and $\ell_{\max} = \lfloor D/2 \rfloor$ is $(SE(3) \times S_n, \mathbb{R}^{3\times n})$ complete.*

The details of the MACE model used for Proposition 8, and the proof of the proposition, are given in Appendix D.2. In Appendix D.3 we show that when message passing is on all points, more iterations can be used instead of a high body order, for example $\nu_{\max} = 2$ with $\log_2(n+1) - 1$ iterations (Corollary 30).

3.3 Improved MACE Completeness for locally regular point clouds

The completeness results of Section 3 require invariants of degree $18n^2 - 18n + 9$, a quantity that grows with the number of points in the cloud. This degree is likely too high for most practical MACE implementations. Moreover, MACE typically applies message passing only within a fixed cutoff radius Batatia et al. (2022). Accordingly, the relevant scale is not n but the maximal number K of points inside a single cutoff ball. In this subsection we isolate a set of properties of a point cloud which make this precise: on the class of clouds satisfying them, a *local* model of MACE type is complete using irreducible features and invariants of degree at most $6dK^2$, a bound independent of n and of the depth of the model. Restricted to such a class, this strengthens the MACE completeness corollary of Section 3. Conversely, on a similar class, we prove a lower bound showing that MACE cannot separate without irreducibles and correlation orders whose total degree is $\sim K^2$.

Below we will fix a cutoff radius $\rho > 0$ and an integer K. In order to focus on applications to MACE, we will set the dimension to $d = 3$. Proofs are given in Appendix E and they are written for general $d \geq 2$, except where the constants are specific to $d = 3$.

Environments. Let $X = (x_1, \dots, x_n) \in \mathbb{R}^{d\times n}$ and write $[n] := \{1, \dots, n\}$. For $i \in [n]$ let

$$\mathcal{B}_1^X(i) := \{\, j \in [n] : \ \|x_j - x_i\| \leq \rho \,\}$$

be the set of indices in the cutoff ball at x_i. Two indices $i, j \in [n]$ are *ρ-adjacent* if $0 < \|x_i - x_j\| \leq \rho$ and the columns indexed by $\mathcal{B}_1^X(i) \cap \mathcal{B}_1^X(j)$ contain $d + 1$ affinely independent points.

Definition 10 (locally regular point clouds)**.** *A matrix $X = (x_1, \dots, x_n) \in \mathbb{R}^{d\times n}$ with $n \geq 2$ is* (ρ, K)-regular *if*

(D) *the columns $x_1, \dots, x_n$ are pairwise distinct;*

(K) $\big|\{\, j \in [n] : \|x_j - y\| \leq \rho \,\}\big| \leq K$ *for every $y \in \mathbb{R}^d$;*

(O) *the ρ-adjacency graph on $[n]$ is connected.*

We say that $X \in \mathbb{R}^{d\times n}$ is (ρ, K)- ***strongly regular****, if X is (ρ, K) regular, and additionally, X does not contain two different ordered subsets of $d + 1$ points which are congruent. Here two ordered subsets $(p_0, \dots, p_d)$ and $(p'_0, \dots, p'_d)$ are congruent if there exist $g \in O(d)$ and $b \in \mathbb{R}^d$ such that $p'_s = gp_s + b$ for all $s = 0, \dots, d$. We denote the collection of (ρ, K) regular and (ρ, K) strongly regular point clouds by $\mathcal{R}_{\rho,K}$ and $\mathcal{R}^S_{\rho,K}$, respectively.*

Hypothesis (K) bounds the local multiplicity from above and is what keeps environments small. Hypothesis (O) bounds it from below and is what makes the environments rigid: connectivity lets information propagate through the cloud, while the overlap requirement built into ρ-adjacency forces the cutoff balls of adjacent points to meet in a set of full affine dimension.

Our main result of this section states that, for (ρ, K) strongly regular point clouds, a single iteration of local MACE with irreducible order $\sim K^2$ is already complete, and that further iterations preserve completeness:

Theorem 11. *Let $K \leq n$ be natural numbers and $\rho > 0$, and let $G = SE(3) \times S_n$. Set*

$$D(j) = 18j^2 - 18j + 9, \qquad D = D(K).$$

Then local MACE with cutoff ρ, $T \geq 1$ iterations, $\nu_{\max} = K - 1$ and $\ell_{\max} = \lfloor D/2 \rfloor$, is $(G, \mathcal{R}^S_{\rho,K})$ complete.

Proof Idea. The main idea is that, after a single MACE iteration, one obtains a set of node features $\{E_i(X),\ i = 1 \dots, n\}$. These features are invariants, so reconstructing an environment from one of them is a separation problem for point clouds of K points, and this is where Theorem 4 enters: it bounds the degree at which invariants separate such clouds by $D = D(K)$, and Corollary 9, which is that theorem combined with the expressivity of one MACE layer, turns the bound into the degree assumption made here. Accordingly $\mathcal{B}_1^X(i)$, the local environment of each node i, can be reconstructed from $E_i(X)$, up to a rotation and a translation. Strong regularity then implies that, whenever the two environments share $d + 1$ affinely independent points (up to congruence), these points can be identified with each other; by the overlap requirement in the definition of ρ-adjacency this is the case for every adjacent pair i, j, so the environments of adjacent nodes can be glued. Since the ρ-adjacency graph is connected, and the intersections contain $d + 1$ affinely independent points, the whole point cloud can be reconstructed uniquely. A full proof is given in the appendix. □

We accompany the above by the following lower bound. In this lower bound, the hyperparameters $\ell_{\max}, \nu_{\max}$ correspond to the maximal irreducible order and maximal correlations used in the MACE iterations, and ν_{read} denotes the correlations used in the final readout stage.

Proposition 12 (proved in Appendix E.2)**.** *Let $K \geq 7$, and let a local MACE model with cutoff ρ, message-passing depth T and hyperparameters $\ell_{\max}, \nu_{\max}, \nu_{\mathrm{read}}$ be $(SE(3) \times S_n, \mathcal{R}_{\rho,K})$ complete on clouds of K points. Then*

$$\nu_{\mathrm{read}} \cdot \ell_{\max} \cdot \sum_{t=1}^{T} \nu_{\max}^t \ \geq\ 2\Big\lfloor \frac{K-1}{2} \Big\rfloor,$$

and if the readout is that of equation 6–equation 7, a function of the invariant channels at each point followed by a sum, then already $\ell_{\max} \sum_{t=1}^{T} \nu_{\max}^{t} \geq 2\lfloor (K-1)/2 \rfloor$.

Proof idea. The counterexample is a local version of the counter-rotating polygons of Theorem 1: for $p, q \geq 3$ we form a family of clouds of $K = p + q + 1$ points in $\mathbb{R}^3$, each consisting of a point at the origin and two rings of p and q points with free phases α, β around a common axis (see Figure 2). The cutoff radius is such that every distance the model can read is constant along the family, the cross-ring distances that do vary all exceeding the cutoff, so only angular resolution is at stake, while the sole invariant $\alpha - \beta$ is defined modulo $2\pi/\mathrm{lcm}(p, q)$. Each spherical harmonic of degree ℓ in a message contributes at most ℓ to the frequency in (α, β) of the features, and products of $\nu_{\max}$ messages add frequencies, which gives the budget on the left. Shifting β by $2\pi/q$ leaves the cloud and each point of the first ring fixed, and a rotation about the axis shifts α and β together, so the output at such a point is constant unless one of its Fourier modes (m_1, m_2) has $|m_1| + |m_2| \geq 2q$; symmetrically, $2p$ for the second ring, and more for the center. A model whose point outputs are all constant separates nothing, and choosing p and q as equal as possible gives $2\min(p, q) = 2\lfloor (K-1)/2 \rfloor$. □

The significance of this proposition is that MACE with the modestly sized hyperparameters usually selected in practice is not complete even under the regularity assumptions. To illustrate this: for $\ell_{\max} = \nu_{\max} = 3$ and the readout of equation 6, the family is not separated once $K \geq 11$ for $T = 1$, $K \geq 39$ for $T = 2$ and $K \geq 119$ for $T = 3$. Figure 2 shows a 12-point configuration whose relative polygon rotations can be distinguished by MACE only if

$$\nu_{\mathrm{read}}\, \ell_{\max} \sum_{i=1}^{T} \nu_{\max}^{i} \;\geq\; 2\left\lfloor \frac{11}{2} \right\rfloor = 10.$$

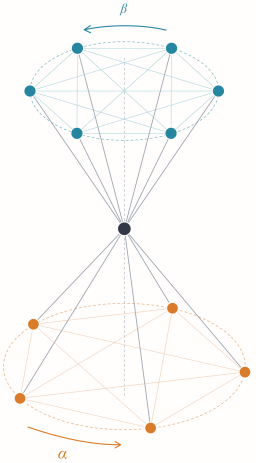


Figure 2: A family of regular point clouds which cannot be separated by mace without high degree/correlation order. See Proposition 12 for details.

4 Conclusion, Limitations and Future Work

In this paper we showed that the degree of separation for point clouds is lower bounded by $\sim n^2$ and upper bounded by $\sim dn^2$. We showed this implies a $\sim n^2$ lower bound and $\sim n^3$ upper bound for weighted graphs, and used the upper bound for point clouds to derive completeness results for TFN and MACE with irreducible order which grows quadratically in n, or under additional assumptions, quadratically in the maximal neighborhood size. We also provide a lower bound for the MACE architecture. To the best of our knowledge these are the first quantitative results of this kind.

In the context of polynomial separation, the remaining gap is that the upper bound exceeds the lower bound by a factor of order d. This is essentially equivalent to closing the gap between the quadratic lower bound and cubic upper bound attained for weighted graphs. In the context of invariant networks there is a more substantial gap: Our results used the fact that TFN and MACE can construct invariant polynomials up to a degree determined by the irreducibles used. However, these models can also express non-polynomial functions. Thus, we do not have tight lower bounds for these models. An interesting open question is to give tight upper and lower bounds on the degree of separation of these models. In particular, are MACE and TFN complete with an irreducible degree that does not depend on n? For regular point clouds Proposition 12 answers it in part: with bounded depth and correlation order the irreducible degree must grow at least linearly in the neighborhood size. Whether linear growth suffices, the question for the class of all point clouds, and the case of MACE architectures with unrestricted neighborhoods, remain open.

AI use statement

We used generative AI tools to edit and improve the writing, check manuscript formatting, and assist in verifying the correctness of our results. We did not use generative AI tools for research

ideation. All AI-assisted suggestions and checks were reviewed by the authors. The authors take full responsibility for the final content of this work, including all claims, proofs, and conclusions.

REFERENCES

Tal Amir, Steven Gortler, Ilai Avni, Ravina Ravina, and Nadav Dym. Neural injective functions for multisets, measures and graphs via a finite witness theorem. *Advances in Neural Information Processing Systems*, 36:42516–42551, 2023.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022. doi: 10.52202/068431-0830. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf.

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.

Zhengdao Chen, Soledad Villar, Lei Chen, and Joan Bruna. On the equivalence between graph isomorphism testing and function approximation with gnns. *Advances in neural information processing systems*, 32, 2019.

Feng Dai and Yuan Xu. *Approximation Theory and Harmonic Analysis on Spheres and Balls*. 01 2013. ISBN 978-1-4614-6659-8. doi: 10.1007/978-1-4614-6660-4.

Harm Derksen and Gregor Kemper. *Computational invariant theory*. Springer, 2015.

Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B*, 99:014104, Jan 2019. doi: 10.1103/PhysRevB.99.014104. URL https://link.aps.org/doi/10.1103/PhysRevB.99.014104.

Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Nanning Zheng, Bin Shao, and Tie-Yan Liu. SE(3) equivariant graph neural networks with complete local frames. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 5583–5608. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/du22e.html.

Nadav Dym and Steven J Gortler. Low-dimensional invariant embeddings for universal geometric learning. *Foundations of Computational Mathematics*, 25(2):375–415, 2025.

Nadav Dym and Haggai Maron. On the universality of rotation equivariant point cloud networks. In *ICLR*, 2021. URL https://openreview.net/forum?id=6NFBvWlRXaG.

Ben Finkelshtein, Chaim Baskin, Haggai Maron, and Nadav Dym. A simple and universal rotation equivariant point-cloud network. In Alexander Cloninger, Timothy Doster, Tegan Emerson, Manohar Kaul, Ira Ktena, Henry Kvinge, Nina Miolane, Bastian Rieck, Sarah Tymochko, and Guy Wolf (eds.), *Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022*, volume 196 of *Proceedings of Machine Learning Research*, pp. 107–115. PMLR, 25 Feb–22 Jul 2022. URL https://proceedings.mlr.press/v196/finkelshtein22a.html.

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

Manfred Göbel. Computing bases for rings of permutation-invariant polynomials. *Journal of Symbolic Computation*, 19(4):285–291, 1995.

Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *arXiv preprint arXiv:1912.12033*, 2019.

Louis Guttman. Enlargement methods for computing the inverse matrix. *Annals of Mathematical Statistics*, 17:336–343, 1946.

Jiaqi Han, Yu Rong, Tingyang Xu, and Wenbing Huang. Geometrically equivariant graph neural networks: A survey. *arXiv preprint arXiv:2202.07230*, 2022.

David Hilbert. Ueber die theorie der algebraischen formen. *Mathematische Annalen*, 36(4):473–534, 1890.

Snir Hordan, Tal Amir, and Nadav Dym. Weisfeiler leman for euclidean equivariant machine learning. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 18749–18784. PMLR / OpenReview.net, 2024a. URL https://proceedings.mlr.press/v235/hordan24a.html.

Snir Hordan, Tal Amir, Steven J. Gortler, and Nadav Dym. Complete neural networks for complete euclidean graphs. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (eds.), *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada*, pp. 12482–12490. AAAI Press, 2024b. doi: 10.1609/AAAI.V38I11.29141. URL https://doi.org/10.1609/aaai.v38i11.29141.

Gregor Kemper. Separating invariants. *Journal of Symbolic Computation*, 44(9):1212–1222, 2009. doi: 10.1016/j.jsc.2008.02.012.

Zian Li, Xiyuan Wang, Yinan Huang, and Muhan Zhang. Is distance matrix enough for geometric deep learning? In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/75f1a165c7561e028c41d42fa6286a76-Abstract-Conference.html.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=KwmPfARgOTD.

Ian G. Macdonald. *Symmetric Functions and Hall Polynomials*. Oxford University Press, 2 edition, 1995.

Christopher Morris, Yaron Lipman, Haggai Maron, Bastian Rieck, Nils M. Kriege, Martin Grohe, Matthias Fey, and Karsten Borgwardt. Weisfeiler and leman go machine learning: The story so far. *Journal of Machine Learning Research*, 24(333):1–59, 2023. URL http://jmlr.org/papers/v24/22-0240.html.

Jigyasa Nigam, Sergey N. Pozdnyakov, Kevin K. Huguenin-Dumittan, and Michele Ceriotti. Completeness of atomic structure representations. *APL Machine Learning*, 2(1), February 2024. ISSN 2770-9019. doi: 10.1063/5.0160740. URL http://dx.doi.org/10.1063/5.0160740.

Emmy Noether. Der endlichkeitssatz der invarianten endlicher gruppen. *Mathematische Annalen*, 77:89–92, 1916.

Arkadij L. Onishchik and Ernest B. Vinberg. *Lie Groups and Algebraic Groups*. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1990. ISBN 978-3-540-50614-0. doi: 10.1007/978-3-642-74334-4.

Sergey N Pozdnyakov and Michele Ceriotti. Incompleteness of graph neural networks for points clouds in three dimensions. *Machine Learning: Science and Technology*, 3(4):045020, nov 2022. doi: 10.1088/2632-2153/aca1f8. URL https://doi.org/10.1088/2632-2153/aca1f8.

Sergey N. Pozdnyakov, Michael J. Willatt, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Phys. Rev. Lett.*, 125: 166001, Oct 2020. doi: 10.1103/PhysRevLett.125.166001. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.166001.

Valentino Delle Rose, Alexander Kozachinskiy, Cristobal Rojas, Mircea Petrache, and Pablo Barceló. Three iterations of (d - 1)-wl test distinguish non isometric clouds of d-dimensional points. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/1e6cf8f77bd8e907f53babcd7664c710-Abstract-Conference.html.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8459–8468. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/sanchez-gonzalez20a.html.

Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8(1): 13890, 2017.

Nimrod Segol and Yaron Lipman. On universal equivariant set networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HkxTwkrKDB.

Nicolas M Thiéry. Algebraic invariants of graphs; a study based on computer exploration. *ACM SIGSAM Bulletin*, 34(3):9–20, 2000.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds, 2018. URL https://arxiv.org/abs/1802.08219.

Hermann Weyl. *The Classical Groups: Their Invariants and Representations*. Princeton University Press, 2 edition, 1946.

Egor Yasinsky. Automorphisms of real del Pezzo surfaces and the real plane Cremona group. *Annales de l'Institut Fourier*, 72(2):831–899, 2022. doi: 10.5802/aif.3460.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander Smola. Deep sets. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf.

A Further Related Work

Completeness and universality in Geometric Deep Learning Understanding the expressive power of machine learning models for point clouds allows researchers to have a benchmark for model comparison. Dym & Maron (2021) characterized the polynomial expressivity of models that utilize high-dimensional tensors, often termed equivariant geometric machine learning models, and proved such models are universal at the limit of tensor order, e.g. Thomas et al. (2018). Subsequent work used tools developed there to prove universality of novel architectures Du et al. (2022); Gasteiger et al. (2021); Finkelshtein et al. (2022).

Universality was also proven for sufficiently high-order geometric graph neural networks (GNNs), which process pairwise invariants and computes high-order correlations among them Hordan et al. (2024a;b); Rose et al. (2023); Li et al. (2023). Yet equivariant models and GNNs are structured differently, with nonlinearities embedded throughout the components of GNNs, in contrast with equivariant models which are composed of linear maps except for the radial functions and readout aggregation, so those results do not carry over to equivariant GNNs. High order GNNs incur high computational cost and poor scalability compared with equivariant models Batatia et al. (2022), leaving quantitative completeness in equivariant models an open question, up to this manuscript.

B Missing proofs

Lemma 2. *Let $V = \mathbb{R}^{d\times n}$ and V_{cent} denote the subspace of point clouds $X \in \mathbb{R}^{d\times n}$ with $X1_n = 0_d$. Then*

$$D(E(d)\times S_n, V) = D(O(d)\times S_n, V_{\text{cent}}) \le D(O(d)\times S_n, V) \tag{1}$$
$$D(SE(d)\times S_n, V) = D(SO(d)\times S_n, V_{\text{cent}}) \le D(SO(d)\times S_n, V) \tag{2}$$

Proof. Throughout, let $H_0 \in \{O(d), SO(d)\}$, and let H denote the corresponding group with translations, i.e. $H = E(d)$ if $H_0 = O(d)$ and $H = SE(d)$ if $H_0 = SO(d)$. We prove

$$D(H\times S_n, V) = D(H_0\times S_n, V_{\text{cent}}) \le D(H_0\times S_n, V),$$

which gives equation 1 and equation 2 simultaneously.

Matrix form of the action. For $\sigma \in S_n$ let P_σ be the permutation matrix with $P_\sigma e_j = e_{\sigma(j)}$. The j-th column of XP_σ^T is $x_{\sigma^{-1}(j)}$, so the action defined in Section 1.2 reads

$$(t, R, \sigma)\cdot X \;=\; R\,(X - t1_n^T)\,P_\sigma^T,$$

and $H_0\times S_n$ is identified with the subgroup $\{(0, R, \sigma)\}$ of $H\times S_n$. Note that $P_\sigma 1_n = 1_n$ and hence also $1_n^T P_\sigma^T = 1_n^T$.

The centering map. Let $C := I_n - \frac{1}{n}1_n 1_n^T$ and define the linear map $c : V \to V$, $c(X) := XC = X - \frac{1}{n}X1_n 1_n^T$. We record four properties.

(i) $c(V) \subseteq V_{\text{cent}}$ and $c(X) = X$ for $X \in V_{\text{cent}}$.

(ii) X and $c(X)$ lie in the same $H\times S_n$-orbit.

(iii) c is equivariant: $c\big((t, R, \sigma)\cdot X\big) = (0, R, \sigma)\cdot c(X)$ for all $(t, R, \sigma) \in H\times S_n$.

(iv) V_{cent} is $H_0\times S_n$-invariant: if $X1_n = 0$ then $RXP_\sigma^T 1_n = RX1_n = 0$.

Claim: $X, Y \in V$ lie in the same $H\times S_n$-orbit if and only if $c(X), c(Y)$ lie in the same $H_0\times S_n$-orbit.

If $Y = (t, R, \sigma)\cdot X$, then $c(Y) = (0, R, \sigma)\cdot c(X)$ by (iii). Conversely, if $c(Y) = (0, R, \sigma)\cdot c(X)$, then $c(X)$ and $c(Y)$ lie in the same $H\times S_n$-orbit, since $H_0\times S_n \subseteq H\times S_n$. By (ii), X lies in the orbit of $c(X)$ and Y lies in the orbit of $c(Y)$, so X and Y lie in the same $H\times S_n$-orbit.

We will repeatedly use that composing a polynomial with a linear map, or restricting it to a linear subspace, yields a polynomial of no larger degree.

We now show that $D(H \times S_n, V) = D(H_0 \times S_n, V_{\text{cent}})$ by proving inequality in both directions *Step 1:* $D(H \times S_n, V) \leq D(H_0 \times S_n, V_{\text{cent}})$. Set $D := D(H_0 \times S_n, V_{\text{cent}})$ and let $X, Y \in V$ lie in different $H \times S_n$-orbits. By the claim, $c(X), c(Y) \in V_{\text{cent}}$ lie in different $H_0 \times S_n$-orbits, so there is $p \in \mathbb{R}[V_{\text{cent}}]_{\leq D}^{H_0 \times S_n}$ with $p(c(X)) \neq p(c(Y))$. Set $q := p \circ c \in \mathbb{R}[V]_{\leq D}$. By (iii), for all $(t, R, \sigma) \in H \times S_n$,

$$q\big((t, R, \sigma) \cdot X\big) = p\big((0, R, \sigma) \cdot c(X)\big) = p(c(X)) = q(X),$$

so q is $H \times S_n$-invariant and separates X from Y. Hence $\mathbb{R}[V]_{\leq D}^{H \times S_n}$ is $(H \times S_n, V)$ complete.

Step 2: $D(H_0 \times S_n, V_{\text{cent}}) \leq D(H \times S_n, V)$. Set $D' := D(H \times S_n, V)$ and let $X, Y \in V_{\text{cent}}$ lie in different $H_0 \times S_n$-orbits. By (i), $c(X) = X$ and $c(Y) = Y$, so by the claim X, Y lie in different $H \times S_n$-orbits of V. Hence there is $q \in \mathbb{R}[V]_{\leq D'}^{H \times S_n}$ with $q(X) \neq q(Y)$. Its restriction $q|_{V_{\text{cent}}}$ is a polynomial on V_{cent} of degree $\leq D'$. It is $H_0 \times S_n$-invariant, because V_{cent} is $H_0 \times S_n$-invariant by (iv) and $H_0 \times S_n \subseteq H \times S_n$. Also $q|_{V_{\text{cent}}}(X) \neq q|_{V_{\text{cent}}}(Y)$. Hence $\mathbb{R}[V_{\text{cent}}]_{\leq D'}^{H_0 \times S_n}$ is $(H_0 \times S_n, V_{\text{cent}})$ complete.

Finally, the inequality $D(H_0 \times S_n, V_{\text{cent}}) \leq D(H_0 \times S_n, V)$ follows directly from the fact that V_{cent} is an invariant subspace of V. □

Proposition 5. *For any natural n, we have that $D(O(n) \times S_n, \mathbb{R}^{n \times n}) = 2D(S_n, \mathcal{H}_n)$. In particular it follows from Theorem 1 and Theorem 4, that for $n \geq 3$,*

$$\frac{1}{2}(n^2/4 - 1) \leq D(S_n, \mathcal{H}_n) \leq \frac{1}{2}\left(5n^3 - 2n^2 - n\right)$$

Proof. We first prove $D(O(n) \times S_n, \mathbb{R}^{n \times n}) \leq 2D(S_n, \mathcal{H}_n)$. Let $X, Y \in \mathbb{R}^{n \times n}$ from different $O(n) \times S_n$ orbits. Then the Gram matrices $X^T X$ and $Y^T Y$ are not in the same S_n orbit, since otherwise, we would have that $X^T X = P^T Y^T Y P$ for some permutation matrix P, implying that X and YP have the same Gram matrix, and thus there exists an orthogonal matrix $R \in O(n)$ such that $X = RYP$, contradicting the assumption that X, Y are not from the same orbit. It follows that there exists an S_n invariant polynomial p of degree $\leq D(S_n, \mathcal{H}_n)$ such that $p(X^T X) \neq p(Y^T Y)$. The mapping $X \mapsto p(X^T X)$ is a $O(n) \times S_n$ invariant polynomial of degree $2\deg(p)$ and thus we have $D(O(n) \times S_n, \mathbb{R}^{n \times n}) \leq 2D(S_n, \mathcal{H}_n)$.

We now prove the converse direction $D(O(n) \times S_n, \mathbb{R}^{n \times n}) \geq 2D(S_n, \mathcal{H}_n)$. Let $A, B \in \mathcal{H}_n$ which are not from the same S_n orbit. There exists some $t > 0$ such that $A + tI_n$ and $B + tI_n$ are both positive definite. It follows that for some $X, Y \in \mathbb{R}^{n \times n}$,

$$A + tI_n = X^T X, B + tI_n = Y^T Y.$$

X, Y are not in the same $O(n) \times S_n$ orbit (for that would imply that A, B are in the same S_n orbit). Therefore there exists an $O(n) \times S_n$ invariant polynomial p of degree $\leq D(O(n) \times S_n, \mathbb{R}^{n \times n})$, such that $p(X) \neq p(Y)$. By the First Fundamental Theorem of Invariant Theory for $O(n)$, any $O(n)$ invariant polynomial is a polynomial in the inner products of columns of X, or in other words, $p(X) = \phi(X^T X)$ for an appropriate polynomial ϕ. The polynomial ϕ is S_n invariant on all positive definite matrices, and since these form an open subset of $\mathcal{H}_n$, it follows that ϕ is S_n invariant on all of $\mathcal{H}_n$. Finally we have that

$$\phi(A + tI_n) = \phi(X^T X) \neq \phi(Y^T Y) = \phi(B + tI_n)$$

and so the polynomial $A \mapsto \phi(A + tI_n)$ separates A from B and has degree

$$\deg(\phi) = \frac{\deg(p)}{2} \leq \frac{D(O(n) \times S_n, \mathbb{R}^{n \times n})}{2}.$$

□

C Upper Bound

C.1 Proofs of elementary statements

Proposition 13. *For $X, Y \in \mathbb{R}^{d \times n}$ we have $\Phi(X) = \Phi(Y)$ if and only if $Y = RX$ for some $R \in \mathrm{O}(d)$.*

Proof. If $Y = RX$ with $R \in \mathrm{O}(d)$, then $Y^TY = X^TR^TRX = X^TX$. Conversely, suppose $X^TX = Y^TY$, and let $x_1, \dots, x_n$ and $y_1, \dots, y_n$ denote the columns of X and Y, so that $\langle x_i, x_j\rangle = \langle y_i, y_j\rangle$ for all i, j. The assignment $\sum_i c_i x_i \mapsto \sum_i c_i y_i$ is a well-defined linear isometry T from $\mathrm{span}(x_1, \dots, x_n)$ onto $\mathrm{span}(y_1, \dots, y_n)$: for every $c \in \mathbb{R}^n$,

$$\Big\| \sum_i c_i y_i \Big\|^2 = c^T Y^T Y c = c^T X^T X c = \Big\| \sum_i c_i x_i \Big\|^2,$$

which gives both well-definedness (take $\sum_i c_i x_i = 0$) and preservation of norms. Since $\ker Z = \ker(Z^TZ)$ for any real matrix Z, both spans have dimension $\mathrm{rank}(X^TX) = \mathrm{rank}(Y^TY)$, hence their orthogonal complements in $\mathbb{R}^d$ have equal dimension as well; extending T by any linear isometry between these complements produces $R \in \mathrm{O}(d)$ with $RX = Y$. □

Proposition 14 (Equivariance of the Gram map)**.** *For all $X \in \mathbb{R}^{d\times n}$, $R \in \mathrm{O}(d)$ and $\tau \in S_n$ we have*

$$\Phi\big((R, \tau) \cdot X\big) \;=\; \tau \cdot \Phi(X).$$

In particular, Φ is $\mathrm{O}(d)$-invariant and intertwines the S_n-actions on $\mathbb{R}^{d\times n}$ and on $\mathcal{H}_n$.

Proof. $(RXP_\tau^T)^T(RXP_\tau^T) = P_\tau X^T R^T R X P_\tau^T = P_\tau(X^TX)P_\tau^T$. □

Lemma 15 (The first r rows determine a rank-r matrix)**.** *Let $C \in \mathcal{H}_n$ be a rank-r symmetric matrix whose leading $r \times r$ block C_{11} is nonsingular. Then the first r rows determine C; that is, writing*

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{12}^T & C_{22} \end{pmatrix}, \qquad C_{22} = C_{12}^T C_{11}^{-1} C_{12}.$$

Proof. Setting $S = C_{22} - C_{12}^T C_{11}^{-1} C_{12}$, block multiplication verifies the congruence

$$C = \begin{pmatrix} I & 0 \\ C_{12}^T C_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} C_{11} & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} I & C_{11}^{-1} C_{12} \\ 0 & I \end{pmatrix},$$

so that $\mathrm{rank}(C) = \mathrm{rank}(C_{11}) + \mathrm{rank}(S)$ (rank additivity of the Schur complement, Guttman (1946)). As both sides of the first equality are r, we get $\mathrm{rank}(S) = 0$, i.e. $S = 0$. Note that positive semidefiniteness is not needed. □

The point of Lemma 15 is that it localizes disagreement between non-equivalent matrices to the first r rows, whose positions number

$$|\mathcal{P}_r| = \sum_{i=1}^{r} (n - i + 1) = rn - \binom{r}{2} = N_r \le N_d, \qquad \mathcal{P}_r := \{(i, j) : 1 \le i \le r,\ i \le j \le n\}.$$

Lemma 16 (Witness lemma)**.** *Let $A, B \in \mathcal{H}_n$ be rank-r matrices in different S_n-orbits, with the leading $r \times r$ block of A nonsingular. Then for every $\sigma \in S_n$ there exists $(i, j) \in \mathcal{P}_r$ such that $(\sigma \cdot B)_{ij} \neq A_{ij}$.*

Proof. Suppose otherwise, so that $\sigma \cdot B$ agrees with A on the first r rows. In particular its leading $r \times r$ block equals that of A and is therefore nonsingular, while $\mathrm{rank}(\sigma \cdot B) = \mathrm{rank}(B) = r$. Lemma 15 thus applies to both A and $\sigma \cdot B$, and since they share the same first r rows they share the same completion: $\sigma \cdot B = A$, contradicting that A and B lie in different orbits. □

Lemma 17 (Row power sums detect rows up to reordering)**.** *For $u \in \mathbb{R}^n$ define*

$$\mathrm{pw}(u) \;=\; \Big(\sum_{l=1}^{n} u_l,\ \sum_{l=1}^{n} u_l^2,\ \dots,\ \sum_{l=1}^{n} u_l^n \Big) \in \mathbb{R}^n,$$

and for $C \in \mathcal{H}_n$ let $\mathrm{pw}_i(C) := \mathrm{pw}(C_{i1}, \dots, C_{in})$, a polynomial map whose components have degree $\le n$ in the entries of C. Then:

(a) $\mathrm{pw}(u) = \mathrm{pw}(v)$ *if and only if u and v agree up to a permutation of their entries;*

(b) $\mathrm{pw}_i(\tau \cdot C) = \mathrm{pw}_{\tau^{-1}(i)}(C)$ *for all* $\tau \in S_n$ *and* $C \in \mathcal{H}_n$.

Proof. (a) Power sums are symmetric in the entries, so agreement up to permutation implies $\mathrm{pw}(u) = \mathrm{pw}(v)$. Conversely, in characteristic zero Newton's identities (Macdonald, 1995, Ch. I, §2) determine the elementary symmetric polynomials $e_1, \ldots, e_n$ of $u_1, \ldots, u_n$ from the power sums $p_1, \ldots, p_n$; hence $\mathrm{pw}(u)$ determines the polynomial $\prod_{l=1}^{n}(t - u_l) = t^n - e_1 t^{n-1} + \cdots + (-1)^n e_n$, whose multiset of roots is $\{u_1, \ldots, u_n\}$. (b) The i-th row of $\tau \cdot C$ is $\left(C_{\tau^{-1}(i),\tau^{-1}(j)}\right)_{j=1}^{n}$; as j runs over $\{1, \ldots, n\}$, so does $\tau^{-1}(j)$, so this row is a reordering of the $\tau^{-1}(i)$-th row of C, and power sums are invariant under reordering. □

We will use the fact that a rank-r Gram matrix contains a nonsingular $r \times r$ *principal* submatrix: writing $C = Z^T Z$ with $Z \in \mathbb{R}^{r \times n}$ and choosing r linearly independent columns of Z, the principal submatrix indexed by those columns is their Gram matrix, which is nonsingular. Principality matters here, because S_n acts by *simultaneous* row and column permutations, so only a principal submatrix can be moved to the leading position. As invariant polynomials are constant on orbits, we may replace A by any element of its orbit, and hence assume that the leading $r \times r$ block of A is nonsingular. No assumption on B is needed.

The *Gram map* $\Phi : \mathbb{R}^{d \times n} \to \mathcal{H}_n$, $\Phi(X) = X^T X$, records the pairwise inner products of the points of the cloud, and its image is the cone $\mathcal{G}_d = \{C \in \mathcal{H}_n : C \succeq 0,\ \mathrm{rank}(C) \le d\}$.

Corollary 18. *Two matrices* $X, Y \in \mathbb{R}^{d \times n}$ *lie in the same* $\mathrm{O}(d) \times S_n$*-orbit if and only if* $\Phi(X)$ *and* $\Phi(Y)$ *lie in the same* S_n*-orbit in* $\mathcal{H}_n$*. Consequently, if* $Q : \mathcal{H}_n \to \mathbb{R}$ *is an* S_n*-invariant polynomial with* $Q(\Phi(X)) \neq Q(\Phi(Y))$*, then* $p := Q \circ \Phi$ *is an* $\mathrm{O}(d) \times S_n$*-invariant polynomial with* $p(X) \neq p(Y)$ *and* $\deg(p) \le 2\deg(Q)$.

Proof. If $Y = (R, \tau) \cdot X$, then $\Phi(Y) = \tau \cdot \Phi(X)$ by Proposition 14. Conversely, if $\Phi(Y) = \tau \cdot \Phi(X) = \Phi(X P_\tau^T)$, then Proposition 13 provides $R \in \mathrm{O}(d)$ with $Y = R(X P_\tau^T) = (R, \tau) \cdot X$. For the second statement: p is $\mathrm{O}(d)$-invariant because Φ is, and S_n-invariant because Q is and Φ intertwines the two S_n-actions (Proposition 14); every entry of $\Phi(X)$ is a polynomial of degree 2 in the entries of X, so $\deg(p) \le 2\deg(Q)$. □

Remark 19. *By Weyl's first fundamental theorem for* $\mathrm{O}(d)$ *(Weyl, 1946, Theorem 2.9.A), this reduction is lossless:* every $\mathrm{O}(d) \times S_n$*-invariant polynomial on* $\mathbb{R}^{d \times n}$ *is a polynomial in the pairwise inner products* $(X^T X)_{ij}$*. For the upper bound below, however, only the elementary direction is used, namely Corollary 18.*

Corollary 18 reduces Theorem 4 to a statement about Gram matrices: it suffices to separate any two matrices $A, B \in \mathcal{G}_d$ lying in different S_n-orbits by an S_n-invariant polynomial of degree $\le dn(2n-1) + N_d(n-1)$ in the Gram entries. We will introduce some preliminary lemmas and claims before embarking on a full proof.

Lemma 20. *If* $\mathrm{rank}(A) \neq \mathrm{rank}(B)$ *then there exists* $k \le n$ *such that* $\mathrm{tr}(A^k) \neq \mathrm{tr}(B^k)$*. Therefore, a polynomial of degree* $\le n$ *can separate* A *and* B.

Proof of this lemma follows from the power sum polynomials being a determining set for the elementary symmetric polynomials in characteristic zero (Newton's identities, (Macdonald, 1995, Ch. I, §2)), combined with the trace of the k-th power of a symmetric matrix being the k-th power sum of its eigenvalues: if $\mathrm{tr}(A^k) = \mathrm{tr}(B^k)$ for all $k \le n$ then A and B have the same characteristic polynomial, hence the same number of nonzero eigenvalues. The power sums are of maximal degree n, and each $\mathrm{tr}(C^k)$ is S_n-invariant since conjugation by a permutation matrix is a similarity. Since rank is likewise constant on orbits, from now on we may assume $\mathrm{rank}(A) = \mathrm{rank}(B) = r \le d$. More generally, if A and B are not similar, then there exists $k \le n$ such that $\mathrm{tr}(A^k) \neq \mathrm{tr}(B^k)$, since symmetric matrices are diagonalizable, so two of them are similar if and only if they have the same characteristic polynomial.

Lemma 21 (Row indicators)**.** *Let* $A, B \in \mathcal{H}_n$ *and* $1 \le i \le n$*. There exists a polynomial* $\delta_i : \mathcal{H}_n \to \mathbb{R}$ *with* $\deg(\delta_i) \le n(2n-1)$ *such that for every* $\tau \in S_n$ *and every* $C \in \{A, B\}$*,*

$$\delta_i(\tau \cdot C) = \begin{cases} 1 & \text{if the } i\text{-th row of } \tau \cdot C \text{ agrees with the } i\text{-th row of } A \text{ up to reordering}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let $\mathcal{N} = \{\mathrm{pw}_j(A) : 1 \le j \le n\} \cup \{\mathrm{pw}_j(B) : 1 \le j \le n\} \subset \mathbb{R}^n$, a set of $m := |\mathcal{N}| \le 2n$ distinct vectors. There is a linear functional $\pi : \mathbb{R}^n \to \mathbb{R}$ that is injective on $\mathcal{N}$: for each pair $v \ne w$ in $\mathcal{N}$, the functionals with $\pi(v) = \pi(w)$ form a proper linear subspace of the dual space $(\mathbb{R}^n)^*$, and a vector space over an infinite field is not a union of finitely many proper subspaces. Let L_i be the univariate Lagrange polynomial of degree $\le m - 1 \le 2n - 1$ taking the value 1 at $\pi(\mathrm{pw}_i(A))$ and 0 at the $m - 1$ remaining points of $\pi(\mathcal{N})$, and set

$$\delta_i \;:=\; L_i \circ \pi \circ \mathrm{pw}_i.$$

Since $\pi \circ \mathrm{pw}_i$ is a polynomial of degree $\le n$ in the entries of C, we get $\deg(\delta_i) \le n(2n-1)$.

Now let $\tau \in S_n$ and $C \in \{A, B\}$. By Lemma 17(b), $\mathrm{pw}_i(\tau \cdot C) = \mathrm{pw}_{\tau^{-1}(i)}(C) \in \mathcal{N}$, so $\delta_i(\tau \cdot C) \in \{0, 1\}$; as π is injective on $\mathcal{N}$, the value is 1 exactly when $\mathrm{pw}_i(\tau \cdot C) = \mathrm{pw}_i(A)$, which by Lemma 17(a) happens exactly when the i-th row of $\tau \cdot C$ agrees with the i-th row of A up to reordering. □

We fix the conventions used below. On $\mathcal{H}_n$ the action is $(\tau \cdot C)_{ij} = C_{\tau^{-1}(i)\tau^{-1}(j)}$, and on $\mathbb{R}^{d\times n}$ it is $(R, \tau) \cdot Z = RZP_\tau^T$ (Proposition 14), where P_τ is the permutation matrix with $P_\tau e_j = e_{\tau(j)}$, so that $P_{\tau\sigma} = P_\tau P_\sigma$. We write $\tau \cdot Z := ZP_\tau^T$ and RZ for the actions of the two factors, so that $(R, \tau) \cdot Z = R(\tau \cdot Z) = \tau \cdot (RZ)$. The j-th column of $\tau \cdot Z$ is $z_{\tau^{-1}(j)}$: this is the action of the Notation section, and it is the convention under which Proposition 14 holds, since $\Phi(\tau \cdot Z)_{ij} = \langle z_{\tau^{-1}(i)}, z_{\tau^{-1}(j)} \rangle = (\tau \cdot \Phi(Z))_{ij}$. Consequently $\tau \cdot (\sigma \cdot Z) = (\tau\sigma) \cdot Z$ on $\mathbb{R}^{d\times n}$, and likewise on $\mathcal{H}_n$. Moreover $\mathrm{rank}(\Phi(Z)) = \mathrm{rank}(Z)$, because $\ker Z = \ker(Z^T Z)$. Finally, for $d \le n$ and $Z \in \mathbb{R}^{d\times n}$ we let $Z_{[d]} \in \mathbb{R}^{d\times d}$ denote the matrix formed by the first d columns of Z. Thus $(\tau \cdot Z)_{[d]}$ has columns $z_{\tau^{-1}(1)}, \ldots, z_{\tau^{-1}(d)}$, and $(\tau \cdot Z)_{[d]}^T (\tau \cdot Z)_{[d]}$ is the leading $d \times d$ block of $\Phi(\tau \cdot Z) = \tau \cdot \Phi(Z)$.

Lemma 22 (Rank-deficient clouds)**.** *Let $X \in \mathbb{R}^{d\times n}$ with $\mathrm{rank}(X) < d$. Then the $\mathrm{SO}(d) \times S_n$-orbit of X equals its $\mathrm{O}(d) \times S_n$-orbit. In particular, if $Y \in \mathbb{R}^{d\times n}$ lies in the $\mathrm{O}(d) \times S_n$-orbit of X, then Y lies in the $\mathrm{SO}(d) \times S_n$-orbit of X.*

Proof. The columns of X span a subspace $U \subseteq \mathbb{R}^d$ of dimension $\mathrm{rank}(X) < d$, so there is a unit vector $u \in \mathbb{R}^d$ orthogonal to U. Let $M := I_d - 2uu^T$ be the reflection in the hyperplane $u^\perp \supseteq U$. Then $M^T M = I_d$, so $M \in \mathrm{O}(d)$; $Mu = -u$ and $Mv = v$ for $v \perp u$, so $\det(M) = -1$; and $Mx = x$ for every $x \in U$, so $MX = X$. Now let $(R, \tau) \in \mathrm{O}(d) \times S_n$. If $\det(R) = 1$ there is nothing to show. If $\det(R) = -1$, then

$$(R, \tau) \cdot X \;=\; RXP_\tau^T \;=\; RMXP_\tau^T \;=\; (RM, \tau) \cdot X, \qquad \det(RM) = 1.$$

Hence every element of the $\mathrm{O}(d) \times S_n$-orbit of X lies in its $\mathrm{SO}(d) \times S_n$-orbit; the reverse inclusion is trivial. □

Lemma 23 (Stabiliser sign)**.** *Let $d \le n$, let $X \in \mathbb{R}^{d\times n}$ have rank d, and set $A := \Phi(X)$. Let $\tau \in S_n$ satisfy $\tau \cdot A = A$.*

- *(a) There is a unique $R_\tau \in \mathrm{O}(d)$ with $\tau \cdot X = R_\tau X$, and $\det(\tau \cdot X)_{[d]} = \det(R_\tau) \det X_{[d]}$.*
- *(b) For every $M \in \mathrm{O}(d)$ we have $\Phi(MX) = A$ and $\det(\tau \cdot MX)_{[d]} = \det(M) \det(R_\tau) \det X_{[d]}$.*
- *(c) If $\det(R_\tau) = -1$, then MX lies in the $\mathrm{SO}(d) \times S_n$-orbit of X for every $M \in \mathrm{O}(d)$ with $\det(M) = -1$.*

Proof. (a) By Proposition 14, $\Phi(\tau \cdot X) = \tau \cdot \Phi(X) = \tau \cdot A = A = \Phi(X)$, so Proposition 13 provides $R_\tau \in \mathrm{O}(d)$ with $\tau \cdot X = R_\tau X$. If also $\tau \cdot X = R'X$ with $R' \in \mathrm{O}(d)$, then $(R_\tau - R')X = 0$, so $R_\tau - R'$ vanishes on the span of the columns of X, which is $\mathbb{R}^d$ because $\mathrm{rank}(X) = d$; hence $R' = R_\tau$. The first d columns of $R_\tau X$ are R_τ times the first d columns of X, so $(\tau \cdot X)_{[d]} = (R_\tau X)_{[d]} = R_\tau X_{[d]}$, and $\det$ is multiplicative.

(b) $\Phi(MX) = \Phi(X) = A$ by Proposition 14. The two actions commute: $\tau \cdot (MX) = MXP_\tau^T = M(\tau \cdot X) = MR_\tau X$ by (a). Hence $(\tau \cdot MX)_{[d]} = MR_\tau X_{[d]}$ and $\det(\tau \cdot MX)_{[d]} = \det(M) \det(R_\tau) \det X_{[d]}$.

(c) By (a), $MX = MR_\tau^{-1}(R_\tau X) = MR_\tau^{-1}(\tau \cdot X) = MR_\tau^{-1} X P_\tau^T = (MR_\tau^{-1}, \tau) \cdot X$, and $\det(MR_\tau^{-1}) = \det(M)/\det(R_\tau) = 1$. □

Proof of Theorem 4. For $1 \le r \le d$ set $c_r := rn(2n-1) + N_r(n-1)$. We show that any two $X, Y \in \mathbb{R}^{d\times n}$ in different $\mathrm{O}(d) \times S_n$-orbits are separated by an $\mathrm{O}(d) \times S_n$-invariant polynomial of degree $\le 2c_d$, and that any two in different $\mathrm{SO}(d) \times S_n$-orbits are separated by an $\mathrm{SO}(d) \times S_n$-invariant polynomial of degree $\le 2c_d + d$. The two bounds of the theorem then follow from the identity $2c_d = 6dn^2 - 4dn - d(d-1)(n-1)$ verified at the end.

Reductions. Let $X, Y \in \mathbb{R}^{d\times n}$ lie in different $\mathrm{O}(d) \times S_n$-orbits, or in different $\mathrm{SO}(d) \times S_n$-orbits, and set $A_0 := \Phi(X)$, $B := \Phi(Y)$. If $\mathrm{rank}(A_0) \ne \mathrm{rank}(B)$, Lemma 20 gives $k \le n$ with $\mathrm{tr}(A_0^k) \ne \mathrm{tr}(B^k)$. As $C \mapsto \mathrm{tr}(C^k)$ is an S_n-invariant polynomial on $\mathcal{H}_n$ of degree k, Corollary 18 shows that $Z \mapsto \mathrm{tr}(\Phi(Z)^k)$ is an $\mathrm{O}(d) \times S_n$-invariant, hence also $\mathrm{SO}(d) \times S_n$-invariant, polynomial of degree $\le 2n$ separating X from Y. So assume $\mathrm{rank}(A_0) = \mathrm{rank}(B) =: r$. Then $r = \mathrm{rank}(X) = \mathrm{rank}(Y) \le \min(d, n)$, and $r \ge 1$, since $\Phi(X) = 0$ forces $X = 0$ and $r = 0$ would give $X = Y = 0$. Choose r linearly independent columns $x_{j_1}, \dots, x_{j_r}$ of X and $\sigma_0 \in S_n$ with $\sigma_0(j_i) = i$ for $i \le r$. The i-th column of $\sigma_0 \cdot X$ is $x_{\sigma_0^{-1}(i)} = x_{j_i}$, so the leading $r \times r$ block of $\Phi(\sigma_0 \cdot X) = \sigma_0 \cdot A_0$ (Proposition 14) is the Gram matrix of $x_{j_1}, \dots, x_{j_r}$, which is nonsingular. Since $\sigma_0 \cdot X$ lies in the S_n-orbit of X, replacing X by $\sigma_0 \cdot X$ changes neither the hypothesis on the pair (X, Y) nor the value at X of any $\mathrm{SO}(d) \times S_n$-invariant polynomial. We may therefore assume that $A := \Phi(X)$ has a nonsingular leading $r \times r$ block A_{11}, where $B = \Phi(Y)$ and $\mathrm{rank}(A) = \mathrm{rank}(B) = r \ge 1$.

Step 1: an S_n-invariant polynomial on $\mathcal{H}_n$. Let $\delta_1, \dots, \delta_r$ be the row indicators of Lemma 21 for the pair (A, B); each has degree $\le n(2n-1)$. For $(i, j) \in \mathcal{P}_r$ let

$$\Lambda_i \;:=\; \{A_{il} : 1 \le l \le n\}$$

be the set of values in the i-th row of A, and let ℓ_{ij} be the univariate Lagrange polynomial with $\ell_{ij}(A_{ij}) = 1$ and $\ell_{ij}(\lambda) = 0$ for all $\lambda \in \Lambda_i \setminus \{A_{ij}\}$; its degree is $\le |\Lambda_i| - 1 \le n - 1$. For $C \in \mathcal{H}_n$ define

$$\chi(C) \;:=\; \prod_{i=1}^{r} \delta_i(C) \prod_{(i,j)\in\mathcal{P}_r} \ell_{ij}\big(C_{ij}\big), \qquad Q(C) \;:=\; \sum_{\tau\in S_n} \chi(\tau \cdot C).$$

The polynomial χ has degree $\le \sum_{i\le r} \deg(\delta_i) + \sum_{(i,j)\in\mathcal{P}_r} \deg(\ell_{ij}) \le rn(2n-1) + N_r(n-1) = c_r$, and since $C \mapsto \tau \cdot C$ is linear, so does Q. For $\sigma \in S_n$ we have $\tau \cdot (\sigma \cdot C) = (\tau\sigma) \cdot C$, hence $\chi(\tau \cdot (\sigma \cdot C)) = \chi((\tau\sigma) \cdot C)$; as $\tau \mapsto \tau\sigma$ is a bijection of S_n, replacing C by $\sigma \cdot C$ merely reindexes the sum, and Q is S_n-invariant.

We claim that for $C \in \{A, B\}$ and $\tau \in S_n$,

$$\chi(\tau \cdot C) \in \{0, 1\}, \qquad \text{and} \qquad \chi(\tau \cdot C) = 1 \iff \tau \cdot C = A. \tag{$*$}$$

Fix such C and τ. If $\delta_i(\tau \cdot C) = 0$ for some $i \le r$, then $\chi(\tau \cdot C) = 0$ (an exact zero annihilates the product), and $\tau \cdot C \ne A$, because $\delta_i(A) = 1$ by Lemma 21. Otherwise $\delta_i(\tau \cdot C) = 1$ for all $i \le r$, so by Lemma 21 every row $i \le r$ of $\tau \cdot C$ is a reordering of row i of A. Then $(\tau \cdot C)_{ij} \in \Lambda_i$ for every $(i, j) \in \mathcal{P}_r$, where ℓ_{ij} takes values in $\{0, 1\}$, with value 1 exactly when $(\tau \cdot C)_{ij} = A_{ij}$. (Off its node set a Lagrange factor is uncontrolled — possibly negative or large — which is why the exact vanishing of a δ-factor is what disposes of the mismatched terms.) Hence $\chi(\tau \cdot C) \in \{0, 1\}$, and $\chi(\tau \cdot C) = 1$ if and only if $(\tau \cdot C)_{ij} = A_{ij}$ for all $(i, j) \in \mathcal{P}_r$, that is, by symmetry, if and only if $\tau \cdot C$ and A have the same first r rows. If $\tau \cdot C = A$ this holds. Conversely, if the first r rows agree, then $\tau \cdot C$ and A share the nonsingular leading block A_{11}, and both have rank r (conjugation by a permutation matrix preserves rank), so Lemma 15 applies to both and yields the same completion: $\tau \cdot C = A$. This proves $(*)$.

Writing $\mathrm{Stab}(A) := \{\tau \in S_n : \tau \cdot A = A\}$, $(*)$ gives

$$Q(A) = |\mathrm{Stab}(A)| \;\ge\; 1, \qquad Q(B) = \big|\{\tau \in S_n :\; \tau \cdot B = A\}\big|,$$

so $Q(B) = 0$ whenever A and B lie in different S_n-orbits (cf. Lemma 16). The polynomial Q is used in Steps 2 and 3, the polynomial χ in Step 4.

Step 2: $\mathrm{O}(d) \times S_n$. Let X, Y lie in different $\mathrm{O}(d) \times S_n$-orbits. By Corollary 18, A and B lie in different S_n-orbits, so $Q(A) \ge 1 > 0 = Q(B)$ by Step 1. By Corollary 18 again, $Q \circ \Phi$ is

an $\mathrm{O}(d) \times S_n$-invariant polynomial of degree $\leq 2c_r \leq 2c_d$ with $Q(\Phi(X)) = Q(A) \neq Q(B) = Q(\Phi(Y))$.

Step 3: $\mathrm{SO}(d) \times S_n$, $r < d$. Let X, Y lie in different $\mathrm{SO}(d) \times S_n$-orbits, with $r < d$. If Y lay in the $\mathrm{O}(d) \times S_n$-orbit of X, Lemma 22, applicable because $\mathrm{rank}(X) = r < d$, would place Y in the $\mathrm{SO}(d) \times S_n$-orbit of X. Hence X, Y lie in different $\mathrm{O}(d) \times S_n$-orbits, and the polynomial $Q \circ \Phi$ of Step 2 separates them. Being $\mathrm{O}(d) \times S_n$-invariant, it is $\mathrm{SO}(d) \times S_n$-invariant, and its degree is $\leq 2c_d$.

Step 4: $\mathrm{SO}(d) \times S_n$, $r = d$. Let X, Y lie in different $\mathrm{SO}(d) \times S_n$-orbits, with $r = d$; then $d \leq n$ and X has rank d. By the conventions above, $X_{[d]}^T X_{[d]} = A_{11}$, so $(\det X_{[d]})^2 = \det A_{11} \neq 0$. Let

$$\kappa(t) \;:=\; \tfrac{1}{2}\Big(1 + \frac{t}{\det X_{[d]}}\Big)$$

be the affine function with $\kappa(\det X_{[d]}) = 1$ and $\kappa(-\det X_{[d]}) = 0$, and define, for $Z \in \mathbb{R}^{d \times n}$,

$$\begin{aligned} p(Z) \;&:=\; \sum_{\tau \in S_n} \kappa\big(\det(\tau \cdot Z)_{[d]}\big)\, \chi\big(\Phi(\tau \cdot Z)\big) \\ &=\; \sum_{\tau \in S_n} \kappa\big(\det(\tau \cdot Z)_{[d]}\big)\, \chi\big(\tau \cdot \Phi(Z)\big), \end{aligned}$$

where the second expression uses $\Phi(\tau \cdot Z) = \tau \cdot \Phi(Z)$ (Proposition 14). Thus p is a polynomial on $\mathbb{R}^{d \times n}$ whose τ-summand is the τ-summand of $Q \circ \Phi$ multiplied by a determinant factor; the latter does not factor through Φ.

Invariance. In the first expression, the τ-summand depends on Z only through $\tau \cdot Z$. For $\sigma \in S_n$ we have $\tau \cdot (\sigma \cdot Z) = (\tau\sigma) \cdot Z$, so the τ-summand of $p(\sigma \cdot Z)$ is the $\tau\sigma$-summand of $p(Z)$, and $p(\sigma \cdot Z) = p(Z)$. For $R \in \mathrm{O}(d)$ we have $\tau \cdot (RZ) = R(\tau \cdot Z)$, hence $\Phi(\tau \cdot RZ) = \Phi(\tau \cdot Z)$ by Proposition 14 and $(\tau \cdot RZ)_{[d]} = R\,(\tau \cdot Z)_{[d]}$, so that

$$p(RZ) \;=\; \sum_{\tau \in S_n} \kappa\big(\det(R)\det(\tau \cdot Z)_{[d]}\big)\, \chi\big(\tau \cdot \Phi(Z)\big).$$

For $\det(R) = 1$ this is $p(Z)$. Hence p is $\mathrm{SO}(d) \times S_n$-invariant. For $\det(R) = -1$ the argument of κ changes sign, so p need not be $\mathrm{O}(d)$-invariant; this is the only place where p can distinguish a cloud from its mirror image.

Values. Let $Z \in \mathbb{R}^{d \times n}$ satisfy $\Phi(Z) = A$. By $(*)$, $\chi(\tau \cdot A) = 1$ for $\tau \in \mathrm{Stab}(A)$ and $\chi(\tau \cdot A) = 0$ otherwise, so

$$p(Z) \;=\; \sum_{\tau \in \mathrm{Stab}(A)} \kappa\big(\det(\tau \cdot Z)_{[d]}\big).$$

For $Z = X$, Lemma 23(a) gives $\det(\tau \cdot X)_{[d]} = \det(R_\tau)\det X_{[d]}$ with $\det(R_\tau) = \pm 1$, so each term is $\kappa(\pm \det X_{[d]}) \in \{0, 1\}$. Fix $M \in \mathrm{O}(d)$ with $\det(M) = -1$, for instance $M = \mathrm{diag}(-1, 1, \ldots, 1)$. By Lemma 23(b), $\Phi(MX) = A$ and $\det(\tau \cdot MX)_{[d]} = -\det(R_\tau)\det X_{[d]}$. Therefore

$$\begin{aligned} p(X) \;&=\; \#\{\tau \in \mathrm{Stab}(A) :\; \det(R_\tau) = +1\} \;\geq\; 1, \\ p(MX) \;&=\; \#\{\tau \in \mathrm{Stab}(A) :\; \det(R_\tau) = -1\}, \end{aligned}$$

where $p(X) \geq 1$ because $\mathrm{id} \in \mathrm{Stab}(A)$ and $R_{\mathrm{id}} = I_d$ by the uniqueness in Lemma 23(a). Finally, if A and B lie in different S_n-orbits, then $\chi(\tau \cdot B) = 0$ for every τ by $(*)$, and $\Phi(Y) = B$ gives $p(Y) = 0$.

Separation. If Y lies outside the $\mathrm{O}(d) \times S_n$-orbit of X, then A and B lie in different S_n-orbits by Corollary 18, and $p(X) \geq 1 > 0 = p(Y)$. Otherwise $Y = (R', \sigma) \cdot X$ with $R' \in \mathrm{O}(d)$ and $\sigma \in S_n$. Every such Y has the form $(R, \sigma) \cdot X$ or $(R, \sigma) \cdot MX$ with $R \in \mathrm{SO}(d)$: take $R = R'$ if $\det(R') = 1$, and otherwise $R = R'M^{-1} \in \mathrm{SO}(d)$, for which $(R, \sigma) \cdot MX = R'M^{-1}MXP_\sigma^T = Y$. The first form is excluded, since Y is not in the $\mathrm{SO}(d) \times S_n$-orbit of X. So $Y = (R, \sigma) \cdot MX$ with $R \in \mathrm{SO}(d)$, and $p(Y) = p(MX)$ by invariance. Moreover Y and MX lie in the same $\mathrm{SO}(d) \times S_n$-orbit, so MX does not lie in the $\mathrm{SO}(d) \times S_n$-orbit of X; by Lemma 23(c), no $\tau \in \mathrm{Stab}(A)$ has $\det(R_\tau) = -1$,

and the value formula gives $p(MX) = 0$. Thus $p(X) \geq 1 > 0 = p(Y)$ in this case too. In either case p is an $\mathrm{SO}(d) \times S_n$-invariant polynomial separating X from Y.

Degree. Since $N_{r+1} - N_r = n - r$, we have $c_{r+1} - c_r = n(2n-1) + (n-1)(n-r) > 0$ for $r < n$, so c_r is increasing in r and $c_r \leq c_d = dn(2n-1) + N_d(n-1)$ for all $r \leq d$; also $2n \leq 2c_1 \leq 2c_d$. Hence the polynomials of the rank reduction and of Steps 2 and 3 have degree $\leq 2c_d$. In Step 4, every entry of $\Phi(Z)$ is quadratic in the entries of Z, so $Z \mapsto \chi(\Phi(\tau \cdot Z))$ has degree $\leq 2c_d$; and κ is affine while $\det(\tau \cdot Z)_{[d]}$, the determinant of the $d \times d$ matrix with columns $z_{\tau^{-1}(1)}, \ldots, z_{\tau^{-1}(d)}$, is a homogeneous polynomial of degree d in the entries of Z. Hence $\deg(p) \leq 2c_d + d$. Using $2N_d(n-1) = 2dn(n-1) - d(d-1)(n-1) = 2dn^2 - 2dn - d(d-1)(n-1)$,

$$\begin{aligned} 2c_d &= 2dn(2n-1) + 2N_d(n-1) \\ &= (4dn^2 - 2dn) + (2dn^2 - 2dn) - d(d-1)(n-1) \;=\; 6dn^2 - 4dn - d(d-1)(n-1). \end{aligned}$$

Therefore $D(\mathrm{O}(d) \times S_n, \mathbb{R}^{d\times n}) \leq 6dn^2 - 4dn - d(d-1)(n-1)$ by Step 2, and $D(\mathrm{SO}(d) \times S_n, \mathbb{R}^{d\times n}) \leq 6dn^2 - 4dn - d(d-1)(n-1) + d$ by Steps 3 and 4. By Lemma 2, $D(E(d) \times S_n, \mathbb{R}^{d\times n}) \leq D(\mathrm{O}(d) \times S_n, \mathbb{R}^{d\times n})$ and $D(SE(d) \times S_n, \mathbb{R}^{d\times n}) \leq D(\mathrm{SO}(d) \times S_n, \mathbb{R}^{d\times n})$, so the same two bounds hold for $E(d)$ and $SE(d)$. Both bounds are at most $6dn^2$, because $d \leq 4dn + d(d-1)(n-1)$ for $n \geq 1$. □

D Constructing invariant polynomials via invariant models

D.1 Tensor Field Networks

Here we add some details on the construction of $SE(3) \times S_n$ invariant polynomials by TFN (Proposition 6). First we discuss our assumptions: the TFN model we refer to is the one defined in Dym & Maron (2021). It differs from the original TFN models only in that it allows the radial functions in the TFN model to be polynomial rather than MLPs. This allows the TFN models to exactly express polynomials. In the MLP version, the word express can be replaced with 'approximate on a compact set', since MLPs can approximate any polynomial on a compact set. This is discussed in detail in Dym & Maron (2021).

In our analysis we assume that TFN produces a permutation invariant global feature by summing over all node features.

Now we explain how Proposition 6 follows from the results in Dym & Maron (2021). The first part of Theorem 2 in that paper, in the special case where W_T is the trivial representation $W_0 = \mathbb{R}$, gives the following result:

Proposition 24 (Taken from Theorem 2 in Dym & Maron (2021))**.** *For all $n, D \in \mathbb{N}$, every polynomial $p : \mathbb{R}^{3\times n} \to \mathbb{R}^n$ which is rotation and translation invariant, and permutation equivariant, can be expressed using TFN with irreducible features of order D, and $2D$ layers.*

It remains only to show that this claim can be translated from equivariant polynomials to invariant polynomials. This is easily done: let $q : \mathbb{R}^{3\times n} \to \mathbb{R}$ be a $SE(3) \times S_n$ invariant polynomial of degree D. Then $p(X) = n^{-1} q(X) 1_n$ is a rotation and translation invariant, and permutation equivariant, polynomial. And therefore p can be expressed as a TFN model of the advertised size. By applying sum pooling we obtain q, that is $q(X) = \sum_{j=1}^n p_j(X)$.

D.2 MACE

Here we give the details regarding the construction of $SE(3) \times S_n$ invariant polynomials by MACE (Proposition 8).

We consider the original MACE Batatia et al. (2022) architecture with the following simplifications: (a) we take a single iteration $t = 1$ (b) we remove from the input the z_i features discussed in MACE (representing the atom numbers) (c) For every point i we perform message passing on all points and not only on points within a given distance threshold. We also make two changes (a) as in the TFN model, we allow the functions $R_{k\ell}$ to be any continuous functions, and (b) following ACE Drautz (2019), in which each factor of a basis function has its own radial index, we use weights $w_{k\tilde{k}\xi}$ that depend on the position ξ of the factor, instead of $w_{k\tilde{k}\ell_\xi}$ of Batatia et al. (2022). This extends MACE:

with $w_{k\tilde{k}\ell_\xi}$, factors of equal order in the same channel coincide, so couplings antisymmetric in them vanish, e.g. the triple product $\det(v, v, v) = 0$ of three equal factors of order 1. In particular, if $w_{k\tilde{k}\xi} = 1$ for $\tilde{k} = k_\xi$ and $w_{k\tilde{k}\xi} = 0$ otherwise, as in the proof below, then equation 4 becomes $\sum_{\boldsymbol{\ell m}} \mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{\ell m}} \prod_{\xi=1}^{\nu} A_{i,k_\xi \ell_\xi m_\xi}$, which has the form of an ACE basis function Drautz (2019), with the channel k_ξ in the role of the radial index of the ξ-th factor. Finally, the Clebsch–Gordan coefficients in equation 4, and in equation 12 below, range over all couplings of representations of $SO(3)$, without the parity selection rule of the $O(3)$-equivariant implementation of Batatia et al. (2022), which makes the model invariant under $X \mapsto -X$; with this rule, the same argument gives $(E(3) \times S_n, \mathbb{R}^{3\times n})$ completeness with $D = 18n^2 - 18n + 6$, since only couplings with $\ell_1 + \cdots + \ell_\nu$ even are then needed.

With these changes, the MACE model is of the form:

$$A_{i,k\ell m} = \sum_{j\neq i} R_{k\ell}(|x_j - x_i|)\, Y^m_\ell \left(\frac{x_j - x_i}{|x_j - x_i|}\right) \tag{3}$$

$$B_{i,\eta_\nu kLM} = \sum_{\boldsymbol{\ell m}} \mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{\ell m}} \prod_{\xi=1}^{\nu} \sum_{\tilde{k}} w_{k\tilde{k}\xi}\, A_{i,\tilde{k}\ell_\xi m_\xi}, \qquad \boldsymbol{\ell m} = (\ell_1 m_1, \ldots, \ell_\nu m_\nu) \tag{4}$$

$$m_{i,kLM} = \sum_{\nu} \sum_{\eta_\nu} W_{kL,\eta_\nu}\, B_{i,\eta_\nu kLM}, \tag{5}$$

$$E_i = \mathrm{MLP}\,(m_{ik00}, k = 1, \ldots, K_{\max}) \tag{6}$$

$$E = \sum_{i=1}^{n} E_i \tag{7}$$

Important hyperparameters of this model are: $\ell_{\max}$, the maximal irreducible degree used throughout the model. $\nu_{\max}$, the maximal 'body order', and $K_{\max}$, the number of 'channels'.

We now explain this in more detail. In equation 3 we define $K_{\max}$ high dimensional features $A_{i,k}, k = 1, \ldots, K_{\max}$ for each point x_i. Each such feature is in $W_0 \oplus W_1 \oplus \ldots \oplus W_{\ell_{\max}}$, and so is indexed by $A_{i,k,\ell,m}$ where ℓ denotes the representation and m is the internal index inside the representation. The functions $R_{k\ell}$ are generally chosen to be some parametric model, and in our analysis we will choose them to be polynomial functions, as will be seen in the proof.

In Equation 4 new features $B_{i,k}$ are obtained by taking tensor products of $\nu \leq \nu_{\max}$ linear combinations of $A_{i,\tilde{k}}$ features, and the using the Clebsch-Gordan coefficients $\mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{lm}}$ to decompose these tensor products back to $W_0^{r_0} \oplus W_1^{r_1} \oplus \ldots \oplus W_{\ell_{\max}}^{r_{\ell_{\max}}}$ (the tensor product could have components in W_ℓ with $\ell > \ell_{\max}$ but these components are discarded). The features $B_{i,\eta_\nu kLM}$ are indexed similarly to the $A_{i,k\ell m}$ features, but now we have an additional index η_ν due to the possible multiplicity $r_\ell > 1$ of the features.

Equation 5 is a linear pooling step over multiplicities and body orders. As can be seen in the next equation, in our setting only the $L = 0, m = 0$ components are of interest in this step.

Equation 6 obtains a rotation invariant feature per node E_i by applying an MLP (Multi-Layer Perceptron) to the $\ell = 0, m = 0$ features, for which the action of $SO(3)$ is trivial, namely $D_\ell(R) = 1$ when $\ell = 0$.

Finally, Equation 7 produces a permutation invariant feature by summing over E_i. Thus $E = E(X)$ is a $SE(3) \times S_n$ invariant function. We now recall and prove Proposition 8:

Proposition 8. *MACE with $\nu_{\max} = \min\{D, n-1\}$ and $\ell_{\max} = \lfloor D/2 \rfloor$ can express any $SE(3)\times S_n$ invariant polynomial of degree $\leq D$.*

In other words, this means that if we choose the MACE hyperparameters to be $\nu_{\max} = \min\{D, n-1\}$ and $\ell_{\max} = \lfloor D/2 \rfloor$, and choose some appropriate $K_{\max} = K_{\max}(D)$, then there exists a choice of the functions R_{kl} and their parameters $w_{k\tilde{k}\eta}, W_{kL\eta_\nu}$ of MACE such that $E(X) = q(X)$.

The proof of the proposition will be based on the following proposition:

Proposition 25. *Let $p : \mathbb{R}^{3\times n} \to \mathbb{R}^n$ be a polynomial with degree at most D that is invariant to rotations and translations, and equivariant to permutations. That is, that the coordinates $p_j(X)$ are rotation and translation invariant, and*

$$(p_1(\tau(X)), \dots, p_n(\tau(X))) = \left(p_{\tau^{-1}(1)}(X), \dots, p_{\tau^{-1}(n)}(X)\right), \quad \forall \tau \in S_n. \tag{8}$$

Then there exists a single layer MACE architecture with $\nu_{\max} = \min\{D, n-1\}$ and $\ell_{\max} = \lfloor D/2 \rfloor$, such that the node features $E_i = E_i(X)$ from the MACE architecture satisfy

$$E_i(X) = p_i(X), \forall i = 1 \dots, n$$

This proposition easily implies the original proposition 8, for if $q : \mathbb{R}^{3\times n} \to \mathbb{R}$ is of degree $\leq D$ and $SE(3) \times S_n$ invariant, then we can apply Proposition 25 for the equivariant polynomial $p = n^{-1}(q, q, \dots, q)$ to obtain a MACE architecture with $E_i(X) = n^{-1}q(X)$, and so $E(X) = q(X)$.

In the proof we will use the following lemma, which shows that a product of many sums over the same points can be written using products of fewer sums.

Lemma 26. *Let $u_1, \dots, u_M \in \mathbb{R}^3$ and let $N \geq M$. For a function $f : \mathbb{R}^3 \to \mathbb{R}$ denote $P(f) = \sum_{j=1}^{M} f(u_j)$. Then for every $f_0, \dots, f_N : \mathbb{R}^3 \to \mathbb{R}$, the product $P(f_0)P(f_1)\cdots P(f_N)$ is a linear combination of products of the form*

$$P\Big(\prod_{s\in c_1} f_s\Big) \cdot P\Big(\prod_{s\in c_2} f_s\Big) \cdots P\Big(\prod_{s\in c_r} f_s\Big),$$

where $c_1, \dots, c_r$ is a partition of $\{0, \dots, N\}$ into $r \leq N$ sets. The coefficients depend only on the partition, and not on the functions f_s, the points u_j or M.

Proof. We first show that

$$\sum_{\sigma\in S_{N+1}} \operatorname{sgn}(\sigma) \prod_c P\Big(\prod_{s\in c} f_s\Big) = 0, \tag{9}$$

where σ is a permutation of $\{0, \dots, N\}$ and the product is over the cycles c of σ (including fixed points). Expanding the sums, the product over the cycles of σ is equal to $\sum_\psi \prod_{s=0}^{N} f_s(u_{\psi(s)})$, where the sum is over all functions $\psi : \{0, \dots, N\} \to \{1, \dots, M\}$ which are constant on the cycles of σ, that is $\psi \circ \sigma = \psi$. Changing the order of summation, the left hand side of equation 9 is equal to

$$\sum_\psi \prod_{s=0}^{N} f_s(u_{\psi(s)}) \sum_{\sigma:\, \psi\circ\sigma=\psi} \operatorname{sgn}(\sigma).$$

Since $N + 1 > M$, for every ψ there exist $s \neq s'$ with $\psi(s) = \psi(s')$. Let τ be the transposition of s and s'. Then $\sigma \mapsto \sigma \circ \tau$ is a bijection of the set $\{\sigma : \psi \circ \sigma = \psi\}$ to itself which changes the sign, so the inner sum is zero. This proves equation 9.

The term of the identity permutation in equation 9 is $P(f_0)P(f_1)\cdots P(f_N)$, and every other permutation has at most N cycles. Moving these terms to the other side we obtain the claim. The coefficient of a partition is minus the sum of the signs of the permutations whose cycles are the sets of the partition, so it depends only on the partition. □

Proof of Proposition 25. The proof is based on the characterization of $SE(3) \times S_n$ equivariant and invariant polynomials in Dym & Maron (2021).

Now fix some $i \in \{1, \dots, n\}$ and consider the i-th coordinate $p_i : \mathbb{R}^{3\times n} \to \mathbb{R}$. The permutation equivariance of p implies (see Segol & Lipman (2020) Theorem 2) that p_i is a linear combination of monomials of the form

$$\sum_{i_2, i_3, \dots, i_r = 1}^{n} x_i^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_r}^{\alpha_r}$$

where $r \geq 1$, each $\alpha_j = (\alpha_{1j}, \alpha_{2j}, \alpha_{3j})$ is a multi-index and the total degree $\sum_{j=1}^{r} \|\alpha_j\|_1$ is at most D. Since p_i is translation invariant, it is unaffected by translating all entry columns by the column x_i, showing that it is in fact a linear combination of expressions of the form

$$\sum_{i_2, i_3, \dots, i_r = 1}^{n} (x_{i_2} - x_i)^{\alpha_2} \cdots (x_{i_r} - x_i)^{\alpha_r} = \left(\sum_{j\neq i} (x_j - x_i)^{\alpha_2}\right) \dots \left(\sum_{j\neq i} (x_j - x_i)^{\alpha_r}\right)$$

Noting that factors with $\alpha_t = 0$ are constants and are absorbed into the coefficients and every remaining factor has degree ≥ 1, so the number of factors in the product above is at most $\sum_t |\alpha_t| \leq D$. Moreover, each of the sums is over the $n-1$ points $x_j - x_i$, $j \neq i$. Therefore, by applying Lemma 26 repeatedly, a product of more than $n-1$ such sums is a linear combination of products of at most $n-1$ sums of the same form, where the new multi-indices are sums of the original α_t, so the total degree does not change. The coefficients of this linear combination do not depend on i. Thus we can assume that the number of factors is at most $\min\{D, n-1\}$.

Next, we leverage the fact that every polynomial $q : \mathbb{R}^3 \to \mathbb{R}$ of degree D can be written as linear combination of polynomials of the form $\|x\|^{2k} Y_\ell^m(x)$ with $\ell + 2k \leq D$ Dai & Xu (2013). As a result, we deduce that every expression of the form $\left(\sum_{j\neq i}(x_j - x_i)^\alpha\right)$ above can be rewritten as a linear combination of polynomials of the form

$$a_{i,k\ell m}(X) = \sum_{j\neq i}^{n} \|x_j - x_i\|^{2k} Y_\ell^m(x_j - x_i)$$

and therefore p_i is a linear combination of polynomials of the form

$$c_{i,\boldsymbol{k},\boldsymbol{l},\boldsymbol{m}}(X) = a_{i,k_1,\ell_1,m_1}(X) \cdot a_{i,k_2,\ell_2,m_2}(X) \cdot \ldots \cdot a_{i,k_\nu,\ell_\nu,m_\nu}(X)$$

where

$$\boldsymbol{k} = (k_1, \ldots, k_\nu), \boldsymbol{l} = (\ell_1, \ldots, \ell_\nu), \boldsymbol{m} = (m_1, \ldots, m_\nu).$$

This gives that each $\sum_j (x_j - x_i)^\alpha$ is homogeneous of degree $|\alpha|$, so its expansion uses only $2k+l = |\alpha|$, therefore every product $c_{i,k,l,m}$ in the expansion of p_i satisfies $\sum_s (2k_s + l_s) \leq D$. Now, let $c_{i,\boldsymbol{k}}(X)$ be the tensor containing all entries $c_{i,\boldsymbol{k},\boldsymbol{l},\boldsymbol{m}}(X)$, and let

$$b_{i,\boldsymbol{k}}(X) = \mathcal{C} c_{i,\boldsymbol{k}}(X)$$

where $\mathcal{C}$ is the linear equivariant invertible map mapping the tensor product domain to a direct sum of irreducibles which we denote by $W = W_0^{r_0} \oplus W_1^{r_1} \oplus \ldots \oplus W_D^{r_D}$ (whose coordinates are the Clebsch-Gordan coefficients).

Since $\sum_s l_s \leq D$, the irreducibles in W are labeled at most by D. We deduce that for appropriate linear $\Lambda_{\boldsymbol{k}} : W \to \mathbb{R}$,

$$p_i(X) = \sum_{\boldsymbol{k}} \Lambda_{\boldsymbol{k}} \left(b_{i,\boldsymbol{k}}(X)\right) \tag{10}$$

Moreover, if p_i is additionally rotation invariant, then an averaging argument (see the proof of Theorem 1 in Dym & Maron (2021)) shows that the $\Lambda_{\boldsymbol{k}}$ are rotation invariant. It follows almost directly from Schur's lemma (see Lemma 5 in Dym & Maron (2021) for details) that $\Lambda_{\boldsymbol{k}}$ only operates on the components of $b_{i,\boldsymbol{k}}$ in the trivial representation $W_0^{r_0}$. Namely, if a general element in the tensor $b_{i,\boldsymbol{k}}$ is indexed by $b_{i,\boldsymbol{k},\eta,\ell,m}$, where η ranges from 1 to r_ℓ, the multiplicity of W_ℓ in W, then

$$\Lambda_{\boldsymbol{k}}\left(b_{i,\boldsymbol{k}}\right) = \sum_{\eta=1}^{r_0} \lambda_{\boldsymbol{k},\eta} b_{i,\boldsymbol{k}\eta 00} \text{ for some vector } \lambda_{\boldsymbol{k}} = (\lambda_{\boldsymbol{k},1}, \ldots, \lambda_{\boldsymbol{k},r_0}) \tag{11}$$

Since $\Lambda_{\boldsymbol{k}}$ reads only the W_0-components of $b_{i,\boldsymbol{k}}$, a term $c_{i,\boldsymbol{k},\boldsymbol{l},\boldsymbol{m}}$ contributes to p_i only if W_0 occurs in $W_{\ell_1} \otimes \cdots \otimes W_{\ell_\nu}$. A tensor product of irreducibles contains no W_J with J larger than the sum of its labels (iterate the Clebsch–Gordan rule $W_a \otimes W_b = \bigoplus_{c=|a-b|}^{a+b} W_c$), so if $W_0 \subseteq W_{\ell_\xi} \otimes \left(\bigotimes_{\xi' \neq \xi} W_{\ell_{\xi'}}\right)$ then $\ell_\xi \leq \sum_{\xi' \neq \xi} \ell_{\xi'}$ for every ξ, thus $2\ell_\xi \leq \sum_\xi \ell_\xi \leq \sum_\xi (2k_\xi + \ell_\xi) \leq D$. All other terms may be dropped from the expansion of p_i, and we assume from now on that $\ell_\xi \leq \lfloor D/2 \rfloor$ and $k_\xi \leq \lfloor D/2 \rfloor$ for every factor.

We now explain how MACE can express p_i. First note that we can express $a_{i,klm}$ using $A_{i,klm}$, by choosing $R_{kl}(|x_j - x_i|) = |x_j - x_i|^{2k+l}$ to obtain

$$A_{i,klm}(X) = \sum_{j\neq i} |x_j - x_i|^{2k+l} Y_l^m\left(\frac{x_j - x_i}{|x_j - x_i|}\right) \overset{(*)}{=} \sum_{j\neq i} |x_j - x_i|^{2k} Y_l^m\left(x_j - x_i\right) = a_{i,klm}(X),$$

where (*) follows from the fact that Y_ℓ^m is a degree ℓ homogeneous polynomial and above $A_{i,klm}$ features have $\lfloor D/2 \rfloor + 1$ channels indexed by $k = 0, \ldots, \lfloor D/2 \rfloor$, with $R_{kl}(r) = r^{2k+l}$ for $l \leq \lfloor D/2 \rfloor$.

Next, we choose

$$w_{\boldsymbol{k}\tilde{k}\xi} = \begin{cases} 1 & \text{if } \tilde{k} = k_\xi \\ 0 & \text{if } \tilde{k} \neq k_\xi \end{cases}, \text{ so that } \sum_{\tilde{k}} w_{k\tilde{k}\xi}\, A_{i,\tilde{k}l_\xi m_\xi} = A_{i,k_\xi l_\xi m_\xi}$$

and we can then obtain that

$$\begin{aligned} B_{i,\boldsymbol{k}} &= \sum_{\boldsymbol{lm}} \mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{lm}} \prod_{\xi=1}^{\nu} \sum_{\tilde{k}} w_{k\tilde{k}\xi}\, A_{i,\tilde{k}l_\xi m_\xi} \\ &= \sum_{\boldsymbol{lm}} \mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{lm}} \prod_{\xi=1}^{\nu} A_{i,k_\xi l_\xi m_\xi} \\ &= \mathcal{C}\left(a_{i,k_1} \otimes a_{i,k_2} \otimes \ldots \otimes a_{i,k_\nu}\right) \end{aligned}$$

where $\mathcal{C}$ is the linear mapping defined by the Clebsch-Gordan coefficients $\mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{lm}}$. The $B_{i,\boldsymbol{k}}$ are indexed by the tuple $\boldsymbol{k} = (k_1, \ldots, k_\nu)$ and there are $K_{\max}(D)$ valid tuples, that is those with $k_i \leq \lfloor D/2 \rfloor$ and $\nu \leq \min\{D, n-1\}$, which are those we iterate over so that $w_{k\tilde{k}\xi}$ has a well-defined index range. The $L = 0$ components of $B_{i,\boldsymbol{k}}$ coincide with those of $b_{i,\boldsymbol{k}}$ used in Equation (11), a fact we use below.

Thus, choosing $W_{\boldsymbol{k}0\eta} = \lambda_{\boldsymbol{k}\eta}$ and $W_{\boldsymbol{k}L\eta} = 0$ for $L > 0$, and choosing the MLP in the definition of E_i to just sum over all coordinates $\boldsymbol{k}$, we obtain

$$p_i(X) \overset{equation\ 10}{=} \sum_{\boldsymbol{k}} \Lambda_{\boldsymbol{k}}\left(b_{i,\boldsymbol{k}}(X)\right) \overset{equation\ 11}{=} \sum_{\boldsymbol{k}} \sum_{\eta=1}^{r_0} \lambda_{\boldsymbol{k},\eta} b_{i,\boldsymbol{k}\eta 00} = E_i(X)$$

□

D.3 MACE WITH MULTIPLE ITERATIONS

So far we considered MACE with a single iteration. We now explain how the results above extend to MACE with $T \geq 1$ iterations. As above, we perform message passing on all points, so there is no underlying graph. In the MACE architecture Batatia et al. (2022) with T iterations, every point x_i has a hidden state $h_i^{(t)}$, where $h_i^{(0)} = 1$, and the hidden states are updated by

$$A^{(t)}_{i,k\ell_3 m_3} = \sum_{\ell_1 m_1, \ell_2 m_2} \mathcal{C}^{\ell_3 m_3}_{\ell_1 m_1, \ell_2 m_2} \sum_{j \neq i} R^{(t)}_{k\ell_1\ell_2\ell_3}(|x_j - x_i|)\, Y^{m_1}_{\ell_1}\left(\frac{x_j - x_i}{|x_j - x_i|}\right) \sum_{\tilde{k}} W^{(t)}_{k\tilde{k}\ell_2}\, h^{(t)}_{j,\tilde{k}\ell_2 m_2} \tag{12}$$

$$h^{(t+1)}_{i,kLM} = \sum_{\tilde{k}} U^{(t)}_{kL,\tilde{k}}\, m^{(t)}_{i,\tilde{k}LM} + \sum_{\tilde{k}} \tilde{U}^{(t)}_{kL,\tilde{k}}\, h^{(t)}_{i,\tilde{k}LM} \tag{13}$$

for $t = 0, \ldots, T-1$, where $B_i^{(t)}$ and $m_i^{(t)}$ are computed from $A_i^{(t)}$ as in equation 4 and equation 5, and the node feature E_i is obtained by applying the MLP to the features $h^{(T)}_{i,k00}$. Note that for $t = 0$ we have $h_j^{(0)} = 1$, so only $\ell_2 = 0$ appears and equation 12 is the same as equation 3. Therefore the single iteration model above is obtained by choosing $T = 1$, $U^{(0)}$ to be the identity and $\tilde{U}^{(0)} = 0$.

Proposition 27. *Let $T \geq 1$. For every MACE model with a single iteration, there exists a MACE model with T iterations and the same $\nu_{\max}$, $\ell_{\max}$ and $K_{\max}$, which has the same node features E_i. In particular, Proposition 8 and Proposition 25 hold for MACE with T iterations. The same is true when the summation in equation 3 and equation 12 is only over the points at distance at most ρ from x_i.*

Proof. In the first iteration we choose the parameters of the single iteration model, $U^{(0)}$ to be the identity and $\tilde{U}^{(0)} = 0$, so that $h_i^{(1)} = m_i^{(0)}$. For $t \geq 1$ we choose $U^{(t)} = 0$ and $\tilde{U}^{(t)}$ to be the identity, so that $h_i^{(t+1)} = h_i^{(t)}$. We obtain that $h_i^{(T)} = h_i^{(1)}$, and therefore the node features E_i are the same as in the single iteration model. The same argument applies when the summation in equation 12 is only over the points at distance at most ρ from x_i. □

In the proof of Proposition 27 the additional iterations are not used. We now show that also when all iterations are used, features with $\ell > \lfloor D/2 \rfloor$ do not contribute to polynomials of degree $\leq D$. As in the proof of Proposition 25, we consider radial functions of the form $R^{(t)}_{k\ell_1\ell_2\ell_3}(r) = c\, r^{2a+\ell_1}$ where $a \geq 0$ is an integer, so that $R(|x|)Y^{m_1}_{\ell_1}(x/|x|) = c|x|^{2a}Y^{m_1}_{\ell_1}(x)$ is a homogeneous polynomial of degree $2a+\ell_1$. We also assume that in each of equation 12, equation 4, equation 5 and equation 13, all the terms in the sum which defines a given coordinate are homogeneous polynomials in X of the same degree. A model whose radial functions are linear combinations of the functions $r^{2a+\ell_1}$, and in particular the model in the proof of Proposition 25, can be brought to this form without changing T and $\nu_{\max}$, by splitting every channel according to the degrees, since the radial functions can be chosen separately for every channel and the weights can select a single term in each of these sums. We say that a coordinate of $A^{(t)}_i, B^{(t)}_i, m^{(t)}_i$ or $h^{(t)}_i$ is of type (e, L) if it is a homogeneous polynomial of degree e and it is a coordinate of a feature in W_L.

Lemma 28. *Consider a MACE model with $T \geq 1$ iterations as above, with no restriction on ℓ, where E_i is a linear combination of the coordinates $h^{(T)}_{i,k00}$. Let $E_i^{\leq D}$ be the sum of the terms in this linear combination whose degree is at most D. Then $E_i^{\leq D}$ does not change if we discard the spherical harmonics with $\ell_1 > \lfloor D/2 \rfloor$ from equation 12, and the coordinates with $\ell > \lfloor D/2 \rfloor$ from all the features $A^{(t)}_i, B^{(t)}_i, m^{(t)}_i$ and $h^{(t)}_i$. In particular, if E_i is a polynomial of degree $\leq D$ then it can be expressed with $\ell_{\max} = \lfloor D/2 \rfloor$.*

Proof. We first note that every coordinate of type (e, L) satisfies $L \leq e$. This holds for $h^{(0)}_i = 1$ which is of type $(0, 0)$. In equation 12, a term which is obtained from a spherical harmonic of degree ℓ_1, a radial function $c\, r^{2a+\ell_1}$ and a coordinate of $h^{(t)}_j$ of type (e_2, ℓ_2) with $\ell_2 \leq e_2$, is of type (e_3, ℓ_3) with $e_3 = 2a + \ell_1 + e_2 \geq \ell_1 + \ell_2 \geq \ell_3$. In equation 4, a product of coordinates of types (e_ξ, ℓ_ξ) with $\ell_\xi \leq e_\xi$ is of type (e, L) with $e = \sum_\xi e_\xi \geq \sum_\xi \ell_\xi \geq L$. Finally, equation 5 and equation 13 do not change the type.

We say that a coordinate of type (e, L) is *useful* if $e + L \leq D$. A useful coordinate satisfies $2L \leq e + L \leq D$ and therefore $L \leq \lfloor D/2 \rfloor$. We claim that a useful coordinate is computed only from useful coordinates, and from spherical harmonics with $\ell_1 \leq \lfloor D/2 \rfloor$. For equation 5 and equation 13 this is clear. For equation 4, if W_L occurs in $W_{\ell_1} \otimes \cdots \otimes W_{\ell_\nu}$ then by the Clebsch–Gordan rule $\ell_\xi \leq L + \sum_{\xi' \neq \xi} \ell_{\xi'}$ for every ξ, and therefore

$$e_\xi + \ell_\xi \leq e_\xi + L + \sum_{\xi' \neq \xi} \ell_{\xi'} \leq L + \sum_{\xi'} e_{\xi'} = e + L \leq D.$$

For equation 12 we have $\ell_2 \leq \ell_1 + \ell_3$ and $\ell_1 \leq \ell_2 + \ell_3$, and therefore

$$e_2 + \ell_2 \leq e_2 + \ell_1 + \ell_3 \leq e_3 + \ell_3 \leq D \quad \text{and} \quad 2\ell_1 \leq \ell_1 + \ell_2 + \ell_3 \leq e_3 + \ell_3 \leq D.$$

It follows by induction on t that the useful coordinates do not change when we discard the coordinates with $\ell > \lfloor D/2 \rfloor$, since these coordinates are not useful. The coordinates $h^{(T)}_{i,k00}$ of degree $\leq D$ are of type $(e, 0)$ with $e \leq D$, so they are useful, and thus $E_i^{\leq D}$ does not change. Finally, if E_i is a polynomial of degree $\leq D$ then the terms of degree larger than D cancel out, so $E_i = E_i^{\leq D}$, and we can choose the coefficients of these terms to be zero. □

We now show that when the summation is over all points, more iterations can be used instead of a higher $\nu_{\max}$. For $\nu \geq 1$ we define $\mu_0 = 0$ and $\mu_t = \nu(\mu_{t-1} + 1)$ for $t \geq 1$, that is $\mu_t = \nu + \nu^2 + \ldots + \nu^t$. In particular $\mu_1 = \nu$, for $\nu = 1$ we have $\mu_t = t$, and for $\nu = 2$ we have $\mu_t = 2^{t+1} - 2$.

Proposition 29. *Let $n \geq 2$ and let $\nu, T \geq 1$ be such that $\mu_T \geq \min\{D, n-1\}$. Then Proposition 8 and Proposition 25 hold also for MACE with T iterations, $\nu_{\max} = \nu$ and $\ell_{\max} = \lfloor D/2 \rfloor$, when the summation in equation 12 is over all points.*

As a corollary of this proposition and Theorem 4, we obtain the following variant of Corollary 9:

Corollary 30. *Let $n \geq 2$ and set $D = 18n^2 - 18n + 9$. Let $\nu, T \geq 1$ be such that $\mu_T \geq n - 1$. Then a MACE model with T iterations, $\nu_{\max} = \nu$ and $\ell_{\max} = \lfloor D/2 \rfloor$, in which the summation in equation 12 is over all points, is $(SE(3) \times S_n, \mathbb{R}^{3\times n})$ complete. In particular, this holds for $\nu_{\max} = n-1$ and $T = 1$, for $\nu_{\max} = 2$ and $T \geq \log_2(n+1) - 1$, and for $\nu_{\max} = 1$ and $T \geq n - 1$.*

Proof of Proposition 29. For $m \geq 0$ and a polynomial $H : (\mathbb{R}^3)^m \to W_L$ of degree $\leq D$ we denote

$$S_i(H) = \sum_{j_1,\ldots,j_m=1}^{n} H(x_{j_1} - x_i, \ldots, x_{j_m} - x_i),$$

and if H is equivariant we say that $S_i(H)$ is an m-point sum centered at x_i (for $m = 0$ these are the constants in W_0). Note that an m-point sum is also an m'-point sum for every $m' \geq m$, since if H does not depend on the last $m' - m$ points then the sum over these points gives the constant $n^{m'-m}$. We prove by induction on t that we can choose the parameters of the first t iterations such that for every i, every μ_t-point sum centered at x_i is a linear combination of the features $h^{(t)}_{i,kLM}$, with coefficients which do not depend on i. For $t = 0$ this holds since $h^{(0)}_i = 1$.

Now assume that this holds for t, and denote $m = \mu_t$. Fix i and denote $u_j = x_j - x_i$, so that $u_i = 0$. Let $\phi : \mathbb{R}^3 \to W_{\ell_1}$ and $H : (\mathbb{R}^3)^m \to W_{\ell_2}$ be equivariant polynomials. Then $\phi(x) = R(|x|)Y_{\ell_1}(x/|x|)$, where Y_{ℓ_1} is the vector of the spherical harmonics of degree ℓ_1 and R is a linear combination of the functions $r^{2a+\ell_1}$, and by the induction hypothesis $S_j(H)$ is a linear combination of the features $h^{(t)}_j$. Therefore in equation 12 we can obtain the features

$$\sum_{j\neq i} \phi(u_j) \otimes S_j(H) = \sum_{j\neq i} \sum_{j_1,\ldots,j_m=1}^{n} \phi(u_j) \otimes H(u_{j_1} - u_j, \ldots, u_{j_m} - u_j) = S_i(\tilde{H}) - S_i(\tilde{H}_0),$$

where

$$\tilde{H}(u, w_1, \ldots, w_m) = \phi(u) \otimes H(w_1 - u, \ldots, w_m - u), \quad \tilde{H}_0(w_1, \ldots, w_m) = \tilde{H}(0, w_1, \ldots, w_m),$$

and the tensor product is decomposed into the components in W_{ℓ_3} using the Clebsch-Gordan coefficients. Here we used $x_{j_s} - x_j = u_{j_s} - u_j$, and the fact that the term of $j = i$ in $S_i(\tilde{H})$ is $S_i(\tilde{H}_0)$. Recall that every equivariant polynomial of degree $\leq D$ in two groups of points is a linear combination of products of equivariant polynomials in each of the groups, decomposed using the Clebsch-Gordan coefficients, since the polynomials in both groups are the tensor product of the polynomials in each group. The same holds for ν groups of points, with the generalized Clebsch-Gordan coefficients of equation 4. Since $(u, w_1, \ldots, w_m) \mapsto (u, w_1 - u, \ldots, w_m - u)$ is an invertible linear map which commutes with the rotations, it follows that every equivariant polynomial $G : (\mathbb{R}^3)^{m+1} \to W_L$ of degree $\leq D$ is a linear combination of polynomials of the form $\tilde{H}$, and thus we obtain $S_i(G) - S_i(G_0)$ for every such G, where $G_0(w_1, \ldots, w_m) = G(0, w_1, \ldots, w_m)$. If G does not depend on u then $S_i(G) = nS_i(G_0)$, so we obtain $(n-1)S_i(G_0)$, and since $n \geq 2$ we obtain all m-point sums. It follows that every $(m+1)$-point sum is a linear combination of the features $A^{(t)}_i$. In equation 4 we obtain products of ν such sums, which are the $\nu(m+1)$-point sums $S_i(H_1 \otimes \cdots \otimes H_\nu)$, where H_ξ depends on the ξ-th group of $m+1$ points. Every equivariant polynomial of degree $\leq D$ in $\nu(m+1)$ points is a linear combination of such products, and therefore every μ_{t+1}-point sum is a linear combination of the features $h^{(t+1)}_i$.

Finally, let p be as in Proposition 25. By the proof of that proposition, p_i is a linear combination of products of at most $\min\{D, n-1\} \leq \mu_T$ sums of the form $\sum_{j\neq i}(x_j - x_i)^\alpha$ with $\alpha \neq 0$, with coefficients which do not depend on i. The term of $j = i$ is zero, so these sums are equal to $\sum_{j=1}^{n}(x_j - x_i)^\alpha$. Therefore $p_i = S_i(H)$ for a polynomial $H : (\mathbb{R}^3)^{\mu_T} \to \mathbb{R}$ of degree $\leq D$ which does not depend on i, and since p_i is rotation invariant we can choose H to be rotation invariant by averaging over the rotations. Thus p_i is a μ_T-point sum, so it is a linear combination of the features $h^{(T)}_{i,k00}$, and we choose the MLP to be this linear combination. In this construction all the polynomials can be chosen to be homogeneous, with a separate channel for every polynomial (if ϕ and H are homogeneous then $\tilde{H}$ is homogeneous of degree $\deg\phi + \deg H$, and $\tilde{H}_0 \neq 0$ only if ϕ is constant), so the assumptions of Lemma 28 hold, and therefore $\ell_{\max} = \lfloor D/2 \rfloor$ is sufficient. Proposition 8 follows as in the case of a single iteration. □

We now give an example which illustrates Proposition 29 for $\nu = 2$ and $T = 2$. For $n \geq 4$, the polynomial in this example cannot be expressed by a single iteration with $\nu_{\max} = 2$ and a linear readout as in Lemma 28.

Example 31. *Fix i and denote $u_j = x_j - x_i$. Let $v_i = \sum_{j\neq i} u_j$ and $M_i = \sum_{j\neq i} u_j u_j^T$, and consider the invariant polynomial*

$$q_i(X) = v_i^T M_i v_i = \sum_{j_1,j_2,j_3\neq i} \langle u_{j_1}, u_{j_2}\rangle\langle u_{j_2}, u_{j_3}\rangle.$$

A single iteration of MACE with $\nu_{\max} = 3$ can express q_i, since it is a linear combination of products of three of the features $a_{i,k\ell m}$. For $n \geq 4$, a single iteration with $\nu_{\max} = 2$ in which E_i is a linear combination of the features $m_{i,k00}$ cannot express q_i: these features are sums of functions which depend on at most two of the points u_j, while q_i contains the monomials of $\langle u_{j_1}, u_{j_2}\rangle\langle u_{j_2}, u_{j_3}\rangle$ with j_1, j_2, j_3 distinct.

We now show that two iterations with $\nu_{\max} = 2$ and $\ell_{\max} = 2$ can express q_i. In the first iteration we compute for every point x_j the feature $v_j = \sum_{j'\neq j}(x_{j'} - x_j)$, using $\nu = 1$, $\ell = 1$ and $R(r) = r$, and we keep the constant feature $h_j^{(0)} = 1$ using $\tilde{U}^{(0)}$. Since the summation is over all points, we have

$$v_j = \sum_{j'=1}^{n}(x_{j'} - x_i) - n(x_j - x_i) = v_i - nu_j.$$

Therefore in the second iteration, choosing $\ell_1 = \ell_2 = 1$ and $R(r) = r$ in equation 12, we obtain the entries of

$$\sum_{j\neq i} u_j \otimes v_j = v_i \otimes v_i - nM_i,$$

and choosing $\ell_2 = 0$, $\ell_1 \in \{0, 2\}$ and $R(r) = r^2$ we obtain the entries of M_i. Using $\nu = 2$ we obtain

$$\langle M_i, v_i \otimes v_i - nM_i\rangle + n\langle M_i, M_i\rangle = v_i^T M_i v_i = q_i(X).$$

Remark 32. *The proof of Proposition 29 uses the fact that the summation is over all points. For $t \geq 1$ the features $A_i^{(t)}$ in equation 12 are computed from the hidden states $h_j^{(t)}$ of the other points, while $h_i^{(t)}$ itself only appears in the linear term of equation 13, so the sums from the previous iterations are centered at the points x_j. In the proof we rewrote them as sums centered at x_i. When the summation in equation 12 is only over the points at distance at most ρ from x_i, in general the neighbors of x_j are different from the neighbors of x_i, and this argument does not apply.*

E LOCAL MODELS: PROOF OF THEOREM 11

Here we prove Theorem 11. We begin by restating the theorem.

Theorem 11. *Let $K \leq n$ be natural numbers and $\rho > 0$, and let $G = SE(3) \times S_n$. Set*

$$D(j) = 18j^2 - 18j + 9, \qquad D = D(K).$$

Then local MACE with cutoff ρ, $T \geq 1$ iterations, $\nu_{\max} = K - 1$ and $\ell_{\max} = \lfloor D/2\rfloor$, is $(G, \mathcal{R}^S_{\rho,K})$ complete.

E.1 PROOF OF THEOREM 11

The equations for a single MACE iteration are those given above, with the only difference that the summation in equation 3 is over the points at distance at most ρ from x_i rather than over all points. In this subsection we call this version MACE, or local MACE, and refer to the option of summing over all points as global MACE. The theorem rests on the following lemma.

Lemma 33. *There exists an instantiation of MACE with $T = 1$, $\nu_{\max} = K - 1$ and $\ell_{\max} = \lfloor D/2\rfloor$, where $D = D(K) = 18K^2 - 18K + 9$, such that, if X, Y are (ρ, K) regular, then*

1. *For all $1 \leq i, k \leq n$, if $E_i(X) = E_k(Y)$ then there exists a rotation $R \in SO(d)$, such that*
$$\{x_j - x_i| \quad \|x_j - x_i\| \leq \rho\} = \{R(y_j - y_k)| \quad \|y_j - y_k\| \leq \rho\}.$$
2. *If $E(X) = E(Y)$ then the node features of X and Y are equal, up to permutation, namely we have the multiset equality*
$$\{E_i(X), i = 1, \ldots, n\} = \{E_i(Y), i = 1, \ldots, n\}$$

Proof of Lemma 33. To prove the first part of the lemma, we note that if we remove from X and Y the points which are at distance more than ρ from x_i and y_k, this does not influence the features $E_i(X)$ and $E_k(Y)$ obtained from the first iteration of MACE. Moreover, the feature $A_{i,000}$ with $R_{00} = 1$ counts the number of points, so point clouds with a different number of points are separated. Accordingly, it is sufficient to consider point clouds X, Y which have the same number of points $n \leq K$ (and $n \geq d+1$ by (O) together with the overlap requirement in the definition of ρ-adjacency), all of them at distance at most ρ from x_i and y_k respectively. To prove this, it is sufficient to show that we can choose an instantiation of global MACE such that, if there exists some point clouds $X, Y \in \mathbb{R}^{d\times n}$ and $i, k \in [n]$ such that

$$\{x_j - x_i| \quad j \in [n]\} \neq \{R(y_j - y_k)| \quad j \in [n]\}, \quad \forall R \in SO(d),$$

then $E_i(X) \neq E_k(Y)$. This follows from our results on global MACE and invariant polynomials.

Indeed, by Theorem 4 there exists an $SO(d) \times S_n$ invariant polynomial $\phi : \mathbb{R}^{d\times n} \to \mathbb{R}$ of degree $\leq D(n) \leq D(K)$ such that

$$\phi(x_1 - x_i, \ldots, x_n - x_i) \neq \phi(y_1 - y_k, \ldots, y_n - y_k).$$

Using this, we can define a permutation equivariant and rotation and translation invariant polynomial $p : \mathbb{R}^{d\times n} \to \mathbb{R}^n$ whose coordinates are given by

$$p_i(Z) = \phi(z_1 - z_i, \ldots, z_n - z_i)$$

By Proposition 25, there exists a global MACE model with the prescribed size, such that $p(Z) = (E_1(Z), \ldots, E_n(Z))$, and then

$$E_i(X) = \phi(x_1 - x_i, \ldots, x_n - x_i) \neq \phi(y_1 - y_k, \ldots, y_n - y_k) = E_k(Y)$$

This concludes the proof of the first part of the claim (note that we in fact proved pairwise separation by a MACE architecture, but since the polynomials of bounded degree have a finite space, the extension to global separation is immediate. The same architecture can be used for all $n \leq K$: we choose the node feature E_i to be the vector of all the invariant features $B_{i,\eta_\nu \boldsymbol{k}00}$ with $\nu \leq K - 1$ from the proof of Proposition 25, which include the feature $A_{i,000}$ above, using a separate channel in equation 5 for every η_ν. Then by equation 10 and equation 11 all the polynomials p_i above are linear combinations of the coordinates of E_i).

The second part of the claim follows from the fact that there exist MLPs such that the mapping $(E_1, \ldots, E_n) \mapsto \sum_{i=1}^n \mathrm{MLP}(E_i)$ is injective as a function of multisets (see Amir et al. (2023)). □

Proof of Theorem 11. We choose the MACE architecture that satisfies the conditions of Lemma 33. For $T > 1$ iterations, by Proposition 27 there exists a MACE architecture with T iterations which has the same node features, so it is sufficient to consider $T = 1$. It remains to show that these conditions imply completeness on the subset of strongly regular point clouds. To do this, it is sufficient to show that, if we are given a multiset of node features

$$\{E_1, \ldots, E_n\}$$

obtained from applying MACE to some strongly regular X, then we can uniquely reconstruct X, up to an $SO(d) \times S_n$ transformation. The lemma immediately allows us to reconstruct from each feature E_i, a subset $X[i]$ of $\leq K$ point of X, up to rigid motion. What remains to show is that there is a unique reconstruction up to a single rigid motion.

We perform this reconstruction as follows: at a first step, we choose some arbitrary node feature and name it E_1. We choose some arbitrary orientation for $X[1]$, the subset inferred from E_1. It now remains to show that the remainder of X is uniquely defined by this choice. This is done iteratively:

First, we go over all E_j: from each E_j we can reconstruct $X[j]$ up to rigid motion. We can then check whether $X[j]$ has a subset of cardinality $d+1$ which is congruent to a subset of $d+1$ points in $X[1]$, which is a finite combinatorial search problem. If so, then by strong regularity these congruent points are in fact identical, and so we know their position. Once these points are fixed, we can deduce where all of the $X[j]$ points are, since by assumption these $d + 1$ points are affine independent, and we know $X[j]$ up to rigid motion. By the overlap requirement in the definition of ρ-adjacency, the reconstruction step will succeed for all j which are adjacent to 1.

In the next step, we repeat this, now comparing each not yet reconstructed set $X[j]$ with all the sets which were already reconstructed. We continue iteratively. Since we assume that the ρ-adjacency graph is connected, this procedure will terminate with the full reconstruction of X. □

E.2 LOWER BOUNDS ON HYPERPARAMETERS

Throughout this subsection a *local MACE model* with cutoff ρ, depth T and hyperparameters $\ell_{\max}, \nu_{\max}, \nu_{\mathrm{read}}$ is the model equation 3–equation 7 iterated T times, with the sum in equation 3 restricted to $x_j \in B(x,\rho) \cap X$ for $x = x_i$ and the neighbour's feature of the previous iteration inserted in it: the feature vector of any $x \in X$ at step t is denoted $h^{(t)}(x)$ and lies in a finite sum of copies of $W_0 \oplus \cdots \oplus W_{\ell_{\max}}$, one iteration is a polynomial of degree at most $\nu_{\max}$ in the aggregated features A, the site output is an invariant polynomial of degree at most ν_{read} in $h^{(T)}(x)$, and the cloud output is an arbitrary function of the multiset of site outputs. The radial functions $R_{k\ell}$ are arbitrary continuous functions, the number of channels is unrestricted, and arbitrary continuous functions may be applied to invariant channels at any stage, so that the MLPs of equation 6 are covered.

The following family is the local form of the counter-rotating polygons used to prove Theorem 1. Its whole point is that all the distances a model can read are constant along it, so that the radial functions carry no information whatsoever and only the angular resolution of the model is at stake.

Definition 34. *Let $p, q \geq 3$. For $\alpha, \beta \in \mathbb{R}$ we define $Y(\alpha,\beta) \subset \mathbb{R}^3$ to be the set containing the origin c together with*

$$a_j = \Big(\tfrac{9}{20}\cos\big(\tfrac{2\pi j}{p}+\alpha\big),\ \tfrac{9}{20}\sin\big(\tfrac{2\pi j}{p}+\alpha\big),\ -\tfrac{3}{5}\Big), \quad b_k = \Big(\tfrac{7}{20}\cos\big(\tfrac{2\pi k}{q}+\beta\big),\ \tfrac{7}{20}\sin\big(\tfrac{2\pi k}{q}+\beta\big),\ \tfrac{3}{5}\Big),$$

for $j \leq p$ and $k \leq q$, and let $\mathcal{Y}_{p,q} := \{Y(\alpha,\beta) :\ \alpha,\beta \in \mathbb{R}\}$, a family of clouds of $n = p+q+1$ points.

Lemma 35. *Every $Y \in \mathcal{Y}_{p,q}$ is $(1,n)$-regular. Its cutoff balls are $\mathcal{B}_1^Y(c) = Y$, $\mathcal{B}_1^Y(a_j) = \{c\} \cup \{a_1,\dots,a_p\}$ and $\mathcal{B}_1^Y(b_k) = \{c\} \cup \{b_1,\dots,b_q\}$. Moreover $Y(\alpha,\beta)$ and $Y(\alpha',\beta')$ lie in the same $SE(3)\times S_n$-orbit if and only if $\alpha - \beta \equiv \alpha' - \beta' \pmod{2\pi/\mathrm{lcm}(p,q)}$, so $\mathcal{Y}_{p,q}$ meets infinitely many orbits.*

Proof. We note that for all choices of indices j, j', k, k' there holds $\|c - a_j\| = \sqrt{(9/20)^2 + (3/5)^2} = 3/4$, $\|c - b_k\| = \sqrt{(7/20)^2 + (3/5)^2} < 3/4$, $\|a_j - a_{j'}\| \leq 9/10$, $\|b_k - b_{k'}\| \leq 7/10$ and $\|a_j - b_k\| \geq 6/5$. This gives the three cutoff balls $\mathcal{B}_1^Y(\cdot)$ as claimed, and (D), (K) with $K = n$ are immediate. For (O), the ball of a_j meets the ball of c in $\{c, a_1, \dots, a_p\}$, which contains three non-collinear points of the plane $z = -3/5$ together with c, hence four affinely independent points, so a_j and c are ρ-adjacent; the same holds for each b_k, and the adjacency graph is connected.

For the second claim, c is the only point of Y at distance $3/4$ from p points of a common plane and $< 3/4$ from q points of another, so any isometry carrying $Y(\alpha,\beta)$ to $Y(\alpha',\beta')$ fixes c and, the two rings having different cardinalities, radii and heights, maps each ring to the corresponding one. It therefore preserves the axis and each ring plane and, being orientation preserving, is the rotation by some angle s about the axis, denoted g_s. Since the ring phases are defined modulo $2\pi/p$ and $2\pi/q$, this says $\alpha' - \alpha \in s + \frac{2\pi}{p}\mathbb{Z}$ and $\beta' - \beta \in s + \frac{2\pi}{q}\mathbb{Z}$, whence $(\alpha'-\beta') - (\alpha-\beta) \in \frac{2\pi}{p}\mathbb{Z} + \frac{2\pi}{q}\mathbb{Z} = \frac{2\pi}{\mathrm{lcm}(p,q)}\mathbb{Z}$. The converse is clear. □

Lemma 36 (frequency budget)**.** *Let a local MACE model be applied to elements of $\mathcal{Y}_{p,q}$, and suppose that the continuous functions it applies to invariant channels are applied only to channels that are constant on $\mathcal{Y}_{p,q}$; by the proof of Theorem 37, this holds automatically whenever $\ell_{\max}\sum_{i=1}^{T}\nu_{\max}^i < 2\min(p,q)$. Then every component of every $h^{(t)}(x)$ is a trigonometric polynomial in (α,β) with Fourier support contained in $\{(m_1,m_2) \in \mathbb{Z}^2 : |m_1| + |m_2| \leq \omega_t\}$, where $\omega_0 = 0$ and $\omega_t = \nu_{\max}(\ell_{\max} + \omega_{t-1})$, so that $\omega_T = \ell_{\max}\sum_{i=1}^{T}\nu_{\max}^i$; and the site outputs have Fourier support in $|m_1| + |m_2| \leq \nu_{\mathrm{read}}\,\omega_T$.*

Proof. By Lemma 35 every pair of points within the cutoff consists of c and a ring point or of two points of the same ring. The length of such a pair does not depend on (α,β), so every $R_{k\ell}$ in equation 3 contributes a constant. Its direction is $g_\alpha \hat{u}$ for a fixed $\hat{u}$ (and where g_s denotes the rotation about the axis of $Y(\alpha,\beta)$ by angle s) when both points belong to ring A or to $\{c\}\cup$ring A, because increasing α by s rotates that part of the cloud rigidly by g_s; and it is $g_\beta\hat{u}$ in the two symmetric cases. Since $Y_\ell^m(g_\alpha\hat{u}) = e^{\mathrm{i}m\alpha}Y_\ell^m(\hat{u})$ for the rotation about the axis, each such factor

has Fourier support $|m_1| \le \ell \le \ell_{\max}, m_2 = 0$ or $|m_2| \le \ell \le \ell_{\max}, m_1 = 0$. Hence each summand of equation 3, being a product of a constant, such a factor and a feature of the previous iteration, has support bounded by $\ell_{\max} + \omega_{t-1}$, and the same holds for the total sum of these terms. Passing to equation 4 multiplies at most $\nu_{\max}$ such objects and supports add under multiplication, giving $\omega_t \le \nu_{\max}(\ell_{\max} + \omega_{t-1})$; equation 5 is linear, and by assumption the continuous functions applied to invariant channels only produce constants. The site output is an invariant polynomial of degree at most ν_{read} in $h^{(T)}(x)$, so the same argument bounds its support by $\nu_{\text{read}}\omega_T$. □

The site outputs are constrained by the symmetries of the family in a way that depends on the site. Write $s_x(\alpha, \beta)$ for the output at the site x. Increasing β by $2\pi/q$ leaves the cloud unchanged and fixes c and every a_j, so s_c and s_{a_j} are $2\pi/q$-periodic in β; increasing α by $2\pi/p$ leaves the cloud unchanged and fixes c but carries a_j to a_{j+1}, so s_c is $2\pi/p$-periodic in α whereas $s_{a_j}(\alpha+2\pi/p, \beta) = s_{a_{j+1}}(\alpha, \beta)$: the multiset of ring outputs is periodic, an individual ring output is not. Finally every site output is unchanged by the rotation g_s, which carries $Y(\alpha, \beta)$ to $Y(\alpha + s, \beta + s)$.

Theorem 37. *Let $p, q \ge 3$. If a local MACE model separates $\mathcal{Y}_{p,q}$ then*

$$\nu_{\text{read}} \cdot \ell_{\max} \cdot \sum_{i=1}^{T} \nu_{\max}^{i} \;\ge\; 2\min(p, q),$$

and if its readout is that of equation 6–equation 7, an arbitrary function of the invariant channels at each point followed by a sum, then already $\ell_{\max} \sum_{i=1}^{T} \nu_{\max}^{i} \ge 2\min(p, q)$. Consequently, on $\mathcal{R}_{\rho,K}$ with $K \ge 7$, completeness requires $\nu_{\text{read}}\ell_{\max} \sum_{i \le T} \nu_{\max}^{i} \ge 2\lfloor (K-1)/2 \rfloor$, and $\ell_{\max} \sum_{i \le T} \nu_{\max}^{i} \ge 2\lfloor (K-1)/2 \rfloor$ for the readout of equation 6.

Proof. Both bounds hold if $\omega_T = \ell_{\max} \sum_{i=1}^{T} \nu_{\max}^{i} \ge 2\min(p, q)$, since $\nu_{\text{read}} \ge 1$, so we may assume that $\omega_T < 2\min(p, q)$. We first check that Lemma 36 then applies. An invariant channel of $h^{(t)}(x)$ is invariant under the rotation g_s and under the shifts of α and β that fix x, exactly as the site outputs above, so, by the argument below, it is constant unless one of its modes has $|m_1| + |m_2| \ge 2\min(p, q)$. By induction on $t \le T$, the continuous functions applied in the first t iterations act on constant channels, so the invariant channels of $h^{(t)}(x)$ have support in $|m_1|+|m_2| \le \omega_t \le \omega_T < 2\min(p, q)$ and are therefore constant as well. Hence the hypothesis of Lemma 36 holds.

If every site output were constant on $\mathcal{Y}_{p,q}$ the cloud output would be constant too, against Lemma 35. Let s_x be a non-constant site output and $(m_1, m_2) \ne (0, 0)$ a mode in its support. Rotation invariance gives $m_1 + m_2 = 0$ at every site. If $x = a_j$, periodicity in β gives $m_2 \equiv 0 \pmod q$, so $|m_1| + |m_2| = 2|m_2| \ge 2q$; if $x = b_k$, symmetrically $|m_1| + |m_2| \ge 2p$; and if $x = c$, both periodicities apply and $m_1 = -m_2$ is a multiple of $\operatorname{lcm}(p, q)$. In all cases $|m_1| + |m_2| \ge 2\min(p, q)$, and Lemma 36 gives the first bound. For the readout equation 6–equation 7, the site output is a function of the invariant channels of $h^{(T)}(x)$, which have support bounded by ω_T and are therefore all constant unless $\omega_T \ge 2\min(p, q)$; a function of constants is constant. For the consequences, Lemma 35 places $\mathcal{Y}_{p,q}$ inside $\mathcal{R}_{\rho,p+q+1}$; take $p = q = (K-1)/2$ if K is odd and $p, q = (K-2)/2,\ K/2$ if K is even, so that $2\min(p, q) = 2\lfloor (K-1)/2 \rfloor$. □

Remark 38 (why the bound is linear)**.** *Along $\mathcal{Y}_{p,q}$ the orbit invariant $\alpha - \beta$ is defined modulo $2\pi/\operatorname{lcm}(p, q)$, and the output at the centre c is accordingly constant unless its frequency reaches $\operatorname{lcm}(p, q)$, quadratic in K for p, q coprime. This does not give a quadratic bound on the model, because the outputs at the ring points are not subject to the α-periodicity individually, only as a multiset, and a nonlinear readout summed over a ring that the symmetry permutes turns a frequency into a multiple of it: for coprime p, q,*

$$\sum_{j=1}^{p} \cos^{p}\!\big(q(\alpha - \beta) + \tfrac{2\pi q j}{p}\big) = \tfrac{p}{2^{\,p-1}} \cos\!\big(pq(\alpha - \beta)\big) + \textit{const}.$$

The sum of per-point MLPs in equation 6–equation 7 does exactly this, so the quadratic bound of Theorem 1 does not transfer to MACE; only the linear bound of Theorem 37, which rests on the periodicity that does fix the ring points, does.

Corollary 39. *A local MACE model whose depth and body orders are bounded independently of K cannot be complete on $\mathcal{R}_{\rho,K}$ with an irreducible degree bounded independently of K: one needs $\ell_{\max} = \Omega(K)$.*

Remark 40 (the range of K)**.** *Two rings of fewer than three points cannot be used in Definition 34: a ring point must be ρ-adjacent to c, which requires four affinely independent points among the points it sees, while its ring seen together with points of the axis spans a plane when the ring has two points and a line when it has one; points of the other ring cannot be added without making the distances depend on $\alpha - \beta$. This is why the family starts at $K = 7$. On the other hand, no counterexample can have $K \leq 3$: if $K \leq 3$, a ball of radius ρ never contains four affinely independent points of the cloud, so no two points are ρ-adjacent, and* (O) *forces the cloud to consist of a single point. Thus the only values of K not covered by the rings are $K = 4, 5, 6$, and the following lemma covers them, and every larger K, for the models that cannot produce a pseudoscalar.*

Lemma 41 (parity)**.** *Consider MACE with the readout equation 6–equation 7 and T iterations, and suppose that $\ell_{\max} = 0$, or $T = 1$ and $\nu_{\max} \leq 2$, or $T = 2$ and $\nu_{\max} = 1$. Then every invariant channel $h^{(T)}_{i,k00}$, and hence the output of the model, takes the same value on X and on $-X$. Consequently such a model is not $(SE(3) \times S_n, \mathcal{R}_{\rho,K})$-complete on clouds of K points for any $K \geq 4$.*

Proof. Under $X \mapsto -X$ every relative vector changes sign and every distance is unchanged, so a product of spherical harmonics $Y_{\ell_1} \cdots Y_{\ell_s}$ of relative directions, weighted by radial functions, is multiplied by $(-1)^{\ell_1+\cdots+\ell_s}$. Call a coordinate of type W_L *natural* if it is multiplied by $(-1)^L$. The channel mixings in equation 4, equation 5, equation 12 and equation 13 are linear and preserve the type, so they preserve naturality. If $\ell_{\max} = 0$ every feature is a function of distances and there is nothing to prove. If $T = 1$, the coordinates of $A^{(0)}$ are natural, and a coordinate of type W_0 of $B^{(0)}$ is either a coordinate of $A^{(0)}$ of type W_0, when $\nu = 1$, or, when $\nu = 2$, a Clebsch–Gordan contraction of two coordinates of the same type W_ℓ, which is multiplied by $(-1)^{2\ell} = 1$; hence every $h^{(1)}_{i,k00}$ is inversion invariant. If $T = 2$ and $\nu_{\max} = 1$, then $B^{(0)} = A^{(0)}$, so every coordinate of $h^{(1)}$ is natural; in equation 12 a coordinate of type W_0 of $A^{(1)}$ contracts Y_{ℓ_1} with a coordinate of $h^{(1)}$ of type W_{ℓ_2}, which requires $\ell_1 = \ell_2$ and is therefore multiplied by $(-1)^{2\ell_1} = 1$; since $B^{(1)} = A^{(1)}$, every $h^{(2)}_{i,k00}$ is inversion invariant.

For the consequence, let X consist of $K \geq 4$ points, not all in a plane, with pairwise distinct mutual distances all at most ρ. Then X and $-X$ satisfy (D), (K) and (O), every ball around a point containing all K points, so both lie in $\mathcal{R}_{\rho,K}$. If $g(-X)$ were a relabelling of X for some $g \in SE(3)$, the isometry $x \mapsto g(-x)$ would map X onto itself preserving all distances; as these are distinct it would fix every point of X, hence, fixing four affinely independent points, it would be the identity, which is impossible since its linear part has determinant -1. So X and $-X$ lie in different orbits and receive the same output. □

The hypotheses of Lemma 41 are sharp for MACE as defined in Subsection D.3: with $\ell_{\max} \geq 1$, a pseudoscalar arises for $T = 1$ and $\nu_{\max} = 3$ from the triple product $1 \otimes 1 \otimes 1 \to 0$, for $T = 2$ and $\nu_{\max} = 2$ from the cross product $1 \otimes 1 \to 1$ in the first iteration followed by the contraction with Y_1 in equation 12, and for $T = 3$ and $\nu_{\max} = 1$ from two consecutive couplings in equation 12. With the parity selection rule discussed in Subsection D.2, instead, every feature carries a parity ± 1, by which it is multiplied under $X \mapsto -X$, and the invariant channels have parity $+1$, so the conclusion of Lemma 41 holds for all $\ell_{\max}$, $\nu_{\max}$ and T.

Remark 42 (small counterexamples in the literature)**.** *Known degeneracies of low body order are local and do not improve on these values. The three-body descriptors of an atomic environment are complete for at most three neighbors and have degenerate pairs as soon as there are four (Pozdnyakov et al., 2020), and degeneracies of the four-body descriptors are known for environments of seven neighbors (Nigam et al., 2024). These are statements about the environment of one atom, and they are mitigated when, as in equation 7, the output sums contributions from every atom; the three-body case is in any event subsumed by Lemma 41, which already gives $K = 4$, the least possible value.*

Corollary 43. *Let MACE with the readout equation 6 be $(SE(3) \times S_n, \mathcal{R}_{\rho,K})$-complete on clouds of K points. Then*

$$\ell_{\max} \;\geq\; \frac{2\lfloor (K-1)/2 \rfloor}{\nu_{\max}} \;\; (T=1), \qquad \ell_{\max} \;\geq\; \frac{2\lfloor (K-1)/2 \rfloor}{\nu_{\max}(1+\nu_{\max})} \;\; (T=2),$$

$$\ell_{\max} \;\geq\; \frac{2\lfloor (K-1)/2 \rfloor}{\nu_{\max}(1+\nu_{\max}+\nu_{\max}^2)} \;\; (T=3).$$

Equivalently, combining Theorem 37 with Lemma 41, the model fails to be complete as soon as $K \geq 4$ if it satisfies the hypotheses of Lemma 41, and otherwise as soon as

$$K \;\geq\; \max\Big\{7,\; 2\Big\lfloor \tfrac{1}{2}\,\ell_{\max} \textstyle\sum_{i=1}^{T} \nu_{\max}^{i} \Big\rfloor + 3\Big\};$$

these thresholds are listed in Table 1 for $0 \leq \ell_{\max}, \nu_{\max} \leq 4$.

$\ell_{\max} \setminus \nu_{\max}$	1	2	3	4
1	$4^{\dagger}/4^{*}/\mathbf{7}$	$4^{\dagger}/\mathbf{9}/\mathbf{17}$	$\mathbf{7}/\mathbf{15}/\mathbf{41}$	$\mathbf{7}/\mathbf{23}/\mathbf{87}$
2	$4^{\dagger}/4^{*}/\mathbf{9}$	$4^{\dagger}/\mathbf{15}/\mathbf{31}$	$\mathbf{9}/\mathbf{27}/\mathbf{81}$	$\mathbf{11}/\mathbf{43}/\mathbf{171}$
3	$4^{\dagger}/4^{*}/\mathbf{11}$	$4^{\dagger}/\mathbf{21}/\mathbf{45}$	$\mathbf{11}/\mathbf{39}/\mathbf{119}$	$\mathbf{15}/\mathbf{63}/\mathbf{255}$
4	$4^{\dagger}/4^{*}/\mathbf{15}$	$4^{\dagger}/\mathbf{27}/\mathbf{59}$	$\mathbf{15}/\mathbf{51}/\mathbf{159}$	$\mathbf{19}/\mathbf{83}/\mathbf{339}$

Table 1: Smallest number K of points per cutoff ball from which local MACE with the readout equation 6–equation 7, irreducible degree $\ell_{\max}$ (rows), correlation order $\nu_{\max}$ (columns) and T iterations is provably not $(SE(3) \times S_n, \mathcal{R}_{\rho,K})$-complete, by Corollary 43. Each cell lists the thresholds for $\mathbf{T = 1 / T = 2 / T = 3}$, in this order. All entries equal to 4 come from Lemma 41: the invariant channels of the model are unchanged under $X \mapsto -X$, so it cannot tell a chiral cloud from its mirror image; 4 is the least possible value, since $\mathcal{R}_{\rho,K}$ contains only one-point clouds when $K \leq 3$ (Remark 40). **Entries marked $4^{\dagger}$ were known before this work**: with $T = 1$ the invariant channels are functions of distances ($\nu_{\max} = 1$) or three-body invariants ($\nu_{\max} = 2$), which are classically inversion invariant. **Entries marked 4^{*} rest on the same known mechanism**, but the observation that two iterations with $\nu_{\max} = 1$ still produce no pseudoscalar is stated here. **Entries in bold are new**: they come from the two-ring family of Definition 34 and Theorem 37, apply to models that can produce pseudoscalars, and grow linearly in $\ell_{\max}$ and like $\nu_{\max}^{T}$ in $\nu_{\max}$.

Remark 44. *Lower bounds of this kind in the learning literature are for one fixed order at a time: the counterexamples of Pozdnyakov & Ceriotti (2022) exhibit pairs of clouds that a given body order cannot tell apart. The rings add a bound that grows with the size of the neighbourhood, by the mechanism of Theorem 1 with the azimuthal frequency in the role of the polynomial degree. Theorem 11 spends an irreducible degree quadratic in K on one iteration; whether this can be brought down to linear is open.*

F Invariant Theory Background

Suppose that G is a group that acts a vector space V. In Invariant Theory one studies polynomial functions on V that are constant on orbits. The origins of Invariant Theory go back to the 19^{th} century. Historically, one often works over $\mathbb{C}$, but many results have been extended to arbitrary fields, including fields with positive characteristic. For our applications we will work over $\mathbb{R}$. Suppose V is an n-dimensional $\mathbb{R}$-vector space. We denote the ring of all polynomial functions $V \to \mathbb{R}$ by $\mathbb{R}[V]$. If $x_1, x_2, \ldots, x_n$ are linearly independent coordinate functions on V, then we can identify V with $\mathbb{R}^n$ and $\mathbb{R}[V]$ is just the polynomial ring $\mathbb{R}[x_1, x_2, \ldots, x_n]$. Suppose that a group G acts on V. For $v \in V$, let $[v] = \{g \cdot v \mid g \in G\}$ be the G-orbit of v. A polynomial invariant is a polynomial function $V \to \mathbb{R}$ that is constant on orbits. The invariant ring $\mathbb{R}[V]^G$ is the set of all polynomial invariants.

Suppose that G is a Lie group and $\rho : G \to \mathrm{GL}(V)$ is an orthogonal representation, i.e., ρ is a continuous homomorphism of groups. Then G acts on V via $g \cdot v = \rho(g)v$ for all $v \in V$ and $g \in G$. If G is compact, then one can choose a basis of V such that $\rho(G)$ is contained in the orthogonal group $\mathrm{O}(n)$.

Theorem 45 (Hilbert (1890))**.** *If G is compact then invariant ring $\mathbb{R}[V]^G$ is generated by finitely many invariants $f_1, f_2, \ldots, f_r$. So every invariant polynomial is of the form $p(f_1, f_2, \ldots, f_r)$ where $p \in \mathbb{R}[y_1, y_2, \ldots, y_r]$ is a polynomial with coefficients in $\mathbb{R}$.*

If the group G is not compact, then the invariant ring may not be finitely generated. The following well-known result can be found in (Onishchik & Vinberg, 1990, Ch. 3, §4, 3°, Thm. 3):

Theorem 46. *Suppose that G is compact, $u, v \in V$ and $[u] \neq [v]$. Then there exists an invariant $f \in \mathbb{C}[V]^G$ with $f(u) \neq f(v)$. If $f_1, f_2, \ldots, f_r$ are the generators from Theorem 45, then $f_i(u) \neq f_i(v)$ for some i.*

We next recollect the definition of completeness from the main text:

Definition 47. *A set $\mathcal{Q} \subseteq \mathbb{R}[V]^G$ is called (G, V)-complete if*

$$\forall q \in \mathcal{Q}\ q(u) = q(v) \quad \Leftrightarrow \quad [u] = [v].$$

Theorem 46 states the sets $\mathbb{C}[V]^G$ and the set $\{f_1, f_2, \ldots, f_r\}$ of generates are both (G, V)-complete.

If G is not compact, then a (G, V)-complete set of invariants may not exist. Take for example the action of $\mathrm{GL}(V)$ on V where $n \geq 1$. There are two orbits in V, namely $\{0\}$ and $V \setminus \{0\}$. Every polynomial invariant is constant on $V \setminus \{0\}$ and by continuity also constant on V. So no polynomial invariant can separate the two orbits. To take into account this situation, we recall the general notion of separating invariants in Invariant Theory:

Definition 48. *A set $\mathcal{Q} \subseteq \mathbb{R}[V]^G$ is called separating if for all $u, v \in V$ we have*

$$\forall q \in \mathcal{Q}\ q(u) = q(v) \quad \Leftrightarrow \quad \forall f \in \mathbb{R}[V]^G\ f(u) = f(v).$$

For compact G, the notion of completeness and separation are equivalent.

Tautologically, $\mathbb{R}[V]^G$ is separating, but it may not be finitely generated. However, we have:

Theorem 49. *There exists a finite set of separating invariants.*

A proof is given in (Derksen & Kemper, 2015, Theorem 2.4.8) (the assumption that the base field is algebraically closed is not necessary there), or see (Kemper, 2009, Theorem 12) for a more general statement. Note that if the set $\mathbb{R}[V]^G$ of all invariants is (G, V)-complete, then any set of separating invariants is (G, V)-complete. In particular, there exists a finite (G, V)-complete set of invariants in that case.

The separating degree $D(G, V)$ is the smallest nonnegative integer d such that the space $K[V]^G_{\leq d}$ of invariants of degree $\leq d$ is a separating set for the action of G on V. In this manuscript we work over $K = \mathbb{R}$. Complexification is introduced below when discussing bounds on the degrees of generators.

If any (G, V)-complete set of invariant polynomials exists, then $D(G, V)$ is also the smallest non-negative integer for which $K[V]^G_{\leq d}$ is (G, V)-complete. We have $D(G, V) \leq \beta(G, V)$ where $\beta(G, V)$ is the smallest nonnegative integer d such that $K[V]^G$ is generated by $K[V]^G_{\leq d}$. Upper bounds for $\beta(G, V)$ have been studied extensively. If G is a compact Lie group and $\rho : G \to \mathrm{GL}(n, \mathbb{R})$ is a real representation, then ρ extends to a complex representation $\rho_{\mathbb{C}} : G_{\mathbb{C}} \to \mathrm{GL}(n, \mathbb{C})$, where $G_{\mathbb{C}}$ is the complexification of G. Now $G_{\mathbb{C}}$ is a complex reductive group. We have $\beta(G, \mathbb{R}^n) = \beta(G_{\mathbb{C}}, \mathbb{C}^n)$ and upper bounds are discussed in (Derksen & Kemper, 2015, §3.2, §3.10, §4.7). Emmy Noether proved the bound $\beta(G, V) \leq |G|$ for finite groups G (Noether (1916)). For permutation actions one has upper bounds for $\beta(G, V)$ that often are sharper. Suppose that G acts on $V = \mathbb{R}^n$ by permuting the basis vectors $e_1, e_2, \ldots, e_n$ of $V = \mathbb{R}^n$. In this case we have

$$\beta(G, V) \leq \max\left\{n_1, n_2, \ldots, n_r, \binom{n_1}{2} + \binom{n_2}{2} + \cdots + \binom{n_r}{2}\right\} \leq \max\left\{n, \binom{n}{2}\right\}, \quad (14)$$

where $n_1, n_2, \ldots, n_r$ are the cardinalities of the orbits for the action of G on $\{e_1, e_2, \ldots, e_n\}$. This follows for example from Göbel's algorithm (see Göbel (1995)).

For weighted graphs with n vertices we consider the action of the symmetric group S_n on the space $\mathcal{H}_n \subseteq \mathbb{R}^{n \times n}$ of symmetric $n \times n$ matrices by conjugation. Thiéry gave in Thiéry (2000) the following upper bound

$$D(S_n, \mathcal{H}_n) \leq \beta(S_n, \mathcal{H}_n) \leq \binom{\binom{n}{2}}{2} - \mu_n = \frac{n^4}{8} + O(n^3),$$

where $\mu_n = 0$ if n is even and $\mu_n = \lceil \frac{3}{4}(n-1) \rceil$ if n is odd. This bound is slightly sharper than the one obtained from (14) with $n_1 = \binom{n}{2}$ and $n_2 = n$. In this paper we improve the upper bound for $D(S_n, \mathcal{H}_n)$ to

$$D(S_n, \mathcal{H}_n) \leq \frac{5n^3}{2} - n^2 - \frac{n}{2} = O(n^3)$$

and also provide a lower bound proportional to n^2.