CONTINUOUS EQUIVARIANT LEARNING WITH LINEARLY MANY FRAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Weighted frame averaging makes an ordinary network equivariant by evaluating it in several input-dependent coordinate systems. How many frames are needed to preserve continuity? For n labeled vectors spanning $\mathbb{R}^d$, with $d \geq 2$ fixed, the minimum worst-case support of a continuous $\mathrm{SO}(d)$- or $\mathrm{O}(d)$-equivariant weighted frame grows linearly in n. The lower bound follows from characteristic classes of Grassmannians. An explicit family of $d(n-d)+1$ probes, adapted from Wronskian rank condensers, gives the upper bound. Determinant weights of order $2k$ preserve C^{2k-1} regularity across probe failures, and this guarantee is sharp. For planar rotations the minimum support is exactly $n-1$. In synthetic force-learning experiments, disagreement between transported predictions produces jumps under hard frame selection, while continuous averaging removes them. At 24 particles and a common update budget, trained Wronskian hard selection reduces force error by 42.2% relative to a Gram-matrix model; averaging over 61 probes reduces the hard model's error by a further 21.4%, using 6.5 times its recorded worker time.

1 INTRODUCTION

Equivariance does not require equivariant layers. One can choose coordinates that transform with the input, apply an ordinary network in them, and transform the prediction back (Puny et al., 2022; Kaba et al., 2023; Ma et al., 2024). For forces on particles, this turns any backbone, such as a Transformer, into a rotation-equivariant model, at one forward pass per choice of coordinates, or frame. The prediction should also be continuous: a force field that jumps under an arbitrarily small displacement is unreliable in simulation and fragile under noise (Pozdnyakov & Ceriotti, 2023; Dym et al., 2024).

Dym et al. (2024) proved obstructions to continuous canonicalization on the entire point-cloud space. For planar rotations, even a finite unweighted frame cannot preserve continuity there. Their remedy is a weighted frame, that is, finitely many choices with input-dependent weights, where a frame may appear or disappear as its weight tends to zero. Their constructions use $n(n-1)\cdots(n-d+2)$ frames for $\mathrm{SO}(d)$ and twice as many for $\mathrm{O}(d)$. On the full-rank domain, how much support is necessary, and can a smaller explicit construction attain the same order? We answer these questions by applying cohomological lower bounds and constructing a linear family of probes.

Let $\mathcal{M}^{\mathrm{fr}}_{d,n} = \{A \in \mathbb{R}^{d\times n} : \operatorname{rank} A = d\}$, whose columns are n vectors spanning $\mathbb{R}^d$ ($n = N-1$ for N centered particles). The group $G \in \{\mathrm{O}(d), \mathrm{SO}(d)\}$ acts freely on it by $A \mapsto RA$. A *weighted frame* assigns to each A a finitely supported probability measure μ_A on G, whose atoms we call frames. It is equivariant if $\mu_{RA} = R_*\mu_A$, and continuous if $\int_G \varphi\, d\mu_A$ depends continuously on A for every continuous $\varphi\colon G \to \mathbb{R}$. For vector outputs, a frame Q gives the transported prediction $Qf(Q^\top A)$: evaluate the backbone in frame coordinates and rotate its output back. Averaging these predictions against μ_A is equivariant and preserves continuity (Dym et al., 2024), with one backbone evaluation per atom. Let $\mathrm{wfr}_G(\mathcal{M}^{\mathrm{fr}}_{d,n})$ be the least k such that some continuous equivariant weighted frame has at most support size k at every input.

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.

Theorem 1 (Linearly many frames are necessary and sufficient)**.** *Let* $n > d \geq 2$ *and* $G \in \{\mathrm{O}(d), \mathrm{SO}(d)\}$. *Then*

$$1 + \left\lfloor \frac{n-d}{2} \right\rfloor \;\leq\; \mathrm{wfr}_G(\mathcal{M}^{\mathrm{fr}}_{d,n}) \;\leq\; d(n-d)+1. \tag{1}$$

An explicit family of $d(n-d)+1$ *fixed linear probes gives the upper bound. For* $d=3$, $\lfloor (n-1)/2 \rfloor \leq \mathrm{wfr}_G(\mathcal{M}^{\mathrm{fr}}_{3,n}) \leq 3n-8$. *For even* d *the lower bound improves to* $n-d+1$, *and* $\mathrm{wfr}_{\mathrm{SO}(2)}(\mathcal{M}^{\mathrm{fr}}_{2,n}) = n-1$. *Thus* $\mathrm{wfr}_G(\mathcal{M}^{\mathrm{fr}}_{d,n}) = \Theta(n)$ *for fixed* d.

The lower bound combines the rational kernel cup-length theorem of Daundkar & García-Calcines (2026) with a characteristic-class calculation for the Stiefel bundle (Section 4.2). It applies to every continuous equivariant weighted frame on full-rank inputs, including the restrictions of robust frames that have finite support at full-rank inputs (Corollary 5). The planar equality follows from a short construction using any $n-1$ of the n columns (Corollary 4).

For $N = 24$ particles in $\mathbb{R}^3$, so $n = 23$, every continuous G-equivariant weighted frame requires at least 11 atoms at some input. Our bank has 61 probes, the $\mathrm{O}(3)$ construction of Dym et al. (2024) has at most 1,012 frames, and the coordinate-minor bank has 1,771 probes (Figure 1). These are worst-case support bounds and candidate counts for the same n-vector representation. Runtime also depends on scoring and evaluating the probes. The lower bound already holds on $AA^\top = I_d$: even perfectly conditioned inputs require this support in the worst case.

Counting charts is not enough. The cited atlas-based constructions and our probe construction blend frames chosen continuously on open charts with invariant weights (Dym et al., 2024; Pozdnyakov & Ceriotti, 2023). The fewest charts is a classical invariant, the Schwarz genus (Schwarz, 1966), but it does not give a lower bound on support: full-rank 3×4 matrices under $\mathrm{O}(3)$ need four charts, yet two frames with weights $\frac{1}{2}$ suffice, because reflections may exchange them (Section 4). We therefore bound frames directly. A weighted frame with at most k atoms restricts to an equivariant map from a Stiefel manifold to the space $\mathcal{P}_{\leq k}(G)$ of measures on G with at most k atoms. Equivalently, it assigns such a measure to each fiber of the Stiefel bundle. Characteristic classes on the Grassmannian vanish after pullback to the Stiefel manifold, but their products remain nonzero on the base. Rational cup length turns these products into support lower bounds. An independent argument compares characteristic-class degree with the dimension of $\mathcal{P}_{\leq k}(G)/G$.

An explicit linear atlas. A probe $P \in \mathbb{R}^{n \times d}$ combines the input vectors into d candidate axes AP. Where these are independent, the polar factor of AP is a frame that transforms with the input. We obtain a covering bank by adapting Wronskian rank condensers (Forbes & Guruswami, 2015; Forbes et al., 2014). Regard the rows of A as d polynomials of degree less than n. The probe built from a polynomial basis and its first $d-1$ derivatives at a node t fails exactly when the Wronskian of

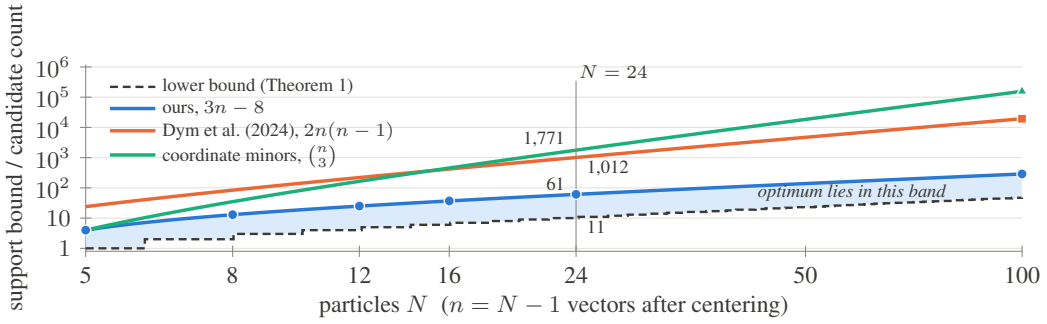

Figure 1: Worst-case support bounds and candidate-bank sizes for $\mathrm{O}(3)$-equivariant weighted frames on $n = N-1$ contrast vectors (logarithmic axes). The minimum worst-case support lies in the shaded band. The other curves count frames in the construction of Dym et al. (2024) and charts in the coordinate-minor atlas. Dots mark the experimental sizes. Each evaluated atom requires one backbone call; these counts are not timings.

these polynomials vanishes at t, and this Wronskian is a nonzero polynomial of degree at most $d(n-d)$. Hence among any $d(n-d)+1$ nodes, at least one gives a valid probe at each full-rank input. Geometrically, the charts are complements of Schubert divisors for osculating flags (Eremenko & Gabrielov, 2002; Sottile, 2011). Weighting each frame by $\det(AP)^2$ makes its contribution vanish where its probe fails, and the average is equivariant and C^1 whenever the backbone is. Weights $\det(AP)^{2k}$ give C^{2k-1} (Section 3). For $\mathrm{O}(d)$ every atlas needs at least $\max\{d, n-d\}+1$ charts (Appendix A.2), so ours uses at most d times the fewest.

Scope. The probes combine the input vectors in a fixed order, so the construction is not permutation equivariant. It suits labeled point sets, such as our particles, skeletal keypoints, or consistently indexed atoms along a molecular trajectory. The lower bound concerns rotations alone and so also applies to frames that are in addition permutation invariant. Whether linearly many such frames suffice is open (Section 7). Our determinant probes require full rank. At rank r, the stabilizer is $\mathrm{O}(d-r)$ for $\mathrm{O}(d)$ and $\mathrm{SO}(d-r)$ for $\mathrm{SO}(d)$. Thus a planar rank-two configuration in $\mathbb{R}^3$ has a trivial rotation stabilizer. Extension through rank loss requires additional analysis, with weak equivariance relevant where stabilizers are nontrivial (Dym et al., 2024). Equivariant architectures and invariant-feature models select no frames, so the lower bound does not concern them.

Experiments. We test the theory on synthetic force fields for labeled particles in $\mathbb{R}^3$ (Section 6). At a regular boundary, hard selection jumps if the two transported predictions disagree (3). In our trained models this disagreement persists: moving the selection boundary moves the observed response, including after training with a consistency penalty. Continuous averaging removes the jump. On five-particle trajectories, averaging over four coordinate charts reduces error at the first chart crossing by 39% and final position error by 15% relative to hard selection. There, the 24-frame construction of Dym et al. (2024), a Gram-matrix invariant model and an equivariant network are more accurate still. With Wronskian probes at $N \in \{5, 8, 12, 16, 24\}$ and a fixed number of updates, continuous averaging reduces error by 13–23% relative to separately trained hard selection over the same probes at every size. At $N = 24$, hard selection already improves on Gram by 42.2%. Averaging improves further, with 54.5% lower error than Gram at 5.9 times its recorded worker time. Gram is more accurate at $N = 5, 8$ and within its shorter observed development budget at $N = 12, 24$.

2 Related work

Frames and canonicalization. Frame averaging (Puny et al., 2022) and canonicalization (Kaba et al., 2023; Ma et al., 2024) evaluate a generic backbone on transformed copies of the input. They have been applied to materials (Duval et al., 2023) and to pretrained models (Mondal et al., 2023), with explicit frames for point clouds (Baker et al., 2024) and frames of minimal size (Lin et al., 2024). The continuity obstructions for unweighted frames depend on the domain (Dym et al., 2024); the full-rank two-frame example above shows why this qualification matters. Probabilistic symmetrization (Kim et al., 2023) defines an exactly equivariant expectation, while finite Monte Carlo evaluation generally gives equivariance in distribution. Stochastic frame averaging (Duval et al., 2023) also uses sampling. We evaluate a finite weighted sum deterministically.

Continuous weighted frames. Dym et al. (2024) proved that continuous canonicalization is impossible for permutations and rotations of point clouds, introduced weighted frames that are only weakly equivariant at inputs with nontrivial stabilizers, and constructed continuous (robust) weighted frames for S_n, $\mathrm{SO}(d)$ and $\mathrm{O}(d)$. Their lower bounds on the number of atoms concern S_n. In $\mathbb{R}^3$ their rotation frames resemble the coordinate-system ensembles of Pozdnyakov & Ceriotti (2023), which average atomistic models over frames spanned by pairs of neighbors, prune frames with adaptive cutoffs, and fall back to an equivariant model on nearly collinear environments. These constructions already provide continuous or smooth weighted symmetrization. We give a globally covering linear probe bank on full-rank inputs and quantify regularity in terms of the determinant power. Their permutation symmetry and treatment of degenerate inputs go beyond our construction. Their candidate counts also need not equal their cost after adaptive pruning.

Regularity of symmetrized models. Zhang et al. (2020) noted discontinuities of canonicalization for sets of vectors. Tahmasebi & Jegelka (2025) characterized when canonicalized models are continuous or differentiable through conditions on the backbone, and Sverdlov et al. (2026) relate

generalization bounds to covering numbers of the canonical image. Isometric canonicalizations attain the covering-number bound of the metric quotient. In our construction the weights compensate for singular probe frames: determinant powers $2k$ preserve C^{2k-1} regularity for every backbone of that class.

Equivariant architectures and invariant features. Equivariant networks (Cohen & Welling, 2016; Thomas et al., 2018; Satorras et al., 2021; Batzner et al., 2022; Batatia et al., 2022; Geiger & Smidt, 2022) and invariant-feature models built from inner products (Villar et al., 2021) select no frames, so our lower bound does not constrain them. Our experiments include both kinds.

Probabilistic sectional category. The Schwarz genus of a continuous map is the fewest open sets carrying local sections (Schwarz, 1966). It underlies Lusternik–Schnirelmann category and topological complexity (Farber, 2003). Analog and distributional sectional category replace sections by continuously varying, finitely supported probability measures (Knudsen & Weinberger, 2024; Jauhari, 2025), extending distributional category and topological complexity (Dranishnikov & Jauhari, 2024). Their general frameworks agree for proper maps in the relevant metric setting (Knudsen & Weinberger, 2026). We apply this identification to principal compact Lie-group bundles. The rational cup-length bound of Daundkar & García-Calcines (2026) supplies the lower bound in Theorem 1. An independent finite-dimensional characteristic-class argument recovers the main lower bounds for $d = 2, 3$. We also record a weaker estimate from Dold's theorem.

3 Averaging over weighted frames

Throughout, $G \in \{\mathrm{O}(d), \mathrm{SO}(d)\}$ acts on $\mathcal{M}^{\mathrm{fr}}_{d,n}$ by $A \mapsto RA$, $\rho\colon G \to \mathrm{O}(V)$ is an orthogonal representation on the finite-dimensional output space, and $f\colon \mathcal{M}^{\mathrm{fr}}_{d,n} \to V$ is a backbone. For forces on N centered particles, A collects $n = N - 1$ fixed orthonormal contrasts of the positions and $\rho(R)$ rotates every force vector (Appendix B.2). A weighted frame $\mu_A = \sum_j w_j(A)\, \delta_{g_j(A)}$ gives the prediction

$$F(A) = \sum_j w_j(A)\, Y_j(A), \qquad Y_j(A) = \rho(g_j(A))\, f\big(g_j(A)^{-1} A\big). \tag{2}$$

Each Y_j evaluates the backbone in frame coordinates and transports its output back. The identity $\mu_{RA} = R_* \mu_A$ gives $F(RA) = \rho(R)F(A)$ by change of variables in the integral, even when the group permutes the atoms. For local gauges, this also follows term by term from $g_j(RA) = Rg_j(A)$ and $w_j(RA) = w_j(A)$. If f is already equivariant, every Y_j equals f, so averaging leaves f unchanged.

Hard selection jumps. Hard selection evaluates only the frame of largest weight. If, along a path $A(t)$, the largest weight passes from frame i to frame j at $t = 0$ and both transported predictions extend continuously to $A(0)$, then

$$\lim_{t \uparrow 0} F_{\mathrm{hard}}(A(t)) - \lim_{t \downarrow 0} F_{\mathrm{hard}}(A(t)) = Y_i(A(0)) - Y_j(A(0)). \tag{3}$$

Thus a regular two-chart crossing produces a jump precisely when the transported predictions disagree. Ordinary force-error training does not explicitly enforce agreement. A finite consistency penalty encourages it but need not make it exact (Section 6.1). Appendix A.7 specifies the crossing and tie assumptions.

Frames from probes. A probe is a fixed matrix $P \in \mathbb{R}^{n \times d}$ with orthonormal columns. It proposes the d axes AP, and on its chart $U_P = \{A : \det(AP) \neq 0\}$ the polar decomposition

$$AP = Q_P(A)\, S_P(A), \qquad Q_P(A) \in \mathrm{O}(d), \quad S_P(A) \succ 0, \tag{4}$$

gives a frame with $Q_P(RA) = RQ_P(A)$. For $\mathrm{O}(d)$ we use $g_P = Q_P$. For $\mathrm{SO}(d)$ we fix a reflection J and use $g_P(A) = Q_P(A)$ where $\det(AP) > 0$ and $g_P(A) = Q_P(A)J$ where $\det(AP) < 0$. Given probes $P_1, \dots, P_m$ whose charts cover $\mathcal{M}^{\mathrm{fr}}_{d,n}$ and an integer $k \geq 1$, we use the determinant weights

$$w_j(A) = \frac{\det(AP_j)^{2k}}{\sum_{\ell=1}^{m} \det(AP_\ell)^{2k}}. \tag{5}$$

They are invariant, vanish exactly where probe j fails, and have a positive denominator because the charts cover. The particle-count study uses $k = 1$. The earlier studies modify the gauges and weights (Table 1).

Theorem 2 (Continuous and smooth averaging)**.** *Let $P_1, \dots, P_m$ be probes whose charts cover $\mathcal{M}^{\mathrm{fr}}_{d,n}$, let $k \geq 1$, and define F by* (2) *and* (5)*, omitting frames of weight zero. Then $A \mapsto \mu_A$ is a continuous equivariant weighted frame with at most m atoms, and for every continuous backbone f the prediction F is continuous and G-equivariant. If f is C^{2k-1}, then so is F. In general F is not C^{2k}, even for a C^∞ backbone.*

Proof outline. Equivariance holds term by term. Near a full-rank input, the canonical inputs remain in a compact set, so a vanishing weight suppresses a frame even when the frame has no limit. For higher derivatives, set $s = \sigma_{\min}(AP_j)$. Derivatives of the polar factor of order r are $O(s^{-r})$, whereas derivatives of the weighted numerator of order $r < 2k$ are $O(s^{2k-r})$. They therefore extend by zero across a singular probe. Appendix A.6 proves these estimates and the extension argument, and proves sharpness for both groups, including every fixed special-orthogonal correction. □

For energy prediction, a C^{2k-1} scalar backbone gives an invariant C^{2k-1} energy and a C^{2k-2} equivariant force $-\nabla E$. Differentiation must include both frames and weights. Our experiments predict forces directly. Continuity alone does not make such forces conservative, and smoothness does not guarantee exact energy conservation by a numerical integrator. The Lipschitz constant depends on the probe coverage margin (Proposition 13).

4 How many frames does continuity require?

4.1 Charts, frames, and sectional category

Whitening, $A \mapsto (AA^\top)^{-1/2}A$ (or precisely the straight-line homotopy from A to its whitening), is an equivariant deformation retraction of $\mathcal{M}^{\mathrm{fr}}_{d,n}$ onto the Stiefel manifold $\mathrm{St}^{\mathrm{row}}_{d,n} = \{U \in \mathbb{R}^{d\times n} : UU^\top = I_d\}$, so both spaces need the same number of frames (Appendix A.2). The quotient of $\mathrm{St}^{\mathrm{row}}_{d,n}$ by $\mathrm{O}(d)$ is the Grassmannian $\mathrm{Gr}_d(\mathbb{R}^n)$ of row spaces. The quotient by $\mathrm{SO}(d)$ is the oriented Grassmannian $\mathrm{Gr}^+_d(\mathbb{R}^n)$. For a free action of a compact Lie group G on a space X, let

$$\mathrm{wfr}_G(X) = \min\big\{k : \exists\, \mu\colon X \to \mathcal{P}_{\leq k}(G) \text{ continuous with } \mu_{gx} = g_*\mu_x\big\}, \tag{6}$$

where $\mathcal{P}_{\leq k}(G)$ is the space of finitely supported probability measures on G with at most k atoms, in the weak topology.

An atlas of local frames, that is, invariant open sets U_i covering X with continuous $g_i\colon U_i \to G$ satisfying $g_i(gx) = gg_i(x)$, blends into a weighted frame through a partition of unity. Hence $\mathrm{wfr}_G(X)$ is at most the fewest charts in an atlas, the Schwarz genus of $X \to X/G$ (Schwarz, 1966). For $\mathrm{O}(d)$ on $\mathcal{M}^{\mathrm{fr}}_{d,n}$ the genus lies between $\max\{d, n-d\}+1$ and $d(n-d)+1$ (Theorem 8). Replacing sections by finitely supported measures gives exactly the minimum number of frames.

Theorem 3 (Topological characterization)**.** *If a compact Lie group G acts freely on a compact metrizable space X, then*

$$\mathrm{wfr}_G(X) = \mathrm{asecat}(X \to X/G) + 1 \leq \mathrm{genus}(X \to X/G),$$

where asecat *is the reduced analog sectional category (Knudsen & Weinberger, 2024).*

Sending a frame g at x to the orbit point $g^{-1}x$ identifies equivariant measures on G with measures supported in single orbits, which preserves the number of atoms and continuity (Appendix A.1). For compact X, the analog sectional category agrees with the distributional one (Dranishnikov & Jauhari, 2024; Knudsen & Weinberger, 2026).

Frames can be fewer than charts. For $\mathrm{O}(3)$ on full-rank 3×4 matrices, the quotient is $\mathrm{Gr}_3(\mathbb{R}^4) \cong \mathbb{R}\mathrm{P}^3$ and

$$\mathrm{genus} = 4, \qquad \mathrm{wfr}_{\mathrm{O}(3)} = 2. \tag{7}$$

The row space of $U \in \mathrm{St}^{\mathrm{row}}_{3,4}$ has a unit normal $q(U)$ given by signed 3×3 minors, with $q(RU) = \det(R)\, q(U)$. Identifying $\mathbb{R}^4$ with the quaternions, the coordinates of iq, jq, kq form a second

orthonormal basis of the row space, and the change of basis $c(U) \in \mathrm{SO}(3)$ satisfies $c(RU) = \det(R)\, R\, c(U)$. So c is a global frame for $\mathrm{SO}(3)$, and $\mu_U = \frac{1}{2}\delta_{c(U)} + \frac{1}{2}\delta_{-c(U)}$ is a continuous $\mathrm{O}(3)$-equivariant weighted frame. Reflections exchange its two atoms. Four charts are necessary, because the generator of $H^1(\mathbb{R}\mathrm{P}^3; \mathbb{F}_2)$ has nonzero cube and vanishes on the Stiefel manifold (Appendix A.8). A lower bound on the number of charts therefore need not give the same lower bound on weighted support.

4.2 Lower bounds from equivariant cohomology

Recall that a weighted frame is a section whose values are finitely supported measures in the fibers of the Stiefel bundle. Write $q\colon \mathrm{St}^{\mathrm{row}}_{d,n} \to \mathrm{Gr}^+_d(\mathbb{R}^n)$ for this bundle with group $\mathrm{SO}(d)$. Its tautological bundle becomes trivial after pullback along q: the rows of a Stiefel matrix supply a basis. Thus its positive-degree characteristic classes lie in $\ker q^*$. The question is how long a nonzero product of these classes can be.

For a graded ideal I, its rational cup length is the largest number of positive-degree elements of I with nonzero product. The rational kernel cup-length theorem (Daundkar & García-Calcines, 2026, Theorem 4.21), together with Theorem 3, gives

$$\mathrm{wfr}_{\mathrm{SO}(d)}(\mathcal{M}^{\mathrm{fr}}_{d,n}) \geq 1 + \mathrm{cl}_{\mathbb{Q}}(\ker q^*). \tag{8}$$

Here the source and base are connected compact manifolds, so the cited fibration theorem applies. If $c = n - d$ and $r = \lfloor c/2 \rfloor$, then $p_1(\gamma_d)^r \neq 0$. To see this, restrict to oriented two-planes by $V \mapsto \mathbb{R}^{d-2} \oplus V$. The first Pontryagin class becomes the square of the Euler class on $\mathrm{Gr}^+_2(\mathbb{R}^{c+2})$, a complex quadric of dimension c. Its Euler class is, up to sign, the hyperplane class, with nonzero c-th power. For every even d, the longer product $e(\gamma_d)^c$ is nonzero: projecting c fixed independent vectors onto a varying oriented d-plane gives two transverse common zeros with equal local signs. Lemma 11 gives the details. Equation (8) therefore yields $1 + \lfloor c/2 \rfloor$ in general and $c + 1$ for even d.

An $\mathrm{SO}(d)$-equivariant pushforward transfers these lower bounds to $\mathrm{O}(d)$. Appendix A.4 gives this map and an independent proof for $d = 2, 3$ from characteristic-class degree (Theorem 10).

Corollary 4 (Planar rotations)**.** *For every $n \geq 2$,*

$$\mathrm{wfr}_{\mathrm{SO}(2)}(\mathcal{M}^{\mathrm{fr}}_{2,n}) = n - 1.$$

Proof. Identify the columns with $z_1, \ldots, z_n \in \mathbb{C}$ and rotations with unit complex numbers. Full row rank implies $S(A) = \sum_{j=1}^{n-1} |z_j|^2 > 0$, since otherwise only the last column could be nonzero. Set

$$\mu_A = \sum_{\substack{1 \leq j < n \\ z_j \neq 0}} \frac{|z_j|^2}{S(A)}\, \delta_{z_j/|z_j|}. \tag{9}$$

Rotating A rotates every atom and preserves every weight. When a selected column tends to zero, its weight tends to zero, so the measure is weakly continuous. Its support has size at most $n - 1$, matching the lower bound. □

The construction distinguishes the last column. It does not give permutation invariance or an extension to every nonzero matrix. On all nonzero $2 \times n$ matrices the minimum support is n: restriction to the Hopf bundle gives the lower bound, and using all n columns in (9) gives the upper bound (Appendix A.4). This explains the domain difference from the n-frame construction of Dym et al. (2024).

Corollary 5 (Robust frames on all inputs)**.** *Let $n > d \geq 2$. Every robust frame of Dym et al. (2024) for $\mathrm{O}(d)$ or $\mathrm{SO}(d)$ on $\mathbb{R}^{d \times n}$, finitely supported at every full-rank input, attains support at least $1 + \lfloor (n-d)/2 \rfloor$ there. For even d this lower bound improves to $n - d + 1$.*

Proof. On the full-rank domain stabilizers are trivial, so robust continuity and weak equivariance reduce to weak continuity and equivariance. If every full-rank input had smaller support, the restriction would contradict Theorem 1. No global support bound is needed in advance. □

5 AN EXPLICIT LINEAR ATLAS

Wronskian rank condensers preserve every d-dimensional subspace under at least one of $d(n-d)+1$ fixed maps (Forbes et al., 2014; Forbes & Guruswami, 2015). Transposing and orthonormalizing these maps gives a probe cover (Appendix A.5). We give the ordinary-Wronskian proof and then use the cover for weighted equivariant averaging. Choose polynomials $\psi_0, \dots, \psi_{n-1}$ with $\deg \psi_r = r$, put $b(t) = (\psi_0(t), \dots, \psi_{n-1}(t))^\top$, and let

$$\widetilde{B}(t) = \big[b(t), b'(t), \dots, b^{(d-1)}(t)\big] \in \mathbb{R}^{n\times d}. \tag{10}$$

Read the rows a_i of A as coefficient vectors of the polynomials $f_i(t) = a_i b(t)$. Entry (i,r) of $A\widetilde{B}(t)$ is $f_i^{(r-1)}(t)$, so $W_A(t) = \det(A\widetilde{B}(t))$ is the Wronskian of $f_1, \dots, f_d$, and the probe at t fails exactly at the zeros of W_A.

Theorem 6 (Universal coverage)**.** *Let $n > d \geq 1$ and $D = d(n-d)$. For distinct nodes $t_1, \dots, t_{D+1} \in \mathbb{R}$, let P_j be obtained by orthonormalizing the columns of $\widetilde{B}(t_j)$. Then*

$$\max_{1\leq j\leq D+1} |\det(AP_j)| > 0 \qquad \textit{for all } A \in \mathcal{M}^{\mathrm{fr}}_{d,n}. \tag{11}$$

Proof. Full row rank makes $f_1, \dots, f_d$ linearly independent, so their Wronskian is a nonzero polynomial. After an invertible row operation, the f_i have distinct degrees $e_1 < \cdots < e_d \leq n-1$. The leading coefficient of the Wronskian is then a nonzero multiple of $\prod_{i<j}(e_j - e_i)$, and its degree is $\sum_i e_i - d(d-1)/2 \leq d(n-d)$ (Lemma 12). A nonzero polynomial of degree at most D cannot vanish at $D+1$ points, and orthonormalizing the columns of $\widetilde{B}(t_j)$ multiplies $\det(A\widetilde{B}(t_j))$ by a nonzero constant. □

Geometrically, probe j fails exactly when the row space of A contains a nonzero polynomial vanishing to order d at t_j, that is, on the Schubert divisor of the osculating flag at t_j. Theorem 6 is the classical fact that $D+1$ such divisors have no common point (Eremenko & Gabrielov, 2002; Sottile, 2011). For instance, with $d=2$, $n=3$ and $b(t) = (1, t, t^2)^\top$, the probes at $t = -1, 0, 1$ cover all full-rank 2×3 matrices. With Theorem 2, this proves the upper bound of Theorem 1.

Comparison and cost. For $\mathrm{O}(d)$ every atlas needs at least $\max\{d, n-d\}+1$ charts, so our $d(n-d)+1$ charts are within a factor d of the fewest possible. The constructions of Dym et al. (2024) cover all inputs with $n(n-1)\cdots(n-d+2)$ frames for $\mathrm{SO}(d)$ and twice as many for $\mathrm{O}(d)$. For fixed $d \geq 3$ our count grows more slowly. For $d=2$, the planar construction (9) is smaller than the $(2n-3)$-probe Wronskian bank. With $m = d(n-d)+1$ dense probes and backbone cost $C_f(N)$, one prediction costs $O(mnd^2 + md^3 + mC_f(N))$. Scoring alone is quadratic in N for fixed d. We use normalized Legendre polynomials and Chebyshev nodes (Appendix A.5). The root count guarantees coverage but gives no quantitative conditioning margin as n grows.

6 EXPERIMENTS

We report three studies on synthetic force prediction for labeled particles in $\mathbb{R}^3$: trajectories through natural chart boundaries, an intervention that moves the selection boundary, and a comparison across particle counts. The first two use the protocols in Appendices B.4 and B.5. The particle-count study uses the cyclic potential and regular-core generator in Appendix B.1, whose ratio of smallest to largest singular value lies in $[0.15, 1]$. Normalized root-mean-square error (NRMSE) is RMS force error divided by RMS target force. The study-specific pairing and bootstrap conventions appear in Appendix B.

6.1 HARD SELECTION PRODUCES JUMPS

With $N = 5$ particles ($n = 4$), we use four coordinate-minor charts. At a regular selection boundary, the hard prediction jumps precisely when the two transported branches disagree (3). We test this mechanism by keeping the trained branches fixed and moving only the selection boundary: seven weight biases place seven boundaries on each of 504 input paths. At each boundary, every rule is evaluated on the same pair of nearby inputs, so only one rule switches there.

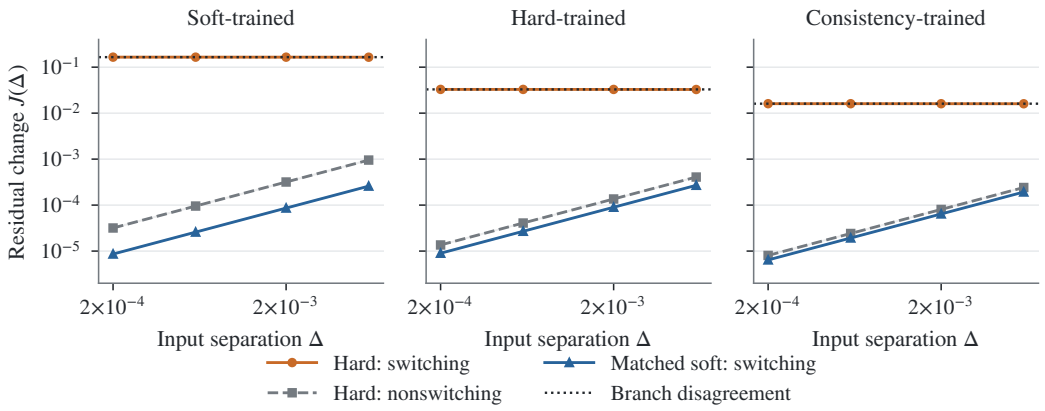

Figure 2: Residual change $J(\Delta)$ as input separation decreases ($N = 5$, four coordinate charts). Panels identify checkpoint training; routing changes while the transported branches remain fixed. Hard switching responses approach direct branch disagreement (dotted lines), whereas nonswitching and matched-soft responses decrease. Shading shows marginal paired 95% intervals.

Figure 2 shows the residual changes over four input separations, from 0.006 to 0.0002. The switching response approaches the directly measured branch disagreement, while nonswitching and matched-soft responses decrease. At separation 0.0006, switching minus mean nonswitching response is 0.0328 for hard-trained models, 0.0160 after consistency training, and 0.165 when hard selection is imposed on soft-trained models. The corresponding matched-soft contrasts have absolute value below 4×10^{-7}. These are contrasts, not the raw responses plotted in the figure (Appendix B.5).

Thus consistency training roughly halves the hard-routing contrast under this protocol but leaves residual disagreement. Its selected penalty coefficient, $\lambda = 1$, is the largest tested. This does not rule out backbones whose transported branches agree. The two-frame construction (7) and the 24-frame construction of Dym et al. (2024) show no response pattern following the moved boundary.

Trajectories through chart boundaries. On $1{,}200$ held-out five-particle trajectories with 302 first chart crossings and eight training seeds, continuous averaging over the four coordinate charts reduces NRMSE at the first crossing by 39.3% (paired 95% interval 34.4–43.9%) and the mean across seeds of median final position error by 15.0% (10.2–19.6%) relative to hard selection over the same charts. The gain is concentrated near chart boundaries, where error falls by 34.4%. Far from them it rises by 2.6%. All rollouts completed, but one rollout under continuous averaging exceeded the position-error threshold. On this small task continuous averaging is not the most accurate method: the 24-frame construction of Dym et al. (2024) (overall NRMSE 0.0060 versus 0.0068), a Gram-matrix model (0.0015) and an e3nn message-passing network (0.0039), each with three seeds, all do better (Table 5).

6.2 Scaling with the number of particles

At $N \in \{5, 8, 12, 16, 24\}$ we train four methods with five paired seeds for $40{,}000$ updates each and evaluate them on $20{,}000$ held-out states. *Wronskian averaging* uses all $3N - 11$ Wronskian probes with determinant weights. *Wronskian hard selection* trains and evaluates only the highest-weight probe, and *coordinate hard selection* selects among the $\binom{N-1}{3}$ coordinate minors. All three use the same particle Transformer, with about 0.53M parameters. The *Gram model*, with about 0.73M parameters, predicts invariant coefficients from inner products of the positions and returns linear combinations of the position vectors. The Gram representation contains the information needed for every continuous $\mathrm{O}(3)$-equivariant vector target on full-rank inputs (Appendix A.9). The comparison concerns learning with these specified architectures.

Wronskian averaging reduces NRMSE by 13–23% relative to separately trained Wronskian hard selection at every size. Every paired interval also favors it over coordinate hard selection (Table 8).

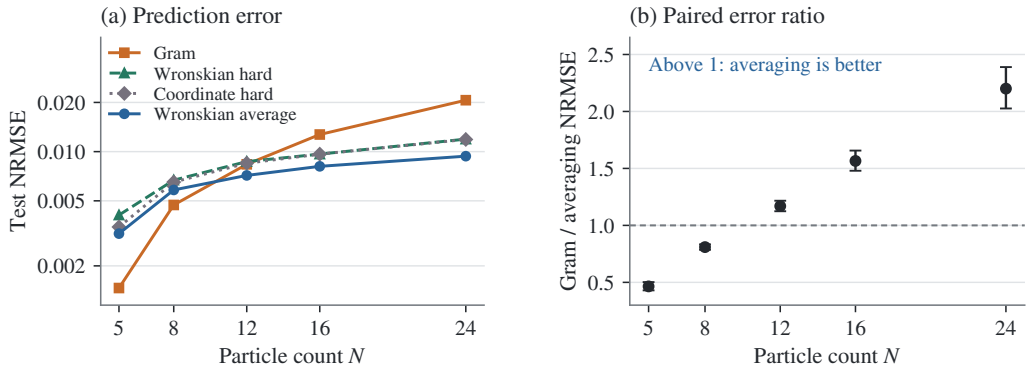

Figure 3: Force error across particle counts. Left: pooled test NRMSE for all four trained methods. Right: Gram error divided by Wronskian-averaging error, with simultaneous two-sided 95% intervals; ratios above one favor averaging. All methods use 40,000 updates. At $N = 24$, averaging takes 5.9 times Gram's recorded worker time and 6.5 times that of Wronskian hard selection.

At $N = 24$, both hard variants already improve on Gram: NRMSE is 0.01193 for Wronskian hard selection, 0.01187 for coordinate hard selection, and 0.02064 for Gram. Averaging lowers NRMSE further to 0.00938, a 21.4% reduction from Wronskian hard selection and 54.5% from Gram. Figure 3 shows all four methods. Gram is more accurate at $N = 5, 8$. Averaging is more accurate at $N = 12, 16, 24$, by 14.5%, 36.1%, and 54.5%.

All 105 selected checkpoints were at update 40,000, so this comparison does not establish convergence or equal-compute performance. At $N = 24$, mean recorded A40 worker time is 12.32 hours for averaging, 2.08 for Gram, and 1.90 for Wronskian hard selection. Within Gram's shorter observed budget, its development error is lower at both reported sizes, $N = 12, 24$ (Appendix B.7).

Keeping the eight largest weights at inference raises averaging's $N = 24$ error by 10.0%, still about half the Gram error. This evaluates 8 frames instead of 61 but scores the entire bank and inherits full-support training cost. Selecting just one frame raises error by 66.7%. This posthoc change differs from training with hard selection. Neither truncation retains the full construction's continuity guarantee (Appendix A.6).

7 Discussion and Open Questions

The linear worst-case support bound leaves several questions open.

Constants. In $\mathbb{R}^3$ the minimum lies between $\lfloor (n-1)/2 \rfloor$ and $3n - 8$. The planar rotation problem has the exact answer $n - 1$, but the three-dimensional bounds leave a constant-factor gap. Can frames whose atoms exchange, as in (7), improve the leading constant over a construction from local gauges?

Permutations and degenerate inputs. The lower bound also applies to frames that are invariant under permutations, but the known continuous permutation-invariant constructions in $\mathbb{R}^3$ use quadratically many frames. Whether linearly many suffice is open. On all of $\mathbb{R}^{d\times n}$, the minimum worst-case support among robust, weakly equivariant frames lies between the restriction lower bound and the $O(n^{d-1})$ construction of Dym et al. (2024). Strictly equivariant measures would have to be Haar measure at the zero matrix, so finite support there requires the weakly equivariant formulation.

Conditioning and compute. Our Lipschitz estimate depends on how far the probes stay from failing simultaneously, and the root count does not control this margin as n grows. Relative cutoffs in the weights, as used by Dym et al. (2024) and Pozdnyakov & Ceriotti (2023), can suppress low-weight frames without breaking continuity. The resulting average cost remains to be measured. A comparison at matched compute remains to be done.

REPRODUCIBILITY STATEMENT

The appendices give proofs, probe parameters, study-specific data generators, models, training rules, endpoints, and interval constructions.

AI USE STATEMENT

LLMs assisted with manuscript restructuring and editing, literature checks, development and checking of mathematical arguments, interpretation of reported results, and preparation of figures from reported values. LLM-based coding agents also substantially wrote and orchestrated the experimental code, including data generation, model implementations, training and evaluation pipelines, and statistical analysis. The authors specified the experimental protocols and decision criteria. The authors are responsible for the mathematical arguments, experimental design, numerical claims, and final text.

REFERENCES

Justin Baker, Shih-Hsin Wang, Tommaso de Fernex, and Bao Wang. An explicit frame construction for normalizing 3D point clouds. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 2456–2473. PMLR, 2024.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

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

Pavle V. M. Blagojević, Aleksandra S. Dimitrijević Blagojević, and Günter M. Ziegler. The topological transversal Tverberg theorem plus constraints. In Gergely Ambrus, Imre Bárány, Károly J. Böröczky, Gábor Fejes Tóth, and János Pach (eds.), *New Trends in Intuitive Geometry*, volume 27 of *Bolyai Society Mathematical Studies*, pp. 49–64. Springer, 2018. doi: 10.1007/978-3-662-57413-3_3. URL https://arxiv.org/abs/1604.02814v2. Preprint: arXiv:1604.02814v2.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pp. 2990–2999, 2016.

Navnath Daundkar and José Manuel García-Calcines. New bounds on distributional sectional category and applications to distributional homotopic distance. arXiv:2608.08298, 2026. URL https://arxiv.org/abs/2608.08298.

Albrecht Dold. Simple proofs of some Borsuk–Ulam results. In *Proceedings of the Northwestern Homotopy Theory Conference*, volume 19 of *Contemporary Mathematics*, pp. 65–69. American Mathematical Society, 1983. doi: 10.1090/conm/019/711043.

Alexander Dranishnikov and Ekansh Jauhari. Distributional topological complexity and LS-category. In Michael Farber and Jesús González (eds.), *Topology and AI: Topological Aspects of Algorithms for Autonomous Motion*, volume 4 of *EMS Series in Industrial and Applied Mathematics*, pp. 363–385. EMS Press, 2024.

Alexandre Duval, Victor Schmidt, Alex Hernández-García, Santiago Miret, Fragkiskos D. Malliaros, Yoshua Bengio, and David Rolnick. FAENet: Frame averaging equivariant GNN for materials modeling. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 9013–9033. PMLR, 2023.

Nadav Dym, Hannah Lawrence, and Jonathan W. Siegel. Equivariant frames and the impossibility of continuous canonicalization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12228–12267, 2024.

Alexandre Eremenko and Andrei Gabrielov. Degrees of real Wronski maps. *Discrete & Computational Geometry*, 28:331–347, 2002.

Michael Farber. Topological complexity of motion planning. *Discrete & Computational Geometry*, 29(2):211–221, 2003.

Michael A. Forbes and Venkatesan Guruswami. Dimension expanders via rank condensers. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics*, pp. 800–814, 2015. doi: 10.4230/LIPIcs.APPROX-RANDOM.2015.800. URL https://arxiv.org/abs/1411.7455. Full version: arXiv:1411.7455.

Michael A. Forbes, Ramprasad Saptharishi, and Amir Shpilka. Hitting sets for multilinear read-once algebraic branching programs, in any order. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pp. 867–875. ACM, 2014. doi: 10.1145/2591796.2591816. URL https://arxiv.org/abs/1309.5668. Full version: Pseudorandomness for Multilinear Read-Once Algebraic Branching Programs, in any Order, arXiv:1309.5668 (2013).

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Andrew M. Gleason. Spaces with a compact Lie group of transformations. *Proceedings of the American Mathematical Society*, 1(1):35–43, 1950. doi: 10.2307/2032430. Erratum, volume 1, number 6, page 826 (1950).

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

Nicholas J. Higham. Computing the polar decomposition—with applications. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1160–1174, 1986.

Dale Husemoller. *Fibre Bundles*. Springer, 3 edition, 1994.

Ekansh Jauhari. On sequential versions of distributional topological complexity. *Topology and its Applications*, 363:109271, 2025.

S.-O. Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 15546–15566, 2023.

Jinwoo Kim, Tien Dat Nguyen, Ayhan Suleymanzade, Hyeokjun An, and Seunghoon Hong. Learning probabilistic symmetrization for architecture agnostic equivariance. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Ben Knudsen and Shmuel Weinberger. Analog category and complexity. *SIAM Journal on Applied Algebra and Geometry*, 8(3):713–732, 2024.

Ben Knudsen and Shmuel Weinberger. On the analog category of finite groups. *Algebraic & Geometric Topology*, 26(4):1585–1596, 2026.

Yuchao Lin, Jacob Helwig, Shurui Gui, and Shuiwang Ji. Equivariance via minimal frame averaging for more symmetries and efficiency. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30042–30079. PMLR, 2024.

George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 60936–60979, 2024.

John W. Milnor and James D. Stasheff. *Characteristic Classes*. Princeton University Press, 1974.

Arnab Kumar Mondal, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Sai Rajeswar, and Siamak Ravanbakhsh. Equivariant adaptation of large pretrained models. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Art B. Owen. The pigeonhole bootstrap. *The Annals of Applied Statistics*, 1(2):386–411, 2007. doi: 10.1214/07-AOAS122.

Sergey N. Pozdnyakov and Michele Ceriotti. Smooth, exact rotational symmetrization for deep learning on point clouds. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J. Smith, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 2021.

Albert S. Schwarz. The genus of a fiber space. *American Mathematical Society Translations, Series 2*, 55:49–140, 1966.

Frank Sottile. *Real Solutions to Equations from Geometry*, volume 57 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2011.

Edwin H. Spanier. *Algebraic Topology*. McGraw-Hill, New York, 1966.

Yonatan Sverdlov, Benjamin Friedman, Snir Hordan, and Nadav Dym. When and how to canonize: A generalization perspective. arXiv:2605.11008, 2026. URL `https://arxiv.org/abs/2605.11008`.

Edward Szpilrajn. La dimension et la mesure. *Fundamenta Mathematicae*, 28(1):81–89, 1937. doi: 10.4064/fm-28-1-81-89.

Behrooz Tahmasebi and Stefanie Jegelka. Regularity in canonicalized models: A theoretical perspective. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 4789–4797. PMLR, 2025.

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

Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. In *Advances in Neural Information Processing Systems*, volume 34, pp. 28848–28863, 2021.

Yan Zhang, Jonathon Hare, and Adam Prügel-Bennett. FSPool: Learning set representations with featurewise sort pooling. In *International Conference on Learning Representations*, 2020.

A THEORY DETAILS AND PROOFS

Throughout this appendix, G is a compact Lie group acting on the left. The matrix domain is $\mathcal{M}_{d,n}^{\mathrm{fr}} = \{A \in \mathbb{R}^{d\times n} : \operatorname{rank} A = d\}$.

A.1 GAUGES AND FIBERWISE PROBABILITY MEASURES

Let $q\colon P \to B = P/G$ be a principal left G-bundle over a paracompact Hausdorff base. A gauge is a continuous map $\gamma\colon P \to G$ satisfying $\gamma(gp) = g\gamma(p)$.

A gauge determines a section by $s([p]) = \gamma(p)^{-1}p$. Conversely, a section determines the unique $\gamma(p)$ satisfying $p = \gamma(p)s(q(p))$. This map is continuous in local bundle coordinates and satisfies the gauge identity. Thus gauges and sections are equivalent.

For an orthogonal representation $\rho\colon G \to \mathrm{O}(V)$, use the convention

$$P \times_G V = (P \times V)/\big((p, v) \sim (gp, \rho(g)v)\big).$$

A continuous equivariant map $F\colon P \to V$ corresponds to the section $[p] \mapsto [p, F(p)]$ of this associated bundle. Canonicalization chooses an orbit representative. Prediction chooses an output over each orbit. An obstruction to the former need not obstruct the latter, as the zero prediction already illustrates.

Measures and their topology. Write $\mathcal{P}(Z)$ for probability measures on Z with the weak topology and $\mathcal{P}_{\le k}(Z)$ for those supported on at most k points. If Z is compact metrizable, then $\mathcal{P}(Z)$ is compact metrizable. Let $\Delta^{k-1} = \{(a_1, \dots, a_k) : a_i \ge 0,\ \sum_i a_i = 1\}$. The map

$$Z^k \times \Delta^{k-1} \longrightarrow \mathcal{P}_{\le k}(Z), \qquad ((z_i), (a_i)) \longmapsto \sum_{i=1}^{k} a_i \delta_{z_i} \tag{12}$$

is a continuous surjection from a compact space to a Hausdorff space. Consequently it is a quotient map. This topology allows atoms to be permuted, to merge, and to disappear when their weights go to zero. A continuous measure therefore need not have continuously labeled atoms.

Theorem 7 (Weighted frames as fiberwise probability measures)**.** *Let $q\colon P \to B$ be a principal left G-bundle with P, B, G compact metrizable. The space of probability measures supported on at most k points in a single fiber of q is homeomorphic over B to*

$$P \times_G \mathcal{P}_{\le k}(G), \qquad (p, \mu) \sim (hp, h_*\mu), \tag{13}$$

where h_* is pushforward by left translation. Sections of this bundle correspond to continuous equivariant maps $\mu\colon P \to \mathcal{P}_{\le k}(G)$. Hence, with reduced analog sectional category,

$$\mathrm{wfr}_G(P) = \operatorname{asecat}(q) + 1.$$

Proof. Let

$$D_k(q) = \{\nu \in \mathcal{P}_{\le k}(P) : q_*\nu = \delta_b \text{ for some } b \in B\}.$$

The condition $q_*\nu = \delta_b$ says that all mass lies in the fiber over b. This defines a compact Hausdorff space, because pushforward is continuous and the Dirac measures form a closed subset of $\mathcal{P}(B)$. For each $p \in P$, the map $\theta_p\colon G \to q^{-1}(q(p))$, $\theta_p(a) = a^{-1}p$, is a homeomorphism. Define

$$T\colon P \times_G \mathcal{P}_{\le k}(G) \longrightarrow D_k(q), \qquad T([p, \mu]) = (\theta_p)_*\mu.$$

The identity $\theta_{hp}(ha) = \theta_p(a)$ makes T well defined. For every continuous test function $u\colon P \to \mathbb{R}$,

$$(p, \mu) \longmapsto \int_G u(a^{-1}p)\, d\mu(a)$$

is continuous, since compactness gives uniform continuity in the parameter p. Thus T is continuous. Freeness and transitivity of the action in each fiber make it bijective. A continuous bijection from a compact space to a Hausdorff space is a homeomorphism.

An equivariant map $p \mapsto \mu_p$ determines the section $q(p) \mapsto (\theta_p)_*\mu_p$, independently of the chosen lift p. Conversely, suppose $b \mapsto \nu_b$ is a section of $D_k(q)$. Let $d(p, z) \in G$ be the unique element such

that $z = d(p, z)^{-1}p$, for $q(p) = q(z)$. The division map d is continuous in bundle coordinates. The measures

$$\mu_p = (d(p, \cdot))_* \nu_{q(p)}$$

vary continuously and satisfy $\mu_{hp} = h_* \mu_p$. These constructions are inverse and preserve support size. The least admissible k is precisely one plus the reduced analog sectional category (Knudsen & Weinberger, 2024). □

For a free action of a compact Lie group on a compact metrizable space, the quotient is a principal bundle (Gleason, 1950, Theorem 3.3). Thus the preceding theorem proves Theorem 3.

From a cover to a measure. For a finite sectional cover $U_i \subset B$ of a numerable principal bundle, let g_i be its local gauges and choose a subordinate partition of unity ϕ_i with $\operatorname{supp} \phi_i \subset U_i$. Then

$$\mu_p = \sum_{i:\phi_i(q(p))>0} \phi_i(q(p))\delta_{g_i(p)}.$$

is equivariant and has support at most the number of cover elements. It is weakly continuous: each continuous test function, bounded on compact G, gives summands that tend to zero where their weights vanish. Consequently $\mathrm{wfr}_G(P) \le \mathrm{genus}(q)$.

Comparison with related definitions. Knudsen & Weinberger (2024) give finitely supported measures the quotient topology from $\coprod_{k\ge1} Z^k \times \Delta^{k-1}$. On each $\mathcal{P}_{\le k}(Z)$ for compact metrizable Z, this agrees with the weak topology (Knudsen & Weinberger, 2026, Lemma 5.3). The Lévy–Prokhorov topology used for distributional sectional category agrees with it here as well. In particular, the analog and distributional sectional categories of our compact principal bundles coincide (Knudsen & Weinberger, 2026, Proposition 5.2).

On a free metrizable domain, stabilizer averaging in the robust-frame definition of Dym et al. (2024) has no effect. Robust continuity and weak equivariance therefore become weak continuity and equivariance. Among frames finitely supported at each input, wfr is the minimum worst-case support in that framework. The finite-support qualification in Corollary 5 is essential: on $\mathrm{O}(d)$ or $\mathrm{SO}(d)$ with $d \ge 2$, the constant Haar probability measure is a robust frame with no atoms, but it has infinite support.

A.2 The matrix bundle and deterministic chart count

Define the row and column Stiefel manifolds by

$$\mathrm{St}^{\mathrm{row}}_{d,n} = \{U \in \mathbb{R}^{d\times n} : UU^\top = I_d\}, \qquad \mathrm{St}^{\mathrm{col}}_{n,d} = \{P \in \mathbb{R}^{n\times d} : P^\top P = I_d\}.$$

Whitening, $U(A) = (AA^\top)^{-1/2}A$, is an equivariant retraction to $\mathrm{St}^{\mathrm{row}}_{d,n}$. Indeed, if $H(A) = (AA^\top)^{1/2}$, then

$$H_t(A) = \big((1-t)H(A) + tI_d\big)U(A), \qquad 0 \le t \le 1,$$

is an equivariant deformation retraction through full-row-rank matrices. The resulting compact quotient bundles are

$$\mathrm{St}^{\mathrm{row}}_{d,n}/\mathrm{O}(d) \cong \mathrm{Gr}_d(\mathbb{R}^n), \qquad \mathrm{St}^{\mathrm{row}}_{d,n}/\mathrm{SO}(d) \cong \mathrm{Gr}^+_d(\mathbb{R}^n).$$

Here Gr^+_d denotes the oriented Grassmannian. For the full-rank domain, these identifications are deformation retractions of the quotients, not homeomorphisms to the Grassmannians.

If i is inclusion and $r = U$ is whitening, then $r \circ i = \mathrm{id}$. Pulling local gauges back along r and restricting them along i gives

$$\mathrm{genus}(\mathcal{M}^{\mathrm{fr}}_{d,n} \to \mathcal{M}^{\mathrm{fr}}_{d,n}/G) = \mathrm{genus}(\mathrm{St}^{\mathrm{row}}_{d,n} \to \mathrm{St}^{\mathrm{row}}_{d,n}/G). \tag{14}$$

The same pullback and restriction preserve weighted support. Together with Theorem 7, this gives

$$\mathrm{wfr}_G(\mathcal{M}^{\mathrm{fr}}_{d,n}) = \mathrm{wfr}_G(\mathrm{St}^{\mathrm{row}}_{d,n}) = \mathrm{asecat}(\mathrm{St}^{\mathrm{row}}_{d,n} \to \mathrm{St}^{\mathrm{row}}_{d,n}/G) + 1. \tag{15}$$

Theorem 8 (Deterministic orthogonal-atlas bounds)**.** *For $n > d \geq 1$,*

$$\max\{d, n-d\} + 1 \leq \operatorname{genus}(\mathrm{St}_{d,n}^{\mathrm{row}} \to \mathrm{Gr}_d(\mathbb{R}^n)) \leq \min\left\{ d(n-d)+1, \binom{n}{d} \right\}. \tag{16}$$

For $n = d$, the genus is one.

Proof. Set $m = n - d$. The fiber of the tautological bundle γ_d over a d-plane is that plane itself. Its pullback to the frame bundle has a specified orthonormal basis in each fiber, so it is trivial. Since $\gamma_d \oplus \gamma_d^{\perp} \cong \mathbb{R}^n$, the pullback of its complement is stably trivial. The Whitney product formula therefore gives

$$w(q^*\gamma_d) = w(q^*\gamma_d^{\perp}) = 1.$$

In particular, every positive-degree Stiefel–Whitney class of either bundle belongs to $\ker q^*$.

Two nonzero products give the lower bound. Projecting a fixed vector of $\mathbb{R}^n$ onto a d-plane defines a section of γ_d. For m independent vectors $v_1, \ldots, v_m$, these sections vanish simultaneously at the unique d-plane V orthogonal to their span. In a graph chart $T\colon V \to V^{\perp}$, the derivative of the combined section is $T \mapsto (T^{\top}v_1, \ldots, T^{\top}v_m)$. Since the v_i form a basis of $V^{\perp}$, this is an isomorphism. The mod–2 zero count gives $\langle w_d(\gamma_d)^m, [\mathrm{Gr}_d(\mathbb{R}^n)] \rangle = 1$. Projecting d independent generic vectors onto the complementary bundle gives, in the same way, $\langle w_m(\gamma_d^{\perp})^d, [\mathrm{Gr}_d(\mathbb{R}^n)] \rangle = 1$, because the unique common zero is the plane spanned by those vectors. These are the usual top Stiefel–Whitney zero counts (Milnor & Stasheff, 1974). Schwarz's kernel cup-length bound now gives both $m + 1$ and $d + 1$ (Schwarz, 1966).

For the coordinate-minor upper bound, let E_I select the columns in a d-element subset $I \subset \{1, \ldots, n\}$. The sets $\{U : \det(UE_I) \neq 0\}$ cover by full row rank, and the polar factor of UE_I gives a continuous gauge on each set. There are $\binom{n}{d}$ such sets. Appendix A.5 supplies the $d(n-d)+1$ Wronskian cover. When $n = d$, the polar factor is a global gauge. □

A.3 Measures with bounded support

Set $\mathcal{Y}_k = \mathcal{P}_{\leq k}(G)$ and $m_G = \dim G$. Choose a faithful orthogonal embedding of G and restrict the Frobenius metric to obtain a bounded bi-invariant metric d_G of Hausdorff dimension m_G. The associated Wasserstein metric W_1 metrizes weak convergence on compact G. For paired atoms,

$$W_1\left(\sum_i a_i \delta_{g_i}, \sum_i b_i \delta_{h_i} \right) \leq \sum_i a_i d_G(g_i, h_i) + \frac{\operatorname{diam} G}{2} \sum_i |a_i - b_i|. \tag{17}$$

Indeed, first move mass a_i from g_i to h_i, and then redistribute the excess mass among the h_i.

Lemma 9 (Geometry of the measure space)**.** *For $k \geq 1$, the space $\mathcal{Y}_k$ is compact metrizable, and*

$$\dim_{\mathrm{cov}}(\mathcal{Y}_k/G) \leq (k-1)(m_G+1).$$

A measure with r support points has finite stabilizer of order dividing r.

Proof. Compactness and metrizability follow from (12). Left translation acts isometrically on measures, so

$$\bar{d}([\mu], [\nu]) = \min_{g \in G} W_1(\mu, g_*\nu)$$

is a metric inducing the orbit-space topology. Translating one support point to the identity gives a Lipschitz surjection

$$G^{k-1} \times \Delta^{k-1} \longrightarrow \mathcal{Y}_k/G,$$

$$((h_2, \ldots, h_k), (a_1, \ldots, a_k)) \longmapsto \left[a_1\delta_e + \sum_{i=2}^{k} a_i \delta_{h_i} \right].$$

Padding with zero weights includes smaller supports, and (17) gives the Lipschitz property. The domain has Hausdorff dimension $(k-1)(m_G+1)$. A Lipschitz map cannot increase Hausdorff dimension, which bounds covering dimension for metric spaces (Szpilrajn, 1937, Theorem 1).

The stabilizer of a measure permutes its r support points faithfully: an element fixing a point g satisfies $hg = g$, hence $h = e$. It is therefore finite, and its free action on the support shows that its order divides r. □

A weaker bound from connectivity. For $G \in \{\mathrm{O}(d), \mathrm{SO}(d)\}$ and $n > d \geq 2$, Dold's obstruction (Dold, 1983, p. 68) gives the weaker estimate

$$\mathrm{wfr}_G(\mathcal{M}^{\mathrm{fr}}_{d,n}) \geq \left\lfloor \frac{n-d}{d(d-1)/2+1} \right\rfloor + 1.$$

Indeed, a planar rotation subgroup of prime order $p > k$ acts freely on $\mathcal{Y}_k$, whose covering dimension is at most $k(m_G+1)-1$ by (12) and (17). The $(n-d-1)$-connectivity of $\mathrm{St}^{\mathrm{row}}_{d,n}$ (Husemoller, 1994) excludes an equivariant map when $k(m_G+1) \leq n-d$. The cohomological arguments below improve this estimate.

A.4 Rational cohomology and characteristic classes

The three lower-bound arguments use different information:

Argument	Input	Result for fixed d
Dold (A.3)	Connectivity and target dimension	Linear order, with weaker constants.
Degree (below)	Degree of one nonzero characteristic class	Recovers the main estimates for $d = 2, 3$.
Cup length (Section 4.2)	Number of factors in a nonzero product	Main estimates, including $n-d+1$ for even d.

The direct degree argument compares the class degree with the orbit-space dimension in Lemma 9.

Theorem 10 (Cohomological degree bound)**.** *Let $n > d \geq 2$, let $m = d(d-1)/2$, and suppose the characteristic-class map $H^D(B\mathrm{SO}(d); \mathbb{Q}) \to H^D(\mathrm{Gr}^+_d(\mathbb{R}^n); \mathbb{Q})$ is nonzero. Then*

$$\mathrm{wfr}_{\mathrm{O}(d)}(\mathcal{M}^{\mathrm{fr}}_{d,n}) \geq \mathrm{wfr}_{\mathrm{SO}(d)}(\mathcal{M}^{\mathrm{fr}}_{d,n}) \geq 1 + \left\lceil \frac{D}{m+1} \right\rceil.$$

One can take $D = 4\lfloor (n-d)/2 \rfloor$, and $D = d(n-d)$ for even d.

We prove this bound using a finite Stiefel manifold in place of the universal bundle. Finite stabilizers make the relevant fibers rationally acyclic through the required degree, so the characteristic class must vanish if that degree exceeds the orbit-space dimension. This gives the same estimates as (8) for $d = 2, 3$ with the classes above; for $d \geq 4$, cup length gives at least as strong a bound.

Lemma 11 (Nonzero characteristic-class powers)**.** *Let $n > d \geq 2$, set $c = n-d$, and let γ_d be the tautological oriented bundle on $\mathrm{Gr}^+_d(\mathbb{R}^n)$. Then*

$$p_1(\gamma_d)^{\lfloor c/2 \rfloor} \neq 0 \quad \textit{in rational cohomology.}$$

For even d, one also has $e(\gamma_d)^c \neq 0$.

Proof. Fix an oriented $(d-2)$-plane and use its orthogonal complement to define

$$\iota \colon \mathrm{Gr}^+_2(\mathbb{R}^{c+2}) \longrightarrow \mathrm{Gr}^+_d(\mathbb{R}^n), \qquad V \longmapsto \mathbb{R}^{d-2} \oplus V.$$

The pullback of γ_d is a trivial bundle plus γ_2. Hence $\iota^* p_1(\gamma_d) = p_1(\gamma_2) = e(\gamma_2)^2$ (Milnor & Stasheff, 1974).

For an oriented orthonormal basis (u, v) of a two-plane, the complex line $[u + iv]$ is independent of that basis. This identifies $\mathrm{Gr}^+_2(\mathbb{R}^{c+2})$ with the smooth quadric

$$Q^c = \{[z_0 : \cdots : z_{c+1}] \in \mathbb{CP}^{c+1} : z_0^2 + \cdots + z_{c+1}^2 = 0\}.$$

Indeed, the real and imaginary parts of an isotropic vector are orthogonal and have equal length, so they recover the oriented plane. For $z = u + iv$, the map $\lambda z \mapsto \mathrm{Re}(\lambda z)$ identifies the tautological complex line, as a real plane, with $\mathrm{span}(u, v)$, up to orientation. Its Euler class is therefore $\pm h$, where h is the hyperplane class restricted to Q^c. Since a quadric has degree two, $\langle h^c, [Q^c] \rangle = 2$. Every power h^a with $0 \leq a \leq c$ is consequently nonzero. Taking $a = 2\lfloor c/2 \rfloor$ proves the Pontryagin assertion; for $d = 2$, taking $a = c$ also proves the Euler assertion.

For general even d, nonvanishing of the Euler power is classical (Blagojević et al., 2018, Lemma 3.3(2)). The following zero count gives its integral evaluation. Choose orthonormal vectors $v_1, \dots, v_c$ and consider the section

$$s(V) = (\pi_V v_1, \dots, \pi_V v_c) \quad \text{of } \gamma_d^{\oplus c},$$

where π_V is orthogonal projection onto V. Its two zeros are the two orientations of $V_0 = \operatorname{span}(v_1, \dots, v_c)^\perp$. In a graph chart $T\colon V_0 \to V_0^\perp$, identify the nearby plane with V_0. The derivative is $T \mapsto (T^\top v_1, \dots, T^\top v_c)$, an isomorphism, so both zeros are transverse.

The oriented Grassmannian is an orientable manifold of dimension dc. The involution reversing the orientation of V reverses the orientations of both V and $V^\perp$. On the tangent space $\operatorname{Hom}(V, V^\perp)$ its orientation sign is $(-1)^{d+c} = (-1)^n$, while on $\gamma_d^{\oplus c}$ it is $(-1)^c$. The two local indices thus differ by the factor $(-1)^d$ and agree when d is even. Consequently,

$$\left\langle e(\gamma_d)^c, [\mathrm{Gr}_d^+(\mathbb{R}^n)] \right\rangle = \pm 2,$$

which proves rational nonvanishing. □

Proof of Theorem 10. Set $X = \mathrm{St}_{d,n}^{\mathrm{row}}$ and $G = \mathrm{SO}(d)$, and suppose there is an equivariant map $\mu\colon X \to \mathcal{Y}_k$. Let $c \in H^D(BG; \mathbb{Q})$ be a class whose image $c(\gamma^{(n)})$ on $\mathrm{Gr}_d^+(\mathbb{R}^n)$ is nonzero, where $\gamma^{(N)}$ denotes the tautological oriented bundle on $\mathrm{Gr}_d^+(\mathbb{R}^N)$. Choose $N \geq \max\{n, D + d + 2\}$ and set $E = \mathrm{St}_{d,N}^{\mathrm{row}}$. This is a compact $(N - d - 1)$-connected manifold with a free G-action (Husemoller, 1994). Appending zero columns gives an equivariant inclusion $\iota_N\colon X \to E$; on quotients it pulls $\gamma^{(N)}$ back to $\gamma^{(n)}$.

The map $x \mapsto (\mu_x, \iota_N(x))$ induces the factorization

$$X/G \xrightarrow{\bar{\mu}} (\mathcal{Y}_k \times E)/G \xrightarrow{\bar{\pi}_2} E/G, \qquad \bar{\pi}_2 \circ \bar{\mu} = \bar{\iota}_N.$$

The other projection, $\bar{\pi}_1\colon (\mathcal{Y}_k \times E)/G \to \mathcal{Y}_k/G$, is a closed surjection between compact metrizable spaces. Its fiber over $[y]$ is E/G_y, where G_y is finite by Lemma 9. Transfer for the finite covering gives $H^*(E/G_y; \mathbb{Q}) \cong H^*(E; \mathbb{Q})^{G_y}$ (Hatcher, 2002, Proposition 3G.1). Thus the fibers have vanishing reduced rational cohomology through degree $D + 1$. They are closed manifolds, so their singular and Čech cohomology agree. The Vietoris–Begle theorem (Spanier, 1966, Theorem 6.9.15) therefore gives

$$\bar{\pi}_1^*\colon \check{H}^D(\mathcal{Y}_k/G; \mathbb{Q}) \xrightarrow{\cong} \check{H}^D((\mathcal{Y}_k \times E)/G; \mathbb{Q}).$$

If $D > (k-1)(m+1)$, the group on the left vanishes by Lemma 9. Indeed, finite open covers then admit refinements whose nerves have dimension less than D, so the defining direct limit for Čech cohomology is zero. Naturality now gives

$$c(\gamma^{(n)}) = \bar{\mu}^* \bar{\pi}_2^* c(\gamma^{(N)}) = 0,$$

a contradiction. On the Grassmannians, Čech and singular cohomology agree. Hence $D \leq (k-1)(m+1)$, proving the support bound, and Lemma 11 supplies the stated degrees.

To compare the two groups, fix a reflection J and define $r\colon \mathrm{O}(d) \to \mathrm{SO}(d)$ by $r(g) = g$ for $\det g = 1$ and $r(g) = gJ$ otherwise. This map is continuous on the two components and commutes with left multiplication by $\mathrm{SO}(d)$. Pushing an orthogonal weighted frame forward by r gives a special-orthogonal weighted frame with no more atoms. Finally, whitening transfers both bounds to the matrix domain. □

The cup-length bound in Theorem 1. The Stiefel bundle and its oriented Grassmannian base are connected compact manifolds for $n > d$. Their distributional and analog sectional categories agree (Knudsen & Weinberger, 2026, Proposition 5.2). The tautological bundle is trivial after pullback, so $p_1(\gamma_d)$ and $e(\gamma_d)$ belong to $\ker q^*$. Lemma 11 gives products of lengths $\lfloor (n-d)/2 \rfloor$ in general and $n - d$ for even d. Equation (8) now yields the lower bounds in Theorem 1; the map r just constructed transfers them to $\mathrm{O}(d)$. Theorems 6 and 2 give its upper bound, and Corollary 4 gives the planar equality.

Well-conditioned inputs and support. The lower bounds already hold on $\mathrm{St}_{d,n}^{\mathrm{row}}$, where all nonzero singular values equal one. They therefore apply to any continuous equivariant weighted frame defined on a neighborhood of that manifold, irrespective of its behavior near rank loss. There is also a local consequence. For any weakly continuous measure map $A \mapsto \mu_A$, the set

$$\{A : |\operatorname{supp} \mu_A| \geq L\}$$

is open for each positive integer L. Indeed, at an input with at least L support points, choose L nonnegative continuous functions on G with disjoint supports and positive integrals against that measure. The integrals remain positive nearby, forcing at least one support point in each of these disjoint sets.

Take L to be the lower bound in Theorem 1. Some Stiefel input requires at least L atoms. For every $\varepsilon > 0$, a nonempty open neighborhood of that input therefore requires at least L atoms and has condition number less than $1 + \varepsilon$.

For example, the PCA frame of Puny et al. (2022) has at most eight atoms for $\mathrm{O}(3)$ and four for $\mathrm{SO}(3)$ on the simple-spectrum domain. With $n = N - 1$, Theorem 1 rules out continuous equivariant extensions with those support bounds to all full-rank centered inputs when $N \geq 20$ and $N \geq 12$, respectively.

The planar domain distinction. The space of all nonzero $2 \times n$ matrices retracts equivariantly to $S^{2n-1} \subset \mathbb{C}^n$. Its rotation quotient is the Hopf bundle over $\mathbb{CP}^{n-1}$. The Euler class of this circle bundle has nonzero $(n-1)$-st power. The same factorization, using $E = S^{2N-1}$ for sufficiently large N, $m = 1$, and $D = 2(n-1)$, gives support at least n. Conversely, (9), with all n columns included, defines a continuous frame there with at most n atoms. The improvement to $n - 1$ on the full-rank domain is possible because every selection of $n - 1$ columns then contains a nonzero column.

A.5 The Wronskian probe cover

Relation to rank condensers. A lossless rank condenser supplies a family $E_j \in \mathbb{R}^{d \times n}$ such that, for every rank-d matrix $M \in \mathbb{R}^{n \times d}$, some $E_j M$ has rank d. Take thin QR, $E_j^\top = P_j R_j$, and set $M = A^\top$. Then

$$E_j A^\top = R_j^\top (A P_j)^\top,$$

so this property is exactly coverage by the probes P_j. Ordinary Wronskians are classical rank condensers in characteristic zero (Forbes et al., 2014, Section 3.1 of the full version). A folded-Wronskian construction of Forbes & Guruswami (2015, Proposition 6.2 and Corollary 6.3) gives the same count. For example, over $\mathbb{R}$ one can take $E(t)_{ab} = (2^a t)^b$, $0 \leq a < d$, $0 \leq b < n$. After removing a power of t, $\det(E(t)M)$ is a nonzero polynomial of degree at most $d(n-d)$, so $d(n-d)+1$ distinct nonzero nodes suffice. This folded bank and our derivative bank need not coincide.

The algebraic coverage principle thus predates this work. Its role here is to provide a fixed linear atlas for equivariant learning. The support lower bounds, the passage from probes to weighted frames, and the sharp determinant-power regularity guarantee address the ensuing topological and analytic questions. We include the short proof to make that passage explicit.

Derivative probes. Let $\psi_0, \dots, \psi_{n-1}$ be degree-graded real polynomials, so $\deg \psi_j = j$, and write $b(t) = (\psi_0(t), \dots, \psi_{n-1}(t))^\top$. As in (10), form the $n \times d$ derivative matrix

$$\widetilde{B}(t) = \big[b(t), b'(t), \dots, b^{(d-1)}(t)\big].$$

Its first d rows form a triangular block with nonzero diagonal, so $\widetilde{B}(t)$ has rank d for every real t.

Lemma 12 (Polynomial Wronskian bound)**.** *If $f_1, \dots, f_d$ are linearly independent real polynomials of degree less than n, their Wronskian is nonzero and has degree at most $d(n-d)$.*

Proof. A constant invertible row operation replaces the family by polynomials with distinct leading degrees $0 \leq e_1 < \cdots < e_d \leq n-1$ and nonzero leading coefficients c_i. It multiplies the Wronskian

by a nonzero constant. For the new family, the coefficient of the highest possible power of t is

$$\left(\prod_i c_i\right) \det\left[(e_i)_{r-1}\right]_{i,r=1}^{d} = \left(\prod_i c_i\right) \prod_{i<j}(e_j - e_i) \neq 0,$$

where $(e)_s = e(e-1)\cdots(e-s+1)$ and $(e)_0 = 1$. The determinant is a Vandermonde determinant because the falling factorials are monic polynomials of successive degrees. Thus

$$\deg W = \sum_i e_i - \frac{d(d-1)}{2} \leq \sum_{r=0}^{d-1}(n-d+r) - \frac{d(d-1)}{2} = d(n-d).$$

□

Proof of Theorem 6. Choose a positive diagonal matrix $\Lambda(t)$ and take thin QR,

$$\widetilde{B}(t)\Lambda(t) = P(t)R(t), \qquad P(t) \in \mathrm{St}_{n,d}^{\mathrm{col}},$$

with positive diagonal in $R(t)$. For a row frame U with rows u_i, set $f_i(t) = u_i b(t)$. The basis coefficient map is an isomorphism, so the f_i are linearly independent. Their Wronskian is $W_U(t) = \det(U\widetilde{B}(t))$, a nonzero polynomial of degree at most $D = d(n-d)$. Hence it cannot vanish at all $D+1$ distinct real nodes t_j. Moreover,

$$\det(U\widetilde{B}(t_j)\Lambda(t_j)) = \det(UP(t_j)) \det R(t_j),$$

and both $\det \Lambda(t_j)$ and $\det R(t_j)$ are positive. Thus the fixed probes $P_j = P(t_j)$ include at least one with $\det(UP_j) \neq 0$ for every U. Whitening transfers coverage to every full-row-rank A. Local polar gauges and determinant weights give an equivariant measure supported on at most $D+1$ atoms. □

The experiments use normalized Legendre polynomials,

$$\psi_k(t) = \sqrt{\frac{2k+1}{2}}\, P_k(t), \qquad k = 0, \ldots, n-1, \tag{18}$$

where P_k is the usual Legendre polynomial. Each derivative column is multiplied by its positive inverse Euclidean norm, followed by unpivoted thin Householder QR with positive diagonal in R. For $K = d(n-d)+1$, the nodes are

$$t_j = \cos\left(\frac{(2j-1)\pi}{2K}\right), \qquad j = 1, \ldots, K. \tag{19}$$

The proof applies to any degree-graded polynomial basis, positive column scalings, and distinct real nodes. Equations (18)–(19), together with the stated QR convention, fix the probes used in the experiments.

A.6 Continuity, Lipschitz dependence, and higher derivatives

Fix a finite family of probes covering the full-rank domain and an integer $k \geq 1$. The determinant scores before and after whitening are

$$s_j(A) = \det(AP_j)^{2k}, \qquad \bar{s}_j(A) = \det(U(A)P_j)^{2k}.$$

Since $A = H(A)U(A)$ with $H(A) = (AA^\top)^{1/2}$, $s_j(A) = \det(H(A))^{2k}\bar{s}_j(A)$. Thus they give identical normalized weights. Their polar factors generally differ:

$$\mathrm{polar}(H(A)U(A)P_j) \neq \mathrm{polar}(U(A)P_j) \quad \text{in general.}$$

The particle-count study uses $k = 1$, the raw polar factor $Q_j(A) = \mathrm{polar}(AP_j)$, and the whitened scores for the weights.

For $G = \mathrm{O}(d)$, put $g_j^G = Q_j$. For $G = \mathrm{SO}(d)$, fix $J \in \mathrm{O}(d)$ with $\det J = -1$ and set

$$g_j^G(A) = \begin{cases} Q_j(A), & \det(AP_j) > 0, \\ Q_j(A)J, & \det(AP_j) < 0. \end{cases}$$

The two sign components of the valid chart are disjoint and open. On each, g_j^G is continuous, belongs to G, and satisfies $g_j^G(RA) = Rg_j^G(A)$ for $R \in G$. Define

$$Y_j(A) = \rho(g_j^G(A))f((g_j^G(A))^{-1}A), \qquad N_j(A) = \det(AP_j)^{2k}Y_j(A) \tag{20}$$

on the valid chart and extend N_j by zero elsewhere.

Continuity and equivariance in Theorem 2. On a valid chart, the gauge law gives $Y_j(RA) = \rho(R)Y_j(A)$, and the determinant score is invariant. If $A_r \to A_0$ while chart j becomes singular, then $\det(A_rP_j)^{2k} \to 0$. The gauges lie in compact G. Canonicalizing inputs in a sufficiently small compact neighborhood of A_0 therefore keeps them in a compact subset of the full-rank domain. Continuity of f makes Y_j bounded there, so $N_j(A_r) \to 0$. This proves that its zero extension is continuous, without assigning a polar factor at a singular matrix. Coverage gives $Z(A) = \sum_j \det(AP_j)^{2k} > 0$, so $F_{\text{soft}} = \sum_j N_j/Z$ is continuous. Equivariance also holds at chart failures because the zero extensions preserve it. The same argument with a continuous test function $\varphi(g_j(A))$ in place of $Y_j(A)$ proves weak continuity of the measure μ_A itself. □

A.6.1 Lipschitz dependence away from rank loss

Proposition 13 (Rank-margin Lipschitz estimate)**.** *Fix $\varepsilon, R > 0$ and let*

$$K_{\varepsilon,R} = \{A : \sigma_{\min}(A) \geq \varepsilon,\ \|A\|_F \leq R\}.$$

If f is bounded and Lipschitz on $K_{\varepsilon/2,R}$, then F_{soft} is Lipschitz on $K_{\varepsilon,R}$. Its Lipschitz constant depends only on ε, R, the backbone bound and Lipschitz constant, the fixed representation, probe bank, and exponent k, and the coverage margin

$$\Gamma_* = \min_{U \in \mathrm{St}_{d,n}^{\text{row}}} \left(\sum_j \det(UP_j)^{2k} \right)^{1/2} > 0.$$

Proof. Compactness of the Stiefel manifold and coverage give $\Gamma_* > 0$. The main issue is that the polar factor becomes sensitive near a singular projection. We show that the determinant weight compensates for this sensitivity. For a differentiable curve of invertible matrices $Z = QS$, polar perturbation gives

$$\|\dot{Q}\|_F \leq \frac{\|\dot{Z}\|_F}{\sigma_{\min}(Z)} \qquad \text{(Higham, 1986).}$$

The fixed-reflection correction has the same derivative bound on each sign component. The representation ρ is smooth with bounded derivative on compact G.

Take $A, B \in K_{\varepsilon,R}$ with $\|A - B\|_F < \varepsilon/2$. Along their line segment, Weyl's inequality gives rank margin at least $\varepsilon/2$, and the Frobenius norm is at most R. Thus the segment and all its canonicalized inputs lie in $K_{\varepsilon/2,R}$; convexity of $K_{\varepsilon,R}$ is not required. For a probe P, the only potentially unbounded term in the derivative of $N(A) = \det(AP)^{2k}Y(A)$ comes from differentiating the polar factor. Its coefficient is bounded because

$$\frac{\det(AP)^{2k}}{\sigma_{\min}(AP)} = \left(\prod_{i=1}^{d-1} \sigma_i(AP)^{2k} \right) \sigma_d(AP)^{2k-1} \leq R^{2kd-1}.$$

The determinant and its derivative are also bounded on the radius-R ball. On each nonsingular subinterval, the canonical input varies smoothly and the backbone is Lipschitz. The product rule holds almost everywhere and, with the polar estimate, gives a uniform bound on the derivative of N.

Along the segment, $\det(AP)$ is a polynomial of degree at most d. It either vanishes identically, in which case $N = 0$, or has finitely many zeros. Integrating the uniform derivative bound on the intervening subintervals and using the continuous zero extension at their endpoints gives a Lipschitz bound across the entire segment.

Finally, on $K_{\varepsilon,R}$,

$$Z(A) = \det(AA^\top)^k \sum_j \det(U(A)P_j)^{2k} \geq \varepsilon^{2kd}\Gamma_*^2.$$

The denominator is a polynomial with bounded derivative on the radius-R ball. Subtracting two quotients yields the Lipschitz bound for close pairs. For pairs at distance at least $\varepsilon/2$, let M bound the backbone. Since $\|F_{\mathrm{soft}}\| \le M$, $\|F_{\mathrm{soft}}(A) - F_{\mathrm{soft}}(B)\| \le (4M/\varepsilon)\|A - B\|_F$. Together these estimates cover all pairs. □

In particular, a locally Lipschitz backbone on the full-rank domain meets the proposition's hypothesis on every compact $K_{\varepsilon/2,R}$: a finite neighborhood cover controls nearby pairs, and boundedness controls pairs a fixed distance apart.

The estimate depends on both the input rank margin ε and the probe coverage margin Γ_*. Coverage ensures $\Gamma_* > 0$ at each fixed n, but the root argument does not bound its decay as n grows. Appendix B.2 describes the numerical checks used for the implemented probes.

A.6.2 Higher derivatives and sharpness

The same compensation extends to higher derivatives. We give the polar estimate first, since it identifies the possible singularity of a branch.

Lemma 14 (Derivatives of the polar factor)**.** *For every integer $r \ge 1$ there is C_r such that, for invertible real $d \times d$ matrices Z,*

$$\|D^r \operatorname{polar}(Z)\| \le C_r \sigma_{\min}(Z)^{-r}.$$

The norm on the left is the operator norm of the multilinear derivative.

Proof. Consider the symmetric matrix

$$H(Z) = \begin{pmatrix} 0 & Z \\ Z^\top & 0 \end{pmatrix}.$$

Its eigenvalues are $\pm\sigma_i(Z)$, and the upper-right block of $\operatorname{sign} H(Z) = H(Z)(H(Z)^2)^{-1/2}$ is $\operatorname{polar}(Z)$. The resolvent formula

$$\operatorname{sign} H = \frac{1}{\pi}\int_0^\infty \left((H - itI)^{-1} + (H + itI)^{-1}\right) dt$$

follows by diagonalizing H. If $s = \sigma_{\min}(Z)$, each resolvent has norm at most $(s^2 + t^2)^{-1/2}$. Its r-th derivative is a sum of products with $r + 1$ resolvents and r perturbation matrices. Differentiation under the integral is valid locally on the invertible locus, and gives

$$\|D^r \operatorname{sign} H\| \le C_r \int_0^\infty (s^2 + t^2)^{-(r+1)/2}\, dt = C_r' s^{-r}.$$

Taking the indicated block proves the assertion. Multiplication by the fixed reflection J gives the same estimate on either sign component of a special-orthogonal probe frame. □

Lemma 15 (Zero extension)**.** *Let $\Omega \subseteq \mathbb{R}^p$ be open, $S \subseteq \Omega$ relatively closed, and $r \ge 0$ an integer. Suppose $N : \Omega \to V$ vanishes on S and is C^r on $\Omega \setminus S$. If $D^j N(x) \to 0$ as $x \to x_0$ through $\Omega \setminus S$ for every $x_0 \in S$ and $0 \le j \le r$, then N is C^r on Ω, with all these derivatives zero on S.*

Proof. We induct on r. The case $r = 0$ is continuity at S. For $r \ge 1$, fix $x_0 \in S$ and $\varepsilon > 0$. Choose a closed ball about x_0 inside Ω on whose complement of S we have $\|DN\| \le \varepsilon$. For a sufficiently small h with $x_0 + h \notin S$, let t_* be the last $t \in [0, 1]$ for which $x_0 + th \in S$. Along the remaining segment, N is continuous up to t_*, vanishes there, and has derivative in the segment parameter bounded by $\varepsilon\|h\|$. Integrating gives $\|N(x_0 + h)\| \le \varepsilon\|h\|$; the same bound holds when $x_0 + h \in S$. Thus $DN(x_0) = 0$. Its zero extension is continuous and satisfies the hypotheses with $r - 1$, so induction completes the proof. □

Smoothness in Theorem 2. Fix a sufficiently small compact neighborhood K of a full-rank input. The set $G \cdot K$ is a compact subset of the full-rank domain. The derivatives of a C^{2k-1} backbone are bounded there, as are all derivatives of the representation on G.

For one probe, write $h(A) = \det(AP)$ and $s(A) = \sigma_{\min}(AP)$. On its valid chart, Lemma 14 and the chain rule give

$$\|D^r Y(A)\| \le C_r s(A)^{-r}, \qquad 0 \le r \le 2k-1,$$

where the case $r = 0$ means a uniform bound. All derivatives of h are bounded on K, and $|h(A)| \le \|AP\|^{d-1} s(A)$. In differentiating h^{2k} a total of a times, at least $2k-a$ factors remain undifferentiated. Hence

$$\|D^a(h^{2k})(A)\| \le C_a s(A)^{2k-a}, \qquad 0 \le a \le 2k-1.$$

The product rule therefore yields

$$\|D^r(h^{2k}Y)(A)\| \le C_r s(A)^{2k-r}, \qquad 0 \le r \le 2k-1. \tag{21}$$

Every one of these derivatives tends to zero at a singular probe.

Lemma 15 now shows that each weighted numerator is C^{2k-1} across its singular set. Summing and dividing by the positive smooth denominator Z preserves this regularity. □

A sharp example. Take $d = 2$, $n = 3$, $G = \mathrm{O}(2)$, scalar output with trivial representation, and the three coordinate probes P_{12}, P_{13}, P_{23}. They cover the full-rank domain. Let $f(B) = B_{13}$, a smooth linear backbone, and set

$$A(t) = \begin{pmatrix} t & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

This matrix has full row rank for every t. For $t \neq 0$, the 12 probe has determinant t, polar factor $\mathrm{diag}(\operatorname{sign} t, 1)$, and transported prediction $\operatorname{sign} t$. The 13 probe is singular, and the 23 probe has determinant -1, a constant polar factor, and transported prediction zero. Thus the entire averaged prediction is

$$F(A(t)) = \frac{t^{2k} \operatorname{sign} t}{1 + t^{2k}}, \qquad F(A(0)) = 0.$$

Its $2k$-th one-sided derivatives are $(2k)!$ and $-(2k)!$, so it is not C^{2k}. This proves sharpness even with a C^∞ backbone.

Sharpness for $\mathrm{SO}(d)$. The orientation correction does not improve the regularity guarantee. For any $d \ge 2$ and fixed $J \in \mathrm{O}(d)$ with $\det J = -1$, take $n = d+1$ and let P_i omit column i. These coordinate probes cover the full-rank domain. For $s \in \{1, -1\}$, set

$$A_s(t) = [\, I_d \;\; te_1 + se_2 \,].$$

Along this path $D_1 := \det(A_s(t)P_1) = (-1)^{d-1}t$, $|D_2| = D_{d+1} = 1$, and all other determinants vanish. Hence

$$F(A_s(t)) = \frac{t^{2k} Y_1(A_s(t)) + Y_2(A_s(t)) + Y_{d+1}(A_s(t))}{t^{2k} + 2},$$

where the last two branches are smooth for a smooth backbone.

The only nonconstant block of the polar factor of $A_s(t)P_1$ is that of $M_s(t) = \left(\begin{smallmatrix} 0 & t \\ 1 & s \end{smallmatrix}\right)$, from $\mathrm{span}(e_1, e_d)$ to $\mathrm{span}(e_1, e_2)$. For $t \neq 0$ it is

$$\operatorname{polar} M_s(t) = \frac{1}{\sqrt{1 + (1+|t|)^2}} \begin{pmatrix} -s \operatorname{sign} t & t + \operatorname{sign} t \\ 1 + |t| & s \end{pmatrix}.$$

Thus the full polar factor has analytic extensions from each side of zero. Its limits $Q_\pm$ agree on $u_s^\perp$ and satisfy $Q_+ = Q_-(I - 2u_s u_s^\top)$, where $u_s = (e_1 - se_d)/\sqrt{2}$ spans $\ker(A_s(0)P_1)$. After the fixed correction, the frame limits $g_\pm$ coincide exactly when $J = I - 2u_s u_s^\top$, for either parity of d. Since $u_1 \perp u_{-1}$, choose s for which they differ.

The matrices $B_\pm = g_\pm^\top A_s(0)$ are distinct because $A_s(0)$ has full rank, and have equal Frobenius norm. With trivial scalar representation and the linear backbone $f(B) = \langle B_+, B \rangle_F$, the one-sided limits of Y_1 therefore differ by

$$\|B_+\|_F^2 - \langle B_+, B_- \rangle_F = \tfrac{1}{2}\|B_+ - B_-\|_F^2 > 0.$$

The $2k$-th one-sided derivatives of $F(A_s(t))$ differ by $(2k)!/2$ times this quantity. Thus F need not be C^{2k} for $\mathrm{SO}(d)$ either.

Smooth cutoffs. The sharpness examples concern determinant-power weights. A different weight rule can give a smooth average for every smooth backbone. Set $a_j(A) = \det(AP_j)^2 / \sum_\ell \det(AP_\ell)^2$ and choose $0 < \tau < 1/m$. Let $\chi\colon \mathbb{R} \to [0,\infty)$ be smooth, zero on $(-\infty, \tau]$, and positive on (τ, ∞). Then

$$\widetilde{w}_j(A) = \frac{\chi(a_j(A))}{\sum_\ell \chi(a_\ell(A))}$$

is defined everywhere on the full-rank domain: some a_j is at least $1/m$. Near a failed probe, its weight vanishes identically, so its weighted contribution extends smoothly by zero. The resulting average is C^∞ for a C^∞ backbone. Relative cutoffs also occur in Dym et al. (2024); Pozdnyakov & Ceriotti (2023). They can suppress valid but low-weight branches; a reduction in average evaluation cost would require a separate measurement. The cubic cutoff used in the first two studies is only C^1 at its transition points.

A.6.3 Omitted mass under guards and truncation

Consider a rejection rule $I_j(A) \in \{0,1\}$ on the same probe bank. Write $\delta = \sum_{j:I_j=0} w_j$ for the omitted mass and, when $\delta < 1$, renormalize the retained weights as $\widetilde{w}_j = I_j w_j/(1-\delta)$. If $\|Y_j\| \le M$, then

$$\|F_{\mathrm{guard}}(A) - F_{\mathrm{soft}}(A)\| \le 2M\delta. \tag{22}$$

Removing the rejected contributions changes the sum by at most $M\delta$. Renormalization changes it by at most another $M\delta$. The same argument applies to deterministic top-k truncation.

This is a pointwise error bound. An input-dependent choice of retained branches can be discontinuous, including at ordering ties, so the continuity theorem for the full blend does not automatically extend to truncation.

A.7 The jump at a hard routing boundary

Suppose $A(t)$ crosses at $t = 0$ between two regular chart scores, their difference has nonzero derivative, no third chart ties, and Y_i, Y_j extend continuously to $A(0)$. Label the sides so hard routing returns $Y_i(A(t))$ for $t < 0$ and $Y_j(A(t))$ for $t > 0$. Then

$$\lim_{t\uparrow 0} F_{\mathrm{hard}}(A(t)) - \lim_{t\downarrow 0} F_{\mathrm{hard}}(A(t)) = Y_i(A(0)) - Y_j(A(0)).$$

This is (3). The tie convention affects only the value at $t = 0$, not the one-sided limits. With a tie rule selecting either of the two maximizers, the prediction is continuous at the crossing if and only if their transported predictions agree. This routing tie occurs where both probes remain invertible.

A.8 Four deterministic charts but two weighted frames

We prove (7) on $\mathrm{St}^{\mathrm{row}}_{3,4}$; whitening transfers both quantities to $\mathcal{M}^{\mathrm{fr}}_{3,4}$. For a row frame U with rows u_1, u_2, u_3, define the signed cofactor normal by

$$q_j(U) = (-1)^{j-1} \det U_{\widehat{j}}, \qquad j = 1, \dots, 4,$$

where $U_{\widehat{j}}$ omits column j. Laplace expansion and Cauchy–Binet give

$$Uq(U)^\top = 0, \qquad \|q(U)\|_2 = 1, \qquad q(RU) = \det(R) q(U).$$

These signs make the ordered basis $(q(U), u_1, u_2, u_3)$ positively oriented, since expansion of its determinant along its first row gives $\sum_j \det(U_{\widehat{j}})^2 = 1$.

Orthogonal complementation identifies $\mathrm{Gr}_3(\mathbb{R}^4)$ with $\mathbb{R}\mathrm{P}^3$. Its four coordinate-minor charts give genus at most four. If $a \in H^1(\mathbb{R}\mathrm{P}^3; \mathbb{F}_2)$ is the generator, then $a^3 \neq 0$ and $q^* a = 0$ for the frame bundle. Schwarz's cup-length bound therefore gives genus four.

To construct two weighted frames, identify $\mathbb{R}^4$ with the quaternions $\mathbb{H}$ in the oriented basis $(1, i, j, k)$. For a unit quaternion q, let $T(q)$ have rows given by the coordinate vectors of iq, jq, kq. Right multiplication by q is an orientation-preserving isometry, so (q, iq, jq, kq) is a positive orthonormal basis. Consequently U and $T(q(U))$ are equally oriented orthonormal bases of $q(U)^\perp$, and

$$c(U) = UT(q(U))^\top \in \mathrm{SO}(3).$$

Since $T(-q) = -T(q)$,

$$c(RU) = \det(R) R\, c(U).$$

It follows that

$$\mu_U = \tfrac{1}{2}\delta_{c(U)} + \tfrac{1}{2}\delta_{-c(U)} \tag{23}$$

is a continuous $\mathrm{O}(3)$-equivariant measure. A reflection exchanges the two transformed atoms. The two maps c and $-c$ are continuous, but neither is an $\mathrm{O}(3)$-equivariant gauge. The weights remain constant. Allowing reflections to exchange the two choices makes their measure equivariant.

Support one would give a continuous equivariant gauge and therefore a global section, contradicting genus four. Thus $\mathrm{wfr}_{\mathrm{O}(3)}(\mathcal{M}^{\mathrm{fr}}_{3,4}) = 2$. For $\mathrm{SO}(3)$, the map c itself is a global gauge. This example isolates the obstruction caused by reflections and shows why deterministic chart lower bounds do not automatically lower-bound weighted support.

A.9 The Gram quotient and equivariant vector outputs

Let

$$\mathcal{S}^+_{n,d} = \{\Sigma \in \mathbb{R}^{n\times n} : \Sigma = \Sigma^\top,\ \Sigma \succeq 0,\ \operatorname{rank}\Sigma = d\}.$$

Theorem 16 (Gram quotient and vector outputs)**.** *The Gram map induces a homeomorphism $\mathcal{M}^{\mathrm{fr}}_{d,n}/\mathrm{O}(d) \cong \mathcal{S}^+_{n,d}$. Every continuous $\mathrm{O}(d)$-equivariant map $F\colon \mathcal{M}^{\mathrm{fr}}_{d,n} \to \mathbb{R}^{d\times m}$ has the form $F(A) = AC(A^\top A)$ for a continuous $C\colon \mathcal{S}^+_{n,d} \to \mathbb{R}^{n\times m}$.*

Proof. An eigendecomposition $\Sigma = V\Lambda V^\top$, with V column-orthonormal and $\Lambda \succ 0$ of size d, gives the preimage $A = \Lambda^{1/2}V^\top$. Thus the induced Gram map is onto.

If $A^\top A = B^\top B$, then $\|Ac\|_2 = \|Bc\|_2$ for all $c \in \mathbb{R}^n$. Consequently $Ac \mapsto Bc$ is a well-defined isometry of $\mathbb{R}^d$, and $B = RA$ for a unique $R \in \mathrm{O}(d)$. This proves injectivity on the quotient.

The induced continuous bijection is proper. Over a compact subset of $\mathcal{S}^+_{n,d}$, the identity $\|A\|_F^2 = \operatorname{tr}(A^\top A)$ bounds the representatives. Any convergent sequence of representatives over that subset has a limit whose Gram matrix still has rank d, so the limit remains in the full-rank domain. Thus the inverse image before taking the quotient is compact, as is its quotient. A proper continuous bijection between these locally compact Hausdorff spaces is a homeomorphism.

Set $\widetilde{C}(A) = A^+F(A)$. The pseudoinverse is continuous at fixed rank and $(RA)^+ = A^+R^\top$, so $\widetilde{C}(RA) = \widetilde{C}(A)$. It descends to a continuous map C on the Gram quotient. Finally, $AA^+ = I_d$ gives $AC(A^\top A) = F(A)$. □

B Experimental methods and results

All three studies use labeled particles in $\mathbb{R}^3$ and predict forces directly. Table 1 distinguishes their atlas implementations. The moving-boundary study reuses trajectory checkpoints and adds consistency-trained models. Appendices B.1–B.3 specify the particle-count protocol, and Appendices B.4–B.5 specify the other two studies. All reported intervals are nominally 95%.

B.1 Particle-count study: synthetic force task

Inputs and targets. An input is a centered, full-row-rank configuration $X = (x_1, \dots, x_N) \in \mathbb{R}^{3\times N}$. Positions are the only supplied numerical features. Each model also has trainable labeled-slot embeddings. The target is a conservative $\mathrm{O}(3)$-equivariant force with zero total force. All predictions are projected to zero total force.

Regular-core generator. Keyed pseudorandom draws produce a centered Gaussian matrix $G \in \mathbb{R}^{3\times N}$. We whiten its rows to obtain $W = (GG^\top)^{-1/2}G$. The three singular scales are proportional to $(1, q_2, q_3)$ and normalized to unit Euclidean norm. For family identifier f, the stratum $h = f \bmod 4$ has endpoints (L_h, U_h). Independent keyed $u_3, u_2, u_r \sim \mathrm{Unif}(0,1)$ determine

$$q_3 = L_h + (U_h - L_h)u_3, \qquad q_2 = q_3 + (1-q_3)u_2, \qquad r = 0.9 + 0.2u_r.$$

Table 1: Atlas implementations in the three studies. Write $A = X_c H$ for fixed orthonormal contrasts of centered positions and $U = (AA^\top)^{-1/2}A$. The first two studies also include Quaternionic-2 and Dym-24 controls, specified in Appendix B.4.

Study	Frames and weights	Backbone and training
Trajectories, $N = 5$	Four coordinate minors; $\mathrm{polar}(UP_j)$; squared-minor scores with cubic cutoff for soft averaging, raw-score maximization for hard routing.	Width-96, three-block Transformer; 30,000 updates; separate soft and hard training.
Moving boundary, $N = 5$	Same frames and cutoff; positive score biases move routing boundaries; hard and soft evaluation share cached transported branches.	Reuses trajectory checkpoints; adds eight consistency-trained models after a three-seed coefficient screen.
Particle counts	$3N - 11$ Wronskian probes; $\mathrm{polar}(AP_j)$; normalized $\det(UP_j)^2$ weights without a cubic cutoff, subject to numerical guards.	Atlas models use a width-128, four-block Transformer; 40,000 updates and five paired seeds at each N. Gram and native baselines use separate architectures.

The four intervals are $[0.15, 0.25]$, $[0.25, 0.40]$, $[0.40, 0.70]$, and $[0.70, 1.00]$. With $a = (1, q_2, q_3)/\|(1, q_2, q_3)\|_2$, the input is $X = \sqrt{N}\, r\, Q\, \mathrm{diag}(a) W$, where $Q \in \mathrm{O}(3)$ has Haar distribution conditional on determinant $(-1)^f$. For a shared family identifier, stratum, quantiles, radius, and base orientation are reused across N, while the centered Gaussian is independently keyed by N.

Force oracle. With cyclic indices, write $d_i^{(a)} = x_{i+a} - x_i$. The fixed potential is

$$V_N(X) = \sum_{i=1}^{N} \sum_{a \in \{1,2\}} \alpha_{i,a} \exp\left(-\frac{\|d_i^{(a)}\|^2}{2\ell_a^2} \right) + \sum_{i=1}^{N} \beta_i \exp\left(-\frac{\|d_i^{(1)}\|^2 + \|d_i^{(2)}\|^2}{2\ell_3^2} \right) \tanh\left(\frac{\langle d_i^{(1)}, d_i^{(2)} \rangle}{\ell_3^2} \right), \tag{24}$$

where $(\ell_1, \ell_2, \ell_3) = (1, 1.4, 1.2)$, $\alpha_{i,a} = 0.85 + 0.30 u_{i,a}$, and $\beta_i = -0.25 + 0.50 v_i$, with fixed keyed uniform draws $u_{i,a}, v_i \in (0, 1)$. Keys omit N, so shared labels and offsets reuse coefficients across sizes. The target is $F_N = -\nabla_X V_N$. Cyclic coefficients make labels consequential, while distances and inner products ensure orthogonal symmetry.

Task validation. A prespecified diagnostic on 20,000 states per N checked force scale, local sensitivity, rank strata, and radial reference predictors. Across $N = 5, 8, 12, 16, 24, 32$, a six-scale radial reference using labeled cyclic edge classes had NRMSE 0.136–0.154; without those classes, error increased from 0.136 to 0.517. All six sizes passed the task checks, and no model was trained at $N = 32$. Different values of N define different prediction tasks, so comparisons of error and cost must be made within each size.

B.2 Numerical realization and validation

For centered X_c, a fixed Helmert contrast gives $A = X_c H \in \mathbb{R}^{3 \times (N-1)}$. Its one-indexed convention is

$$H_{rk} = \begin{cases} [k(k+1)]^{-1/2}, & r \le k, \\ -k[k(k+1)]^{-1/2}, & r = k+1, \\ 0, & r > k+1, \end{cases} \qquad k = 1, \ldots, N-1. \tag{25}$$

Since $H^\top H = I$ and $HH^\top = I - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$, we have $AA^\top = X_c X_c^\top$ and $X_c = AH^\top$. Thus A has the same singular values as X_c, and the constructions apply with $d = 3$ and $n = N - 1$. The backbone uses labeled particle coordinates while probes act on the contrast columns. The ideal scores,

Table 2: Posthoc geometric conditioning, 20,000 regenerated states per particle count. Quantiles use linear interpolation. c is the largest smallest singular value among probes, Z the score denominator, and κ_* the condition number of the determinant-maximizing whitened probe. These are empirical values, not worst-case bounds.

N	Probes	$\min c$	$q_{.01}(c)$	Median c	$q_{.01}(Z)$	Median Z	$q_{.99}(\kappa_*)$
5	4	0.0724	0.1567	0.6660	0.03907	0.9984	6.382
8	13	0.0581	0.1675	0.4353	0.02611	0.2974	6.813
12	25	0.0694	0.1690	0.3529	0.01576	0.1215	6.733
16	37	0.0984	0.1661	0.3114	0.01186	0.06642	6.293
24	61	0.1065	0.1574	0.2653	0.00699	0.02959	6.090

weights, frames, and prediction are

$$U = (AA^\top)^{-1/2}A, \qquad a_j = \det(UP_j)^2, \qquad w_j = \frac{a_j}{\sum_\ell a_\ell},$$
$$AP_j = Q_jS_j, \qquad \Phi_j(X) = Q_j h(Q_j^\top X_c), \qquad \widehat{F}_{\text{ideal}} = \sum_j w_j\Phi_j. \tag{26}$$

The polar factor uses AP_j, while its score uses U. Whitening multiplies every squared determinant by the same factor $\det(AA^\top)^{-1}$, so the normalized weights agree with (5) at $k = 1$. Thus (26) is the construction of Theorem 2 before numerical guards.

At $N = 5$, the Wronskian bank and the four coordinate-minor probes both have size four, but they are different banks. The condition-number bound for coordinate hard selection does not transfer to Wronskian selection. Indeed, for coordinate probes E_I, Cauchy–Binet gives

$$\sum_{|I|=3} \det(UE_I)^2 = \det(UU^\top) = 1.$$

The largest minor therefore has determinant magnitude at least $\binom{N-1}{3}^{-1/2}$. Every singular value of UE_I is at most one, so the selected minor obeys

$$\kappa(UE_{I_*}) \le \sqrt{\binom{N-1}{3}}.$$

This equals two when $N = 5$.

Empirical conditioning of the Wronskian bank. To measure coverage separately from learned predictions, we performed a posthoc diagnostic on 20,000 states per N, regenerated from the recorded final-test family identifiers with the generator in Appendix B.1 and the serialized binary64 probes. For each whitened input U, define

$$c(U) = \max_j \sigma_{\min}(UP_j), \qquad Z(U) = \sum_j \det(UP_j)^2, \qquad j_* = \arg\max_j \det(UP_j)^2.$$

The probe attaining $c(U)$ need not be the determinant-maximizing j_*. Table 2 summarizes both statistics and the denominator Z. Median c and Z decrease with N. These sampled values do not bound coverage over the Stiefel manifold or measure every low-weight branch. Moreover, the predictor uses AP_j, for which $\kappa(AP_j) \le \kappa(A)\kappa(UP_j)$ on an invertible probe.

Floating-point guards. A chart is retained in binary64 only if its SVD is finite, $\sigma_{\max} > 0$, $\sigma_{\min}/\sigma_{\max} > 64\varepsilon_{64} \approx 1.42 \times 10^{-14}$, and its orthogonality, reconstruction, symmetry, and positive-semidefiniteness residuals are at most 10^{-12}. For a matrix $M = AP_j$ passing the singular-value test, let Q be its computed polar factor and $B = Q^\top M$. The four residuals are

$$r_{\text{orth}} = \|Q^\top Q - I\|_F, \qquad r_{\text{rec}} = \frac{\|M - QB\|_F}{\max\{\|M\|_F, \tau_0\}},$$
$$r_{\text{sym}} = \frac{\|B - B^\top\|_F}{\max\{\|B\|_F, \tau_0\}}, \qquad r_{\text{psd}} = \frac{\max\{0, -\lambda_{\min}((B + B^\top)/2)\}}{\max\{\|B\|_2, \tau_0\}},$$

where $\tau_0 = 2^{-1022}$ is the smallest positive normal binary64 number. Here the first three norms are Frobenius norms, and the PSD denominator uses the operator norm.

With retention indicator r_j, omitted mass $\eta = \sum_j (1-r_j)w_j$, and $1-\eta > 0$, the numerical prediction is

$$\widetilde{w}_j = \frac{r_j w_j}{1-\eta}, \qquad \widehat{F}_{\mathrm{guard}} = \sum_j \widetilde{w}_j \Phi_j. \tag{27}$$

An unresolved selected chart raises an error. The continuity theorem applies to the ideal map; discrete guards are part of its numerical implementation.

On all full-rank samples in the pretraining validation cohort, denominators resolved, the numerical checks met tolerance, and guard-omitted mass was zero. During training, frames and weights were computed in binary64 without gradient tracking and cast to the float32 predictor. Active weights were checked for normalization at tolerance 10^{-6}. The training records do not retain an aggregate diagnostic of mass omitted by the guards. The posthoc omitted-weight records measure deliberate top-k truncation instead. Since the experiments learn forces directly, gradients of frames and weights with respect to positions are not used in training. Computing forces as gradients of an averaged energy would require differentiating through both.

B.3 Models, training, and statistical analysis

Table 3: Methods in the particle-count study. The three trained frame methods share the particle Transformer specified below. Parameter counts vary with N through the slot embeddings. Figure labels full-soft, hard-trained, and native denote Wronskian averaging, Wronskian hard selection, and the spherical-harmonic network, respectively.

Method	Frames evaluated per input	Parameters
Wronskian averaging	All resolved Wronskian frames from the $3N-11$ probes	0.532–0.534M
Wronskian hard selection	The Wronskian frame of largest weight	0.532–0.534M
Coordinate hard selection	The coordinate frame of largest minor score	0.532–0.534M
Gram model	None; invariant coefficients carried by X	0.733–0.734M
Spherical-harmonic network ($N = 12$)	None; equivariant message passing	0.502M
Evaluation-only variants of Wronskian-averaging checkpoints		
Posthoc hard selection	The Wronskian frame of largest weight	—
Top-k truncation	The k largest-weight Wronskian frames, renormalized	—

Wronskian averaging evaluates resolved frames in chunks of eight. All frame methods score the complete bank before selection. Posthoc hard selection and top-k truncation change only the evaluation of an averaging checkpoint. The following architectures are specific to the particle-count study.

Particle Transformer. A linear map $\mathbb{R}^3 \to \mathbb{R}^{128}$ of each position is added to a learned 128-dimensional slot embedding. Each of four blocks applies LayerNorm before four-head self-attention and before a $128 \to 256 \to 128$ feedforward map with GELU, with a residual connection around each sublayer. Dropout is zero. A final LayerNorm and linear $128 \to 3$ map produce each force, followed by projection to zero total force. All chart methods share this backbone architecture and paired initialization and data-order seeds.

Gram coefficient model. Write $G = X^\top X$. A learned 32-dimensional slot embedding concatenated with G_{ii} initializes a 128-dimensional node state h_i by a linear map. Each of four blocks computes

$$m_i = N^{-1/2} \sum_j M(h_i, h_j, G_{ij}, G_{ii}, G_{jj}), \qquad h_i \leftarrow h_i + U(\mathrm{LayerNorm}(m_i)).$$

Here M is a $259 \to 256 \to 128$ MLP and U a $128 \to 256 \to 128$ MLP, each with SiLU between its linear layers; blocks have separate parameters. A shared $257 \to 256 \to 1$ SiLU MLP applied to (h_i, h_j, G_{ij}) gives C_{ij}. The output is XC, projected to zero total force. This model learns invariant coefficients using all pairwise inner products; it has no spatial frame selection.

Native equivariant model. The $N = 12$ model uses three complete-graph message-passing blocks with hidden representation $24 \times 0e + 18 \times 1o + 12 \times 2e$ in e3nn notation. Slot embeddings initialize the even scalars; the other channels start at zero. Messages use spherical harmonics of degrees $0, 1, 2$ and an eight-feature radial basis without a distance cutoff, scaled by $1/(1 + r^2)$. Blocks have irrep-wise RMS normalization, equivariant linear self-connections, SiLU scalars, and sigmoid gates on vector and tensor channels. The readout is one $1o$ vector per node, projected to zero total force.

Training and checkpoint selection. At each N, the fixed calibration, training, development, checkpoint-validation, and final-test cohorts contain 20,000, 100,000, 10,000, 10,000, and 20,000 states, respectively. Calibration, training, and development geometries were inherited from engineering work; model weights, optimizer states, seeds, checkpoint decisions, and pilot metrics were not. Each run used one NVIDIA A40.

Both this study and the trajectory study use AdamW, effective batch size 512 with microbatches of 128, weight decay 10^{-5}, gradient clipping at 1, a 2,000-update warmup, and cosine decay to zero. Training is float32, with mixed precision and TF32 disabled and detached binary64 frame preprocessing. Table 4 gives the study-specific choices.

Table 4: Training and checkpoint selection. Both studies run to their fixed update budget without early stopping. Checkpoint validation is separate from the development split in the particle-count study.

Setting	Particle-count study	Trajectory study
Updates	40,000	30,000
Learning rate	3×10^{-4} for every method	3×10^{-4}; radial e3nn uses .01, selected from $\{.001, .003, .01\}$.
Validation	Development every 500 updates; 16 candidate checkpoints at $2500, 5000, \ldots, 40000$.	Every 1,000 updates.
Selection	Minimum finite NRMSE on the entire checkpoint-validation split; earlier checkpoint wins exact ties.	Replace the incumbent when normalized validation NRMSE improves by more than 10^{-6}.

The 100 main runs (four methods, five sizes, five seeds) and five native runs at $N = 12$ all selected update 40,000. Relative convergence beyond this budget is therefore unknown. Initialization and minibatch-order streams are keyed by both N and seed index. Within a size, the three Transformer methods share initialization and data order; other architectures have distinct initialization keys but the same data order. Across sizes, both streams change. This is the dependence structure used below.

Pooled NRMSE. For method m and particle count N, the reported estimate is

$$\widehat{\mathrm{NRMSE}}_{m,N} = \left(\frac{\sum_{s=1}^{5} \sum_i \|\widehat{F}_{m,N,s,i} - F_{N,i}\|_F^2}{5 \sum_i \|F_{N,i}\|_F^2} \right)^{1/2}. \tag{28}$$

Squared errors are pooled over model seeds and final-test geometries before taking the square root.

Paired bootstrap. The analysis uses 20,000 PCG64 replicates with the recorded seed and linear quantiles. At each N, five model seed indices were multinomial-resampled, independently across N and with pairing preserved across methods. Within each of the four rank strata, 5,000 geometry families were resampled; the geometry count draw was shared across methods and coupled values of N. Separately resampling model and geometry factors follows the crossed-factor structure of the pigeonhole bootstrap (Owen, 2007); within-size method pairing and cross-size geometry coupling preserve the shared factors. Every replicate recomputes (28) from resampled squared-error and target-energy sums before forming contrasts. Intervals condition on the realized training dataset, oracle, architectures, and protocol. With only five independent training realizations per method and N, they are approximate uncertainty summaries for nonlinear NRMSE contrasts, not finite-sample coverage guarantees.

For a prespecified family of contrasts $\theta_1, \dots, \theta_J$, let $\hat{\theta}_j$ be the reported estimate and θ^*_{bj} its value in bootstrap replicate b. Define

$$s_j = \mathrm{sd}_{b=1}^{20000}(\theta^*_{bj}),$$
$$q = Q_{0.95}\left(\max_{1\le j\le J}\left|\frac{\theta^*_{bj}-\hat{\theta}_j}{s_j}\right|\right), \qquad I_j = [\hat{\theta}_j - qs_j, \hat{\theta}_j + qs_j]. \tag{29}$$

The standard deviation uses divisor 19999; Q_p is the linearly interpolated empirical quantile. The primary one-sided bounds use the following percentile-style lower envelope:

$$q_- = Q_{0.95}\left(\max_j \frac{\hat{\theta}_j - \theta^*_{bj}}{s_j}\right), \qquad L_j = \hat{\theta}_j - q_- s_j. \tag{30}$$

This chooses a common standardized offset so that approximately 95% of the joint bootstrap vectors lie coordinatewise above the lower endpoints. For one component, $L = Q_{0.05}(\theta^*)$; for several, the maximum accounts for simultaneous downward excursions. This calibrates the lower tail of the bootstrap distribution, unlike a basic error-inversion bound, which uses the opposite centered tail. Zero-variance components have point intervals and are omitted from the maximum. Each five-size contrast family has its own critical value; the 13 posthoc top-k contrasts form a separate family. Individual method intervals and the single native-equivariant contrast use the 2.5th and 97.5th percentiles directly. The primary slope is equal-weight OLS of R_N against $\log N$, refitted in every replicate, with the 5th percentile as its one-sided lower bound.

For the five R_N, the critical values are $q = 2.44219$ and $q_- = 2.13354$.

B.4 Original seven-method trajectory study

Force task and initial conditions. This study uses a different potential from (24). For unit particle masses, its energy is

$$V(X) = \frac{1}{4}\sum_{i<j} a_{ij}(\|x_i - x_j\|^2 - \ell_{ij}^2)^2 + 0.05 \sum_{i<j<k} (\langle x_j - x_i, x_k - x_i\rangle - c_{ijk})^2, \tag{31}$$

and the force is $-\nabla V$. The reference configuration is

$$X_0 = \begin{pmatrix} .92 & -.48 & -.73 & .07 & .22 \\ .10 & .95 & -.35 & -.80 & .10 \\ 0 & .15 & .85 & -.75 & -.25 \end{pmatrix}.$$

Here ℓ_{ij}^2 and c_{ijk} are its squared distances and indicated inner products. The ten a_{ij}, in lexicographic order of pairs, are fixed NumPy PCG64 uniform draws on $[0.6, 1.2]$.

For independent centered standard Gaussian 3×5 matrices Z_x, Z_v, $a \sim \mathrm{Unif}(0.9, 1.1)$, and Haar $Q \in \mathrm{O}(3)$, set

$$X(0) = Q(aX_0 + \sigma_x Z_x), \qquad \dot{X}(0) = Q\sigma_v Z_v.$$

The three physical regimes use $(\sigma_x^{\min}, \sigma_x^{\max}, \sigma_v)$ equal to $(.08, .30, .15)$, $(.12, .38, .20)$, and $(.15, .45, .25)$, with σ_x uniform on its interval. Initial states must have $\kappa(XH) \le 3$ and minimum pair distance at least 0.4. Velocity Verlet uses step 0.000625 for horizon 10, storing the 100 noninitial states at intervals of 0.1. A trajectory is accepted only if all integration states are finite, full rank, have $\kappa(XH) \le 10$, and minimum pair distance at least 0.25. Numerical rank uses threshold $10^{-10}\max\{\sigma_{\max}, 1\}$. There are at most 50 proposals per trajectory slot. Each regime contributes 1,000 training, 100 validation, and 400 test trajectories. Training uses 32 fixed, approximately equally spaced stored states per trajectory.

Models and training. The shared chart backbone has width 96, three Transformer blocks, four heads, feedforward width 192, GELU, and zero dropout, with the same pre-normalization and position/slot embedding scheme described above. It has 225,699 parameters. The original Gram model instead flattens the ordered 5×5 Gram matrix, projects its 25 entries to width 96, applies four residual LayerNorm–MLP blocks with hidden width 384 and SiLU, and reads out 25 coefficients after

a final LayerNorm. It returns $X_c C$, centered over output particles. The radial e3nn model uses the construction above with hidden representation $16 \times 0e + 12 \times 1o + 8 \times 2e$ and 223,628 parameters.

Soft and hard atlas both use polar factors of the four row-whitened coordinate minors UP_j. Soft atlas uses the cubic cutoff $\tau = 10^{-6}$ in (33) with zero bias. Hard atlas maximizes the raw score $\det(UP_j)^2$, before applying any cutoff, with lowest-index tie breaking. In contrast, the particle-count study uses polar factors of AP_j. Augmented raw trains the same backbone on independent Haar orthogonal transformations. Training and checkpoint selection follow Table 4; the learning-rate screen is specific to the radial e3nn baseline.

Crossings and rollouts. Crossings are strict exchanges of the two largest raw squared-minor scores, bracketed between integrator states. Bisection on the interpolated input locates the equality to tolerance 10^{-10}. A qualifying crossing has third-score gap at least 10^{-4} and a stable selected chart on each side over a 0.02 time window. Events lie in $[0.5, 8]$, events within 0.05 are deduplicated, and the first qualifying event is used. The routing margin is the largest score minus the second largest; near-boundary states have margin at most 0.03, and far-boundary states have margin greater than 0.2.

Learned-force Verlet rollouts start 0.5 before the crossing and end 2 after it, with the same integration step. Final position error is centered RMSD without Procrustes alignment. A numerical failure is a nonfinite position, velocity, or force, or centered RMS radius exceeding ten times that of X_0. The composite event is the first such failure or sampled centered RMSD exceeding 0.25, sampled every 0.1 over the 2.5-unit rollout. Event incidence averages the seed–trajectory indicators; these crossed observations are not independent replicates.

Evaluation and uncertainty. The original $N = 5$ study evaluated fixed checkpoints on 1,200 trajectory-disjoint test trajectories, including 302 qualifying first chart crossings. Soft atlas, Quaternionic-2, and hard atlas each used eight independent training seeds. Dym-24, augmented raw, Gram, and e3nn radial message passing used three each. The primary soft–hard comparison used all eight paired training seeds. Comparisons with three-seed methods used the three corresponding hard-atlas seeds. A 10,000-replicate crossed bootstrap independently resampled paired seeds and trajectories within physical regime, keeping methods paired. Intervals are marginal, unadjusted for multiplicity, and use linearly interpolated 2.5th and 97.5th percentiles. For force-error reductions, each replicate computes NRMSE within each selected seed, averages over seeds, and forms the ratio of these means. The comparisons with three-seed methods have fewer independent training replicates.

Quaternionic-2 implements the explicit two-atom $\mathrm{O}(3)$ weighted frame on 3×4 matrices. For the columns a_i of the unwhitened reduced matrix A, each ordered pair $i \neq j$ gives a Gram–Schmidt frame and its reflection partner. Each partner has weight

$$\frac{1}{2}\frac{r_i}{\sum_\ell r_\ell}\frac{b_{ij}}{\sum_\ell b_{i\ell}}, \qquad r_i = \|a_i\|, \quad b_{ij} = \|a_i \times a_j\|,$$

after setting r_i to zero at or below $64\varepsilon_{64}\max\{1, \|A\|_F\}$ and b_{ij} to zero at or below $64\varepsilon_{64}\max\{1, \|A\|_F^2\}$. A pair with zero first-level weight is inactive; a positive first-level weight requires a positive area denominator. The twelve ordered pairs and their reflection partners give at most 24 atoms. This comparator uses numerical zero thresholds, not the atlas's cubic cutoff.

The soft–hard improvement is concentrated near selection boundaries: NRMSE falls by 34.42% (paired interval 26.88–40.59%) near the boundary, but increases by 2.64% (0.47–4.92%) far from it. The near-minus-far reduction is 37.05 percentage points (28.83–44.24). Final position error falls by 15.03% (10.17–19.62%), using the mean across seeds of within-seed median RMSD. These comparisons do not separate continuity, averaging, and learned branch specialization.

Every method completes all rollouts. The sole composite event is a soft rollout whose sampled RMSD exceeds 0.25 at elapsed time 2.3, among $8 \times 302 = 2{,}416$ seed–trajectory pairs (incidence $1/2416$). Lower position error therefore does not imply fewer adverse events. Gram and radial e3nn attain lower errors than soft atlas on this task (Table 5).

Numerical limitations. Full-support recombination of soft top-4 and Dym top-24 predictions retains nominal-bound flags, with maximum discrepancies 2.493×10^{-6} and 2.290×10^{-6}. The recombination code renormalizes float32 weights even at full support, while the reference uses their

Table 5: Original trajectory study. NRMSE and completion are arithmetic means over model seeds. Calls are logical branch or backbone evaluations. The last column averages each seed's median final centered-position RMSD among completed rollouts, without Procrustes alignment.

Method	Seeds	Calls	Overall NRMSE	First-crossing NRMSE	Completion	Mean seed-median RMSD
Soft atlas	8	4	0.00676	0.0110	1.000	0.0143
Quaternionic-2	8	2	0.00817	0.0145	1.000	0.0173
Hard atlas	8	1	0.00747	0.0181	1.000	0.0169
Dym-24	3	24	0.00603	0.00844	1.000	0.0121
Augmented raw	3	1	0.00920	0.0128	1.000	0.0193
Gram	3	1	0.00153	0.00270	1.000	0.00193
e3nn radial MP	3	1	0.00388	0.00628	1.000	0.00548

original values; retained mass can also round above one, making the nominal omitted-mass bound negative.

B.5 Moving-boundary intervention

Paths and selection rules. Starting from a configuration $A_0 = X_{c,0}H$ in a separate cohort generated by (31) and the three regimes above, set

$$A_p(t) = A_0 R_{ij}(t), \qquad X_{c,p}(t) = A_p(t)H^\top, \tag{32}$$

where $R_{ij}(t) \in \mathrm{SO}(4)$ is a right Givens rotation. It acts on contrast columns, not on spatial coordinates, and preserves $AA^\top$. Choose its sign convention so the signed cofactor vector satisfies

$$q_i(t) = q_i(0)\cos t + q_j(0)\sin t, \qquad q_j(t) = -q_i(0)\sin t + q_j(0)\cos t.$$

The other two cofactors are fixed. Each accepted path comes from a different source trajectory; candidates are ordered by a geometry-only hash before model predictions are computed.

The study uses 504 physical paths: three regimes, six unordered chart pairs, and 28 paths per regime–pair stratum. It evaluates achieved input separations 0.006, 0.002, 0.0006, and 0.0002, with 0.0006 declared primary. Seven biases $\delta \in \{-0.18, -0.12, -0.06, 0, 0.06, 0.12, 0.18\}$ define seven selection rules, also called routers. Each rule has a boundary at a different location along a path. We evaluate all seven rules at all seven locations, giving 49 combinations for each path and model seed.

For row-orthonormal $U \in \mathbb{R}^{3\times 4}$, define the signed cofactors and their squared scores by

$$q_j(U) = (-1)^{j-1}\det U_{\widehat{j}}, \qquad s_j = q_j(U)^2, \qquad \sum_j s_j = 1.$$

For cutoff $\tau = 10^{-6}$, let

$$c_\tau(s) = \begin{cases} 0, & s \le \tau, \\ 3z^2 - 2z^3, & \tau < s < 2\tau, \quad z = (s-\tau)/\tau, \\ 1, & s \ge 2\tau. \end{cases}$$

For a centered bias vector b, effective scores and soft weights are

$$e_j = s_j c_\tau(s_j), \qquad \alpha_j^{(b)} = \frac{e_j \exp(b_j)}{\sum_\ell e_\ell \exp(b_\ell)}. \tag{33}$$

For controlled chart pair (i, j), set $b_i = \delta$, $b_j = -\delta$, and all other entries to zero. Hard selection takes the lowest-index maximizer of $e_j \exp(b_j)$. Matched averaging (matched soft in the tables) combines the same cached transported predictions using $\alpha^{(b)}$. On the active plateau of the cutoff, the boundary solves

$$g_\delta(t) = \log q_i(t)^2 - \log q_j(t)^2 + 2\delta = 0. \tag{34}$$

The implementation scans 1,025 points in $[-0.4, 0.4]$ and locates roots by bracketed float64 rootfinding, to equality tolerance 2×10^{-12}. It retains paths only if all seven roots and their six nonswitching controls exist at every separation. Required bounds are $|g'_\delta| \ge 0.1$, biased third-score gap at least 0.01, minimum pair distance 0.25, oracle force norm at most 100, $\kappa(A) \le 10$, and displacement from the base at most one third of its Frobenius norm.

Table 6: Moving-boundary contrasts at separation 6×10^{-4}. Each row measures the switching router's response minus the average response of the six nonswitching routers on the same inputs. "Matched soft" is an evaluation rule, not a training condition. The marginal paired percentile intervals are unadjusted across the six rows.

Checkpoint training	Evaluation	Estimate	Marginal paired 95% interval
Soft atlas	Moved hard	0.1650	$[0.1553, 0.1763]$
Soft atlas	Matched soft	-1.53×10^{-7}	$[-2.95, -0.20] \times 10^{-7}$
Ordinary hard	Moved hard	0.03279	$[0.03091, 0.03475]$
Ordinary hard	Matched soft	3.00×10^{-7}	$[2.32, 3.80] \times 10^{-7}$
Hard + consistency	Moved hard	0.01603	$[0.01511, 0.01700]$
Hard + consistency	Matched soft	7.65×10^{-8}	$[4.10, 12.87] \times 10^{-8}$

Residual change and paired contrast. Let s index model seeds and p paths. For the boundary of selection rule δ and input separation Δ, let $t^{\pm} = t^{\pm}_{p\delta}(\Delta)$ denote the two path parameters on either side of that boundary with

$$\Delta = \|X_{c,p}(t^+) - X_{c,p}(t^-)\|_F.$$

Write $R_{sp,\delta'}(t) = \widehat{F}_{sp,\delta'}(X_p(t)) - F(X_p(t))$ for the residual evaluated with rule δ'. Its response at the boundary of rule δ is

$$J_{sp,\delta' \leftarrow \delta}(\Delta) = \frac{\|R_{sp,\delta'}(t^+) - R_{sp,\delta'}(t^-)\|_F}{F_{\mathrm{RMS,cal}}}, \qquad F_{\mathrm{RMS,cal}} = 4.6036908809. \tag{35}$$

This dimensionless change in prediction residual is normalized by force scale, without division by input separation. The contrast is

$$T_{sp\delta}(\Delta) = J_{sp,\delta \leftarrow \delta}(\Delta) - \frac{1}{6} \sum_{\delta' \neq \delta} J_{sp,\delta' \leftarrow \delta}(\Delta). \tag{36}$$

Its first term uses the rule that switches between the two inputs; the second averages the six rules that do not switch there. The comparison therefore holds both inputs and all transported predictions fixed.

We average the seven boundary contrasts within each path, the 28 paths within each of the 18 regime–pair strata, and then the strata and eight model seeds, with equal weights at each stage. The 10,000-replicate paired bootstrap independently resamples eight seeds and 28 paths within each regime–pair stratum, using linear percentile intervals. All selection rules, boundary locations, aggregation rules, and input separations remain together within their clusters.

For soft-trained models at the primary separation, the switching hard response is 0.16513 and direct transported-branch disagreement is 0.16519, a 0.035% relative mismatch using unrounded values. Figure 2 shows the former approaching the latter as separation contracts, while nonswitching and matched-soft responses shrink. The tiny matched-soft contrast in Table 6 is the difference of two shrinking responses, rather than either response.

Consistency training. All branch outputs in this study include the zero-total-force projection $\Pi(B) = B - \frac{1}{5}(B\mathbf{1})\mathbf{1}^\top$. Write $\Phi_j(x) = \Pi(Q_j f(Q_j^\top X_c))$ for these projected transported predictions. With zero-bias soft weights, the training penalty is

$$\overline{\Phi}(x) = \sum_j \alpha_j(x)\Phi_j(x), \qquad L_{\mathrm{cons}} = \frac{\sum_{x \in \mathcal{B}} \sum_j \alpha_j(x) \|\Phi_j(x) - \overline{\Phi}(x)\|_F^2}{|\mathcal{B}|\, \mathbb{E}_{\mathrm{cal}} \|F\|_F^2}. \tag{37}$$

Here $\mathcal{B}$ is a training batch and $F_{\mathrm{RMS,cal}} = (\mathbb{E}_{\mathrm{cal}}\|F\|_F^2)^{1/2}$ is the same normalization as in (35). The validation screen uses a different quantity. Let $\mathcal{C}$ be the set of geometry-defined unbiased validation first crossings, with switching pair $i(x), j(x)$ at x. For each tuning seed s, it computes

$$D_s = \frac{1}{|\mathcal{C}| F_{\mathrm{RMS,cal}}} \sum_{x \in \mathcal{C}} \|\Phi_{s,i(x)}(x) - \Phi_{s,j(x)}(x)\|_F.$$

Table 7: Primary comparison across particle counts. Intervals for the five R_N values are simultaneous two-sided 95% intervals. The last column gives the simultaneous one-sided lower bound.

N	Full-soft pooled NRMSE	Gram pooled NRMSE	R_N [two-sided interval]	One-sided lower bound
5	0.003150	0.001464	-0.7665 $[-0.8447, -0.6884]$	-0.8348
8	0.005828	0.004717	-0.2115 $[-0.2409, -0.1821]$	-0.2372
12	0.007152	0.008363	0.1565 $[0.1171, 0.1959]$	0.1220
16	0.008128	0.012725	0.4483 $[0.3917, 0.5048]$	0.3989
24	0.009380	0.020638	0.7886 $[0.7061, 0.8711]$	0.7165

Each qualifying crossing has equal weight. This is the norm of the two-branch disagreement at a crossing, averaged over crossings; it is not the four-branch weighted variance in the training penalty. A three-seed development screen selected coefficient $\lambda = 1$ from $\{0, 0.01, 0.03, 0.1, 0.3, 1\}$ as follows. A coefficient is eligible if its three-seed mean validation NRMSE is at most 1.05 times the contemporaneous zero-control mean and each seed's NRMSE is at most 1.10 times its paired zero-control value. Among eligible coefficients, the screen minimizes the mean of D_s over the three tuning seeds; coefficients within 1% of the minimum are tied, and the smallest wins. Final training uses eight paired seeds and inherits the trajectory optimizer and validation rule. The selected coefficient is the grid's upper endpoint; the resulting hard contrast remains positive (Table 6).

At the primary separation, Quaternionic-2 and Dym-24 controls had residual changes 3.17×10^{-5} and 1.84×10^{-5}, respectively, with no response pattern following the moved wall. In rollouts, consistency training reduced trajectory error among rollouts completed by both methods. Its hard-routed predictions had exactly the same completion, failure, composite-event incidence, and restricted time as ordinary hard-trained hard routing. Ordinary hard routing minus the soft evaluation of those same hard-trained checkpoints had composite-event difference -6.51×10^{-5}, with interval $[-3.89 \times 10^{-4}, 0]$.

B.6 Results across particle counts

Primary endpoint and decision rule. For the pooled estimates in Equation (28), define

$$R_N = \log\left(\frac{\widehat{\mathrm{NRMSE}}_{\mathrm{Gram,N}}}{\widehat{\mathrm{NRMSE}}_{\mathrm{soft,N}}}\right). \tag{38}$$

Positive R_N favors full-soft. The prespecified joint criterion requires positive one-sided lower bounds for both the slope of R_N against $\log N$ and R_{24}, with the latter simultaneous across the five sizes (Appendix B.3).

The slope is 0.9873, with two-sided percentile interval $[0.936, 1.036]$ and one-sided lower bound 0.9451. Both primary lower bounds are positive. The fit summarizes $N = 5$–24, rather than an asymptotic law. All five paired seeds agree with the within-size ordering in Table 7.

Hard-routing contrasts. Every simultaneous interval in Table 8 favors averaging, and all five paired seeds agree in each comparison. Posthoc hard selection changes averaging and branch use at a fixed checkpoint; its contrast is not an isolated effect of continuity.

At $N = 12$, native-equivariant minus full-soft NRMSE is 0.01096 (paired percentile interval $[0.01046, 0.01143]$), with five positive seed directions. This concerns the specified spherical-harmonic architecture under the common learning rate; the trajectory study screens its radial baseline's learning rate separately. It does not compare native equivariant architectures as a class.

Posthoc truncation. At $N = 24$, top-8 has NRMSE 10.0% above full-soft and remains below all three hard-selection conditions and Gram (Figure 4). It inherits full-support training cost. Changes in the retained set can introduce discontinuities, so the continuity theorem does not apply.

Table 8: Additive NRMSE contrasts with Wronskian averaging, multiplied by 10^3 for display. Positive values favor averaging. Brackets give simultaneous two-sided 95% intervals within each five-size contrast family. Posthoc hard uses averaging-trained checkpoints; the other columns use separately trained hard models.

N	Posthoc hard	Wronskian hard	Coordinate hard
5	11.430 [10.663, 12.197]	0.941 [0.796, 1.086]	0.309 [0.204, 0.414]
8	3.239 [3.051, 3.427]	0.862 [0.666, 1.057]	0.683 [0.442, 0.923]
12	3.884 [3.659, 4.109]	1.539 [1.169, 1.908]	1.344 [0.939, 1.749]
16	4.634 [4.432, 4.836]	1.527 [1.146, 1.909]	1.524 [1.206, 1.842]
24	6.253 [5.940, 6.565]	2.547 [2.173, 2.921]	2.490 [1.969, 3.011]

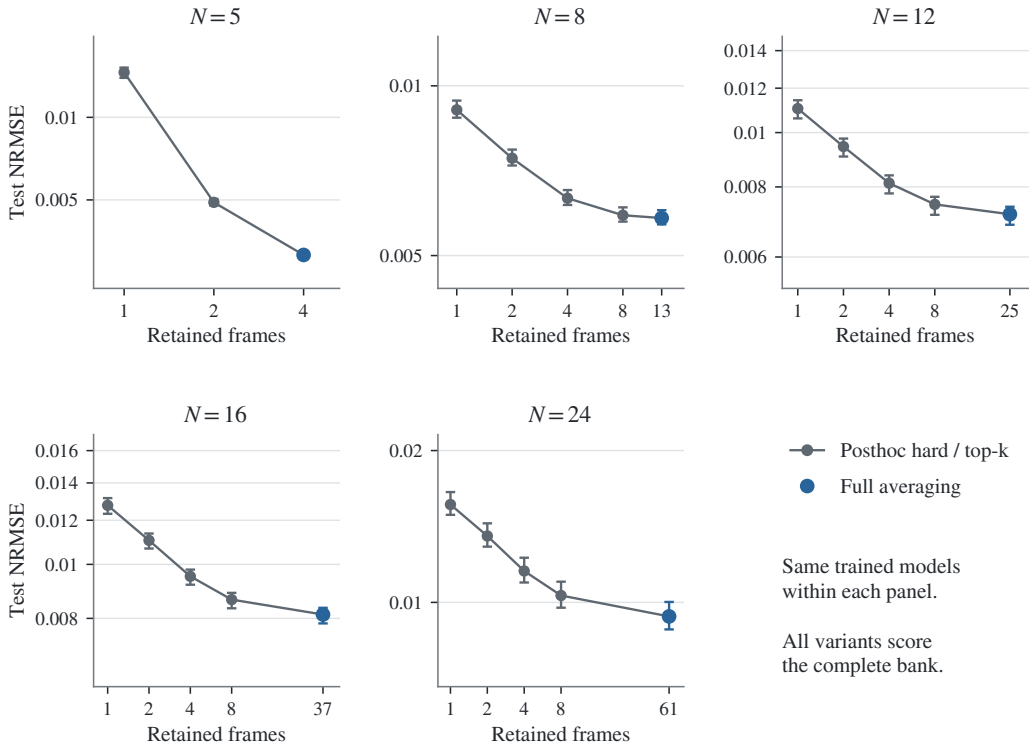

Figure 4: Within-task posthoc truncation, with marginal intervals. Top-k retains and renormalizes the k largest weights of each full-soft-trained checkpoint. At $N = 5$, top-4 is full support and top-8 is not distinct. Every plotted condition scores the complete Wronskian bank before selecting the retained branches.

B.7 Compute records

For m dense probes, forming AP_j, scoring, and transforming particle inputs and outputs cost $O(mnd^2)$, polar decompositions cost $O(md^3)$, and the backbone calls cost $O(mC_f(N))$. The probes are constructed once. With d fixed and $m = \Theta(N)$, the scoring/transport cost is $O(N^2)$; a fixed-width dense-attention backbone costs $O(N^2)$ per call and hence $O(N^3)$ over the full bank. Chunking bounds working memory but does not change this arithmetic count. Top-k evaluation still incurs full-bank scoring before reducing the number of backbone calls.

Figure 5 uses every 500-update development observation for all five full-soft and Gram seeds at $N = 12, 24$. Cumulative worker time includes training, development evaluation, checkpoint writes, and mirror waits within each attempt. The separate medians of time and NRMSE are descriptive summaries; they neither establish convergence nor provide a matched-time final-test comparison.

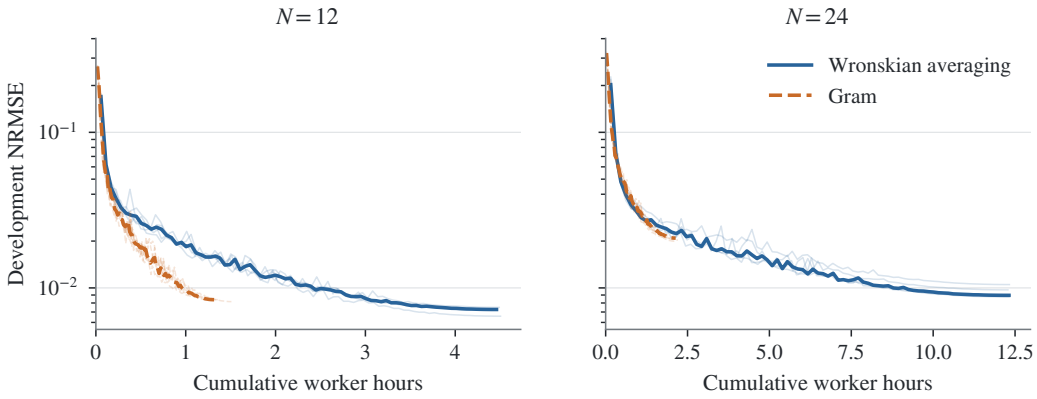

Figure 5: Development NRMSE against cumulative A40 worker time at fixed N. Faint curves show all five seeds. Thick curves join the separate medians of time and error at each shared 500-update observation, without interpolation or smoothing; they need not be individual training runs. Each panel concerns one task; these are development observations.

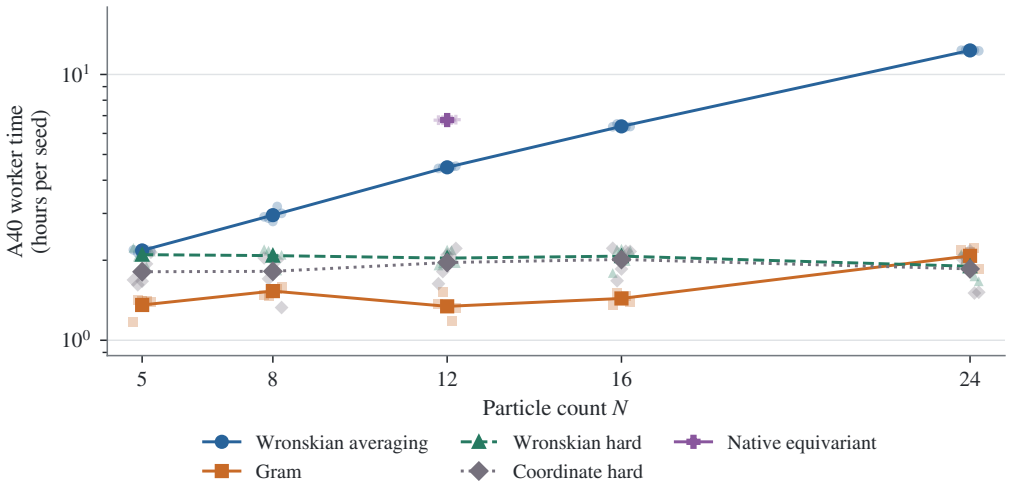

Figure 6: Recorded A40 worker time per seed, including evaluation, checkpoint writes, and waits. Faint points show all five runs; larger markers are arithmetic means.

At Gram's final median worker times, 1.316 hours at $N = 12$ and 2.137 at $N = 24$, its median development errors are 0.00840 and 0.02088. The latest averaging observations at no greater median time are at 1.279 and 2.007 hours, with errors 0.01578 and 0.02287. These nearby recorded observations favor Gram within its shorter budget; we do not interpolate errors to an identical time.

Figure 6 reports all runs. At $N = 24$, mean worker times are 12.32, 2.08, 1.898, and 1.856 hours for full-soft, Gram, Wronskian hard, and coordinate hard, respectively. Full-soft uses 5.93 times Gram's time and 6.49 times Wronskian hard's time, from unrounded records.

Full-soft's lower final error comes with substantially greater computation. The uniform posthoc A40 inference-timing protocol covers hard and top-k evaluation but lacks full-soft latency.