NAILONG: TRANSFER LAWS FOR SHARED LIE GENERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Two operator families can share a continuous symmetry and still differ in the residual that breaks it. The method extracts the shared Lie algebra of infinitesimal generators and freezes it on the target. A jointly equivariant predictor is capped by the group average; a trunk plus a free residual is already outside that class, so the trunk is a comparison, not a corollary of the ceiling. With the shared generators known in the control chart, freezing them and fitting the residual drops matched-design target OLS risk from $(s{+}r)n^2\sigma^2/N$ to $rn^2\sigma^2/N$. The algebra can be recovered from a finite source by identifying the span, aligning it to the target chart, and freezing it. On noisy generator OLS that procedure approaches the oracle residual rate at $\hat{\kappa}{=}3.72{\pm}0.31$, against oracle 3.03 and scratch 6.01. On a separate labeled spherical field probe, generator regression followed by a freeze of Φ reaches residual-of-Φ $R^2{=}0.87$ on the lab axis; a jointly equivariant residual stops at the group-average fraction 0.034.

1 INTRODUCTION

A rotating fluid on the sphere can keep the same $\mathrm{SO}(3)$ action across two operator families while the residual changes—Coriolis, topography, or a jet. A jointly equivariant predictor of the target cannot beat the equivariant fraction of the residual (Theorem 2): 0.034 on the spherical GRF probe, and 0.08 on linearised spherical shallow water. That bound is on the whole map remaining equivariant. A source-trained jointly equivariant trunk stores the group average $P_G T_{\mathrm{src}} = \Phi + P_G R_{\mathrm{src}}$, which equals Φ only when the source residual is orthogonal to every equivariant map. An equivariant trunk plus an unrestricted residual is already outside that class; Residual Pathway Priors already use that composition (Finzi et al., 2021a). Whether freezing the extracted algebra beats freezing a source network and adapting a free head is a comparison, not a corollary of the ceiling.

What the method extracts is the shared Lie algebra of infinitesimal generators. If that algebra is frozen inside Φ and a general residual is learned, target OLS risk drops from $s{+}r$ generators to r (Theorem 1). For an exact freeze, the variance reduction depends on the number of known generator coefficients. The brackets decide which s-plane that count applies to: a commuting triple, or a spectral factor of the same observations, need not be the shared plane, and the freeze then stays at a floor independent of N (Theorem 3).

Figure 1 and Algorithm 1 summarize identification, alignment, and freezing.

Contributions.

1. A jointly equivariant predictor is capped; a free residual is not. The population optimum over all jointly equivariant maps is $P_G T_{\mathrm{src}} = \Phi + P_G R_{\mathrm{src}}$, which equals Φ only if $P_G R_{\mathrm{src}} = 0$. The spherical probes illustrate this class restriction; they do not establish an advantage over trunk-plus-free-head transfer.
2. Residual rate is a dimension count, and the wrong plane floors. A correct freeze collapses $\kappa_{\mathrm{scr}}{=}s{+}r$ to $\kappa_{\mathrm{frz}}{=}r$. Under the non-compensating residual assumptions, a wrong or commuting freeze leaves an N-independent floor.
3. The span is recoverable. Identify, align, and freeze approach the oracle rate from finite-source noisy generator OLS ($\hat{\kappa}{=}3.72{\pm}0.31$ at $(512, 128)$; oracle 3.03, scratch 6.01).

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.

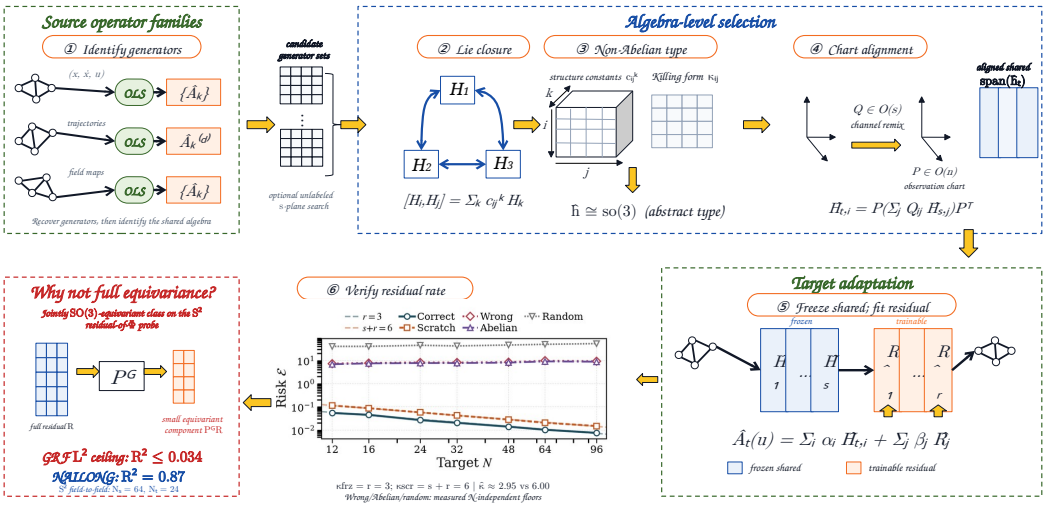

Figure 1: Identify, align, and freeze. Source estimation recovers candidate generators; when channels are unlabeled, a type-check keeps the simple factor; alignment places that span on the target; freeze locks it inside Φ and fits only the residual.

2 Related Work

Neural operators. Fourier Neural Operators and relatives (Li et al., 2021; Kovachki et al., 2023; Lu et al., 2021) learn maps between function spaces. Group-equivariant FNOs (Helwig et al., 2023) and steerable layers (Weiler & Cesa, 2019; Kondor & Trivedi, 2018) encode a known group in the weights. Equivariance then restricts the hypothesis class; it does not by itself guarantee statistical or optimization optimality. Lie-algebra canonicalization (Shumaylov et al., 2025) uses given infinitesimal actions to align inputs for pre-trained models, while physics-aligned canonical FNOs (Xu et al., 2026) jointly learn frame alignment and prediction using known PDE symmetries. Transfer and multi-task operator learning (Goswami et al., 2022; Subramanian et al., 2023) share trunks or fine-tune those layers. Residual Pathway Priors (Finzi et al., 2021a) already let an equivariant path coexist with an unrestricted residual; that composition is not a new transfer object. These approaches motivate separating a symmetry prior from the residual that violates it; they do not imply that every transfer architecture has a prescribed group.

Generator discovery. Otto et al. (2025) unify enforce/discover/promote tasks via Lie derivatives; Hu et al. (2025), Gabel et al. (2023; 2025), Dehmamy et al. (2021), and Markov & Hvatov (2026) recover infinitesimal generators. Lie Neurons and relatives (Lin et al., 2024; Kim et al., 2026; Shutty & Wierzynski, 2023; Mironenco & Forré, 2024) build equivariant heads from structure constants or decompositions; Li et al. (2025) connect latent symmetry mixtures to sample efficiency. These methods recover generators, or they build a head that uses an algebra.

Transfer theory. Domain-adaptation theory relates transfer to distribution discrepancy (Ben-David et al., 2010), while shared-representation theory also tracks task diversity (Tripuraneni et al., 2020). Written on generator residuals, that relatedness is the shared Lie span. Equivariant training (Cohen & Welling, 2016; Bronstein et al., 2021; Finzi et al., 2021b) uses a known group; it does not select the group from data. This paper takes the recovered algebra as the relatedness and asks which freeze is worth transferring once the residual changes.

3 SETUP

3.1 OPERATOR FAMILIES AND SHARED GENERATORS

Let $\mathcal{H}$ be a finite-dimensional observation space of dimension n (e.g., Galerkin coefficients). A continuous group action $\rho : \mathcal{G} \times \mathcal{H} \to \mathcal{H}$ induces infinitesimal generators

$$A_k \;=\; \frac{\partial}{\partial \theta_k}\Big|_{\theta=0} \rho(\exp(\theta X_k), \cdot) \in \mathfrak{gl}(\mathcal{H}).$$

We work with a Lie algebra $\mathfrak{g} \subset \mathfrak{gl}(n)$ of dimension $s{+}r$, decomposed as

$$\mathfrak{g} \;=\; \mathfrak{h} \;\oplus\; \mathfrak{w}, \qquad \dim \mathfrak{h} = s, \quad \dim \mathfrak{w} = r, \tag{1}$$

where $\mathfrak{h}$ is shared across source and target families and $\mathfrak{w}$ is target-specific. Equation (1) is the embedded dynamical span that is frozen. A shared continuous symmetry supplies a Lie subalgebra, not an arbitrary s-plane in $\mathfrak{gl}(n)$. Knowing the abstract type—for instance that the shared factor is $\mathfrak{so}(3)$—does not name the embedded basis; freeze uses the subspace in the shared chart. $\mathfrak{so}(3)$ is the smallest compact simple algebra with nontrivial brackets. For commuting normal generators, joint spectral information can identify a common eigenbasis; recovering the generator span still requires sufficiently informative observations.

When the freeze applies. Three conditions are enough. Source and target share an embedded Lie span $\mathfrak{h} \subset \mathfrak{gl}(n)$. That span can be placed on the residual sample: the target chart is identifiable, or it can be aligned. The residual class D cannot absorb an arbitrary mismatch of the shared span.

Linear generator response model. To expose the residual-dimension law, consider

$$Y \;=\; \sum_{k=1}^{s+r} Z_k A_k^\star \;+\; E, \qquad \mathbb{E}[E] = 0, \quad \mathbb{E}[E_{ij}^2] = \sigma^2, \tag{2}$$

with design coordinates $Z = (Z_1, \dots, Z_{s+r})$ and unknown generators $A_k^\star$ in a fixed basis of $\mathfrak{g}$. Write $A(u) = \sum_k u_k A_k^\star$ for a control u. Scratch estimates all $s{+}r$ generators; a *correct freeze* of the shared algebra estimates only the r residual generators after plugging in $\mathfrak{h} = \operatorname{span}\{A_1^\star, \dots, A_s^\star\}$.

3.2 RISK AND SAMPLE COMPLEXITY

The primary risk is the control-averaged prediction risk

$$\mathcal{E}(\widehat{H}, \widehat{R}) \;:=\; \mathbb{E}_{u\sim\mu} \big\| A(u) - \widehat{A}(u) \big\|_F^2, \tag{3}$$

for a test law μ with second moment Σ_μ. The generator laws and Figures 2–4 use this risk; field experiments use the separate score defined in Theorem 2. We study

$$\kappa \;:=\; \frac{N}{n^2\sigma^2}\, \mathbb{E}[\mathcal{E} \mid \text{design}] \tag{4}$$

in the variance-dominated regime. Bias floors appear when $\mathcal{E}$ does not vanish with N.

Assumption 1 (Matched empirical design)**.** *Noise entries are i.i.d. with variance σ^2, independent of the design; estimation is OLS on vectorized responses. The test law equals the empirical training measure $\mu = \frac{1}{N}\sum_{\ell=1}^{N} \delta_{u^{(\ell)}}$, and the empirical Gram $\bar{\Sigma}_N = \frac{1}{N}\sum_\ell u^{(\ell)} u^{(\ell)\top}$ is positive definite.*

Population isotropy $\mathbb{E}[uu^\top] = I$ is *not* assumed for the exact identities; Corollary 2 gives a PE sandwich when the test law differs from the design.

4 TRANSFER LAWS

If the shared span is given, matched-design OLS risk is a dimension count.

Theorem 1 (Residual-rate transfer law)**.** *Under Assumption 1, for the risk (3),*

$$\mathbb{E}[\mathcal{E}_{\mathrm{scr}} \mid \mathrm{design}] = \frac{(s{+}r)\, n^2\, \sigma^2}{N}, \qquad \kappa_{\mathrm{scr}} = s{+}r, \tag{5}$$

$$\mathbb{E}[\mathcal{E}_{\mathrm{frz}} \mid \mathrm{design}] = \frac{r\, n^2\, \sigma^2}{N}, \qquad \kappa_{\mathrm{frz}} = r, \tag{6}$$

whenever the freeze uses the true shared generator tuple in the observed control coordinates. In particular $\mathbb{E}[\mathcal{E}_{\mathrm{frz}} \mid \mathrm{design}]/\mathbb{E}[\mathcal{E}_{\mathrm{scr}} \mid \mathrm{design}] = r/(s{+}r)$*. The identities use only the dimension of that span: an* $O(s)$ *remix of the true shared subspace matches residual rate (Section 6); Lie brackets are not an ingredient of the OLS algebra.*

Proof sketch. Vectorizing (2), matched empirical-design OLS has conditional covariance $(\sigma^2/N)\bar{\Sigma}_N^{-1} \otimes I_{n^2}$; averaging the prediction risk over $\mu = \widehat{\mu}_N$ cancels $\bar{\Sigma}_N^{-1}$ and yields $mn^2\sigma^2/N$ with $m = s{+}r$ (scratch) or $m = r$ (correct freeze). Full proof: Appendix B.1. □

Knowing only an abstract algebra or an uncalibrated basis is insufficient for this identity: the frozen matrices must reproduce the shared response for the observed controls. A basis remix therefore requires the inverse remix of the control coordinates. The n^2 factor counts unrestricted matrix entries in each fitted generator; imposing a smaller output slot changes that variance count. Theorem 3 instead isolates a population bias under a restricted residual class.

The generator variance law and the following projection bound address different model classes: an equivariant trunk plus an unrestricted residual lies outside the jointly equivariant class.

Theorem 2 (Equivariant misspecification)**.** *Let* $G = \mathrm{SO}(3)$ *act unitarily on a Hilbert space of fields and jointly on controls. For any input–control law* μ *let* $\bar{\mu} = \int_G g_\sharp \mu \, dg$ *be its Haar symmetrization (*$\bar{\mu} = \mu$ *when* μ *is already* G*-invariant), and write* L^2 *for* $L^2(\bar{\mu})$*. Let* Φ *be jointly* G*-equivariant transport,* T *a target field-to-field map, and* $R = T - \Phi$*. Residual-of-*Φ $R^2(E) := 1 - \|E - T\|_{L^2}^2/\|R\|_{L^2}^2$ *equals* 0 *at* $E = \Phi$ *and* 1 *at* $E = T$*. The Reynolds operator* P_G *is the orthogonal projection in this* L^2 *onto jointly equivariant maps, so every jointly equivariant* E *satisfies* $R^2(E) \le \|P_G R\|_{L^2}^2/\|R\|_{L^2}^2$*. The population minimizer over all jointly equivariant maps is* $P_G T_{\mathrm{src}} = \Phi + P_G R_{\mathrm{src}}$ *in* L^2*. A trained trunk attains it only with adequate capacity and optimization. This equals* Φ *if and only if* $P_G R_{\mathrm{src}} = 0$*, and equals* T_{src} *if and only if* R_{src} *is itself equivariant. On a multiplicity-free spherical-harmonic representation, linear intertwiners are irrep scalars (Hall, 2015).*

A freeze of the wrong s-plane does not inherit Theorem 1. The floor statements require a residual slot that cannot absorb a shared-plane mismatch.

Assumption 2 (Direct-sum splitting)**.** *There is an orthogonal decomposition* $V = V_H \oplus V_R$ *such that shared generators act on* V_H *and annihilate* V_R*, while residual generators act on* V_R *and annihilate* V_H*. Write*

$$\mathcal{S}_R := \big\{ X \in \mathfrak{gl}(n) : X(V_H) = \{0\},\ X(V_R) \subset V_R \big\} \cong \mathfrak{gl}(V_R)$$

for the residual slot*. A residual estimator is* model-class *if* $\widehat{R}_j \in \mathcal{S}_R$ *for all* j *(true residuals* R_j *lie in* $\mathcal{S}_R$ *by the splitting).*

Theorem 3 (Wrong-freeze bias floor)**.** *Assume Assumption 2 and model-class residual estimators. Let* $\widetilde{H}$ *be a frozen shared tuple with*

$$\delta_0^2 := \inf_{Q \in O(s)} \sum_{i=1}^{s} \Big\| \widetilde{H}_i - \sum_{k=1}^{s} H_k Q_{ki} \Big\|_F^2 > 0, \tag{7}$$

and assume the mismatch is shared-slot supported*: after a Procrustes gauge realizing (7),* $\Delta H_i := H_i - \widetilde{H}_i$ *satisfies* $\Delta H_i(V_H) \subset V_H$ *and* $\Delta H_i(V_R) = 0$ *(hence* $\Pi_R(\Delta H_i) = 0$*, where* Π_R *projects onto* $\mathcal{S}_R$*). Let* μ *be any test law with finite second moment* $\Sigma = \mathbb{E}[uu^\top]$ *on* $u = (\alpha, \beta)$ *satisfying* $\Sigma_{\alpha\alpha} := \mathbb{E}[\alpha\alpha^\top] \succeq \gamma_\alpha I_s$ *(arbitrary cross-blocks* $\Sigma_{\alpha\beta}, \Sigma_{\beta\beta}$*, including* $\beta \not\equiv 0$*). Then the model-class population risk after optimal residual fitting satisfies*

$$\inf_{\widehat{R} \in \mathcal{S}_R^r} \mathcal{E}(\widetilde{H}, \widehat{R}) \;=\; \mathrm{tr}\big(\Sigma_{\alpha\alpha}\, M\big) \;\ge\; \gamma_\alpha\, \delta_0^2, \tag{8}$$

where $M_{ik} := \langle \Delta H_i, \Delta H_k \rangle_F$*, and consequently* $\mathcal{E}(\widetilde{H}, \widehat{R}) \geq \gamma_\alpha \delta_0^2$ *for every model-class* $\widehat{R}$*. In particular the floor is independent of* N *and of* $\mathrm{Cov}(\alpha, \beta)$*.*

Proof sketch. Under Assumption 2, decompose $\mathfrak{gl}(n) = \mathcal{S}_R \oplus \mathcal{S}_R^\perp$ in the Frobenius product. Model-class residuals live in $\mathcal{S}_R$, so $A(u) - \widehat{A}(u)$ splits into an $\mathcal{S}_R^\perp$ piece $\sum_i \alpha_i \Pi_\perp(\Delta H_i)$ (independent of β and of $\widehat{R}$) and an $\mathcal{S}_R$ piece amenable to residual fitting. Shared-slot support forces $\Pi_R(\Delta H_i) = 0$ and $\Pi_\perp(\Delta H_i) = \Delta H_i$, so the irreducible piece is $\mathbb{E}\| \sum_i \alpha_i \Delta H_i \|_F^2 = \mathrm{tr}(\Sigma_{\alpha\alpha} M)$ and the $\mathcal{S}_R$ piece can be driven to zero by $\widehat{R}^\star = R$. Loewner gives $\mathrm{tr}(\Sigma_{\alpha\alpha} M) \geq \gamma_\alpha \,\mathrm{tr}(M) = \gamma_\alpha \delta_0^2$. (Contrast: unconstrained residuals in $\mathfrak{gl}(n)$ have Schur-complement floor $\mathrm{tr}(\Sigma_{\alpha|\beta} M)$, which can vanish if α is linearly predictable from β; Appendix B.7.) Full proof: Appendix B.7, Theorem 7(ii). □

On a shared-channel probe $\beta = 0$, the same floor holds for every residual estimator (Appendix B.7). In the n=6 block embedding of Appendix B.3, every spectral or Ad-natural map of a commuting triple and a shared drift returns a commuting triple, so Theorem 3 supplies the floor; no Borel function of those observables selects a faithful $\mathfrak{so}(3)$ completion on every compatible world.

On fields, a residual class D hides a wrong Φ exactly when $\Phi - \widetilde{\Phi}$ already lies in D.

Theorem 4 (Residual-class compensation)**.** *Let* Φ *be the residual-free shared propagator,* T *the target field-to-field map,* $R = T - \Phi$*, and* P_D *the* L^2 *projector onto a closed linear residual class* D*. Correct and wrong freezes plus a residual in* D *attain*

$$\|T - E_{\mathrm{cor}}\|_{L^2}^2 = \|(I - P_D)R\|_{L^2}^2, \qquad \|T - E_{\mathrm{wrong}}\|_{L^2}^2 = \|(I - P_D)(R + \Phi - \widetilde{\Phi})\|_{L^2}^2. \tag{9}$$

Thus $E_{\mathrm{wrong}} = E_{\mathrm{cor}}$ *iff* $\Phi - \widetilde{\Phi} \in D$ *on the data measure. Proof: Appendix E.16.*

5 Method: Identify → Align → Freeze

Algorithm 1 NAILONG estimator: identify, align, freeze

1: **Source.** Misspecified linear velocity/snapshot OLS $\dot{x} \approx \big(A_0 + \sum_i \alpha_i \widehat{S}_i\big) x$ (no quadratic column).

2: **Identify.** If channels are unlabeled, type-check k-subsets and keep the unique simple 3-plane (Killing rank 3) under Assumption 2; else take the $\mathfrak{so}(3)$ factor. Algebraic type is a check of the abstract algebra.

3: **Align.** Procrustes / intertwiner on k target generators, disjoint from the residual sample.

4: **Freeze.** Freeze $\widehat{\mathfrak{h}}$ inside Φ; fit a residual in class D on N target samples.

Identify. Source OLS sees $s{+}r$ generators and may lack channel labels. When channels are unlabeled, the step type-checks k-subsets and keeps the unique simple 3-plane. When the three controls are labeled, as in the released S^2 scripts, it is ordinary generator regression with a skew-symmetric projection; the barcode search is not run. The type-check tests the abstract algebra; it does not name an embedded basis.

Align. That plane is placed on k held-out target generators by Procrustes, or by an intertwiner when the observation chart itself is a conjugacy. Those samples are disjoint from the residual sample, so the freeze is not scored on the data that chose the gauge.

Freeze. The aligned span is locked inside Φ, and only a residual in class D is fit.

Field scoring. Field probes never observe $A(u)$. They run velocity or snapshot OLS, then score residual-of-Φ R^2 on the field map.

6 Experiments

Setup. Unless noted: $s{=}r{=}3$, $n{=}6$, $\sigma{=}0.08$, ambient algebra $\mathfrak{so}(3) \oplus \mathfrak{w}$. Designs are matched isotropic; we report $\hat{\kappa} = N\,\mathcal{E}/(n^2\sigma^2)$ and large-N floors. The finite-source cell uses 16 repetitions. Additional scans are in Appendix F.

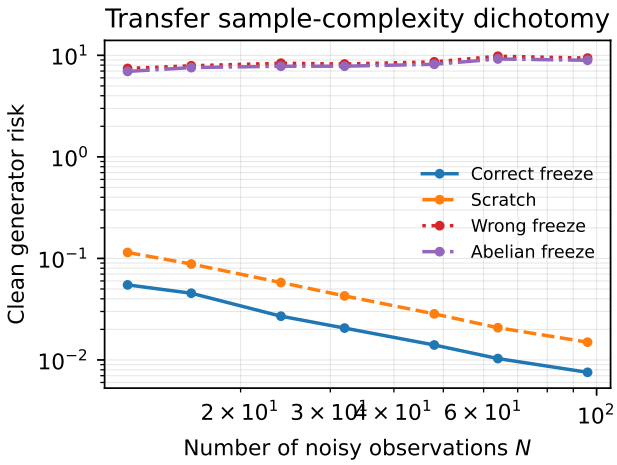

Figure 2: Main dichotomy under matched designs (s=r=3, n=6, σ=0.08). Correct freeze has residual complexity ($\hat{\kappa} \approx 2.95$); scratch has ambient complexity ($\hat{\kappa} \approx 6.00$); wrong/Abelian/spectral freezes leave floors with near-zero log-log slope.

6.1 The residual-rate law is sharp

Figure 2 shows the predicted split: correct freeze $\hat{\kappa} \approx 2.95$, scratch $\hat{\kappa} \approx 6.00$. Wrong and Abelian freezes stay essentially flat. Both decaying curves have slope near -1; their constants distinguish r from s+r (Theorem 1).

A same-span $O(s)$ remix of the true shared subspace, with controls transformed consistently, matches $\hat{\kappa} \approx 2.96$; a random s-plane of the same dimension does not (Appendix F.10). Dimension sweeps and unmatched-design traces appear in Appendix F.

6.2 A finite source approaches the residual rate

Theorem 1 assumes the shared span. Algorithm 1 recovers it. On $D_1 = \mathfrak{so}(3) \oplus \text{Heisenberg} \to D_2 = \mathfrak{so}(3) \oplus \text{toral}$ (Figure 3), the estimated span is type-checked, aligned, and frozen. At $(N_{\text{src}}, N_{\text{tgt}}) = (512, 128)$ the estimated freeze attains $\hat{\kappa}$=3.72±0.31 against oracle 3.03 and scratch 6.01. The excess tracks the source Procrustes gap (Corollary 7). This panel uses labeled source controls and the known shared embedding as its alignment reference; the unknown-chart experiment is reported separately in Appendix E.4. For a fixed imperfect source estimate, target variance decreases with N_{tgt} while the uncompensated source error remains. Thus a value near r at one target size is not an exact residual-rate identity for arbitrarily large target samples. Figure 3 varies source size to distinguish reducing this bias from fitting the target residual more accurately. The same picture holds when the source never observes $A(u)$ directly, when the six OLS channels are unlabeled, and when the target chart is an unknown conjugacy: the recovered freeze stays near residual rate, and the wrong plane does not (Appendices C.3 and E.4).

6.3 Wrong, Abelian, and spectral freezes stay at a floor

A freeze of the wrong s-plane, or of a commuting or spectral substitute, does not inherit residual rate (Theorem 3). In Figure 2 those curves are flat, while the correct freeze decays with slope near -1. Table 1 and Figure 4 quantify these misspecifications. On the same source matrices, identify, align, and freeze yield $\hat{\kappa}$=3.74, against 536 (PCA), 1399 (skew-SVD), and 1513 (DMD).

The flat curves are consistent with uncompensated misspecification. Theorem 3 proves this behavior under its support and residual-class assumptions; the full-trunk row additionally omits target residual adaptation. The same-data spectral experiments compare concrete PCA, skew-SVD, and DMD estimators; the impossibility theorem in Appendix B.3 concerns the narrower class of observables specified there.

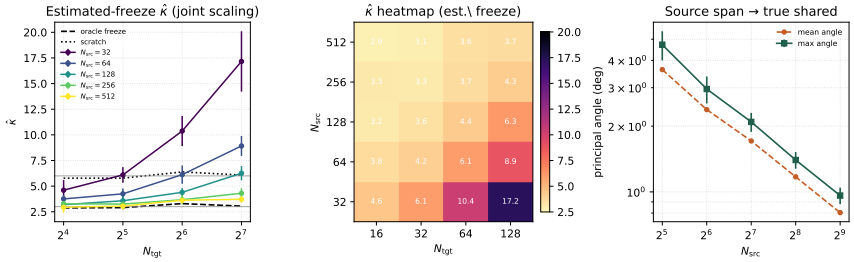

Figure 3: Finite-source to target transfer. Estimated freeze approaches residual $\kappa{=}r$ as N_{src} grows; oracle ($r{=}3$) and scratch ($s{+}r{=}6$) are references. Right: principal angles of the recovered shared span shrink with N_{src}.

Table 1: Misspecified freezes under matched designs. Mean $\hat{\kappa}$ is the residual-rate proxy; $\mathcal{E}_{128}$ is large-N generator risk.

Estimator	Access	mean $\hat{\kappa}$	$\mathcal{E}_{128}$	slope
Correct freeze	true $\mathfrak{so}(3)$ span	2.95	$5.2{\times}10^{-3}$	-1.03
Scratch	none	5.88	$1.1{\times}10^{-2}$	-1.02
Skew-SVD spectral	same $\{A(u)\}$ as Alg 1	—	7.59	$\sim$0
Abelian impostor T	commuting triple	—	8.72	$\sim$0
Wrong s-plane	Heisenberg as shared	—	9.20	$\sim$0
Full-trunk freeze	source gens, no residual	—	6.53	$\sim$0

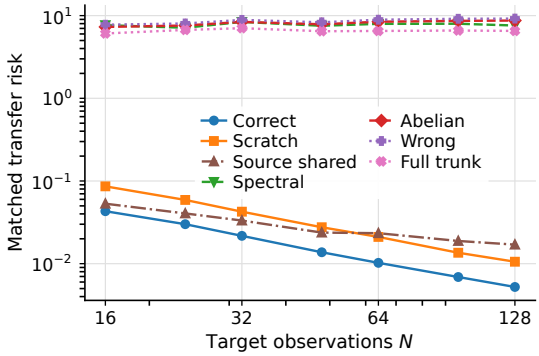

Figure 4: Wrong, Abelian, and spectral freezes remain at an N-independent risk; the exact shared-generator freeze attains the oracle residual rate in this panel.

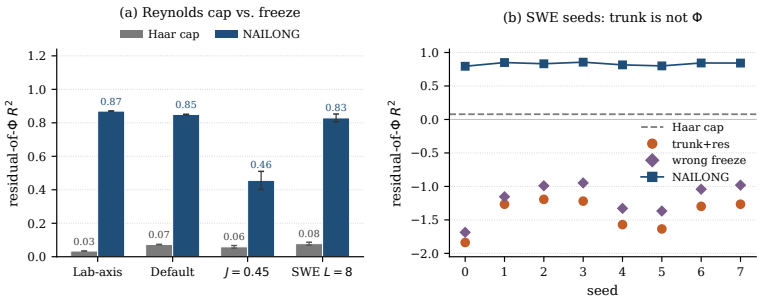

Figure 5: S^2 residual-of-Φ R^2. (a) Haar ceiling versus an unconstrained residual after a freeze of Φ. (b) SWE seeds. The lab-axis gap is 0.87 against 0.034 (Table 10).

The stronger generator baselines also benefit from shared structure. Over 24 seeds, bracket-aware freeze reaches $\hat{\kappa} = 3.30 \pm 0.14$, the Lie-Neuron head 3.38 ± 0.15, and competitive fine-tuning 3.87 ± 0.16, compared with oracle 3.01 ± 0.14 and scratch 6.01 ± 0.20 (Table 7). These results show that the variance reduction is accessible to several structure-aware estimators; it is not an exclusive property of a particular optimizer.

6.4 A JOINTLY EQUIVARIANT PREDICTOR IS CAPPED ON FIELDS

Physical angular momentum on spherical-harmonic coefficients is a faithful $\mathfrak{so}(3)$. Figure 5 scores residual-of-Φ R^2 on spherical fields.

Residual-of-Φ R^2 is 0 at $E = \Phi$ and 1 at $E = T$. The Haar column is the Reynolds ceiling $\|P_G R\|^2_{L^2} / \|R\|^2_{L^2}$ of Theorem 2: it is the most any jointly equivariant map can explain of the residual. Every Wigner or G-FNO predictor that remains jointly equivariant obeys the ceiling; reaching the source projection $\Phi + P_G R_{\mathrm{src}}$ additionally requires adequate capacity and optimization. On the lab-axis probe the ceiling is 0.034; the identified freeze of Φ plus an unconstrained residual reaches 0.87. Panel (b) repeats the gap across SWE seeds: the jointly equivariant predictor stays near the Haar line, and a wrong freeze stays on the floor. Those numbers measure the Reynolds class; they do not decide the trunk-plus-free-head comparison, and on this labeled probe the identified freeze is ordinary generator regression with an antisymmetry constraint. A linear residual class D yields identical fitted predictors for two freezes exactly when $\Phi - \widetilde{\Phi} \in D$ (Theorem 4). Equal scalar risks alone do not imply this condition. The implementation comparison also changes source supervision and control representation: generator OLS uses analytic velocity pairs and integrates the ordered control path, whereas the Wigner baseline uses endpoint supervision and control mean and standard deviation. Consequently, its remaining gap after adding a free residual does not isolate the choice of transferred object.

6.5 APPROXIMATE SYMMETRY STILL HAS A FLOOR

Shared generators need not sit exactly on the algebra. Figure 6 perturbs the frozen span by $\widehat{\mathfrak{h}} = (1-\delta)\mathfrak{h}^\star + \delta M$ for random skew M. That path is a bias diagnostic; for $\delta > 0$ it generally leaves the Lie-subalgebra manifold.

Along Ad-Givens conjugacies that preserve the abstract $\mathfrak{so}(3)$ algebra, the Procrustes gap in Theorem 3 is $\delta_0(\delta)^2 = 8(1 - \cos\delta)$, so the shared-channel floor is

$$F(\delta) \;=\; 8\gamma_\alpha\,(1 - \cos\delta) \tag{10}$$

(γ_α=1 on the isotropic probes). Large-N manifold floors track (10) (0.160 vs. 0.160 at δ=0.2; 0.621 vs. 0.632 at δ=0.4). The interpolated path floors higher and loses bracket closure for every δ>0 (Appendix C). When the finite-sample risk decomposes into this bias plus the full-entry residual OLS variance, freeze has risk $F(\delta) + rn^2\sigma^2/N$. It beats scratch only when $F(\delta) < sn^2\sigma^2/N$. A fixed source mismatch eventually outweighs the variance saved by freezing.

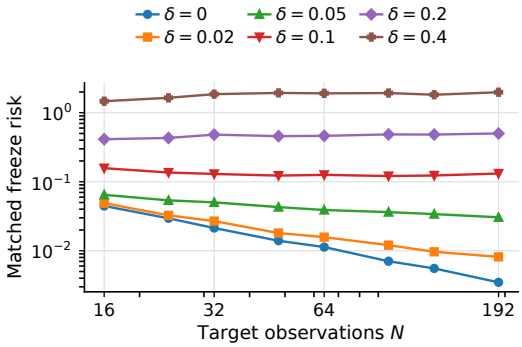

Figure 6: Approximate symmetry: freeze risk vs. N as shared perturbation δ grows.

6.6 The rate survives Galerkin projection

The residual-rate split is not special to n=6 matrix probes. After a Fourier–Galerkin truncation the same law holds on the coefficient ODE: correct freeze $\hat{\kappa} \approx 3$, scratch $\hat{\kappa} \approx 6$, at $n \in \{9, 16, 24, 32\}$ (Table 2; Appendix Figure 20). Wrong and Abelian freezes stay flat.

Table 2: Mean $\hat{\kappa}$ vs. Galerkin dimension (target residual = toral).

n	correct	scratch	wrong	Abelian
9	3.01	6.03	1.12×10^{3}	1.06×10^{3}
16	3.02	6.00	3.67×10^{2}	3.49×10^{2}
24	2.99	6.01	1.64×10^{2}	1.55×10^{2}
32	3.00	6.01	9.38×10^{1}	8.90×10^{1}

Log-log slopes for correct and scratch stay near -1 at every n; wrong and Abelian slopes stay near 0. The decrease in wrong-freeze $\hat{\kappa}$ across rows should be read with its n^2 normalization: it does not establish a decreasing unnormalized bias. Here s and r remain fixed while the observation dimension grows, so the test concerns the matrix-entry scaling of the same generator model.

7 Discussion

The generator experiments separate the variance saved by an exact freeze from the bias introduced by source estimation and alignment. The finite-source value $\hat{\kappa}$=3.72±0.31 approaches, but does not equal, the oracle rate. The field gap 0.87 versus 0.034 concerns a different question: how much a jointly equivariant class misses when the residual breaks symmetry. It establishes a hypothesis-class limitation, whereas a competitive transfer claim requires matching source supervision, control information, residual capacity, and target samples. Algebraic selection is tested in the generator experiments but is not executed in the labeled spherical field path.

8 Conclusion

Under matched-design OLS, correctly calibrated shared generators reduce target estimation from $s+r$ generators to r. An incorrect freeze introduces a persistent bias when the residual class cannot compensate for it. Together, the transfer and projection identities specify what must be known, aligned, and left adaptable; the field probes illustrate the separate cost of enforcing equivariance on a symmetry-breaking target.

Reproducibility Statement

The anonymous code repository is available at https://anonymous.4open.science/status/Nailong-FABC. The assumptions and proofs supporting the theoretical results are given in Appendices A–B. The experiment sections and Appendix F document the data-generating models, estimator settings, repeated runs, and compute resources. Appendix G lists commands for the theoretical checks and experimental probes; the codebase includes dependency specifications, experiment scripts, tests, and recorded JSON/TXT results. These materials specify the reported protocols, including the distinct supervision and control representations used in the field comparisons.

AI Use Statement

Consistent with ICLR 2027 policy, this statement does not count toward the page limit. Large language models assisted with language editing, figure preparation, literature-metadata checks, and consistency checks of mathematical statements and experimental descriptions. The empirical values are taken from the accompanying experiment records; this editorial revision introduced no new experimental results. The authors are responsible for the final mathematical claims, references, and interpretation of the results.

References

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1):151–175, 2010. URL https://doi.org/10.1007/s10994-009-5152-4.

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

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pp. 2990–2999, 2016. URL https://proceedings.mlr.press/v48/cohenc16.html.

Nima Dehmamy, Robin Walters, Yanchen Liu, Dashun Wang, and Rose Yu. Automatic symmetry discovery with Lie algebra convolutional network. In *Advances in Neural Information Processing Systems*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/148148d62be67e0916a833931bd32b26-Abstract.html.

Marc Finzi, Gregory Benton, and Andrew Gordon Wilson. Residual pathway priors for soft equivariance constraints. In *Advances in Neural Information Processing Systems*, 2021a. URL https://proceedings.nips.cc/paper_files/paper/2021/hash/fc394e9935fbd62c8aedc372464e1965-Abstract.html.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International Conference on Machine Learning*, 2021b. URL https://proceedings.mlr.press/v139/finzi21a.html.

Alex Gabel, Victoria Klein, Riccardo Valperga, Jeroen S. W. Lamb, Kevin Webster, Rick Quax, and Efstratios Gavves. Learning Lie group symmetry transformations with neural networks. In *Proceedings of the 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*, volume 221, pp. 50–59, 2023. URL https://proceedings.mlr.press/v221/gabel23a.html. arXiv:2307.01583.

Alex Gabel, Rick Quax, and Efstratios Gavves. Type-II neural symmetry detection with Lie theory. *Scientific Reports*, 15:33500, 2025. doi: 10.1038/s41598-025-17098-8. URL https://doi.org/10.1038/s41598-025-17098-8.

Somdatta Goswami, Katiana Kontolati, Michael D Shields, and George Em Karniadakis. Deep transfer operator learning for partial differential equations under conditional shift. *Nature Machine Intelligence*, 4(12):1155–1164, 2022. doi: 10.1038/s42256-022-00569-2. URL https://doi.org/10.1038/s42256-022-00569-2.

Brian C Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer, 2 edition, 2015. URL https://doi.org/10.1007/978-3-319-13467-3.

Jacob Helwig, Xuan Zhang, Cong Fu, Jerry Kurtin, Stephan Wojtowytsch, and Shuiwang Ji. Group equivariant Fourier neural operators for partial differential equations. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 12907–12930, 2023. URL https://proceedings.mlr.press/v202/helwig23a.html.

Lexiang Hu, Yikang Li, and Zhouchen Lin. Explicit discovery of nonlinear symmetries from dynamic data. In *International Conference on Machine Learning*, 2025. URL https://proceedings.mlr.press/v267/hu25o.html.

Chankyo Kim, Sicheng Zhao, Minghan Zhu, Tzu-Yuan Lin, and Maani Ghaffari. Equivariant neural networks for general linear symmetries on Lie algebras. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/63778.

Anthony W Knapp. *Lie Groups Beyond an Introduction*. Birkhäuser, 2 edition, 2002. URL https://link.springer.com/book/9780817642594.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, pp. 2747–2755, 2018. URL https://proceedings.mlr.press/v80/kondor18a.html.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural Operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Haoran Li, Chenhan Xiao, Muhao Guo, and Yang Weng. Latent mixture of symmetries for sample-efficient dynamic learning. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://papers.neurips.cc/paper_files/paper/2025/hash/9c84902266de123fccfdf84329905bbd-Abstract-Conference.html.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier Neural Operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Tzu-Yuan Lin, Minghan Zhu, and Maani Ghaffari. Lie neurons: Adjoint-equivariant neural networks for semisimple Lie algebras. In *International Conference on Machine Learning*, pp. 30529–30545, 2024. URL https://proceedings.mlr.press/v235/lin24aa.html.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. URL https://doi.org/10.1038/s42256-021-00302-5.

Ilya Markov and Alexander Hvatov. Data-driven discovery of Lie algebra generators via Fréchet-invariant learning. In *International Conference on Mathematics in Artificial Intelligence (MathAI)*, 2026. URL https://openreview.net/forum?id=UNJ2Z8pvxH.

Mircea Mironenco and Patrick Forré. Lie group decompositions for equivariant neural networks. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/61202bb341e7e0a6026ea134a5057abf-Abstract-Conference.html.

Samuel E. Otto, Nicholas Zolman, J. Nathan Kutz, and Steven L. Brunton. A unified framework to enforce, discover, and promote symmetry in machine learning. *Journal of Machine Learning Research*, 26(248):1–83, 2025. URL https://jmlr.org/papers/v26/24-1315.html.

Zakhar Shumaylov, Peter Zaika, James Rowbottom, Ferdia Sherry, Melanie Weber, and Carola-Bibiane Schönlieb. Lie algebra canonicalization: Equivariant neural operators under arbitrary Lie groups. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=7PLpiVdnUC.

Noah Shutty and Casimir Wierzynski. Computing representations for Lie algebraic networks. In *Proceedings of the 1st NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, pp. 1–21, 2023. URL https://proceedings.mlr.press/v197/shutty23a.html.

Shashank Subramanian, Peter Harrington, Kurt Keutzer, Wahid Bhimji, Dmitriy Morozov, Michael Mahoney, and Amir Gholami. Towards foundation models for scientific machine learning: Characterizing scaling and transfer behavior. In *Advances in Neural Information Processing Systems*, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/e15790966a4a9d85d688635c88ee6d8a-Abstract-Conference.html.

Nilesh Tripuraneni, Michael Jordan, and Chi Jin. On the theory of transfer learning: The importance of task diversity. In *Advances in Neural Information Processing Systems*, 2020. URL https://papers.nips.cc/paper/2020/hash/59587bffec1c7846f3e34230141556ae-Abstract.html.

Maurice Weiler and Gabriele Cesa. General E(2)-equivariant steerable CNNs. In *Advances in Neural Information Processing Systems*, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/45d6637b718d0f24a237069fe41b0db4-Abstract.html.

Jiaxiao Xu, Changhong Mou, Yeyu Zhang, and Fengxiang He. Physics-aligned canonical equivariant Fourier Neural Operator under symmetry-induced shifts. *arXiv preprint arXiv:2605.18606*, 2026. URL https://arxiv.org/abs/2605.18606.

A Standing Notation for the Proofs

Let $V = \mathbb{R}^n$ and fix integers $s, r \geq 1$ with $m := s+r$. Shared and residual generators are tuples $H = (H_1, \ldots, H_s)$ and $R = (R_1, \ldots, R_r)$ in $\mathfrak{gl}(n)$. Controls are $u = (\alpha, \beta) \in \mathbb{R}^s \times \mathbb{R}^r$, and

$$A(u) = \sum_{i=1}^{s} \alpha_i H_i + \sum_{j=1}^{r} \beta_j R_j. \tag{11}$$

We observe i.i.d. pairs $(u^{(\ell)}, \widetilde{A}^{(\ell)})_{\ell=1}^N$ with

$$\widetilde{A}^{(\ell)} = A\big(u^{(\ell)}\big) + \Xi^{(\ell)}, \qquad \mathrm{vec}(\Xi^{(\ell)}) \sim \mathcal{N}\big(0, \sigma^2 I_{n^2}\big), \tag{12}$$

independent of the design. For an estimator $(\widehat{H}, \widehat{R})$ the population generator risk is

$$\mathcal{E}(\widehat{H}, \widehat{R}) := \mathbb{E}_{u\sim\mu} \big\| A(u) - \widehat{A}(u) \big\|_F^2, \tag{13}$$

where μ is a test control law with second-moment $\Sigma_\mu = \mathbb{E}[uu^\top]$.

Vectorization. Stack $g := (\mathrm{vec}(G_1), \ldots, \mathrm{vec}(G_m)) \in \mathbb{R}^{mn^2}$ for $G = (H, R)$. Then $a(u) := \mathrm{vec}(A(u)) = U(u)\, g$ with

$$U(u) := \big[u_1 I_{n^2} \;\big|\; \cdots \;\big|\; u_m I_{n^2} \big] \in \mathbb{R}^{n^2 \times mn^2},$$

and the Gram identity

$$U(u)^\top U(u) = (uu^\top) \otimes I_{n^2} \tag{14}$$

holds. Write $\bar{\Sigma}_N := \frac{1}{N} \sum_{\ell=1}^N u^{(\ell)} u^{(\ell)\top}$ for the empirical second-moment matrix.

Scratch OLS. If $\bar{\Sigma}_N \succ 0$, the scratch estimator admits the closed form

$$\hat{g}_{\mathrm{scr}} = G_N^{-1} \sum_{\ell=1}^{N} U\big(u^{(\ell)}\big)^\top \tilde{a}^{(\ell)}, \qquad G_N = \big(N\bar{\Sigma}_N\big) \otimes I_{n^2}. \tag{15}$$

Correct freeze is the same OLS on residual channels only, after subtracting the frozen shared contribution, with residual Gram $(N\bar{\Sigma}_N^{RR}) \otimes I_{n^2}$.

B Complete Proofs

B.1 Proof of Theorem 1

Assume the Gaussian model (12) and Assumption 1: $\bar{\Sigma}_N \succ 0$ and the test law equals the empirical training measure $\mu = \frac{1}{N}\sum_\ell \delta_{u^{(\ell)}}$. The identities below are conditional on the design; they do not require population isotropy $\mathbb{E}[uu^\top] = I$.

Step 1: Error covariance. From (12) and (15),

$$\hat{g}_{\mathrm{scr}} - g = G_N^{-1} \sum_{\ell=1}^{N} U\big(u^{(\ell)}\big)^\top \xi^{(\ell)}, \qquad \xi^{(\ell)} := \mathrm{vec}(\Xi^{(\ell)}).$$

Conditioned on the design, the summands are independent, mean zero, with

$$\mathbb{E}\big[U(u)^\top \xi\xi^\top U(u) \,\big|\, u\big] = \sigma^2 U(u)^\top U(u) = \sigma^2 (uu^\top) \otimes I_{n^2}.$$

Hence

$$\mathrm{Cov}(\hat{g}_{\mathrm{scr}} \,|\, \mathrm{design}) = \sigma^2 G_N^{-1} = \frac{\sigma^2}{N} \bar{\Sigma}_N^{-1} \otimes I_{n^2}. \tag{16}$$

Step 2: Risk as a trace. For any fixed control u,

$$\big\|A(u) - \widehat{A}(u)\big\|_F^2 = \big\|U(u)(\hat{g} - g)\big\|_2^2 = (\hat{g} - g)^\top \big((uu^\top) \otimes I_{n^2}\big)(\hat{g} - g).$$

Taking expectation over the estimator error,

$$\mathbb{E}\big[\|A(u) - \widehat{A}(u)\|_F^2 \,\big|\, \mathrm{design}\big] = \mathrm{tr}\big(\mathrm{Cov}(\hat{g} \,|\, \mathrm{design})\,\big((uu^\top) \otimes I_{n^2}\big)\big). \tag{17}$$

Averaging over the matched empirical measure and using (16),

$$\mathbb{E}[\mathcal{E} \,|\, \mathrm{design}] = \mathrm{tr}\left(\frac{\sigma^2}{N}(\bar{\Sigma}_N^{-1} \otimes I_{n^2})\,(\bar{\Sigma}_N \otimes I_{n^2})\right) = \frac{\sigma^2}{N}\mathrm{tr}(I_{mn^2}) = \frac{mn^2\sigma^2}{N}.$$

Since $m = s{+}r$, this is the scratch identity.

Step 3: Correct freeze. The residual problem is identical with parameter count r and design Gram $(N\bar{\Sigma}_N^{RR}) \otimes I_{n^2}$. Repeating Steps 1–2 with $m \leftarrow r$ yields $\mathbb{E}[\mathcal{E}_{\mathrm{frz}} \,|\, \mathrm{design}] = rn^2\sigma^2/N$.

Step 4: Kappa. Dividing by $n^2\sigma^2/N$ gives $\kappa_{\mathrm{scr}} = s{+}r$ and $\kappa_{\mathrm{frz}} = r$.

Unbiased linear estimators. Under the Gaussian linear model, OLS is the UMVUE of g. Thus no unbiased linear estimator of the full parameter can beat the scratch identity under matched design, and no unbiased linear residual estimator can beat the freeze identity when H is known.

Corollary 1 (Exact sample ratio)**.** *Under Theorem 1, reaching conditional risk ε requires $N_{\mathrm{scr}} = (s{+}r)n^2\sigma^2/\varepsilon$ versus $N_{\mathrm{frz}} = rn^2\sigma^2/\varepsilon$, so $N_{\mathrm{frz}}/N_{\mathrm{scr}} = r/(s{+}r)$.*

Corollary 2 (PE sandwich)**.** *Assume $\gamma I_m \preceq \bar{\Sigma}_N \preceq \Gamma I_m$ and the test law satisfies $\gamma_\mu I_m \preceq \Sigma_\mu \preceq \Gamma_\mu I_m$. Then*

$$\frac{(s{+}r)n^2\sigma^2}{N} \cdot \frac{\gamma_\mu}{\Gamma} \;\le\; \mathbb{E}[\mathcal{E}_{\mathrm{scr}} \,|\, \mathrm{design}] \;\le\; \frac{(s{+}r)n^2\sigma^2}{N} \cdot \frac{\Gamma_\mu}{\gamma}.$$

An identical statement holds for correct freeze with $s{+}r$ replaced by r.

Proof. From (16)–(17) with population Σ_μ,

$$\mathbb{E}[\mathcal{E} \,|\, \mathrm{design}] = \frac{\sigma^2}{N} n^2\, \mathrm{tr}(\bar{\Sigma}_N^{-1}\Sigma_\mu).$$

For SPD matrices, $\mathrm{tr}(\bar{\Sigma}_N^{-1}\Sigma_\mu)$ is bounded above by $\gamma^{-1}m\Gamma_\mu$ and below by $\Gamma^{-1}m\gamma_\mu$. □

B.2 Proof of Theorem 3

Assumption 2 is the direct-sum domain construction used in the experiments (shared triad slot vs. residual slot). For a frozen shared tuple $\widetilde{H}$ define

$$\delta_0^2 := \inf_{Q \in O(s)} \sum_{i=1}^{s} \Big\| \widetilde{H}_i - \sum_{k=1}^{s} H_k Q_{ki} \Big\|_F^2. \tag{18}$$

Proof of Theorem 3. Theorem 3 is Theorem 7(ii) (shared-slot / model-class case). The $\beta = 0$ identity below is Corollary 4. Under Assumption 2, take any control law with $\mathbb{E}[\alpha\alpha^\top] \succeq \gamma_\alpha I_s$ and $\beta = 0$ a.s. Then

$$A(u) - \widehat{A}(u) = \sum_{i=1}^{s} \alpha_i (H_i - \widetilde{H}_i),$$

so residual fitting cannot compensate. Hence

$$\mathcal{E}(\widetilde{H}, \widehat{R}) = \mathbb{E}\Big\| \sum_{i=1}^{s} \alpha_i (H_i - \widetilde{H}_i) \Big\|_F^2 = \operatorname{tr}\big(\mathbb{E}[\alpha\alpha^\top]\, M\big)\,,$$

where $M_{ik} := \langle H_i - \widetilde{H}_i,\ H_k - \widetilde{H}_k \rangle_F$ (after rotating channels so that $Q = I$ realizes the infimum in (18)). Therefore

$$\mathcal{E} \geq \gamma_\alpha \operatorname{tr}(M) \geq \gamma_\alpha \delta_0^2.$$

The floor is independent of $\widehat{R}$ and of N. □

B.3 Proof of Theorem 5

Theorem 5 (Bracket-blind insufficiency for $\mathfrak{so}(3)$)**.** *In the n=6 block embedding of Appendix B.3, let $\mathfrak{h}^\star$ be a faithful embedded $\mathfrak{so}(3)$ triad and let T be the embedded commuting diagonal family on the same slot, sharing a spectral drift A_0. A map π from (A_0, T) to a freeze is* spectral *if $\pi(A_0, T)$ lies in the joint functional calculus of (A_0, T), and* Ad-natural *if $\pi(gA_0g^{-1}, gTg^{-1}) = \mathrm{Ad}_g \pi(A_0, T)$. Every spectral or Ad-natural π returns a commuting triple, so $\delta_0(\pi(A_0, T); \mathfrak{h}^\star) > 0$ on this embedding and Theorem 3 supplies an N-independent floor on shared-channel probes. No Borel function of (A_0, T) selects a faithful $\mathfrak{so}(3)$ completion on every world compatible with these observables. Thus bracket-blind Abelian/spectral proxies cannot attain residual-rate freeze.*

Definition 1 (Observation model for Theorem 5)**.** *Observable data are the shared spectral drift A_0 and a commuting triple $T = (T_1, T_2, T_3)$ on the triad slot (jointly diagonalizable). Bracket structure constants of a non-Abelian completion are* not *observed. A map π from (A_0, T) to a freeze is* spectral *if $\pi(A_0, T)$ lies in the joint functional calculus of (A_0, T) (polynomials, SVD/DMD modes, and other functions of the commuting family). It is* Ad-natural *if $\pi(gA_0g^{-1}, gTg^{-1}) = \mathrm{Ad}_g \pi(A_0, T)$ for all g in the ambient group.*

Proof. **Construction.** Take $n = 6$, triad indices $\{0, 1, 2\}$, residual indices $\{3, 4, 5\}$. Let $H_i = \iota_{\mathrm{triad}}(L_i)$ where $\{L_i\}$ is the standard $\mathfrak{so}(3)$ basis, and let $T_i = \iota_{\mathrm{triad}}(E_{ii})$ for the 3×3 diagonal matrix units (an Abelian impostor of dimension three—not a one-dimensional Cartan torus). Let $A_0 = \mathrm{diag}(\rho - \nu k^2)$ be a shared Fourier diffusion drift.

The true shared factor H satisfies $[H_1, H_2] = H_3$ (cyclic), while every bracket of T vanishes. No $Q \in O(3)$ makes $\sum_i \|T_i - \sum_k H_k Q_{ki}\|_F = 0$, so $\delta_0(T; H) > 0$. Theorem 3 therefore yields a floor for freezing T on shared-channel probes. Freezing H reduces to Theorem 1 with residual dimension $r = 3$.

Spectral and Ad-natural proxies fail on this world. The joint functional calculus of the commuting family (A_0, T) is spanned by polynomials in (A_0, T_1, T_2, T_3) and is therefore Abelian. Any spectral π returns a commuting triple, hence $\delta_0(\pi(A_0, T); H) > 0$ on the given faithful embedding. An Ad-natural π is determined by the conjugacy class of (A_0, T); that class contains the Cartan/diagonal family and does not contain a faithful $\mathfrak{so}(3)$ triad, so π cannot land on H.

Fibre / minimax. The pair (A_0, T) is compatible with a continuum of non-Abelian completions (any so(3) that shares the same Cartan and spectral drift). A Borel function of (A_0, T) is constant on this

fibre, so it cannot equal the true shared triad in every compatible world. The obstruction is for maps of (A_0, T); additional non-commuting observations (brackets, velocity cross-terms, or identified H) lie outside this statement. □

Skew-SVD, PCA on $\mathrm{vec}\, A(u)$, and DMD on the same source trajectories as Algorithm 1 still return $\delta_0 > 0$ (Appendix E.7): the obstruction is selection of the embedded plane.

Lemma 1 (Lie closure selects the transferable plane)**.** *If $\mathfrak{h} \subset \mathfrak{gl}(n)$ is the Lie algebra of a continuous matrix group action, then $[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{h}$. Theorem 1 therefore applies to any freeze of the correct shared span, but a shared continuous symmetry can only supply a Lie subalgebra—not an arbitrary s-plane in $\mathfrak{gl}(n)$.*

Proof of Lemma 1. The Lie algebra of a matrix Lie group is closed under commutators by definition (Hall, 2015). An s-dimensional linear subspace of $\mathfrak{gl}(n)$ that fails bracket closure cannot be the infinitesimal algebra of a continuous group action, so it cannot be the transferable object of a shared continuous symmetry even if Theorem 1 would apply to a freeze of that span. □

Proposition 1 (Selection uniqueness)**.** *Under Assumption 2 with $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{w}$, $[\mathfrak{h}, \mathfrak{w}] = 0$, $\mathfrak{h} \cong \mathfrak{so}(3)$ and $\mathfrak{w}$ Heisenberg or toral, $\mathfrak{h}$ is the unique simple three-dimensional subalgebra of $\mathfrak{g}$ (Killing form of rank three; nondegenerate factor $\mathfrak{h}$). Combined with Theorems 1 and 3, a three-plane freeze attains residual rate $\kappa{=}r$ if and only if it is this simple factor: the barcode is the residual-rate selector under this splitting.*

Proof of Proposition 1. Let $\mathfrak{s} \subset \mathfrak{g}$ be a simple three-dimensional subalgebra and consider the projection $\pi_{\mathfrak{h}}|_{\mathfrak{s}}$. Its kernel is an ideal of $\mathfrak{s}$. It cannot equal $\mathfrak{s}$, since the nilpotent or Abelian algebra $\mathfrak{w}$ contains no simple three-dimensional subalgebra; hence it is zero. Dimension then makes the projection an isomorphism onto $\mathfrak{h}$, so $\mathfrak{s}$ is the graph of a homomorphism $\phi : \mathfrak{h} \to \mathfrak{w}$. The kernel of ϕ is again an ideal. If ϕ were nonzero, simplicity would make it injective, contradicting solvability of $\mathfrak{w}$; therefore $\phi = 0$ and $\mathfrak{s} = \mathfrak{h}$. Equivalently, mixed coordinate planes have degenerate Killing form, while the Heisenberg and Abelian planes have zero Killing form. The direct-sum Killing form is thus negative definite on $\mathfrak{h}$ and zero on $\mathfrak{w}$, with radical exactly $\mathfrak{w}$. Under any $\mathrm{GL}(6)$ basis change its matrix changes by congruence, so the radical and nondegenerate factor remain intrinsic. These are standard Cartan–Killing facts (Hall, 2015). The transfer content is the characterization: mixed or solvable 3-planes have $\delta_0 > 0$ and therefore the floor of Theorem 3, so the unique residual-rate freeze of Theorems 1 and 3 is $\mathfrak{h}$. The barcode computes that selector. □

B.4 Proof of Theorem 6

Theorem 6 (Residual-rate selector)**.** *Let $A(u) = \sum_i \alpha_i H_i + \sum_j \beta_j R_j$ and let D be a linear residual class. A freeze $\widetilde{H}$ plus optimal residual OLS in D has the same fitted predictor as the correct freeze if and only if the propagator mismatch lies in D (Theorem 4). For generator regression with $R \in D$, positive-definite shared second moment, and a residual class that cannot absorb a nonzero shared mismatch, the unique residual-rate s-plane is $\mathrm{span}(H)$ up to $O(s)$ remix. The characterization does not use $[\mathfrak{h}, \mathfrak{w}] = 0$: residual rate still isolates the identified shared span among s-planes that D cannot repair.*

Write $\Delta_i := H_i - \widetilde{H}_i$ and $\Delta(u) := \sum_i \alpha_i \Delta_i$. After freezing $\widetilde{H}$, residual OLS in a linear class D is the L^2 projection of $A(u) - \sum_i \alpha_i \widetilde{H}_i = R(u) + \Delta(u)$ onto D, where $R(u) = \sum_j \beta_j R_j$. Theorem 4 (or the Schur calculation of Appendix B.7) gives exact residual match iff $\Delta \in D$ on the data measure. If D is the model-class slot $\mathcal{S}_R$ of Assumption 2 and the mismatch is shared-slot supported, then $\Delta \notin D$ whenever $\delta_0(\widetilde{H}; H) > 0$, so the unique residual-rate s-plane is $\mathrm{span}(H)$ up to $O(s)$ remix. If residuals depend linearly only on β, their matrices are unrestricted, and $\mathbb{E}[\alpha\beta^\top] = 0$, the same obstruction is the shared-channel floor of Corollary 4. Neither case uses $[\mathfrak{h}, \mathfrak{w}] = 0$: residual rate isolates $\mathrm{span}(H)$ among s-planes that D cannot repair, and Algorithm 1 freezes the identified span.

Equality of fitted predictors is stronger than equality of their scalar risks. In general, let $a = (I - P_D)R$ and $b = (I - P_D)\Delta$. The risk difference is $2\langle a, b\rangle + \|b\|^2$, which may vanish with $b \neq 0$. The risk-level selector follows when $R \in D$, since then $a = 0$ and the excess is $\|b\|^2$. It is not an unrestricted equivalence for an arbitrary misspecified D.

B.5 Proof of Theorem 2

Proof of Theorem 2. Write $G = \mathrm{SO}(3)$ and let ρ be a unitary representation on a Hilbert space V of fields, acting jointly on controls by the adjoint action. Let μ be any Borel probability on the input–control space and $\bar{\mu} = \int_G g_\sharp \mu \, dg$ its Haar symmetrization, so $\bar{\mu}$ is G-invariant and $\bar{\mu} = \mu$ whenever μ itself is invariant. L^2 throughout is $L^2(\bar{\mu})$ with the pointwise inner product of V. The Reynolds operator

$$P_G R(u, \alpha) = \int_G \rho(g)^{-1} R(\rho(g)u,\ g \cdot \alpha)\, dg$$

is self-adjoint and idempotent on this L^2: for jointly equivariant E, the change of variables $(u, \alpha) \mapsto g \cdot (u, \alpha)$ together with unitarity of ρ and G-invariance of $\bar{\mu}$ give $\langle R - P_G R,\ E\rangle_{L^2(\bar{\mu})} = 0$. Hence P_G is the orthogonal projection onto the jointly equivariant class, $\|R - E\|^2_{L^2} \geq \|R - P_G R\|^2_{L^2}$, and residual-of-$\Phi$ $R^2(E)$ cannot exceed $\|P_G R\|^2_{L^2} / \|R\|^2_{L^2}$. The reported spherical ceilings are computed on this Haar risk (`haar_residual_ceiling`; isotropic GRF laws are already invariant). Haar averaging is classical. The projector is the obstruction for every jointly equivariant predictor: a source trunk freezes $\Phi + P_G R_{\mathrm{src}}$ whenever that Haar component is nonzero.

If E_{src} is any jointly equivariant estimator of a source map $T_{\mathrm{src}} = \Phi + R_{\mathrm{src}}$, then $E_{\mathrm{src}} = P_G T_{\mathrm{src}} = \Phi + P_G R_{\mathrm{src}}$ in $L^2(\bar{\mu})$, since $P_G \Phi = \Phi$. This equals Φ iff $P_G R_{\mathrm{src}} = 0$, and equals T_{src} iff R_{src} is jointly equivariant. The open interval $0 < \|P_G R_{\mathrm{src}}\| < \|R_{\mathrm{src}}\|$ is the generic mixed case: the trunk is neither Φ nor T_{src}. A source-trained equivariant trunk plus a later residual adapter is therefore a freeze of $\Phi + P_G R_{\mathrm{src}}$. When $P_G R_{\mathrm{src}} \neq 0$ this is a wrong Φ and Theorem 4 applies; when $P_G R_{\mathrm{src}} = 0$ the trunk *is* Φ, and the remaining obstruction is staying inside the jointly equivariant class.

On the multiplicity-free real spherical-harmonic representation through degree L, $V = \bigoplus_{\ell=0}^{L} V_\ell$ with $\dim V_\ell = 2\ell + 1$ and pairwise inequivalent irreps. Schur's lemma therefore says that a *linear* intertwiner is an irrep scalar $E = \bigoplus_\ell \lambda_\ell I_{V_\ell}$ (Hall, 2015). The Frobenius-nearest such map is the block-trace projection $P_{\mathrm{eq}}(R)|_{V_\ell} = \mathrm{tr}(R|_{V_\ell})/(2\ell + 1)\, I_{V_\ell}$, which is also the Haar average $g \mapsto \rho(g) R \rho(g)^{-1}$. The lab-axis Jacobian of Appendix E.10 then has $\eta^2 / \|R\|^2_F = 0.792$ (ceiling 0.208). The same linear map, evaluated in L^2 against GRF fields with spectrum $(1 + \ell)^{-2.5}$, has ceiling 0.034: low-ℓ mass sees the lab axis, not the high-ℓ damping that dominates unweighted Frobenius. A split-sample Haar estimate of P_G on the nonlinear field-to-field residual agrees (0.024 at 48 rotations; `results/sphere_operator_transfer.json`). Irrep-gated Wigner/G-FNO-style layers (invariant norm gates plus control-conditioned rotations) are jointly equivariant (Weiler & Cesa, 2019; Helwig et al., 2023), as is any equivariant fine-tune of them, so they cannot exceed this ceiling. Freezing the generator algebra and fitting an unconstrained residual does not restrict to P_G, so Theorem 1 applies to that residual slot. □

B.6 Proof of Proposition 2

Proposition 2 (Barcode vs. embedded transfer)**.** *For semisimple $\mathfrak{h}$, the Killing form and structure tensor determine the* abstract *isomorphism type. A discrete pin-down of structure constants selects a basis of the abstract algebra up to the residual stabilizer of that pin-down; for $\mathfrak{so}(3)$ the Killing spectrum is fully degenerate, so the barcode alone does* not *fix an embedded basis in $\mathfrak{gl}(n)$. Concrete freeze transfer additionally assumes a shared observation chart (same Galerkin/PCA coordinates, or a known intertwiner) and transfers the embedded subspace* $\mathrm{span}(\mathfrak{h}) \subset \mathfrak{gl}(n)$.

Let $\mathfrak{h}$ be a real Lie algebra of dimension s with basis $E = (E_1, \ldots, E_s)$ and structure constants $[E_i, E_j] = \sum_k c_{ij}^k E_k$. The Killing form has coordinate matrix

$$K_{ij} = \sum_{a,b} c_{ia}^b c_{jb}^a = \mathrm{tr}(\mathrm{ad}_{E_i} \circ \mathrm{ad}_{E_j}).$$

Lemma 2 (Basis change)**.** *If $E' = EG$ with $G \in \mathrm{GL}(s)$, then ${c'}_{ab}^{\ d} = \sum_{i,j,k} G_{ia} G_{jb} c_{ij}^k (G^{-1})_{dk}$ and $K' = G^\top K G$.*

Lemma 3 (Ad-invariance)**.** *For $P \in \mathrm{GL}(n)$ and $\mathfrak{h} \subset \mathfrak{gl}(n)$, $\kappa(PXP^{-1}, PYP^{-1}) = \kappa(X, Y)$ for all $X, Y \in \mathfrak{h}$. Intrinsic Killing/structure data do not determine the embedding.*

Definition 2 (Abstract barcode)**.** *Assume $\mathfrak{h}$ is semisimple. An abstract barcode is the structure-tensor orbit under $\mathrm{GL}(s)$, together with the Killing form transformed by congruence. The orbit classifies the*

algebra up to isomorphism. The ordinary eigenvalues of K alone are not invariant under a general change of basis.

Proof of Proposition 2. **Abstract type.** Semisimplicity $\Rightarrow$ K nondegenerate; the structure tensor modulo $\mathrm{GL}(s)$ is an isomorphism invariant (Hall, 2015; Knapp, 2002). Thus barcodes match iff the abstract algebras are isomorphic.

Degeneracy for $\mathfrak{so}(3)$. In any Killing-orthonormal frame, $K \propto -I_3$, so the Killing spectrum has multiplicity three and does not reduce the continuous $O(3)$ basis gauge. Moreover, every oriented Killing-orthonormal frame has the same structure constants $\pm\varepsilon_{ijk}$; lexicographic minimization of c therefore does *not* select a unique embedded basis. A discrete signed-permutation cleanup can at best match a prescribed abstract tensor, not a concrete matrix realization.

Embedded transfer. Inequivalent representations of the same abstract semisimple algebra share barcodes but need not share an intertwiner in $\mathfrak{gl}(n)$. Hence freeze transfer requires a shared observation chart (or known intertwiner) and uses the embedded subspace $\mathrm{span}(\mathfrak{h})$, which is basis-invariant. When a shared chart is fixed, subspace Procrustes after an abstract type-check yields the freeze used in the experiments; numerical barcode stability under conjugacy (Appendix Figure 18) is the type-check. □

Compact example $\mathfrak{so}(3)$. Every oriented Killing-orthonormal frame has structure constants $c^3_{12} = c^1_{23} = c^2_{31} = 1$ (and skew partners). This common abstract tensor does not pin the continuous embedded basis; the freeze uses the shared-chart subspace.

B.7 Wrong-freeze risk under general joint test covariance

This subsection records the full mixed-covariance development behind the upgraded main-text Theorem 3, for an arbitrary second-moment law $\Sigma = \mathbb{E}[uu^\top]$ on $u = (\alpha, \beta) \in \mathbb{R}^s \times \mathbb{R}^r$. We keep the standing notation of Appendix A: the Frobenius inner product $\langle X, Y\rangle_F := \mathrm{tr}(X^\top Y)$ and norm $\|X\|_F := \sqrt{\langle X, X\rangle_F}$ on $\mathfrak{gl}(n)$, and the population generator risk

$$\mathcal{E}(\widetilde{H}, \widehat{R}) := \mathbb{E}_{u\sim\mu}\big\|A(u) - \widehat{A}(u)\big\|_F^2, \qquad \widehat{A}(u) = \sum_{i=1}^{s} \alpha_i \widetilde{H}_i + \sum_{j=1}^{r} \beta_j \widehat{R}_j. \tag{19}$$

Throughout, μ is any Borel probability on $\mathbb{R}^m$ with finite second moment $\Sigma := \mathbb{E}[uu^\top] \in \mathbb{R}^{m\times m}$, written in blocks

$$\Sigma = \begin{pmatrix} \Sigma_{\alpha\alpha} & \Sigma_{\alpha\beta} \\ \Sigma_{\beta\alpha} & \Sigma_{\beta\beta} \end{pmatrix}, \qquad \Sigma_{\alpha\alpha} \in \mathbb{R}^{s\times s},\ \Sigma_{\beta\beta} \in \mathbb{R}^{r\times r},\ \Sigma_{\alpha\beta} = \Sigma_{\beta\alpha}^\top \in \mathbb{R}^{s\times r}. \tag{20}$$

No Gaussianity of μ is used: every identity below depends on μ only through Σ. (The observation noise in (12) is irrelevant for the population bias floor; set $\sigma = 0$ or pass to the $N \to \infty$ limit of residual OLS.)

Frozen mismatch. As in (18), write

$$\delta_0^2 := \inf_{Q\in O(s)} \sum_{i=1}^{s} \Big\| \widetilde{H}_i - \sum_{k=1}^{s} H_k Q_{ki} \Big\|_F^2 \tag{21}$$

and fix a Procrustes gauge $Q_\star$ attaining the infimum (compactness of $O(s)$). Replace $(\widetilde{H}, \alpha)$ by $(\widetilde{H}Q_\star^\top,\ Q_\star\alpha)$; the law of u transforms by the block-orthogonal $\mathrm{diag}(Q_\star, I_r)$, which preserves the Loewner lower bounds we use on $\Sigma_{\alpha\alpha}$ and on Schur complements. After this gauge fix, write

$$\Delta H_i := H_i - \widetilde{H}_i, \qquad M_{ik} := \langle \Delta H_i, \Delta H_k\rangle_F, \qquad \mathrm{tr}(M) = \sum_{i=1}^{s} \|\Delta H_i\|_F^2 = \delta_0^2.$$

B.7.1 POPULATION RISK FOR A FIXED RESIDUAL TUPLE

Lemma 4 (Quadratic risk identity). *For any frozen $\widetilde{H}$ and any residual tuple $\widehat{R} = (B_1, \dots, B_r)$,*

$$\mathcal{E}(\widetilde{H}, \widehat{R}) = \big\langle \Sigma,\ \Gamma(\Delta H,\ R - B) \big\rangle_F := \sum_{a,b=1}^{m} \Sigma_{ab} \big\langle G_a - \widehat{G}_a,\ G_b - \widehat{G}_b \big\rangle_F, \tag{22}$$

where $G = (H, R)$, $\widehat{G} = (\widetilde{H}, B)$, and the Gram Γ is the $m \times m$ matrix of Frobenius inner products of generator errors. Equivalently, with $\Delta R_j := R_j - B_j$ and $M^{HH}_{ik} = \langle \Delta H_i, \Delta H_k \rangle_F$, $M^{HR}_{ij} = \langle \Delta H_i, \Delta R_j \rangle_F$, $M^{RR}_{jl} = \langle \Delta R_j, \Delta R_l \rangle_F$,

$$\mathcal{E} = \big\langle \Sigma_{\alpha\alpha}, M^{HH} \big\rangle_F + 2 \big\langle \Sigma_{\alpha\beta}, M^{HR} \big\rangle_F + \big\langle \Sigma_{\beta\beta}, M^{RR} \big\rangle_F. \tag{23}$$

Proof. Expand $\|A(u) - \widehat{A}(u)\|_F^2$ by bilinearity of $\langle \cdot, \cdot \rangle_F$ and take $\mathbb{E}$. Only second moments of u appear. □

B.7.2 UNCONSTRAINED RESIDUAL CLASS: SCHUR COMPLEMENT FLOOR

Here $\widehat{R}_j$ ranges freely over $\mathfrak{gl}(n)$ (the ambient OLS class used by the residual learner).

Proposition 3 (Optimal unconstrained residual). *Assume $\Sigma_{\beta\beta} \succ 0$. The unique minimizer of $B \mapsto \mathcal{E}(\widetilde{H}, B)$ over $(\mathfrak{gl}(n))^r$ is*

$$B^\star_k = R_k + \sum_{i=1}^{s} L_{ki}\, \Delta H_i, \qquad L := \Sigma_{\beta\beta}^{-1} \Sigma_{\beta\alpha} \in \mathbb{R}^{r \times s}. \tag{24}$$

The minimal risk equals the Schur-weighted mismatch Gram

$$\boxed{\mathcal{E}^\star_{\mathrm{unc}}(\widetilde{H}) := \inf_B \mathcal{E}(\widetilde{H}, B) = \big\langle \Sigma_{\alpha|\beta},\ M \big\rangle_F = \mathrm{tr}\big(\Sigma_{\alpha|\beta}\, M\big),} \tag{25}$$

where

$$\Sigma_{\alpha|\beta} := \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta} \Sigma_{\beta\beta}^{-1} \Sigma_{\beta\alpha} \tag{26}$$

is the Schur complement of $\Sigma_{\beta\beta}$ in Σ (equivalently, $\mathrm{Cov}(\alpha - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\beta)$ when $\mathbb{E}u = 0$).

Proof. Write $Z(u) := \sum_i \alpha_i \Delta H_i + \sum_j \beta_j R_j$, so $\mathcal{E} = \mathbb{E}\|Z(u) - \sum_k \beta_k B_k\|_F^2$. Differentiating in the Fréchet sense in direction $H \in \mathfrak{gl}(n)$ at coordinate k yields the normal equation

$$\mathbb{E}\Big[\beta_k \Big(Z(u) - \sum_j \beta_j B_j \Big)\Big] = 0 \quad (k = 1, \dots, r),$$

i.e.

$$\sum_{j=1}^{r} \Sigma_{\beta\beta,kj}\, B_j = \sum_{i=1}^{s} \Sigma_{\beta\alpha,ki}\, \Delta H_i + \sum_{j=1}^{r} \Sigma_{\beta\beta,kj}\, R_j.$$

Since $\Sigma_{\beta\beta} \succ 0$ this is equivalent to (24). Substituting $\Delta R_k = -\sum_i L_{ki} \Delta H_i$ into the error gives

$$A(u) - \widehat{A}(u) = \sum_{i=1}^{s} \big(\alpha - L^\top \beta\big)_i \Delta H_i, \qquad L^\top = \Sigma_{\alpha\beta} \Sigma_{\beta\beta}^{-1}.$$

Taking the second-moment yields (25). Uniqueness follows from strict convexity of $B \mapsto \mathcal{E}$ on the finite-dimensional space $(\mathfrak{gl}(n))^r$ whenever $\Sigma_{\beta\beta} \succ 0$. □

Remark 1 (Oblique residual regression). *The map $\beta \mapsto \sum_k \beta_k B^\star_k$ is the $L^2(\mu; \mathfrak{gl}(n))$ oblique regression of the frozen mismatch $\sum_i \alpha_i \Delta H_i$ onto the residual control channels. Compensation is a statement about* control correlation*, not about the Lie bracket. If $\Sigma_{\alpha\beta} = 0$, then $L = 0$, $B^\star = R$, and $\mathcal{E}^\star_{\mathrm{unc}} = \mathrm{tr}(\Sigma_{\alpha\alpha} M)$: residual learning cannot cancel shared mismatch when $\alpha \perp \beta$ in L^2. If $\Sigma_{\alpha|\beta} = 0$ (perfect linear predictability of α from β), the unconstrained floor vanishes even for $\delta_0 > 0$.*

Corollary 3 (Loewner lower bound, unconstrained). *If $\Sigma_{\alpha|\beta} \succeq \gamma_{\alpha|\beta} I_s$ with $\gamma_{\alpha|\beta} > 0$, then*

$$\mathcal{E}^\star_{\mathrm{unc}}(\widetilde{H}) \;\geq\; \gamma_{\alpha|\beta}\, \delta_0^2. \tag{27}$$

In particular, under isotropic probing $\Sigma = I_m$ one has $\Sigma_{\alpha|\beta} = I_s$, hence $\mathcal{E}^\star_{\mathrm{unc}} \geq \delta_0^2$.

Proof. $\mathrm{tr}(\Sigma_{\alpha|\beta} M) \geq \gamma_{\alpha|\beta}\, \mathrm{tr}(M) = \gamma_{\alpha|\beta}\delta_0^2$ because $M \succeq 0$. □

B.7.3 Direct-sum residual slot: irreducible action on V_H

Assumption 2 supplies an orthogonal splitting $V = V_H \oplus V_R$ with

$$H_i(V_R) = 0,\; H_i(V_H) \subset V_H, \qquad R_j(V_H) = 0,\; R_j(V_R) \subset V_R. \tag{28}$$

Equip $\mathfrak{gl}(n)$ with the Frobenius product and define the *residual slot*

$$\mathcal{S}_R := \big\{X \in \mathfrak{gl}(n) : X(V_H) = \{0\},\; X(V_R) \subset V_R\big\} \cong \mathfrak{gl}(V_R). \tag{29}$$

Its orthogonal complement is

$$\mathcal{S}_R^\perp = \big\{X : X_{RR} := P_R X|_{V_R} = 0\big\}, \tag{30}$$

where P_H, P_R are the orthogonal projectors onto V_H, V_R. Equivalently, in blocks relative to $V_H \oplus V_R$,

$$X = \begin{pmatrix} X_{HH} & X_{HR} \\ X_{RH} & X_{RR} \end{pmatrix}, \qquad \mathcal{S}_R = \big\{X : X_{HH} = X_{HR} = X_{RH} = 0\big\}.$$

$$\|X\|_F^2 = \|X_{HH}\|_F^2 + \|X_{HR}\|_F^2 + \|X_{RH}\|_F^2 + \|X_{RR}\|_F^2.$$

Let $\Pi_R : \mathfrak{gl}(n) \to \mathcal{S}_R$ be the orthogonal projector ($X \mapsto$ its RR block) and $\Pi_\perp := I - \Pi_R$.

Definition 3 (Model-class residual). *A residual estimator is* model-class *if $\widehat{R}_j \in \mathcal{S}_R$ for all j (equivalently: residual fitting is constrained to the residual slot of Assumption 2). The true residual generators satisfy $R_j \in \mathcal{S}_R$ by (28).*

Lemma 5 (Block orthogonality of the risk). *For any model-class $\widehat{R}$ and any $\widetilde{H}$,*

$$\mathcal{E}(\widetilde{H}, \widehat{R}) = \mathbb{E}\Big\| \sum_{i=1}^s \alpha_i\, \Pi_\perp(\Delta H_i)\Big\|_F^2 + \mathbb{E}\Big\| \sum_{i=1}^s \alpha_i\, \Pi_R(\Delta H_i) + \sum_{j=1}^r \beta_j (R_j - \widehat{R}_j)\Big\|_F^2. \tag{31}$$

In particular the first summand is independent of $\widehat{R}$ and of $(\Sigma_{\alpha\beta}, \Sigma_{\beta\beta})$.

Proof. Decompose $A - \widehat{A}$ into $\mathcal{S}_R \oplus \mathcal{S}_R^\perp$ components. Since $\widehat{R}_j, R_j \in \mathcal{S}_R$, the $\mathcal{S}_R^\perp$ component equals $\sum_i \alpha_i \Pi_\perp(\Delta H_i)$. Frobenius orthogonality of the two summands and taking $\mathbb{E}$ yield (31). □

Theorem 7 (Wrong-freeze floor under general Σ: direct-sum / model-class). *Assume Assumption 2 and Definition 3. Let $\widetilde{H}$ be a frozen shared tuple with $\delta_0 > 0$ after the Procrustes gauge above, and let*

$$M^\perp_{ik} := \big\langle \Pi_\perp(\Delta H_i),\, \Pi_\perp(\Delta H_k)\big\rangle_F, \qquad M^R_{ik} := \big\langle \Pi_R(\Delta H_i),\, \Pi_R(\Delta H_k)\big\rangle_F, \tag{32}$$

so that $M = M^\perp + M^R$ and $\mathrm{tr}(M^\perp) + \mathrm{tr}(M^R) = \delta_0^2$.

1. (Population risk after optimal residual fitting.) *If $\Sigma_{\beta\beta} \succ 0$, the model-class minimizer is*

$$\widehat{R}_k^\star = R_k + \sum_{i=1}^s L_{ki}\, \Pi_R(\Delta H_i), \qquad L = \Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}, \tag{33}$$

and the minimal risk decomposes as

$$\boxed{\mathcal{E}^\star_{\mathrm{slot}}(\widetilde{H}) = \underbrace{\mathrm{tr}\big(\Sigma_{\alpha\alpha}\, M^\perp\big)}_{\text{irreducible shared mismatch on } \mathcal{S}_R^\perp} + \underbrace{\mathrm{tr}\big(\Sigma_{\alpha|\beta}\, M^R\big)}_{\text{Schur-compensable } RR \text{ leakage}}}. \tag{34}$$

2. (Non-compensation on V_H.) *If the frozen mismatch is supported on the shared slot in the sense* $\Pi_R(\Delta H_i) = 0$ *for all* i *(equivalently:* $\widetilde{H}_i(V_H) \subset V_H$, $\widetilde{H}_i(V_R) = 0$*, and likewise for* H_i*), then* $M^R = 0$, $M^\perp = M$*, and*

$$\mathcal{E}^\star_{\mathrm{slot}}(\widetilde{H}) = \mathrm{tr}\big(\Sigma_{\alpha\alpha}\, M\big) \;\geq\; \gamma_\alpha\, \delta_0^2 \tag{35}$$

whenever $\Sigma_{\alpha\alpha} \succeq \gamma_\alpha I_s$*,* for every $\Sigma_{\alpha\beta}, \Sigma_{\beta\beta}$ *(including* $\beta \not\equiv 0$*). Residual channels cannot cancel shared-slot mismatch on* V_H*: the first summand of (31) never sees* β*.*

3. (Overlap / leakage.) *If* $\Pi_R(\Delta H_i) \neq 0$ *(wrong freeze leaks into the residual slot), the compensable part is exactly the Schur term* $\mathrm{tr}(\Sigma_{\alpha|\beta} M^R)$*; the irreducible floor is* $\mathrm{tr}(\Sigma_{\alpha\alpha} M^\perp)$*, with*

$$\mathcal{E}^\star_{\mathrm{slot}}(\widetilde{H}) \;\geq\; \lambda_{\min}(\Sigma_{\alpha\alpha})\; \mathrm{tr}(M^\perp). \tag{36}$$

Proof. Lemma 5 splits the risk. The second summand is an unconstrained residual problem *inside* $\mathcal{S}_R \cong \mathfrak{gl}(V_R)$ for the projected mismatch $\Pi_R(\Delta H)$ and true residual R. Proposition 3 applied inside $\mathcal{S}_R$ yields (33)–(34).

For (ii), shared-slot support means $\Pi_R(\Delta H_i) = 0$, so $M^R = 0$ and (34) collapses to $\mathrm{tr}(\Sigma_{\alpha\alpha} M)$. The Loewner bound is as in Theorem 3. Independence of $(\Sigma_{\alpha\beta}, \Sigma_{\beta\beta})$ is immediate from Lemma 5: the $\mathcal{S}_R^\perp$ error equals $\sum_i \alpha_i \Delta H_i$ and contains no β.

For (iii), drop the nonnegative Schur term in (34). □

Remark 2 (Scope under general joint covariance)**.** *Under Assumption 2, the matched-design construction places both the true shared generators and the standard wrong/Abelian/spectral freezes in the triad slot, so* $\Pi_R(\Delta H_i) = 0$*. Theorem 7(ii) therefore strengthens Theorem 3: the floor* $\gamma_\alpha \delta_0^2$ *persists for* every *joint law with* $\Sigma_{\alpha\alpha} \succeq \gamma_\alpha I_s$*, including* $\beta \not\equiv 0$*. Residual channels may compensate a shared mismatch only if either (a) residual fitting is unconstrained in* $\mathfrak{gl}(n)$ and $\Sigma_{\alpha\beta} \neq 0$ *(Proposition 3), or (b) the wrong freeze leaks into* $\mathcal{S}_R$ *so that* $M^R \neq 0$ *(Theorem 7(iii)). Shared-slot wrong freezes under model-class residual fitting remain at the floor.*

Remark 3 (Operators on V_H vs. V_R)**.** *Identify endomorphisms with blocks. A model-class residual correction* $\sum_j \beta_j \widehat{R}_j$ *is identically zero as a map* $V_H \to V$ *and as a map* $V \to V_H$*. Hence for every control* u *and every* $x_H \in V_H$*,*

$$\big(A(u) - \widehat{A}(u)\big)x_H = \sum_{i=1}^{s} \alpha_i\, (\Delta H_i) x_H = \sum_{i=1}^{s} \alpha_i\, (\Delta H_i)_{HH} x_H$$

whenever ΔH_i *is shared-slot supported. No choice of* $\widehat{R}$ *and no correlation between* β *and* α *can alter this identity on* V_H*. Passing to* $\|\cdot\|_F^2$ *and taking* $\mathbb{E}$ *recovers (35).*

B.7.4 Shared-channel corollary and mixed experimental probes

Corollary 4 (Shared-channel probes)**.** *If* $\beta = 0$ *a.s. and* $\Sigma_{\alpha\alpha} \succeq \gamma_\alpha I_s$*, then for every residual estimator (model-class or unconstrained),* $\mathcal{E}(\widetilde{H}, \widehat{R}) \geq \gamma_\alpha \delta_0^2$*. This recovers the shared-channel corollary of the upgraded main-text Theorem 3.*

Proof. Under $\beta = 0$, $\widehat{A}(u) = \sum_i \alpha_i \widetilde{H}_i$ regardless of $\widehat{R}$, so $\mathcal{E} = \mathrm{tr}(\Sigma_{\alpha\alpha} M) \geq \gamma_\alpha \delta_0^2$. □

Corollary 5 (Mixed probing measures used in experiments)**.** *Let* $\mu = (1-\varepsilon)\mu_{\mathrm{sh}} + \varepsilon \mu_{\mathrm{res}}$ *be a mixture of a shared-channel law* μ_{sh} *(supported on* $\beta = 0$*) with* $\mathbb{E}_{\mu_{\mathrm{sh}}}[\alpha\alpha^\top] \succeq \gamma_{\mathrm{sh}} I_s$ *and a residual-channel law* μ_{res} *(supported on* $\alpha = 0$*) with* $\mathbb{E}_{\mu_{\mathrm{res}}}[\beta\beta^\top] \succeq \gamma_{\mathrm{res}} I_r$*, for some* $\varepsilon \in [0,1]$*. Then*

$$\Sigma_{\alpha\alpha} = (1-\varepsilon)\, \mathbb{E}_{\mu_{\mathrm{sh}}}[\alpha\alpha^\top], \qquad \Sigma_{\alpha\beta} = 0, \qquad \Sigma_{\alpha|\beta} = \Sigma_{\alpha\alpha}.$$

Consequently:

1. *under model-class residuals and shared-slot wrong freezes,* $\mathcal{E}^\star_{\mathrm{slot}} \geq (1-\varepsilon)\gamma_{\mathrm{sh}}\, \delta_0^2$*;*
2. *under unconstrained residuals,* $\mathcal{E}^\star_{\mathrm{unc}} \geq (1-\varepsilon)\gamma_{\mathrm{sh}}\, \delta_0^2$ *as well, since* $\Sigma_{\alpha\beta} = 0$ *precludes Schur compensation.*

In particular, isotropic product probes with $\Sigma = \mathrm{diag}(\Sigma_{\alpha\alpha}, \Sigma_{\beta\beta})$ never allow residual cancellation of shared-slot mismatch; only nonzero cross-covariance $\Sigma_{\alpha\beta}$ can.

Proof. Linearity of second moments under mixtures gives the block formulas; $\alpha\beta^\top = 0$ μ_{sh}-a.s. and $\alpha\beta^\top = 0$ μ_{res}-a.s., hence $\Sigma_{\alpha\beta} = 0$. Apply Theorem 7(ii) and Corollary 3. □

Corollary 6 (Finite-sample residual OLS)**.** *Under the Gaussian observation model (12) with $\Sigma_{\beta\beta} \succ 0$, let $\widehat{R}_N$ be the residual OLS estimator after freezing $\widetilde{H}$ (unconstrained in $\mathfrak{gl}(n)$). Then*

$$\mathbb{E}\big[\mathcal{E}(\widetilde{H}, \widehat{R}_N)\big] = \mathcal{E}^\star_{\mathrm{unc}}(\widetilde{H}) + \frac{\sigma^2\, n^2}{N}\ \mathrm{tr}\big(\Sigma_{\beta\beta}^{-1}\Sigma^{\mathrm{test}}_{\beta\beta}\big),$$

when the test law equals the training law (matched design) with residual Gram $\Sigma^{\mathrm{test}}_{\beta\beta} = \Sigma_{\beta\beta}$; more generally the variance term is $(\sigma^2 n^2/N)\ \mathrm{tr}(\bar{\Sigma}_N^{RR,-1}\Sigma^{\mathrm{test}}_{\beta\beta})$. The bias floor $\mathcal{E}^\star_{\mathrm{unc}}$ is N-independent; only the OLS variance vanishes as $N \to \infty$.

Proof. Decompose $\widehat{R}_N = B^\star + (\widehat{R}_N - B^\star)$. Matched-design residual OLS is unbiased for $B^\star$ with conditional covariance $(\sigma^2/N)(\bar{\Sigma}_N^{RR})^{-1} \otimes I_{n^2}$ on $\mathrm{vec}(B)$, as in Appendix B.1. Cross terms vanish in expectation; the variance contribution is the freeze identity of Theorem 1 with parameter count r (here the "signal" includes the Schur correction absorbed into $B^\star$). □

Summary of regimes.

Regime	**Population floor after optimal $\widehat{R}$**
$\beta = 0$ (Corollary 4)	$\mathrm{tr}(\Sigma_{\alpha\alpha} M) \geq \gamma_\alpha \delta_0^2$
Unconstrained $\widehat{R}$, general Σ	$\mathrm{tr}(\Sigma_{\alpha\vert\beta} M)$ (Schur)
Model-class $\widehat{R}$ + Ass. 2, shared-slot (Theorem 3/7(ii))	$\mathrm{tr}(\Sigma_{\alpha\alpha} M) \geq \gamma_\alpha \delta_0^2$ *even if* $\beta \not\equiv 0$
Model-class + RR leakage	$\mathrm{tr}(\Sigma_{\alpha\alpha} M^\perp) + \mathrm{tr}(\Sigma_{\alpha\vert\beta} M^R)$

)

B.8 Source-error correction to the residual-rate law

Corollary 7 (Source error enters residual rate)**.** *Fix the frozen tuple $\widetilde{H}$, independently of target noise, in an aligned control chart. Let Δ have rows $\mathrm{vec}(H_i - \widetilde{H}_i)^\top$, and let $Z_\alpha \in \mathbb{R}^{N\times s}$ and $Z_\beta \in \mathbb{R}^{N\times r}$ be the observed control designs, with Z_β of full column rank. For unrestricted matrix-entry residual OLS and matched empirical prediction risk, write $P_\beta = Z_\beta(Z_\beta^\top Z_\beta)^{-1}Z_\beta^\top$. Then*

$$\mathbb{E}[\mathcal{E} \mid \widetilde{H}, Z] = \frac{rn^2\sigma^2}{N} + B_N(\widetilde{H}), \qquad B_N(\widetilde{H}) = \frac{\|(I - P_\beta)Z_\alpha\Delta\|_F^2}{N}. \tag{37}$$

If $Z_\alpha^\top Z_\beta = 0$ and $Z_\alpha^\top Z_\alpha/N = I_s$, then $B_N = \|\Delta\|_F^2$. In a Procrustes-aligned gauge this equals δ_0^2, so

$$\kappa(\widetilde{H}) = r + \frac{N\delta_0^2}{n^2\sigma^2}.$$

For a general finite design, B_N replaces δ_0^2 in this identity. Population independence alone does not make the empirical cross-Gram vanish.

Proof. After subtracting the frozen shared response, the vectorized target response is $Z_\beta R + Z_\alpha\Delta + E$. Residual OLS fits $Z_\beta R + P_\beta Z_\alpha\Delta + P_\beta E$, so the clean prediction error is $(I - P_\beta)Z_\alpha\Delta - P_\beta E$. The two terms are orthogonal in sample space. Since the noise has independent entries of variance σ^2 and $\mathrm{tr}(P_\beta) = r$, its expected squared Frobenius norm is $rn^2\sigma^2$. Dividing by N proves the first identity. The stated Gram conditions give $P_\beta Z_\alpha = 0$ and $\|Z_\alpha\Delta\|_F^2/N = \|\Delta\|_F^2$. □

Theorem 3 gives the separate population bound $B(\widetilde{H}) \geq \gamma_\alpha \delta_0^2$ for slot-constrained residuals with orthogonal shared mismatch. Its variance count is different: if each residual matrix is fitted in a fixed q-dimensional output slot, the matched variance is $rq\sigma^2/N$, with $q = \dim \mathcal{S}_R = (\dim V_R)^2$ for that slot. The full-entry experiments use $q = n^2$ and the finite-design projection above.

Empirical source-error fit. Script `run_source_error_kappa_check.py` records per-replicate δ_0^2 on a compact matched panel. With $s = r = 3$, $n = 6$, $\sigma = 0.08$, $N_{\mathrm{src}} \in \{32, 64, 128, 256, 512\}$, $N_{\mathrm{tgt}} \in \{32, 128\}$, and 12 repetitions, the population-isotropic bias proxy is strongly associated with the measured excess: Pearson correlation between $\hat{\kappa} - r$ and $N_{\mathrm{tgt}}\delta_0^2/(n^2\sigma^2)$ is 0.987 (Spearman 0.922). Using $\hat{\kappa}_{\mathrm{est}} - \hat{\kappa}_{\mathrm{oracle}}$ gives Pearson 0.994 (Spearman 0.993). The empirical through-origin slope is 1.004 for $\hat{\kappa} - r$ and 0.996 oracle-centered, matching the leading bias law rather than merely monotonicity. At the hardest cell $(N_{\mathrm{src}}, N_{\mathrm{tgt}}) = (512, 128)$, $\hat{\kappa}_{\mathrm{est}} = 3.818 \pm 0.236$, $\hat{\kappa}_{\mathrm{oracle}} = 3.196 \pm 0.235$, $\delta_0^2 = 0.00115 \pm 0.00011$, and the predicted excess is 0.637 ± 0.061. Along $N_{\mathrm{tgt}} = 128$, δ_0^2 shrinks $0.02287 \to 0.00115$ while $\hat{\kappa}$ improves $15.77 \to 3.82$, consistent with decreasing source bias. The finite-design identity above uses the projected mismatch, whereas this recorded fit uses the population-isotropic proxy. Full outputs are in `results/source_error_kappa.{json,txt}`.

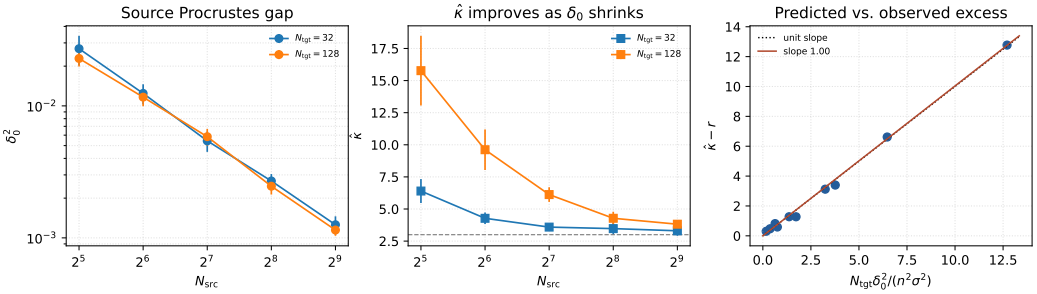

Figure 7: Source-error correction to positive transfer. Left: δ_0^2 shrinks with source sample size. Center: matched-design $\hat{\kappa}$ approaches the residual rate as the freeze gap shrinks. Right: observed excess $\hat{\kappa} - r$ against the predicted $N_{\mathrm{tgt}}\delta_0^2/(n^2\sigma^2)$.

C Lie-manifold approximate symmetry

C.1 Approximate symmetry on the Lie-subalgebra manifold

Real symmetries are imperfect, but a useful mismatch model should remain inside the class of objects the theory freezes: Lie subalgebras isomorphic to $\mathfrak{so}(3)$ (or faithful matrix representations thereof). The interpolated path

$$\widehat{\mathfrak{h}} = (1 - \delta)\mathfrak{h}^\star + \delta M, \qquad M \text{ random skew}, \tag{38}$$

is an empirical bias diagnostic only: for $\delta > 0$ it generally exits the manifold of Lie subalgebras, so δ conflates representation error with model-class exit. Analytic $F(\delta)$ under (38) is therefore not a Lie-algebraic quantity.

Manifold construction (Ad-conjugacy). Let $\{H_i\}_{i=1}^s$ be the Killing-normalized embedded $\mathfrak{so}(3)$ triad of the direct-sum domains (Appendix domains; $\|H_i\|_F^2 = 2$, $\langle H_i, H_j\rangle_F = 0$ for $i \neq j$). Fix a Givens rotation $R(\delta) \in \mathrm{SO}(n)$ that rotates one shared support axis toward an orthogonal residual-slot axis by angle δ, and set

$$\widetilde{H}_i \;=\; \mathrm{Ad}_{R(\delta)} H_i \;=\; R(\delta)\, H_i\, R(\delta)^\top. \tag{39}$$

Then $\widetilde{\mathfrak{h}} := \mathrm{span}\{\widetilde{H}_i\}$ is again a Lie algebra isomorphic to $\mathfrak{so}(3)$ for every δ: conjugacy preserves brackets. We take δ itself as the geometric mismatch—equivalently, the nonzero principal angles between $\mathrm{span}(\mathfrak{h}^\star)$ and $\mathrm{span}(\widetilde{\mathfrak{h}})$ are (δ, δ) (one angle remains 0 for this single-plane tilt).

Procrustes gap. Recall Theorem 3's mismatch

$$\delta_0^2 \;=\; \inf_{Q\in O(s)} \sum_{i=1}^{s} \Big\| \widetilde{H}_i - \sum_{k=1}^{s} H_k Q_{ki} \Big\|_F^2. \tag{40}$$

Proposition 4 (Closed-form δ_0 for Ad-Givens $\mathfrak{so}(3)$)**.** *For $0 \le \delta \le \pi/2$, under the Killing-normalized embedding above and the Ad-Givens construction (39), the principal angles between* $\mathrm{span}(\mathfrak{h}^\star)$ *and* $\mathrm{span}(\widetilde{\mathfrak{h}})$ *are* $(0, \delta, \delta)$*, and*

$$\delta_0(\delta)^2 \;=\; 8\,(1-\cos\delta). \tag{41}$$

In particular $\delta_0(0) = 0$ and $\delta_0(\pi/2)^2 = 8$.

Proof sketch. Frobenius-orthogonal generators with $\|H_i\|_F^2 = 2$ scale the chordal identity $\delta_0^2 = 2\sum_j(1-\cos\theta_j)$ (valid for orthonormal frames) by 2, yielding $\delta_0^2 = 4\sum_j(1-\cos\theta_j)$. The single-plane Ad tilt produces $\theta = (0, \delta, \delta)$, hence (41). Direct Procrustes SVD on the vectorized bases confirms the same closed form to machine precision. □

Population floor. For a shared-channel probe, the squared generator mismatch gives the risk directly, including the cross-support entries introduced by conjugacy. Under isotropic probes the bound is sharp.

Proposition 5 (Analytic freeze floor on the manifold)**.** *Assume $0 \le \delta \le \pi/2$, Assumption 2, and an isotropic shared-channel probe $\mathbb{E}[\alpha\alpha^\top] = \gamma_\alpha I_s$, $\beta = 0$ a.s. For the oracle freeze of $\widetilde{H}$ from (39),*

$$F(\delta) \;:=\; \mathcal{E}(\widetilde{H}, \widehat{R}) \;=\; \gamma_\alpha\, \delta_0(\delta)^2 \;=\; 8\gamma_\alpha\,(1-\cos\delta), \tag{42}$$

for every residual estimator $\widehat{R}$. The same population floor is attained under matched isotropic designs with independent residual channels ($\alpha \perp \beta$), provided the shared and residual controls are centered, because the cross second moment then vanishes.

Proof. Under $\beta = 0$ a.s. and Assumption 2, $A(u) - \widehat{A}(u) = \sum_i \alpha_i (H_i - \widetilde{H}_i)$ regardless of $\widehat{R}$. Hence

$$\mathcal{E} = \mathrm{tr}\big(\mathbb{E}[\alpha\alpha^\top]\, M\big) = \gamma_\alpha\, \mathrm{tr}(M),$$

where $M_{ik} = \langle H_i - \widetilde{H}_i,\; H_k - \widetilde{H}_k \rangle_F$. For (39) the Procrustes optimizer is $Q = I$, so $\mathrm{tr}(M) = \delta_0(\delta)^2$ and (42) follows from Proposition 4. With $\alpha \perp \beta$ isotropic, population residual OLS recovers the true residual generators and leaves the same shared mismatch, so the matched-design floor coincides. □

Corrected crossover. Scratch risk under matched empirical design is $(s+r)n^2\sigma^2/N$ (Theorem 1). For the full-entry OLS comparison with additive bias and residual variance (exact for empirical block-orthogonal designs), freeze risk is $F(\delta) + rn^2\sigma^2/N$. Hence the crossover is:

$$F(\delta) \;<\; \frac{sn^2\sigma^2}{N} \quad\Longleftrightarrow\quad N \;<\; N_\star(\delta) \;:=\; \frac{sn^2\sigma^2}{8\gamma_\alpha(1-\cos\delta)}. \tag{43}$$

For $\delta = 0$, $F = 0$ and the freeze retains residual rate for all N. As δ grows, $N_\star$ shrinks; beyond the crossover, scratch dominates. This replaces an appeal to a merely measured floor for the non-Lie path (38).

Experiment. On Domain D2 (s=r=3, n=6, σ=0.08, γ_α=1), we freeze Ad-Givens $\widetilde{H}$ and the old path (38) on identical matched designs (Figure 8). Noise-free shared-channel Monte Carlo matches $F(\delta)$ within sampling error; large-N noisy floors track $8(1-\cos\delta)$; every Ad-Givens freeze remains bracket-closed, while the old path has positive bracket residual for every $\delta > 0$. Thus the analytic floor is a Lie-manifold statement, and the old non-Lie curve is retained only as an ablation.

Note on small-δ finite-N inflation. At δ=0.05, the analytic bias $F \approx 0.01$ is comparable to residual variance $rn^2\sigma^2/N$ on the reported grid, so the noisy large-N proxy sits slightly above $F(\delta)$. Noise-free shared-channel Monte Carlo recovers F to relative error $< 1\%$ at every δ (including 0.05). For $\delta \ge 0.1$ the bias dominates and relative floor error stays below 10%.

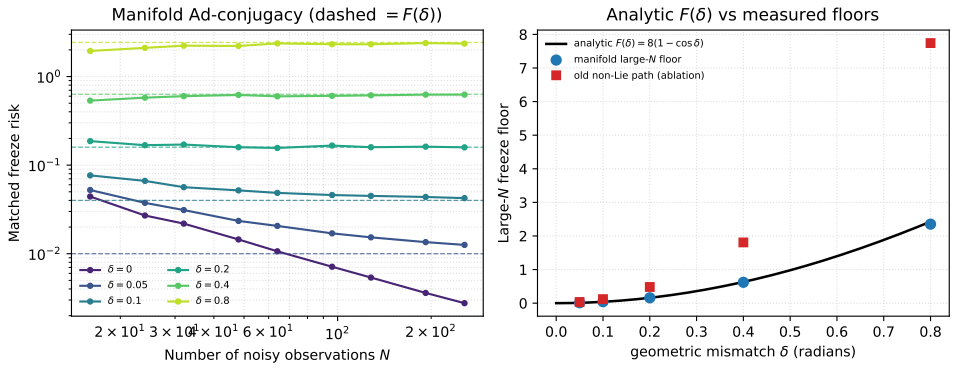

Figure 8: Manifold approximate symmetry. Left: matched freeze risk vs. N for Ad-Givens mismatch δ, with dashed analytic floors $F(\delta) = 8(1 - \cos\delta)$. Right: measured large-$N$ floors vs. analytic $F(\delta)$, with the old non-Lie interpolation path as ablation.

Table 3: Manifold Ad-Givens floors vs. analytic $F(\delta) = 8(1 - \cos\delta)$ (γ_α=1; large-N mean over $N \geq 128$). Old-path floors are ablation only (exit Lie class for $\delta > 0$). Source: `results/approx_symmetry_manifold.json`.

δ	$F(\delta)$ analytic	manifold large-N floor	old-path floor
0	0	0.0039 (slope ≈ -1)	0.0038
0.05	0.0100	0.0138	0.0338
0.1	0.0400	0.0437	0.120
0.2	0.1595	0.160	0.482
0.4	0.6315	0.621	1.81
0.8	2.426	2.35	7.74

C.2 Finite-source identification and target transfer

The matched-design panel freezes the *oracle* shared factor of Domain 1. Here we close the loop with **identify** $\rightarrow$ **align** $\rightarrow$ **freeze** and report joint $(N_{\mathrm{src}}, N_{\mathrm{tgt}})$ curves (`scripts/run_source_target_transfer.py`; Figure 3).

Setup. Domains are as in the main text: source $D_1 = \mathfrak{so}(3) \oplus$ Heisenberg, target $D_2 = \mathfrak{so}(3) \oplus$ toral, with shared dimension s=3, residual dimension r=3, and state dimension n=6. Observation noise is i.i.d. Frobenius Gaussian with σ=0.08.

1. **Identify (source).** Draw N_{src} controls and observe noisy instantaneous generators $\widetilde{A}(u) = A(u) + \sigma\Xi$ on D_1 (matched design of Theorems 1, 3, and 5). Recover channel-labeled generators by OLS (scratch OLS; linear generator regression in `identify.py`). Extract the estimated shared block $\widehat{H} \in \mathfrak{gl}(n)^s$.
2. **Type-check.** Soft $\mathfrak{so}(3)$ barcode via Killing orthonormalization and structure-constant matching (`canonicalize_so3`). Noisy OLS triads need not be exactly bracket-closed; on failure we keep the raw OLS triad and record the barcode miss.
3. **Align (shared chart).** Map the estimated triad into the target control chart by subspace Procrustes on vectorized generators: $\min_{Q \in \mathrm{O}(s)} \|V_{\widehat{H}}\, Q - V_{\mathrm{ref}}\|_F$. The reference is the concrete $\mathfrak{so}(3)$ embedding shared by D_1 and D_2. *Assumption:* source and target already share a matrix chart for the shared factor; Procrustes removes residual basis gauge so frozen generators match target α-channels. It does *not* discover a shared chart from scratch.
4. **Freeze and transfer (target).** Freeze the aligned $\widehat{H}$ on D_2; learn residual generators from N_{tgt} noisy target observations; evaluate matched empirical design risk $\widehat{\mathcal{E}} =$

$\frac{1}{N_{\mathrm{tgt}}}\sum_\ell \|A(u_\ell) - \widehat{A}(u_\ell)\|_F^2$ on the target training design, and the complexity proxy $\hat{\kappa} = \widehat{\mathcal{E}}\, N_{\mathrm{tgt}}/(n^2\sigma^2)$.

Estimators.

- **Estimated-correct freeze:** freeze source-identified, chart-aligned $\widehat{H}$; learn residual.
- **Oracle freeze:** freeze the true shared generators ($\kappa \to r$).
- **Scratch:** learn all $s{+}r$ generators on the target ($\kappa \to s{+}r$).
- **Wrong freeze:** freeze the source Heisenberg residual as if it were shared (bias floor).
- **Raw estimated freeze:** freeze OLS $\widehat{H}$ without Procrustes (channel labels already match; nearly identical to aligned when the chart is shared).

Principal angles between $\mathrm{span}(\widehat{H})$ and the true shared span are recorded versus N_{src} as a source-recovery diagnostic.

Joint scaling. Under an oracle shared freeze, $\hat{\kappa} \to r{=}3$; under scratch, $\hat{\kappa} \to s{+}r{=}6$. With a *finite-source* estimate the target risk decomposes schematically as

$$\widehat{\mathcal{E}} \;\approx\; \underbrace{\mathbb{E}_\alpha \big\|(\widehat{H} - H)\,\alpha\big\|_F^2}_{\text{source bias floor}} + \Theta\big(r\, n^2\sigma^2/N_{\mathrm{tgt}}\big),$$

so

$$\hat{\kappa} \;\approx\; r \;+\; \frac{N_{\mathrm{tgt}}}{n^2\sigma^2}\,\mathbb{E}_\alpha \big\|(\widehat{H} - H)\,\alpha\big\|_F^2.$$

Consequently $\hat{\kappa} \to r$ as $N_{\mathrm{src}} \to \infty$ (bias $\to 0$), while for fixed finite N_{src} the bias term inflates $\hat{\kappa}$ as N_{tgt} grows—exactly the joint scaling in Figure 3.

Measured numbers (σ=0.08, 16 reps). Grid $N_{\mathrm{src}} \in \{32, 64, 128, 256, 512\}$, $N_{\mathrm{tgt}} \in \{16, 32, 64, 128\}$. Oracle mean $\hat{\kappa}$=3.03, scratch mean $\hat{\kappa}$=6.01. At the hardest cell $(N_{\mathrm{src}}, N_{\mathrm{tgt}}) = (512, 128)$, estimated-correct freeze yields $\hat{\kappa}{=}3.72 \pm 0.31$ (within 1 of r=3). At fixed N_{tgt}=128, estimated $\hat{\kappa}$ falls with source size: 17.2 (N_{src}=32) → 8.9 (64) → 6.3 (128) → 4.3 (256) → 3.7 (512). Principal angles between $\mathrm{span}(\widehat{H})$ and the true shared chart shrink as $4.72^\circ \to 0.96^\circ$ (max angle) over the same N_{src} range; type-check success rate rises from 0 to 0.97. Full tables: `results/source_target_transfer.{json,txt}`.

Identification model. Shared generators are recovered by linear generator regression with observed control channels. The barcode is a type-check on a noisy triad; Procrustes then aligns the triad into a shared chart.

C.3 Trajectory source identification

Trajectory source identification. For the source stage we replace oracle matrix observations by controlled trajectories. Each source trajectory supplies an initial condition, labeled controls $u_t = (\alpha_t, \beta_t)$, states x_t, and noisy velocity observations $\dot{x}_t = A(u_t)x_t + \xi_t$; the regression target is never the matrix $A(u_t)$. We solve the linear velocity-matching system $\dot{x}_t \approx (\sum_i \alpha_{t,i} S_i + \sum_j \beta_{t,j} R_j)x_t$, extract the $s = 3$ shared block, apply the $\mathfrak{so}(3)$ barcode type-check, Procrustes-align to the shared chart, and freeze on the target under matched generator risk.

N_{src}	traj-freeze $\hat{\kappa}$	oracle $\hat{\kappa}$	scratch $\hat{\kappa}$	wrong $\hat{\kappa}$	max angle
64	3.92±0.50	3.11	6.21	5106.5	1.09°
128	3.45±0.33	3.08	6.05	5012.4	0.72°
256	3.21±0.36	3.03	6.06	5176.1	0.51°
512	3.19±0.46	3.10	6.14	5216.1	0.35°

At $N_{\mathrm{src}} = 512$, trajectory-freeze has $\hat{\kappa} = 3.19$ versus oracle 3.10 and scratch 6.14; wrong and Abelian freezes remain floor-level.

D NEURAL-OPERATOR TRAJECTORY TRANSFER

D.1 NEURAL-OPERATOR TRANSFER ACROSS COEFFICIENT FAMILIES

Motivation. The Galerkin coefficient OLS probes in the main text are structurally matched to Eq. (2): supervision is on instantaneous generators $A(u)$, not on held-out solution maps. A natural request is something closer to *operator learning*—a map from inputs (initial fields / controls) to trajectory snapshots—while still testing the shared-subalgebra transfer claim. We therefore run a minimal neural-operator (NO) transfer panel on the same controlled bilinear families, deliberately keeping sizes CPU-tractable.

Families. Source and target are Fourier–Galerkin coefficient ODEs on $\mathbb{R}^n$ with $n{=}6$ (triad modes 0:2, residual slot 3:5):

$$\dot{x} = \Big(A_0 + \sum_{i=1}^{3} \alpha_i S_i + R(\beta)\Big)x,$$

where $\{S_i\} \cong \mathfrak{so}(3)$ is shared across families, A_0 is the diffusion drift, and the residual differs: source uses a Heisenberg-like shear residual; target uses a toral (diagonal) reaction residual. Controls (α_t, β_t) are piecewise constant; ground-truth trajectories use the matrix exponential.

Operator task and split. Each sample is an *operator query* $(x_0, \{\alpha_t, \beta_t\}_{t=1}^{T}) \mapsto \{x_t\}_{t=0}^{T}$ with horizon $T{=}6$. Training uses N_{tgt} target trajectories; evaluation is *population-style*: mean trajectory MSE on a held-out set of 48 $(x_0, \mathrm{control})$ pairs drawn independently of training (never reused across estimators within a seed). We report mean$\pm$std over 5 seeds. Source pretraining uses $N_{\mathrm{src}}{=}96$ trajectories from the Heisenberg family; the learned shared head is what we transfer (not an oracle embedding of $A(u)$).

Architecture. We use a bilinear generator network with a neural residual (SG-NRO):

$$A_\theta(\alpha, \beta) = A_0 + \sum_{i=1}^{3} \alpha_i S_i + R_\varphi(\beta), \qquad R_\varphi(\beta) = \mathrm{reshape}\big(W_2 \tanh(W_1\beta + b_1) + b_2\big) \in \mathbb{R}^{n\times n}.$$

Parameters are fit by next-step *velocity matching* on observed snapshots ($\hat{v} = A_\theta x$ vs. finite-difference velocities), with analytic SGD gradients; test rollouts use $\exp$m. This is operator/trajectory learning: the model never sees oracle $A(u)$ matrices.

Protocols.

1. **Scratch:** train $\{S_i\}$ and R_φ from scratch on the target.
2. **Correct freeze:** freeze the source-learned shared head $\{S_i\}$, adapt only R_φ on the target.
3. **Wrong freeze:** freeze the source Heisenberg residual generators in the shared slots; adapt R_φ.
4. **Abelian freeze:** freeze commuting diagonal impostors on the triad; adapt R_φ.
5. **Full fine-tune:** initialize from the source model; train all parameters on the target.

Results. Figure 9 and Table 4 summarize held-out rollout MSE. At small $N_{\mathrm{tgt}}{=}12$, correct shared transfer beats scratch ($\approx 1.7\times10^{-4}$ vs. $\approx 6.6\times10^{-4}$); wrong and Abelian freezes sit two orders of magnitude higher and do *not* exhibit residual-rate collapse as N grows (floors $\sim 10^{-2}$ at $N{=}48$). Scratch catches up once N_{tgt} is large, as predicted by a DOF story: with enough target data the shared head is re-learnable. Fine-tuning is competitive with correct freeze but does not erase the wrong/Abelian floors.

On this bilinear coefficient family, transferring a source-learned $\mathfrak{so}(3)$ head reduces small-N trajectory risk relative to scratch; algebraically wrong freezes do not. Supervision is trajectory-level rather than known-channel OLS on $A(u)$. Code: `nailong/op_transfer.py`, `scripts/run_no_transfer_experiment.py`.

Table 4: Held-out rollout MSE (mean±std, 5 seeds) for the SG-NRO transfer panel.

Estimator	N=12	N=24	N=48
Correct freeze	$1.70\pm0.36\times10^{-4}$	$1.16\pm0.30\times10^{-4}$	$0.84\pm0.18\times10^{-4}$
Scratch	$6.58\pm1.99\times10^{-4}$	$1.88\pm1.07\times10^{-4}$	$0.80\pm0.13\times10^{-4}$
Fine-tune	$2.38\pm0.45\times10^{-4}$	$1.33\pm0.58\times10^{-4}$	$0.74\pm0.10\times10^{-4}$
Wrong freeze	$2.21\pm0.28\times10^{-2}$	$1.20\pm0.15\times10^{-2}$	$0.93\pm0.17\times10^{-2}$
Abelian freeze	$2.17\pm0.48\times10^{-2}$	$1.18\pm0.26\times10^{-2}$	$0.89\pm0.28\times10^{-2}$

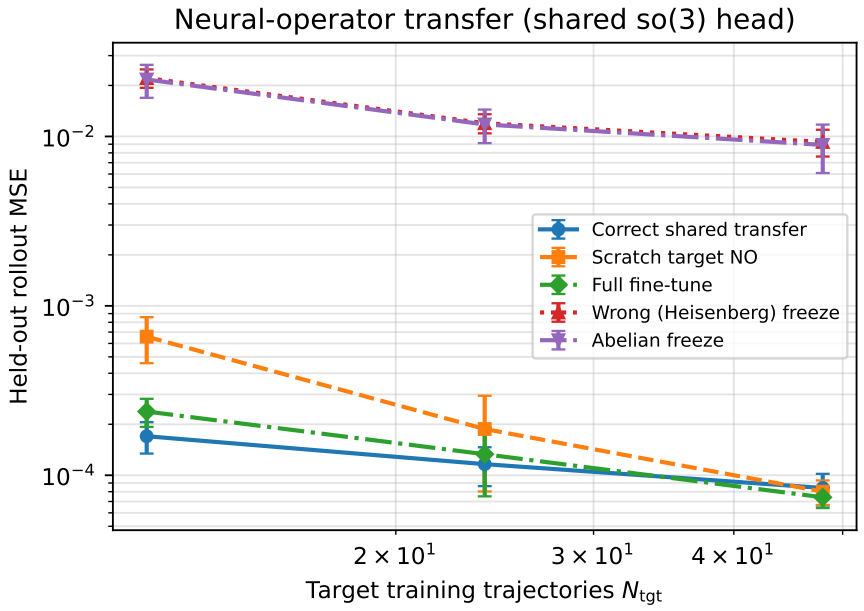

Figure 9: NO-style transfer: held-out trajectory risk vs. target sample size. Correct shared-head transfer dominates scratch at small N; wrong/Abelian freezes retain high floors.

D.2 Harder SG-NRO panel (n=9, horizon 16, OOD populations)

Appendix D uses a tiny bilinear toy (n=6, horizon 6, matched train/test laws). Here the same estimator is scored on a harder operator probe (`scripts/run_no_transfer_harder.py`; Figure 10; `results/no_transfer_harder.txt`):

- Galerkin coefficient dimension n=9 (triad + residual slots; unused modes carry only A_0 drift);
- rollout horizon T=16;
- held-out IC/control populations (train: Gaussian IC + Gaussian controls; test: Laplace IC + uniform controls at larger scale);
- $N_{\mathrm{tgt}} \in \{8, 16, 32, 64\}$, N_{src}=120, 5 seeds.

Architecture and estimators match Appendix D (source-learned shared head freeze vs. scratch / wrong / Abelian / full fine-tune).

Results. At small N, correct shared-head freeze retains a sample-complexity win over scratch (≈3.2× at N=8: $4.51\pm0.42\times10^{-3}$ vs. $1.44\pm0.17\times10^{-2}$; ≈2.6× at N=16). Wrong and Abelian freezes remain floor-level (∼0.07–0.16) and do not collapse with N. As a DOF story predicts, scratch catches up by N=64 (scratch/correct ratio ≈1.02; gap does *not* persist). Fine-tune tracks correct freeze throughout.

Table 5: Harder SG-NRO held-out population rollout MSE (mean±std, 5 seeds; n=9, horizon 16, OOD test laws).

Estimator	N=8	N=16	N=32	N=64
Correct freeze	$4.51\pm0.42\times10^{-3}$	$2.13\pm0.38\times10^{-3}$	$1.69\pm0.48\times10^{-3}$	$1.50\pm0.45\times10^{-3}$
Scratch	$1.44\pm0.17\times10^{-2}$	$5.56\pm2.23\times10^{-3}$	$1.96\pm0.62\times10^{-3}$	$1.54\pm0.27\times10^{-3}$
Fine-tune	$4.51\pm0.12\times10^{-3}$	$2.27\pm0.52\times10^{-3}$	$1.49\pm0.29\times10^{-3}$	$1.36\pm0.30\times10^{-3}$
Wrong / Abelian	~1.4–1.6×10^{-1}	~0.9–1.0×10^{-1}	~0.7–0.8×10^{-1}	$\sim0.7\times10^{-1}$

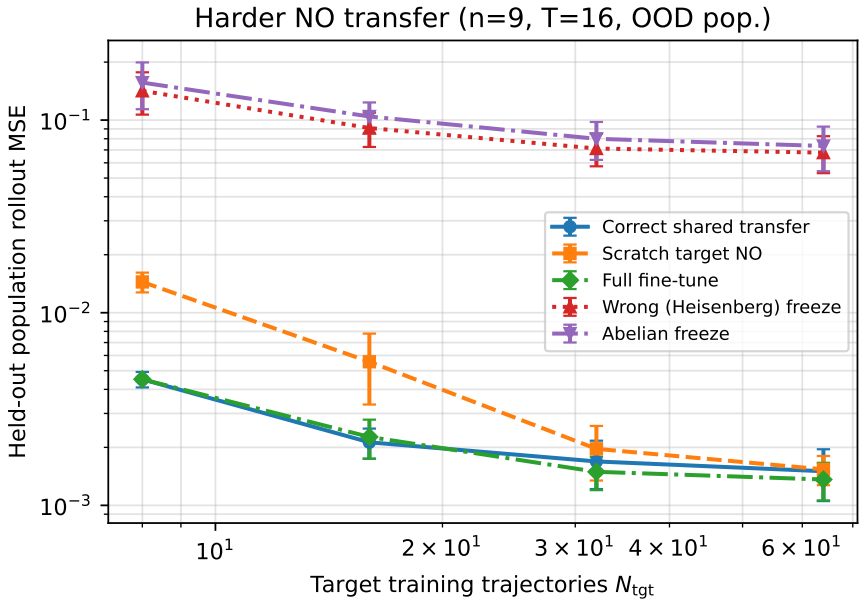

Figure 10: Harder NO-style transfer (n=9, horizon 16, held-out IC/control laws). Correct shared-head transfer beats scratch at small N; wrong/Abelian freezes retain floors; scratch catches up at large N.

The small-N dichotomy survives a longer horizon, larger coefficient dimension, and a train/test population mismatch. Correct beats scratch at N=8; wrong and Abelian freezes remain worse than scratch.

D.3 Field-level operator transfer

The same SG-NRO model is trained in Galerkin coefficient space and scored as a field map. For each held-out rollout we reconstruct $u(\theta)$ from the predicted and true coefficients using `field_from_coefficients`, then average MSE over forecast times $t = 1, \ldots, T$ and a 64-point grid. The source residual is Heisenberg-like, the target residual is toral, and the shared head is the same non-Abelian $so(3)$ triad.

At $N = 8$, correct freeze gives field MSE 2.19×10^{-2} versus scratch 8.52×10^{-2}, a scratch/correct ratio of $3.89\times$. Wrong and Abelian freezes remain much worse (ratios $37.6\times$ and $46.0\times$ versus correct at the same N). Thus the dichotomy holds for field maps, not just the coefficient coordinates.

Table 6: Held-out field MSE for SG-NRO transfer (mean±std over 5 seeds; $n = 9$, horizon 16, $N_{\text{src}} = 120$).

Estimator	$N = 8$	$N = 16$	$N = 32$
Correct freeze	$2.19 \times 10^{-2} \pm 1.98 \times 10^{-3}$	$1.03 \times 10^{-2} \pm 1.81 \times 10^{-3}$	$8.22 \times 10^{-3} \pm 2.31 \times 10^{-3}$
Scratch	$8.52 \times 10^{-2} \pm 1.11 \times 10^{-2}$	$3.05 \times 10^{-2} \pm 1.28 \times 10^{-2}$	$9.85 \times 10^{-3} \pm 3.11 \times 10^{-3}$
Fine-tune	$2.18 \times 10^{-2} \pm 5.63 \times 10^{-4}$	$1.10 \times 10^{-2} \pm 2.49 \times 10^{-3}$	$7.26 \times 10^{-3} \pm 1.41 \times 10^{-3}$
Wrong freeze	$8.24 \times 10^{-1} \pm 1.91 \times 10^{-1}$	$5.39 \times 10^{-1} \pm 1.17 \times 10^{-1}$	$4.25 \times 10^{-1} \pm 8.57 \times 10^{-2}$
Abelian freeze	$1.01 \times 10^{0} \pm 2.77 \times 10^{-1}$	$6.66 \times 10^{-1} \pm 1.34 \times 10^{-1}$	$5.07 \times 10^{-1} \pm 1.30 \times 10^{-1}$

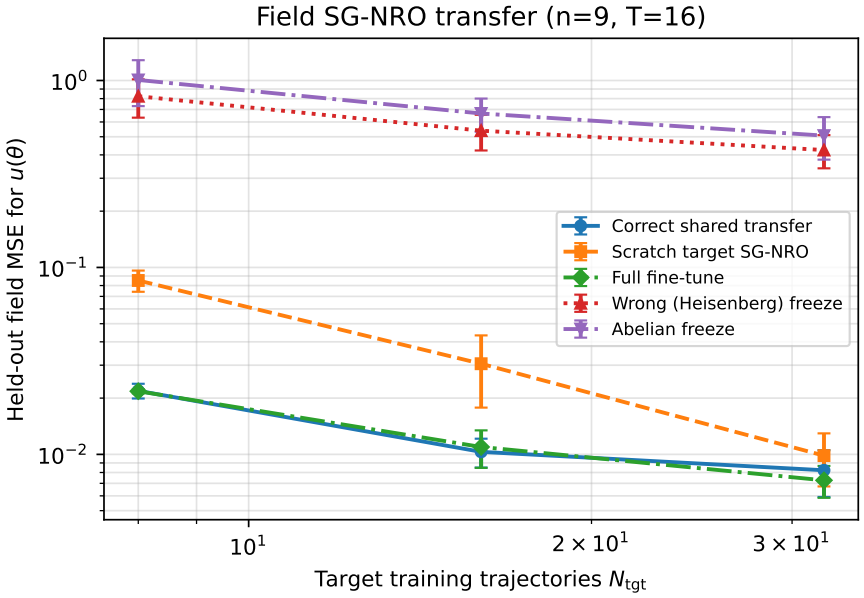

Figure 11: Field-level transfer after reconstructing $u(\theta)$ from coefficient rollouts. Correct shared-head transfer is smaller at N=8, while wrong and Abelian freezes remain separated from the non-Abelian transfer.

E Stronger baselines with uncertainty

E.1 Stronger baselines with seed uncertainty

The panel below adds bracket-aware and Lie-Neuron heads on the same designs and reports seed-level uncertainty for every estimator.

Seed setup. Unless noted: s=r=3, n=6, σ=0.08, $N \in \{16, 24, 32, 48, 64, 96, 128\}$, matched isotropic designs on $D_2 = \mathfrak{so}(3) \oplus \text{toral}$ with source $D_1 = \mathfrak{so}(3) \oplus \text{Heisenberg}$. We draw n_{seed}=24 independent seeds ($\text{seed}_i = 10007\, i$). Source pretraining uses N_{src}=256 at the same σ. For each seed we record the N-grid of matched risks, the seed-mean $\hat{\kappa}$, and the log–log slope of risk vs. N; tables report mean±std across seeds and a normal 95% CI on the mean ($\pm 1.96\,\text{se}$).

Hyperparameters.

- **Bracket-aware freeze.** OLS source shared generators $\widehat{S}$, skew-symmetrize, Killing / structure-constant projection to nearest $\mathfrak{so}(3)$ (direct `canonicalize_so3` when the span closes; otherwise GL(3) remix minimizing $\|\hat{c} - c_{\mathfrak{so}(3)}\|$ after bracket completion), then Procrustes scale-match back to the OLS estimate and freeze on D_2 while learning the residual by OLS.

Table 7: Strong baselines with uncertainty (24 seeds; mean $\hat{\kappa}$ over N, then mean±std across seeds). Bracket-aware / Lie-Neuron / competitive methods *succeed* when the source shares $\mathfrak{so}(3)$; spectral / Abelian / wrong / full-trunk still floor.

Estimator	mean $\hat{\kappa}$	slope
Correct freeze	3.01 ± 0.14	-0.97 ± 0.09
Scratch	6.01 ± 0.20	-0.99 ± 0.06
Bracket-aware freeze	3.30 ± 0.14	-0.89 ± 0.08
Lie-Neuron head	3.38 ± 0.15	-0.90 ± 0.08
Finetune shared (no projection)	3.68 ± 0.15	-0.80 ± 0.07
Competitive fine-tune	3.87 ± 0.16	-0.95 ± 0.07
Spectral proxy	$(2.20 \pm 0.23) \times 10^3$	$+0.04 \pm 0.20$
Abelian impostor	$(2.19 \pm 0.09) \times 10^3$	$+0.06 \pm 0.10$
Wrong freeze	$(2.32 \pm 0.10) \times 10^3$	$+0.06 \pm 0.10$
Full-trunk freeze	$(1.65 \pm 0.05) \times 10^3$	-0.02 ± 0.08

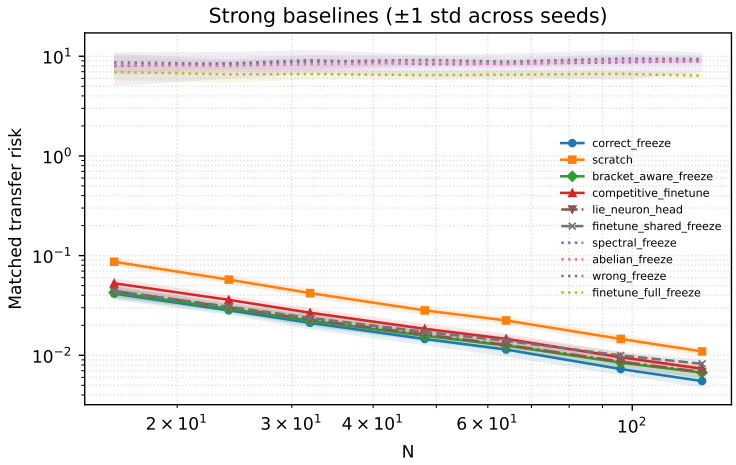

Figure 12: Strong baselines: mean matched risk vs. N with ± 1 std bands across 24 seeds. Bracket-aware and Lie-Neuron heads track residual rate; spectral / full-trunk floors persist.

- **Competitive fine-tune.** Initialize from the bracket-projected source shared block. On the target, closed-form ridge-OLS jointly fits (i) a GL(s) remix of the frozen shared span and (ii) a rank-2 skew out-of-span dictionary (6 fixed random elementary skew matrices) with ridge λ_{oos}=0.1 on out-of-span coefficients only; residual generators are unpenalized.
- **Lie-Neuron head (toy).** Killing-normalized $\mathfrak{so}(3)$ chart from the projected source shared block; shared-side degrees of freedom are an SO(3) remix (Rodrigues / $\mathfrak{so}(3)$ exponential) plus a global scale, trained by Adam (120 steps, lr=5×10^{-2}) with free residual OLS-initialized generators (residual lr 8×10^{-2}).
- **Reference estimators.** Correct freeze, scratch, source-shared freeze without projection, spectral skew-SVD proxy, Abelian impostor, wrong freeze, and full-trunk freeze — all under the same seeds.

Bracket-aware and Lie-Neuron baselines recover the residual-rate trend: bracket-aware freeze attains $\hat{\kappa}$=3.30 ± 0.14 (95% CI $[3.24, 3.36]$), and Lie-Neuron / competitive fine-tune likewise beat scratch with slopes ≈ -1. Spectral proxies and Abelian impostors remain at $\hat{\kappa} \sim 10^3$ with near-zero slope; full-trunk freeze without residual adaptation stays floor-level; competitive fine-tune reaches 3.87 versus oracle 3.01, consistent with the rank-2 out-of-span adapter (Appendix E.3). Bracket-aware $\hat{\kappa}$ mildly inflates with N under source estimation noise (seed-mean slope -0.89).

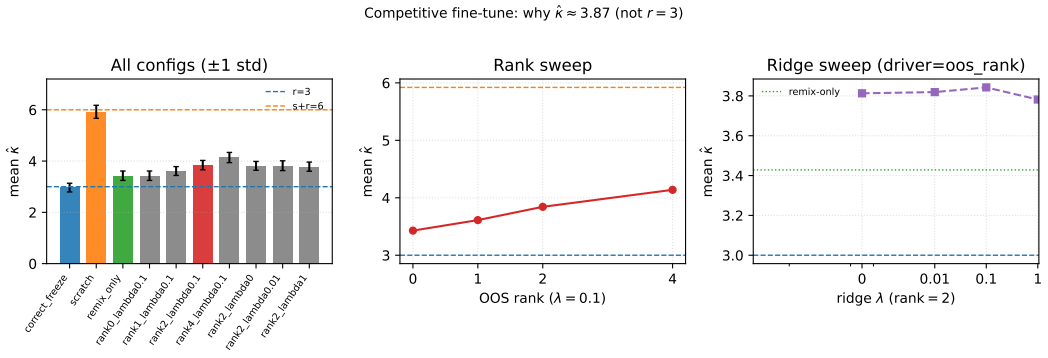

Figure 13: Competitive FT ablations. Left: all configs. Center: OOS rank at λ=0.1 (monotonic lift toward scratch). Right: ridge at rank 2 (flat). Driver: out-of-span rank, not ridge.

E.2 Target-only $\mathfrak{so}(3)$-constrained scratch

The structure-aware heads above still consume a source $\mathfrak{so}(3)$. Here the learner sees only target samples: OLS all $s{+}r$ generators, project the labeled shared 3-plane onto the nearest $\mathfrak{so}(3)$, and re-fit the residual (`scripts/run_target_only_so3.py`; 16 seeds, σ=0.08, same N-grid as Table 7).

Table 8: Target-only $\mathfrak{so}(3)$ scratch vs. oracle / source freeze / unstructured scratch (16 seeds; mean $\hat{\kappa}$ over N, then mean±std across seeds).

Estimator	mean $\hat{\kappa}$
Oracle freeze (true $\mathfrak{so}(3)$)	2.96 ± 0.11
Source $\mathfrak{so}(3)$ freeze	3.27 ± 0.11
Target-only $\mathfrak{so}(3)$ scratch	4.40 ± 0.17
Unstructured scratch	6.00 ± 0.23

The Lie constraint on the target 3-plane beats scratch and does not replace a source freeze: $\hat{\kappa}$ sits between residual rate and ambient rate because a finite-N target OLS plane, even after Killing projection, is a noisier chart than a source-identified $\mathfrak{so}(3)$. (`results/target_only_so3.json`).

E.3 Why competitive fine-tune lands at $\hat{\kappa} \approx 3.87$, not r=3

The structure-aware baseline panel reports competitive fine-tune at $\hat{\kappa}$=3.87 ± 0.16 versus oracle correct freeze $\hat{\kappa}$=3.01 ± 0.14 and scratch 6.01 ± 0.20 (Sec. E.1). The closed-form estimator jointly fits a GL(s) remix of the bracket-projected source shared span and a ridge-penalized rank-2 skew out-of-span (OOS) dictionary. We ablate those knobs to locate the bias.

Setup. Same transfer setting as Appendix E.1: $s{=}r{=}3$, $n{=}6$, σ=0.08, $N \in \{16, 24, 32, 48, 64, 96, 128\}$, source N_{src}=256, n_{seed}=16. Oracles: correct freeze and scratch. Competitive variants: remix-only (OOS rank 0); rank $\in \{0, 1, 2, 4\}$ at panel λ_{oos}=0.1; $\lambda \in \{0, 0.01, 0.1, 1\}$ at panel rank 2. Adam step count is unused (closed-form solver).

- **Remix-only recovers near-residual rate.** With the OOS dictionary removed, $\hat{\kappa}$=3.43±0.18 (gap ≈ 0.46 above oracle 2.97), consistent with finite-N_{src} bracket-projection error on the source shared span—not with a broken remix. Theory's GL(s)-within-chart claim holds once out-of-span DOF are shut off.
- **OOS rank is the knob that closes the gap to 3.87.** At λ=0.1, mean $\hat{\kappa}$ rises 3.43 → 3.61 → 3.84 → 4.14 for ranks 0, 1, 2, 4 ($\Delta\kappa(0{\to}2){=}{+}0.41$). Panel rank 2 reproduces the structure-aware baseline figure (≈ 3.84 vs. 3.87). Each added skew direction is an extra

Table 9: Competitive fine-tune ablations (16 seeds; mean $\hat{\kappa}$ over N, then mean±std). Remix-only sits near residual rate; raising OOS rank monotonically inflates $\hat{\kappa}$. Ridge at fixed rank 2 barely moves the estimate.

Config	rank	λ_{oos}	mean $\hat{\kappa}$
Correct freeze	—	—	2.97 ± 0.17
Scratch	—	—	5.92 ± 0.26
Remix-only (no OOS)	0	—	3.43 ± 0.18
Rank sweep	1	0.1	3.61 ± 0.17
Panel default	2	0.1	3.84 ± 0.18
Rank sweep	4	0.1	4.14 ± 0.20
Ridge sweep	2	0	3.81 ± 0.17
Ridge sweep	2	0.01	3.82 ± 0.19
Ridge sweep	2	1	3.78 ± 0.18

shared-side degree of freedom that the target fit can use, inflating effective complexity above r.

- **Ridge does not explain the inflation.** At fixed rank 2, $\lambda \in \{0, 0.01, 0.1, 1\}$ spans only $\Delta\kappa \approx 0.06$; even λ=1 leaves $\hat{\kappa} \approx 3.78$. Penalizing OOS coefficients cannot undo the capacity introduced by the dictionary itself under this design.
- **The estimator is closed form.** The competitive fit is ridge-OLS; optimization therefore does not explain the gap.

The $3 \to 3.87$ excess is primarily the rank-2 out-of-span skew adapter, with a smaller residual from noisy source projection. Remix-only is the fair "structure-aware competitive" comparison to oracle freeze; the panel's OOS term buys robustness to embedding noise at the cost of a visible $\hat{\kappa}$ tax.

E.4 Unknown orthogonal intertwiner (shared chart up to conjugacy)

Appendix C.2 assumes a *shared matrix chart*: source and target embed the shared factor in the same coordinates, up to an $O(s)$ channel remix removed by subspace Procrustes. Here we treat a shared chart that is unknown up to an **unknown orthogonal intertwiner** $P \in \mathrm{O}(n)$ acting by conjugacy $X \mapsto PXP^\top$ on target generators/observations (`scripts/run_unknown_intertwiner_transfer.py`; Figure 14; `results/unknown_intertwiner_transfer.json`).

What "unknown intertwiner" means under direct-sum embeddings. Domains remain $D_1 = \mathfrak{so}(3) \oplus \text{Heisenberg}$ and $D_2 = \mathfrak{so}(3) \oplus \text{toral}$ on $V = V_H \oplus V_R$ with $\dim V_H = \dim V_R = 3$ (Assumption 2). We sample a single unknown P and replace the target by its conjugate $\mathrm{Ad}_P(D_2)$. Two conjugacy classes are reported:

- **Block-orthogonal.** $P = \mathrm{blkdiag}(P_H, P_R)$ with $P_H, P_R \in \mathrm{O}(3)$. This preserves the splitting $V_H \oplus V_R$. For the standard shared $\mathfrak{so}(3)$ factor, Ad_{P_H} acts on $\mathrm{span}(H)$ as an $O(s)$ channel remix: principal angles between source and target shared spans remain ≈ 0, so *subspace Procrustes* (not full conjugation) is the correct alignment.
- **Full Haar.** $P \sim \mathrm{Haar}(\mathrm{O}(n))$ mixes shared and residual supports. Abstract barcodes still match, but $\mathrm{span}(\mathrm{Ad}_P H)$ is a different embedded subspace of $\mathfrak{gl}(n)$; channel Procrustes alone cannot recover a transferable freeze.

Inequivalent representations with the same barcode need not share an intertwiner (Proposition 2); the full mode is the concrete instance of that gap.

Estimators. Source identification is as in Appendix C.2 (OLS generators, soft $\mathfrak{so}(3)$ barcode check) but *without* assuming the target chart. Target observations use $\mathrm{Ad}_P(D_2)$ with noise σ=0.08.

(i) **Oracle known P.** Freeze $\mathrm{Ad}_P(H^\star)$; learn residual. Then $\hat{\kappa} \to r$=3.

(ii) **Estimated P / Procrustes.** Estimate $(P, Q) \in \mathrm{O}(n) \times \mathrm{O}(m)$ from source vs. target OLS generator sets by alternating channel Procrustes and Riemannian steps on $\mathrm{O}(n)$ (`est_conjugation`), and separately report pure $O(s)$ `channel_procrustes`.

(iii) **Barcode-only / no alignment.** Accept the abstract type-check; freeze the source shared triad in the source chart. This freeze remains at a floor.

(iv) **Scratch.** Learn all $s{+}r$ generators on the target. Then $\hat{\kappa} \to s{+}r{=}6$.

When alignment recovers residual rate vs. when it fails.

- **Block-orthogonal:** oracle $\hat{\kappa} \approx 3$, scratch ≈ 6; channel Procrustes and estimated conjugation recover residual rate ($\hat{\kappa} \to 3$ up to a finite-source floor); barcode-only floors ($\hat{\kappa} \gg 3$) despite near-zero principal angles—the failure is channel mislabeling, not subspace mismatch.
- **Full Haar:** oracle/scratch rates unchanged; estimated conjugation recovers $\hat{\kappa} \to 3$ (finite-source floor); channel Procrustes *fails* (bias floor, $\hat{\kappa}$ thousands); barcode-only floors with large principal angles ($\sim 80°$).

Measured numbers (σ=0.08, 12 reps). Grid $N_{\mathrm{src}} \in \{64, 128, 256, 512\}$, $N_{\mathrm{tgt}} \in \{32, 64, 128\}$.

- **Block.** Oracle mean $\hat{\kappa}$=3.01, scratch 5.94. At $(N_{\mathrm{src}}, N_{\mathrm{tgt}}) = (512, 128)$: channel Procrustes $\hat{\kappa}{=}3.76 \pm 0.46$, est. conjugation 3.96 ± 0.55, barcode-only 5773 (floor). Est. $\hat{\kappa}$ at N_{tgt}=128 falls with source size: $9.25\,(64) \to 6.17\,(128) \to 4.78\,(256) \to 3.96\,(512)$.
- **Full.** Oracle mean $\hat{\kappa}$=2.96, scratch 5.99. At $(512, 128)$: est. conjugation $\hat{\kappa}{=}3.96 \pm 0.29$ (recovers), channel Procrustes 4535 (fails), barcode-only 6394 (floors; max angle $\approx 81°$). Est. conjugation curve at N_{tgt}=128: $9.23\,(64) \to 6.09\,(128) \to 4.83\,(256) \to 3.96\,(512)$.

Full tables: `results/unknown_intertwiner_transfer.{json,txt}`.

Disjoint alignment budget. Source identify uses N_{src}=512 only. Procrustes consumes k target generators that are never reused for residual OLS; the residual sample is a disjoint N=128. At k=16, freeze $\hat{\kappa}{=}4.76{\pm}1.23$ versus scratch $5.71{\pm}0.68$; at k=32, $4.11{\pm}0.74$ versus $5.97{\pm}0.76$ (`results/align_budget.json`). Chart alignment is a $k{\sim}10$ generator cost; the residual-rate payoff is on the disjoint target sample.

E.5 Unmatched train/test designs

Corollary 2 gives $\mathbb{E}[\mathcal{E} \mid \text{design}] = \frac{\sigma^2 n^2}{N} \mathrm{tr}(\bar{\Sigma}_N^{-1} \Sigma_\mu)$ on the estimated channels (full u for scratch; residual β for a correct freeze). Script `run_population_design.py` draws independent train/test Gaussians with $\mathrm{cond} \in \{1, 8, 64\}$ and $N \in \{32, 64, 128\}$. Across the unmatched train/test grid, the ratio of Monte Carlo risk to the trace formula has mean absolute error 0.037 (`results/population_design.json`). Matched $\hat{\kappa}$ is the isotropic special case $\mathrm{tr}(\bar{\Sigma}_N^{-1} \Sigma_\mu) = m$.

E.6 Overlap interpolation of Assumption 2

Script `run_overlap_interpolation.py` leaks shared generators onto V_R and residual generators onto V_H with parameter $\eta \in [0, 1]$. At η=0 the construction is the orthogonal direct-sum of Assumption 2; $\eta > 0$ leaks supports, so the unique-simple-factor argument of Proposition 1 no longer applies. With independent (α, β): unconstrained correct freeze stays at $\hat{\kappa}{\approx}2.90$–$3.19$; slot residuals break as soon as the true residual leaves $\mathcal{S}_R$ ($\hat{\kappa}{=}0.72{\to}661.8$); wrong freeze floors at every η ($\hat{\kappa}{\sim}3{\times}10^3$). With $\mathrm{corr}(\alpha, \beta){=}0.85$, unconstrained wrong freeze Schur-compensates (1072 vs. slot 2960 at η=0) but does not attain r (`results/overlap_interpolation.json`).

E.7 Same-observation spectral, PCA, and DMD proxies

Theorem 5 is an insufficiency construction for the commuting pair (A_0, T). Script `run_same_data_spectral.py` gives the same *noncommuting* source matrices $\{A(u)\}$ (and the same trajectories) to Algorithm 1 and to bracket-blind proxies. At $(N_{\mathrm{src}}, N_{\mathrm{tgt}}) =$

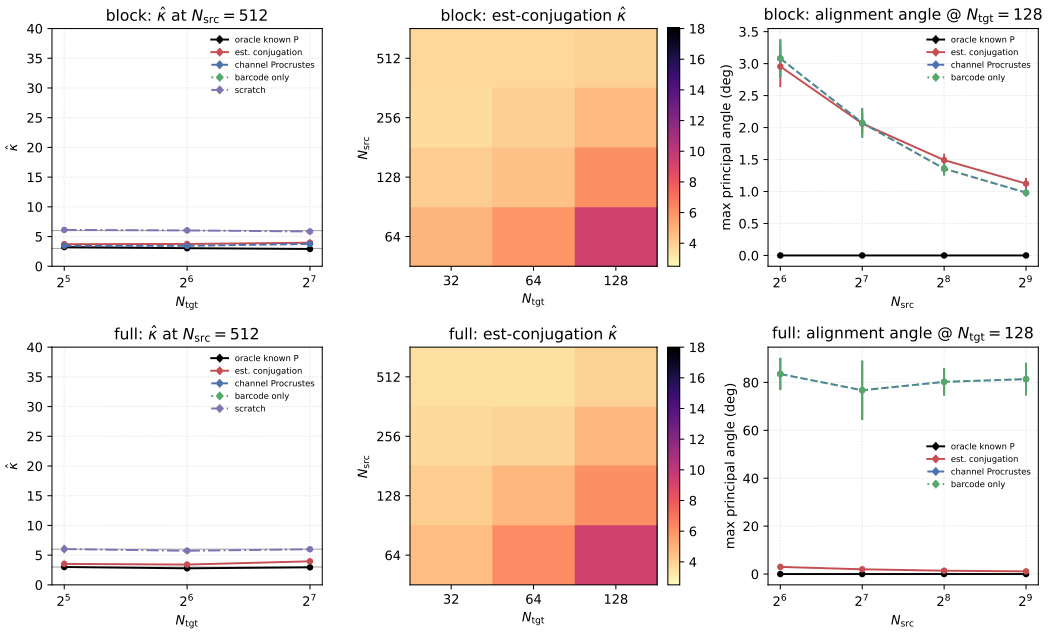

Figure 14: Unknown orthogonal intertwiner transfer. Top: block-orthogonal P (direct-sum preserving). Bottom: full Haar $P \in \mathrm{O}(n)$. Left: $\hat{\kappa}$ vs. N_{tgt} at largest N_{src}. Center: est-conjugation $\hat{\kappa}$ heatmap. Right: max principal angle of the frozen shared span vs. N_{src}. Channel Procrustes recovers only in the block mode; barcode-only floors in both.

$(256, 64)$: Alg 1 OLS $\hat{\kappa}$=3.74 (δ_0=0.003), trajectory identify 3.17, oracle 3.09, scratch 6.01; skew-SVD 1399 (δ_0=5.63, max angle 89°), PCA 536, DMD 1513, commuting triple 2205 (`results/same_data_spectral.json`).

E.8 Identify-stage supervision ablations

Script `run_identify_ablation.py` records the information used by each extractor (N_{src}=256, N_{tgt}=64): labeled channels + Procrustes $\hat{\kappa}$=3.84; unlabeled barcode search over $\binom{6}{3}$ OLS generators 3.76; no Procrustes (already shared chart) 3.83; shuffled channels 1182; wrong s=2 freeze 535; contaminated 2+1 shared/residual tuple 1068 (`results/identify_ablation.json`). Labeled OLS in a shared chart is the finite-source number reported in the main text; the barcode search recovers the $\mathfrak{so}(3)$ factor without channel labels.

E.9 Genuine SO(3) on spherical harmonics

Script `run_sphere_so3_transfer.py` realizes L_x, L_y, L_z as the physical angular-momentum action on real spherical-harmonic coefficients of degree $\leq L_{\max}$, not as a designated 3-mode block. Source residual mixes neighboring ℓ-shells; target residual is toral in the SH chart. At $L_{\max}$=4 (n=25): oracle $\hat{\kappa}$=3.04, Alg 1 4.25, scratch 6.06, one-step field MSE 2.0×10^{-4} / 2.5×10^{-4} / 3.4×10^{-4}; skew-SVD 8512; a Wigner/Schur residual (irrep multipliers) $\hat{\kappa}$=397. Resolution lift by representation theory ($L_{\max}$=3→5) keeps freeze $\hat{\kappa}$=2.98 vs. scratch 5.95 (`results/sphere_so3_transfer.json`). The probe is physical angular momentum on spherical-harmonic coefficients through $L_{\max}$. The field-to-field operator version (never observing $A(u)$) is Appendix E.10.

E.10 Field-to-field SO(3) operator transfer on S^2

The coefficient probe of Appendix E.9 still regresses generator matrices. Script `run_sphere_operator_transfer.py` learns maps $\mathcal{G} : (u_0, \alpha(\cdot)) \mapsto u_T$ on SH coefficients

of $\partial_t u = \nu\Delta_{S^2} u + \alpha(t)\cdot(Lu)$ + lab-axis residual. The shared propagator Φ is the residual-free SO(3)+Laplace flow; the scientifically correct score is residual-of-Φ R^2 (0 if a method equals Φ, 1 if it recovers the axis-breaking residual). Identification is misspecified linear vector-field OLS (the source residual is not in that model class). Wigner is jointly equivariant: a control-conditioned Wigner rotation plus irrep gates, so it *can* represent Φ and *cannot* represent a lab axis. NAILONG freezes identified L inside Φ and fits an affine residual head (the field-level analogue of residual OLS). At $L_{\max}$=3, N_{tgt}=24:

N_{src}	Φ	scratch	Wigner	W-FT	NAILONG	Wigner R^2	W-FT R^2
16	0.0150	0.0265	0.0179	0.0153	0.0067	−0.42	−0.04
64	0.0150	0.0265	0.0178	0.0153	0.0054	−0.41	−0.04
chirp (64)	0.0155	0.0481	0.0241	—	0.0058	−1.41	—

Oracle freeze is 0.0046 (R^2=0.91); source identification relative error is 0.043 / 0.025 at N_{src}=16 / 64, not machine zero. The $\mathfrak{so}(3)$ functor lifts to $L_{\max}$=4 (NAILONG RMSE 0.0039, R^2=0.90) while truncated scratch/Wigner do not (0.029 / 0.018, R^2<0). Theorem 2 supplies the equivariant cap: Frobenius–Schur on the lab-axis Jacobian is 0.208; the GRF L^2 ceiling is 0.034; a split-sample Haar estimate of the Reynolds operator on the nonlinear residual is 0.024. NAILONG R^2=0.87 exceeds all three; jointly equivariant Wigner is at −0.41. Wigner-FT (fit Φ, then adapt only irrep gates) is −0.04 at N_{src}=64, still below the Haar cap 0.024. A source-trained Wigner trunk scores −0.81; adding an unconstrained residual on the raw target field reaches 0.06 / 0.49 at N_{tgt}=8 / 24, while NAILONG is 0.64 / 0.86 (oracle 0.67 / 0.90; `results/trunk_residual.json`). The GRF ceiling 0.034 is $\|P_G R\|^2_{L^2}/\|R\|^2_{L^2} \neq 0$, so the trunk is $\Phi + P_G R_{\text{src}}$ (Theorem 2). Equivariant fine-tunes remain in the Reynolds class (Theorem 2). (`results/sphere_operator_transfer.json`).

E.11 Residual-class capacity sweep

Theorem 4 says a richer residual class D can hide a wrong freeze when $\Phi - \widetilde{\Phi} \in D$. Script `run_residual_capacity.py` freezes the same source Wigner trunk, the identified Φ, a spectral-radius-matched random skew 3-plane, and the oracle $\mathfrak{so}(3)$, then fits one D on N_{tgt}=24 (matched `run_trunk_residual.py`: $L_{\max}$=3, $N_{\text{src,trunk}}$=48). Heads: bias only; random features of width 16/64 with ridge 10^{-2}; affine OLS.

D	trunk+res	wrong+res	NAILONG	oracle
bias	−0.85	−70	−0.02	−0.01
rf16	0.46	−144	0.84	0.80
rf64	0.65	−235	0.88	0.91
affine	0.49	−433	0.86	0.90

Affine recovers the published trunk+res 0.49 / NAILONG 0.86. Width 64 is the most generous adapter: the trunk rises to 0.65 and still sits below NAILONG 0.88. The lab-axis mismatch is not a constant, so a bias head fails for every freeze. The random skew plane is not in any of these D (Theorem 4), so wrong freeze stays a large-negative floor. (`results/residual_capacity.json`).

E.12 Raw-field vorticity on T^2

Script `run_kolmogorov_transfer.py` maps 32×32 vorticity to vorticity. No generator matrix and no Φ are given to the learner. Source and target share viscosity ν=0.03 and the Jacobian; they differ by localized force bumps at opposite points and a real-space drag on the target. Identification is a one-step Fourier magnitude OLS for ν on source fields (high-k modes; $\hat{\nu}$=0.0276). The translation-equivariant linear class is a gain-only Fourier multiplier (P_G for $G = T^2$ through the origin). NAILONG freezes the identified heat kernel and fits the same residual as the trunk. Scratch is a target-only Fourier multiplier with bias. Four seeds; residual-of-Φ R^2 against the true heat flow.

The localized force is almost Haar-orthogonal: $\|P_G R\|^2_{L^2}/\|R\|^2_{L^2}$=0.036±0.007. Theorem 2 then says a gain-only source trunk should already sit near Φ: it does (0.100±0.006). A pointwise scale-and-bias residual—the class that can represent drag plus a bump—reaches 0.222±0.024 on that trunk and 0.167±0.020 on the identified heat kernel. Target-only Fourier scratch is −0.310 because drag is

not a multiplier; the same pointwise residual on scratch is 0.219. On this T^2 vorticity map the trunk sits near Φ. (`results/kolmogorov2d.json`).

E.13 Raw-field identify, unknown chart, operator baselines

Script `run_full_chain_operator.py` is the same S^2 field map $(u_0, \alpha(\cdot)) \mapsto u_T$ as Appendix E.10, now as a closed operator chain. The learner sees lat–lon grids (synthesize then analyze); L and Φ are not OLS inputs. The target field is observed after an unknown Euler chart of radius 0.85; control-channel labels stay in the lab frame, so the identified 3-plane is an adjoint remix. Source identify uses N_{src}=48 raw trajectories; chart recovery is Procrustes against N_{align}=32 target snapshots—the remix Q is never given. Three seeds; $L_{\max}$=3; N_{tgt}=24. Residual-of-Φ R^2 in the observed chart.

Identify relative error is 0.045; aligned-plane error is 0.048. Scratch FNO, source FNO + full fine-tune, and a rank-4 adapter on the frozen source FNO stay negative (-2.55 ± 0.41 / -2.54 ± 0.38 / -7.02 ± 2.02). The source Wigner trunk—identified L, source chart—scores -1.72 ± 0.79 on the rotated target.

D	trunk+res	wrong+res	NAILONG	oracle
bias	-1.90	-66	-0.00	0.01
lora4	-3.08	-282	0.60	0.70
affine	-2.93	-301	0.89	0.92
rf16	-1.25	-118	0.83	0.83
rf64	-1.81	-187	0.91	0.94

Affine NAILONG is 0.890 ± 0.029 (oracle 0.925 ± 0.015). The same residual class on the source trunk stays below 0 at every width: unknown-chart mismatch is not in D. A spectral-radius-matched wrong 3-plane stays a large-negative floor after every widening of D (Theorem 4). (`results/full_chain_operator.json`).

E.14 Barotropic vorticity on S^2

Scripts `run_barotropic_transfer.py` (CPU) and `run_barotropic_gpu.py` evolve the controlled barotropic vorticity equation

$$\partial_t \zeta = \nu \Delta_{S^2} \zeta + \alpha(t) \cdot (L\zeta) - J(\psi, \zeta + f + h) - r\zeta - \gamma\, \mathrm{axis} \odot \zeta, \qquad \Delta\psi = \zeta,$$

in the real-SH chart through $L_{\max}$=8 (d=81; Gauss–Legendre $\times$ equispaced grid 18×36). The spherical Jacobian $J(\psi, \zeta) = \partial_\lambda \psi\, \partial_\mu \zeta - \partial_\mu \psi\, \partial_\lambda \zeta$ is $\mathrm{SO}(3)$-equivariant. Planetary vorticity $f \propto \mu$, a Y_2^{2c} ridge, and lab-axis drag break the group. Identification is misspecified linear vector-field OLS on independent source snapshots, or velocity OLS on short source rollouts; both omit the quadratic column. Φ is the residual-free $\mathrm{SO}(3)$+Laplace flow. NAILONG freezes identified L inside Φ and fits a residual head. The quadratic grid dictionary is a compensating class in the sense of Theorem 4: $\Phi - \widetilde{\Phi}$ can lie in D. A structured residual head (Jacobian summands, axis drag, linear friction) drops the generic pairwise products; on this residual it still compensates when J is small, because the linear breaking terms themselves sit in D. Wigner is the same jointly equivariant SH operator as in Appendix E.10. Default amplitudes match the CPU probe of `results/barotropic_transfer.json`. The GPU panel (`results/barotropic_gpu/`, Haar 64) at N_{tgt}=128, T=24, N_{src}=1536 is

	Haar cap	Wigner	scratch	affine	wrong	NAILONG
J=0.45 (8 seeds)	0.059 ± 0.007	-0.07	-0.05	-0.51	0.34 ± 0.05	0.46 ± 0.05
J=0.34 (8 seeds)	0.071 ± 0.008	-0.05	0.00	-0.26	0.46 ± 0.04	0.50 ± 0.04
Default (8 seeds; J=0.08)	0.073	-0.03	0.00	0.48	0.84	0.85

Oracle freeze matches NAILONG on the reported rows (0.46 / 0.50 / 0.85). The same 8 seeds with a structured head:

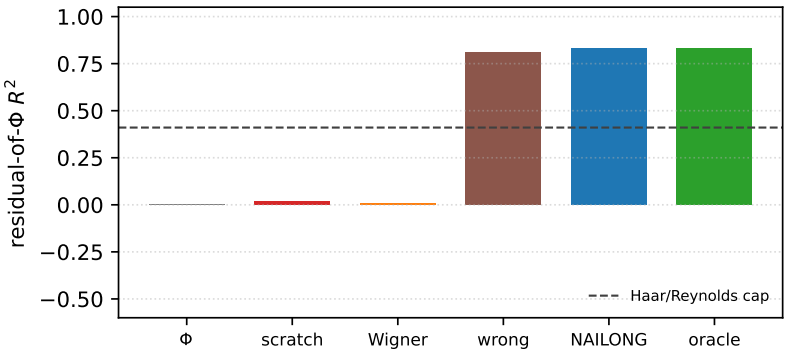

Figure 15: Barotropic residual-of-Φ R^2 at $L_{\max}$=8 (CPU probe). The dashed line is a Haar/Reynolds estimate; the GPU panel uses 64 rotations and reports cap 0.073.

	Haar	wrong-quad	NAILONG-quad	wrong-struct	NAILONG-struct
J=0.45	0.059	0.34±0.05	0.46±0.05	0.36±0.05	0.45±0.04
Default (J=0.08)	0.073	0.84	0.85	0.80	0.81

Default tests the Reynolds cap: both dictionaries compensate a wrong Φ, and jointly equivariant Wigner / scratch sit at that cap (Theorem 2). Wigner-FT on seed 0 stays at the Wigner score (default 0.006, J=0.45 −0.03). A source-trained Wigner trunk plus a raw residual (N_{src}=1536) is 0.821 on default seed 0 (NAILONG 0.829, wrong 0.815) and 0.43±0.06 at J=0.45 over 8 seeds (NAILONG 0.46±0.05, wrong 0.34±0.05; NAILONG exceeds the trunk on that row). The trunk is $\Phi + P_G R_{\text{src}}$ (Haar 0.059 at J=0.45, Theorem 2); identify still selects the algebra on the J-heavy row (`results/barotropic_gpu/trunk_j12_summary.json`). At J energy 0.34 quadratic almost returns compensation (0.50 vs. 0.46); at J=0.45 identify separates for both heads. Raising the $J(\psi, \zeta)$ share (weaker Coriolis / ridge / axis drag) makes affine floor. Trajectory identify at N_{src}=1536 is within 0.002 of the snapshot OLS. The probe is a truncated 2D Euler / GFD discretization at $L_{\max}$=8.

E.15 Spherical shallow water with a G-FNO trunk

Script `run_swe_transfer.py` evolves linearised spherical shallow water in the same real-SH chart as Appendix E.14. The state is (ζ, δ, η) (vorticity, divergence, height). The shared propagator Φ is residual-free gravity waves

$$\partial_t \zeta = \nu \Delta \zeta + \alpha \cdot (L\zeta), \qquad \partial_t \delta = \nu \Delta \delta - g \Delta \eta + \alpha \cdot (L\delta), \qquad \partial_t \eta = -H\delta + \alpha \cdot (L\eta).$$

Axial Coriolis $f = 2\Omega\mu$ and an axisymmetric topography source in the continuity equation break SO(3). Identification is misspecified linear vector-field OLS on source vorticity. NAILONG freezes identified L inside Φ and fits an affine residual on the concatenated triple. The source-trained linear spherical G-FNO (Helwig-style per-ℓ gains, shared across m) estimates $P_G T_{\text{src}} = \Phi + P_G R_{\text{src}}$ (Theorem 2). A source-trained trunk plus a raw target residual is the same map with an unconstrained adapter. At $L_{\max}$=3, N_{tgt}=24 (`results/swe/cpu_L3.json`): Haar 0.00, so the linear trunk is a Φ slot and the −2.14 is an adapter failure on that truncation; NAILONG 0.92; wrong freeze 0.06. The $L_{\max}$=8 GPU panel (8 seeds, N_{tgt}=128; `results/swe/summary.json`) is Haar 0.078±0.009, source Wigner / G-FNO trunk+res −1.41±0.24, wrong freeze −1.19±0.26, scratch −4.46±0.29, NAILONG 0.83±0.02 (oracle 0.83). NAILONG exceeds the Reynolds cap and the source trunk in that panel. Figure 5 collects the field-level scores.

E.16 When an unconstrained residual erases a wrong freeze

Theorem 3 is a floor for *model-class* residuals in the direct-sum slot. Field-to-field heads are often richer: a dictionary D of maps $u \mapsto \widehat{R}(u)$ is fit after a frozen propagator $\widetilde{\Phi}$. Theorem 4 is L^2 residual OLS. Write P_D for the L^2 projector onto D, and $E_{\text{cor}} = \Phi + P_D R$, $E_{\text{wrong}} = \widetilde{\Phi} + P_D(T - \widetilde{\Phi})$ for

Table 10: Field-to-field S^2 residual-of-Φ R^2 vs. the Haar cap (Theorem 2).

Probe	Haar	Wigner	scratch	affine	wrong	NAILONG
Lab-axis L=3	0.034	−0.41	−2.12	—	—	0.87
J-heavy (8 seeds)	0.059	−0.07	−0.05	−0.51	0.34	0.46
Default (8 seeds)	0.073	−0.03	0.00	0.48	0.84	0.85
SWE L=8 (8 seeds)	0.078	—	−4.46	—	−1.19	0.83

the two population projection maps. The identity also holds on the empirical measure, with equality interpreted at the sampled inputs.

Proof of Theorem 4. Write $\Delta := \Phi - \widetilde{\Phi}$. Then $T - \widetilde{\Phi} = R + \Delta$ and

$$E_{\text{wrong}} \;=\; \widetilde{\Phi} + P_D(R+\Delta) \;=\; \Phi - \Delta + P_D R + P_D\Delta \;=\; \Phi + P_D R + (P_D - I)\Delta.$$

Subtracting from $T = \Phi + R$ gives $T - E_{\text{wrong}} = (I - P_D)(R + \Delta)$. The correct-freeze residual is $(I - P_D)R$. These coincide iff $(I - P_D)\Delta = 0$, i.e. $\Delta \in D$. □

Barotropic S^2 rows. The default barotropic residual has Jacobian energy 0.08. Both the quadratic grid dictionary and the thinner structured dictionary (Jacobian summands, axis drag, linear friction) leave $\Phi - \widetilde{\Phi}$ nearly in D: wrong≈NAILONG for either head (0.85 vs. 0.84 quadratic; 0.81 vs. 0.80 structured; 8 seeds). Raising the Jacobian share to 0.45 pushes part of Δ out of both classes and identify separates (0.46 vs. 0.34 quadratic; 0.45 vs. 0.36 structured). The selector is the residual's J share, not the extra pairwise products. A residual class hides a wrong Φ exactly when $\Phi - \widetilde{\Phi}$ already lies in D; Appendix B.7 records the generator-OLS Schur form of the same identity.

F Extended Experimental Details

F.1 Synthetic setup and OLS implementation

Generators are orthonormal bases of $\mathfrak{so}(3)\oplus\mathfrak{w} \subset \mathfrak{gl}(n)$ under the Frobenius inner product. Responses follow (2)/(12). Implemented OLS uses the efficient control-linear form $\mathrm{vec}(Y) \approx ZW$ with $W \in \mathbb{R}^{m\times n^2}$, which is algebraically equivalent to Kronecker OLS but avoids materializing $n^2 \times mn^2$ designs (critical for $n = 32$ Galerkin runs).

F.2 Exact constant recovery

With matched design, Monte Carlo OLS recovers $\kappa_{\text{scr}} \approx 5.975$ and $\kappa_{\text{frz}} \approx 2.981$ against targets 6 and 3 (relative error $< 1\%$). Script: `scripts/run_theory_constant_check.py`.

F.3 Dimension sweep

Table 11: Dimension sweep: empirical $\hat{\kappa}$ vs. theory (tol 0.35, N=64, σ=0.08).

s	r	n	$\hat{\kappa}_{\text{scr}}$	s+r	$\hat{\kappa}_{\text{frz}}$	r
1	1	2	1.97	2	1.06	1
2	2	4	4.11	4	2.14	2
3	3	5	6.18	6	2.98	3
3	1	4	4.06	4	1.08	1
1	3	4	4.25	4	3.18	3

F.4 Noise sweep

Across $\sigma \in \{0.02, 0.05, 0.08, 0.12\}$ at fixed N=48: $\kappa_{\text{correct}} \approx 2.94$ (CV 0.017), $\kappa_{\text{scratch}} \approx 6.00$ (CV 0.019).

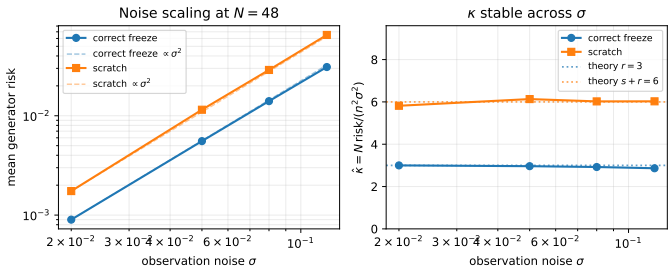

Figure 16: Risk scales as σ^2 (log-log slopes $\approx 1.98/2.02$) while κ stays near r and $s+r$.

F.5 Bias-floor sweep

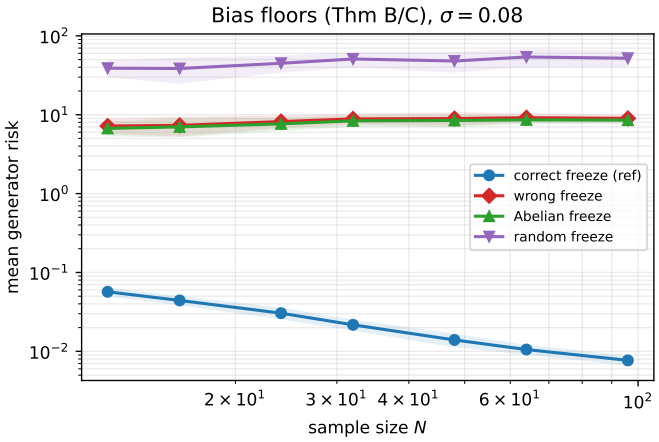

Figure 17: Bias floors for wrong/Abelian/random freezes versus decaying correct freeze.

Wrong, Abelian, and random same-dimension freezes stay at $O(1)$ risk while correct freeze decays with the residual-rate slope.

F.6 Barcode type-check

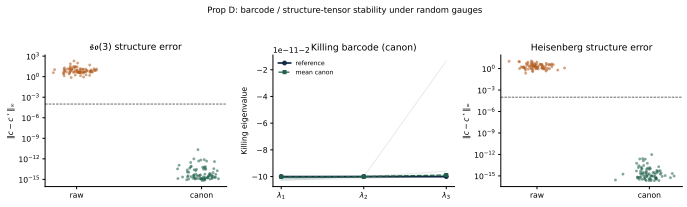

Figure 18: Structure-constant error with vs. without Cartan–Killing canonicalization under random gauges.

Over 80 random conjugacies for $\mathfrak{so}(3)$: raw max error 1.94×10^{2}, canonical max error 2.20×10^{-11}; Killing barcode distance after canon $< 10^{-10}$. Heisenberg canonicalization likewise succeeds (max error $< 10^{-12}$).

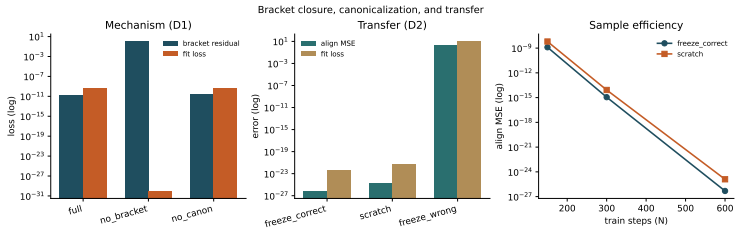

Figure 19: Bracket closure is required for algebraic meaning; correct freeze dominates scratch/wrong on transfer.

Table 12: Multi-seed ablation metrics (steps 600).

Case	fit	bracket res.	align	rollout
Full	3.6×10^{-10}	1.5×10^{-11}	1.5×10^{-11}	1.6×10^{-13}
No bracket	7.6×10^{-31}	1.01	≈ 0	≈ 0
No canon	4.4×10^{-10}	2.1×10^{-11}	1.9×10^{-11}	2.7×10^{-13}
Transfer freeze correct	≈ 0	≈ 0	≈ 0	≈ 0
Transfer scratch	≈ 0	≈ 0	≈ 0	≈ 0
Transfer freeze wrong	9.53	5.7×10^{-3}	1.59	1.8×10^{-2}

F.7 Multi-seed ablations

Fit-alone without bracket closure yields near-zero regression loss but bracket residual ≈ 1, so the latent matrices do not define a Lie algebra and are not transferable as shared structure. On this latent-generator panel, multi-GPU seeds on $8\times$A800 reproduce freeze alignment $\ll$ wrong alignment (6×10^{-22} vs. 1.59).

F.8 Galerkin PDE construction

We use the controlled 1D periodic reaction–advection–diffusion skeleton

$$\partial_t u = \nu\,\partial_{\theta\theta} u + \rho\, u + \text{controlled mode-mixing actuation.}$$

After real Fourier truncation to n modes, the coefficient ODE is bilinear:

$$\dot{x} = \Big(A_0 + \sum_i \alpha_i S_i + \sum_j \beta_j R_j\Big) x,$$

with shared $S_i \cong \mathfrak{so}(3)$ on a designated triad and residual R_j differing by domain (Heisenberg-like shear vs. toral reaction). Figure 20 reports the dichotomy on this truncation ($\kappa_{\text{correct}} \approx 2.97$, $\kappa_{\text{scratch}} \approx 5.92$).

F.9 Approximate-symmetry floors

Freeze $\widehat{\mathfrak{h}} = (1-\delta)\mathfrak{h}^\star + \delta M$ for random Frobenius-normalized skew M. Figure 6 is in the main text; Table 13 reports the floors for completeness.

F.10 DOF isolation / unstructured oracle

We freeze (i) the true shared basis, (ii) an $O(s)$ remix of the same span with consistently transformed controls, (iii) a random skew s-plane in $\mathfrak{gl}(n)$, and (iv) a random s-plane in $\mathfrak{g}$ dominated by residual directions. Mean $\hat{\kappa}$: correct 2.96, same-span remix 2.96, scratch 5.92, random $\mathfrak{gl}(n)$ plane 2.20×10^3, ambient wrong plane 2.44×10^3 (`results/unstructured_oracle.json`). Same-span remix matches residual rate; a random s-plane of the same dimension does not. Under Assumption 2, Proposition 1 identifies that span with the simple $\mathfrak{so}(3)$ factor; Lemma 1 requires a Lie algebra when the shared object is a continuous symmetry.

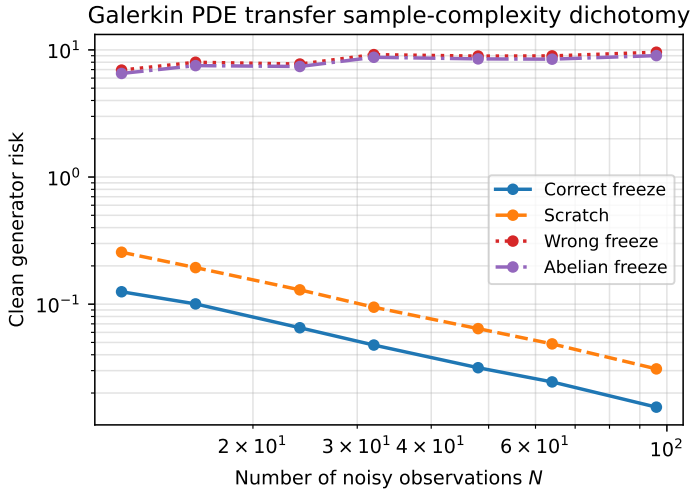

Figure 20: Galerkin PDE attainment: the same dichotomy on projected coefficient dynamics.

Table 13: Approximate-symmetry floors and log-log slopes.

δ	large-N freeze floor	log-log slope
0	0.0054	-1.01
0.02	0.0099	-0.72
0.05	0.034	-0.29
0.1	0.125	-0.07
0.2	0.490	$+0.07$
0.4	1.91	$+0.09$

F.11 Misspecified-freeze estimators

Estimators under identical matched designs:

- **Correct freeze:** freeze source canonical $\mathfrak{so}(3)$; learn residual.
- **Scratch:** learn all $s{+}r$ generators on the target.
- **Wrong freeze:** freeze Heisenberg residual of the source as if it were shared.
- **Abelian freeze:** freeze the embedded commuting diagonal triple of Theorem 5 (dimension three), not a one-dimensional Cartan torus.
- **Spectral freeze:** average observed generators, take dominant skew-SVD factors as a bracket-blind proxy. Consumes the same $\{A(u)\}$ as Alg 1; not restricted to Definition 1.
- **PCA / DMD:** top-s principal matrices, or DMD generator skew-SVD on the same trajectories as trajectory identify.
- **Finetune shared:** scratch-fit on the source ($N_{\mathrm{src}}{=}256$), freeze the estimated shared block, learn residual on the target.
- **Full-trunk freeze:** freeze all source-fitted generators with no residual adaptation.

Large-N risks appear in Table 1; spectral proxies on the same observations appear in Appendix E.7.

F.12 Compute notes

Algebraic OLS probes, the released S^2 field-to-field script, and Galerkin κ sweeps run on CPU. The barotropic multi-seed and Jacobian-share panel (`run_barotropic_gpu.py`) used 8×NVIDIA A100-PCIE-40GB. The latent-generator multi-seed panel (`run_long_ablation_panel.py` / `run_stage_c_gpu_long_scale.py`) used a remote host with 8×NVIDIA A800-SXM4-80GB.

G REPRODUCIBILITY

Theorems 1, 2, 3, 4 and 5 and Propositions 1, 2 have complete proofs in Appendix B (Theorem 2 in §B.5). Experiment scripts and raw JSON/TXT logs ship with the code release.

```
python scripts/run_theory_constant_check.py
python scripts/run_sample_complexity_law.py
python scripts/run_dimension_sweep.py
python scripts/run_noise_floor_sweep.py
python scripts/run_barcode_stability.py
python scripts/run_long_ablation_panel.py
python scripts/run_pde_dichotomy.py
python scripts/run_stage_c2_approx_symmetry.py
python scripts/run_stage_c3_method_negatives.py
python scripts/run_stage_c1_large_galerkin.py
python scripts/run_stage_c_gpu_long_scale.py
python scripts/run_unstructured_oracle.py
python scripts/run_population_design.py
python scripts/run_overlap_interpolation.py
python scripts/run_same_data_spectral.py
python scripts/run_identify_ablation.py
python scripts/run_sphere_so3_transfer.py
python scripts/run_sphere_operator_transfer.py
python -m pytest -q
```