AVERAGE ACCURACY DOES NOT CERTIFY TOPOLOGY IN LEARNED HAMILTONIANS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learned electronic Hamiltonians are commonly evaluated by errors in matrix elements, band energies, or occupied subspaces, but when do these metrics certify topology? We show that no strictly positive finite-p integrated-error threshold universally certifies topology over unrestricted smooth gapped families: a topology-changing discrepancy can be confined to arbitrarily small support while retaining finite pointwise amplitude. Conversely, topologically distinct, well-defined rank-r occupied bundles must attain subspace infidelity at least $1/r$ somewhere, motivating uniform, gap-aware control. Learned four-band models realize this mismatch: lower band MAE ranks a family with $0/15$ correct topologies ahead of one with $15/15$ correct. In realistic settings, a structured perturbation changes the Chern number of a 32-orbital $NaRuO_2$ Wannier Hamiltonian at 8.074 meV near-Fermi band MAE, while a released neural Hamiltonian undergoes gap collapse that invalidates a raw topological comparison. Equal-budget loop-structured sampling shows no stable advantage. In an exploratory 10-meV band-matched study, all 276 evaluations valid at both endpoints preserve topology, and no target-valid plane becomes newly unsafe. Together, these results expose a certification gap between predictive accuracy and topological fidelity without implying that such failures are prevalent under generic small spectral perturbations.

1 INTRODUCTION

Learned electronic Hamiltonians increasingly support downstream calculations of band structures, wavefunctions, and topology (Li et al., 2022; Gong et al., 2023; Zhong et al., 2026). They are commonly evaluated by errors in Hamiltonian matrices, band energies, or occupied subspaces, although such metrics need not align with downstream fidelity (Li et al., 2025; Yin et al., 2025a). Topology raises a sharper question: *when does predictive accuracy certify preservation of the reference topological class?*

We show that average accuracy does not provide such a certificate. A topologically decisive discrepancy can be confined to an arbitrarily small region of the Brillouin zone even at fixed spectral gap. In the exact construction of Fig. 1, mean projector infidelity vanishes as the texture localizes while the Chern discrepancy remains fixed. Learned four-band models realize the same mismatch: a family with lower median band MAE is topologically incorrect in $15/15$ cases, whereas a diagnostic family with higher MAE is correct in $15/15$. Thus spectral accuracy can both miss topological error and reverse model ranking for a topological objective.

The underlying distinction is between integrated and uniform fidelity. Over unrestricted smooth gapped families, no strictly positive finite-p integrated-error threshold universally certifies topology. Conversely, topologically distinct, well-defined rank-r occupied bundles must satisfy

$$\sup_k \left[1 - \frac{1}{r}\operatorname{Tr}(P_*(k)P_\theta(k))\right] \geq \frac{1}{r}.$$

Topology-changing error can therefore be arbitrarily localized, but not uniformly small. This yields conservative sufficient guarantees: uniform projector distance below one implies bundle isomorphism, while uniform Hamiltonian error below half the reference gap preserves a gapped class-A interpolation. Finite observations require additional coverage and an independently justified regularity bound.

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.

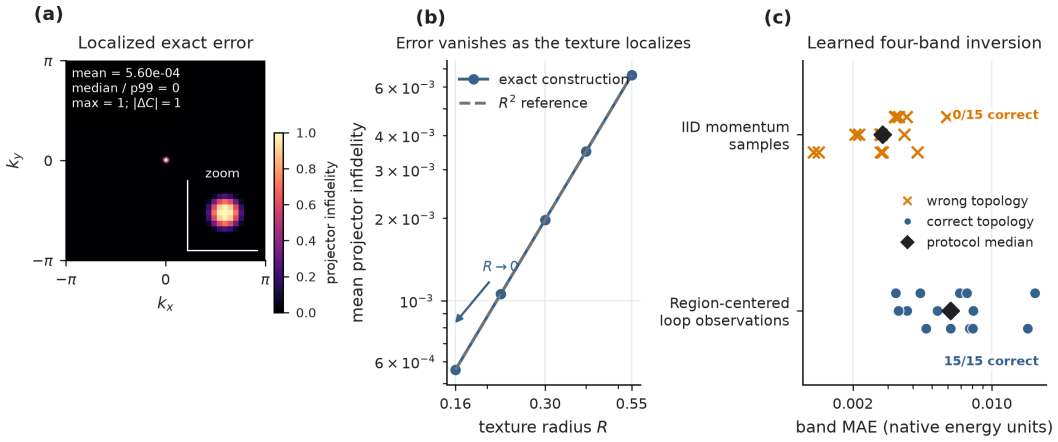


Figure 1: **Average accuracy does not certify topological fidelity. (a)** A localized two-band mismatch has mean infidelity 5.60×10^{-4} while its maximum remains 1 and the Chern discrepancy remains fixed. **(b)** Mean infidelity scales as R^2 as the texture localizes, without changing the topology. **(c)** Learned four-band models with lower median band MAE (0.002813) are correct in $0/15$ cases, whereas region-centered diagnostic models have higher MAE (0.006211) and correct $15/15$.

The distinction also appears in realistic Hamiltonians. A structured perturbation changes the occupied Chern number of a 32-orbital spin–orbit-coupled $NaRuO_2$ Wannier Hamiltonian at 8.074 meV near-Fermi band MAE. In a released Uni-HamGNN prediction for $Ca_3Si_4Ir_4$, the direct gap collapses from 1.546 meV to 6.17×10^{-4} meV, making the raw Z_2 disagreement a validity failure rather than a valid topological mismatch.

These failures are possible but need not be prevalent. Across 216 model–setting instances, equal-budget loop-structured finite-point queries show no stable advantage over IID, stratified-IID, or random-path sampling. Moreover, in a 10-meV band-matched study over 432 material–mechanism–plane evaluations, all 276 evaluations valid at both endpoints preserve topology, with no target-valid plane becoming newly unsafe; the remaining 156 are already invalid at the reference endpoint. The empirical conclusion is therefore a *certification gap*, not widespread topological failure at small spectral error. Our contributions are:

- **Average accuracy is not a topology certificate.** We establish the finite-p obstruction and demonstrate its consequence in learned multiband Hamiltonians.
- **Topological mismatch leaves a uniform witness.** Different rank-r bundles require pointwise subspace infidelity at least $1/r$, connecting the phenomenon to uniform projector and gap-preserving Hamiltonian guarantees.
- **Finite-point evidence requires more than sampling geometry.** Sparse accuracy becomes certifying only with coverage and justified regularity; the tested loop-structured schemes show no stable equal-budget advantage.
- **Realistic evidence separates sensitivity, validity, and prevalence.** We find structured multiorbital sensitivity and a released-model validity failure, while the 10-meV material study yields no valid failures in the tested cohort.

2 From Prediction Accuracy to Topological Fidelity

2.1 Hamiltonians, Occupied Subspaces, and Accuracy

Let $X = \mathbb{T}^d$ be the compact Brillouin torus, equipped with normalized Haar measure $\mu(X) = 1$, and $H_*, H_\theta : X \to \mathrm{Herm}(m)$ the reference and learned Bloch Hamiltonians. Matrix differences are always taken in a common continuous, periodic orthonormal orbital representation and a common normalized momentum coordinate identification. For a generalized eigenproblem $Hu = \varepsilon Su$,

we assume that reference and prediction share the same continuous, periodic, uniformly positive-definite overlap matrix $S(k)$, and use the common Löwdin representation $\widetilde{H} = S^{-1/2} H S^{-1/2}$.

We consider an occupied subspace of fixed rank r, with spectral projectors $P_j(k) = P_r[H_j(k)]$, $j \in \{*, \theta\}$. Whenever topology is compared, both endpoints are assumed to have a positive uniform occupied–unoccupied gap,

$$\Delta_j = \inf_{k \in X} \left[\lambda_{r+1}(H_j(k)) - \lambda_r(H_j(k))\right] > 0, \qquad j \in \{*, \theta\}, \tag{1}$$

so that P_j defines a continuous rank-r bundle. We use the occupied-projector convention, e.g. $C = \mathrm{Ch}_1(P_{\mathrm{occ}})$.

The accuracy measures of interest aggregate pointwise error over momentum space. We write

$$\mathcal{L}_p(H_\theta, H_*) = \left(\int_X \|H_\theta(k) - H_*(k)\|^p \, \mathrm{d}k\right)^{1/p}, \quad f(k) = 1 - \frac{1}{r}\,\mathrm{Tr}[P_*(k) P_\theta(k)],$$
$$\bar{f} = \int_X f(k)\, \mathrm{d}k.$$

for $1 \le p < \infty$, where integration is with respect to the normalized Haar measure and the Hamiltonian norm is the operator norm unless stated otherwise. Here $f(k) = r^{-1} \sum_i \sin^2 \theta_i(k)$ in terms of principal angles between the occupied subspaces. We also report band MAE over the stated band window in each experiment. These are measures of predictive accuracy; the question studied here is whether small values are sufficient to imply equality of occupied topological classes, $[P_\theta] = [P_*]$.

2.2 Topological fidelity and validity

A prediction is *topologically faithful* when the reference and predicted occupied bundles are both well defined and belong to the same relevant topological class. In class A, the certificate establishes bundle isomorphism; equality of a particular invariant such as the Chern number is only one consequence. Class-AII comparisons additionally require the same protecting time-reversal structure.

We distinguish three outcomes: *valid agreement*, when both endpoints are well defined and topologically agree; *valid mismatch*, when both are well defined but topologically differ; and *invalid prediction*, when the predicted Hamiltonian does not support a reliable invariant, e.g., because the occupied–unoccupied gap closes or the numerical invariant fails to converge. A raw difference between discretized Chern or Z_2 labels is therefore not by itself evidence of a topological mismatch.

Our evaluation order is *validity* $\rightarrow$ *fidelity* $\rightarrow$ *certification*. Projector-space certificates compare already valid occupied bundles, whereas Hamiltonian-space certificates can use uniform error relative to a spectral gap to preserve validity along the comparison path.

3 Accurate Hamiltonians Can Have the Wrong Topology

A localized topological blind spot. Consider the flattened two-band family $H_R(k) = n_R(k) \cdot \sigma$, $\|n_R(k)\|_2 = 1$, whose nontrivial texture is confined to a disk of radius R, and a constant trivial comparator H_0. Both have spectrum $\{-1, +1\}$ and gap 2, but their occupied Chern numbers differ by one. Because the projector discrepancy occupies only $O(R^2)$ area, its mean infidelity decreases quadratically while its maximum remains order one. At $R = 0.16$, Fig. 1(a,b) gives mean infidelity 5.60×10^{-4}, fitted scaling exponent 2.00, and maximum 1; even the median and, for the smallest textures, the 99th percentile are zero. Thus average accuracy can improve simply by localizing a topology-changing error. Section 4 formalizes this obstruction for every finite p.

The mismatch persists in multiband occupied subspaces. The construction extends to a four-band system by adding a trivial spectator pair and smooth periodic orbital mixing. The target and comparator have identical pointwise eigenvalues $(-1.7, -1, +1, +1.7)$, yet differ by one unit of occupied Chern number. At the smallest tested radius, their mean rank-two subspace infidelity is 2.80×10^{-4}, while the maximum remains approximately 0.49 and the band error is numerically negligible. Pointwise spectral agreement therefore does not determine the topology of a multiband occupied bundle.

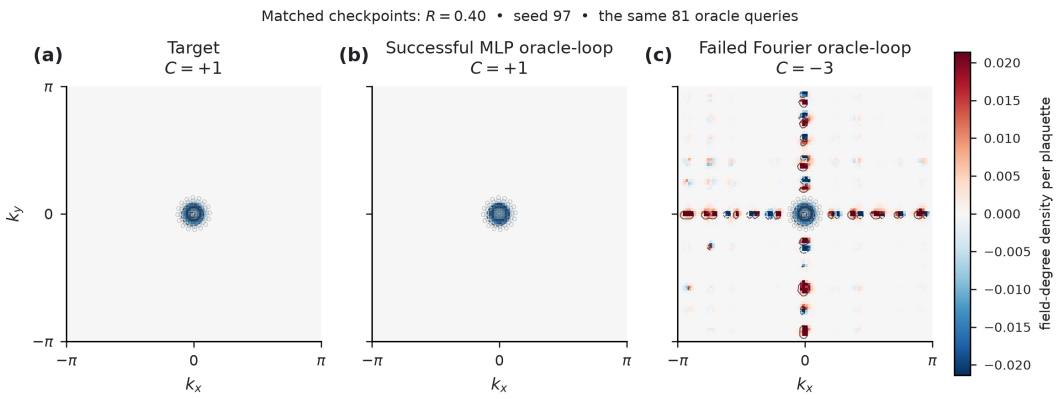


Figure 2: **The same finite observations permit different global extensions.** For the same $R = 0.40$ target, seed, and 81 region-centered queries, the periodic MLP recovers the target topology whereas the Fourier model develops additional winding in unsampled regions. All maps share the same normalization; the region-centered loops are an oracle diagnostic, not a general acquisition strategy.

Conventional metrics can rank learned models incorrectly. We next train matched four-band models with the same architecture and label budget, changing only the placement of 81 additional observations: either IID momentum samples or region-centered loops. The latter use the known texture location and serve only as a diagnostic comparison. As shown in Fig. 1(c), IID models have lower median band MAE, 0.002813, yet recover the correct occupied Chern number in $0/15$ cases. Region-centered models have higher median band MAE, 0.006211, but are correct in $15/15$ cases. Both families remain gapped. Band MAE therefore reverses the family-level ordering with respect to topological correctness in this controlled setting.

Finite observations leave the global extension underdetermined. Finite observations constrain occupied eigenspaces only at queried momenta; the extension between them remains model dependent. In matched two-band examples with the same 81 region-centered observations, a periodic MLP recovers the target topology while a Fourier-feature model can generate additional winding in unsampled regions. Across localized targets, IID sampling yields correct topology in $1/15$ MLP and $3/15$ Fourier fits, while region-centered observations yield $15/15$ and $8/15$, respectively. Some Fourier failures add extra integer winding rather than simply becoming trivial. Thus observing the relevant region does not uniquely determine the global bundle, nor does loop sampling provide a universal remedy. Full breakdowns appear in Appendices E and D.

4 From Average Error to Uniform Certificates

Topology-changing errors can be arbitrarily localized, yet topologically distinct, well-defined occupied bundles cannot remain uniformly close. Let $P_0, P_1 : X \to \mathrm{Proj}_r(\mathbb{C}^m)$ be continuous equal-rank projectors on the compact Brillouin zone X; when $P_j = P_r[H_j]$, each H_j is assumed to have a positive uniform $r/(r+1)$ gap. Thus "topological mismatch" refers to two valid occupied bundles of different topology, not gap closure or numerical invalidity. The conditions below are sufficient accuracy-based certificates, not an exhaustive characterization of all possible certification strategies.

4.1 Finite-p accuracy does not universally certify topology

The localized construction in Sec. 3 extends to every finite p.

Proposition 1 (No positive finite-p threshold)**.** *For every $1 \le p < \infty$ and $\varepsilon > 0$, there exist smooth periodic two-band Hamiltonians $H_0, H_1 : \mathbb{T}^2 \to \mathrm{Herm}(2)$, both with pointwise spectrum $\{-1, +1\}$, such that*

$$\|H_1 - H_0\|_{L^p} < \varepsilon, \qquad \|P_1 - P_0\|_{L^p} < \varepsilon,$$

while $\mathrm{Ch}_1(P_1) \neq \mathrm{Ch}_1(P_0)$.

A smooth degree-one texture confined to a disk of radius R gives the construction: the Hamiltonian and projector discrepancies are supported on area $O(R^2)$, so their L^p norms scale as $O(R^{2/p})$, while the gap and Chern discrepancy remain fixed. Hence no strictly positive threshold depending only on a finite-p norm can certify topology over unrestricted smooth gapped families. Exact zero error still implies equality, and additional uniform regularity can change what an integrated norm implies. The full construction is given in Appendix B.

4.2 Topological mismatch requires a pointwise witness

For the subspace infidelity $f(k) = 1 - \frac{1}{r}\operatorname{Tr}[P_0(k)P_1(k)]$, principal angles give $f(k) \le \|P_1(k) - P_0(k)\|_{\mathrm{op}}^2 \le r f(k)$.

Proposition 2 (Uniform projector certificate)**.** *If*

$$\sup_{k\in X} \|P_1(k) - P_0(k)\|_{\mathrm{op}} < 1,$$

then the image bundles of P_0 and P_1 are isomorphic. Consequently, if two well-defined bundles are not isomorphic, then

$$\sup_k \|P_1(k) - P_0(k)\|_{\mathrm{op}} = 1, \qquad \sup_k f(k) \ge \frac{1}{r}.$$

The first statement follows from the Kato–Nagy direct rotation (Kato, 1995); the second is its contrapositive together with Eq. equation 4.2. Classical subspace-rotation bounds provide closely related perturbative control (Davis & Kahan, 1970). The key interpretation is: *topological mismatch can be arbitrarily localized, but it cannot be uniformly small in projector space*. For rank one the required witness is $\sup f = 1$, and for rank two $\sup f \ge 1/2$. Numerical refinement of the learned rank-two failures is consistent with this bound (Appendix C).

4.3 Uniform Hamiltonian control gives a gap certificate

Projector control assumes both bundles are already globally valid. Hamiltonian-space control can guarantee that validity as well. Let

$$\Delta_* = \inf_k \left[\lambda_{r+1}(H_*(k)) - \lambda_r(H_*(k))\right] > 0, \qquad \delta = \sup_k \|H_\theta(k) - H_*(k)\|_{\mathrm{op}}.$$

Proposition 3 (Uniform Hamiltonian certificate)**.** *If $\delta < \frac{\Delta_*}{2}$, then the interpolation $H_s = (1-s)H_* + sH_\theta$, $s \in [0,1]$, remains gapped between bands r and $r+1$. The corresponding occupied bundles therefore have the same class-A topology.*

Indeed, Weyl's inequality (Bhatia, 1997) gives a gap along the interpolation of at least $\Delta_* - 2s\delta > 0$. The condition is sufficient rather than necessary and can be conservative, for example for perturbations dominated by an identity shift. Fixed-Fermi-level and time-reversal-equivariant extensions are given in Appendix B; for Z_2 certification, the same protecting time-reversal structure must be preserved along the comparison.

4.4 Sparse observations require coverage and regularity

Uniform certificates concern the full Brillouin zone. For a finite nonempty query set Q, define its fill distance $h(Q) = \sup_{k\in X} \min_{q\in Q} d(k,q)$. If the Hamiltonian error $E = H_\theta - H_*$ and the subspace-infidelity field f have certified global Lipschitz constants L_E and L_f, respectively, then

$$\|E\|_\infty \le \max_{q\in Q} \|E(q)\|_{\mathrm{op}} + L_E h(Q), \qquad \sup_k f(k) \le \max_{q\in Q} f(q) + L_f h(Q).$$

Thus

$$\max_{q\in Q} \|E(q)\|_{\mathrm{op}} + L_E h(Q) < \frac{\Delta_*}{2}$$

is a sufficient Hamiltonian certificate, while $\max_{q\in Q} f(q) + L_f h(Q) < \frac{1}{r}$ certifies bundle equivalence *conditional on global validity of the projector families*. The Hamiltonian route can protect the gap itself; the projector route cannot detect an unsampled loss of validity.

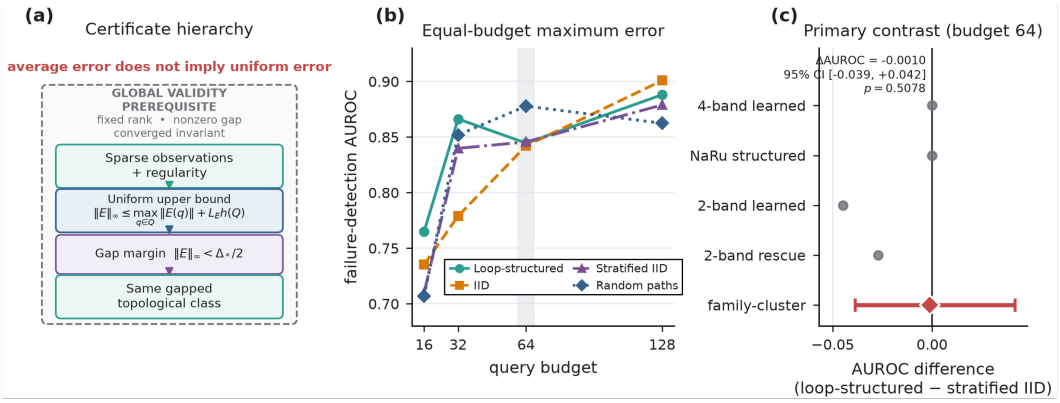


Figure 3: **Sparse point observations require more than loop geometry. (a)** Finite-sample certification requires observed error, coverage, and justified regularity. **(b)** The best equal-budget protocol varies with $B \in \{16, 32, 64, 128\}$. **(c)** At $B = 64$, loop-structured and stratified-IID sampling are nearly identical in AUROC (0.8445 vs. 0.8456; $\Delta = -0.0010$, CI $[-0.039, 0.042]$, $p = 0.5078$).

Crucially, L_E and L_f must be independently justified global upper bounds; this is the standard geometry underlying scattered-data coverage arguments (Wendland, 2005). We do not estimate certified values for the learned models studied here, so these relations are analytical sufficient conditions rather than empirical risk scores.

4.5 Finite point sets are not universally certifying

The regularity term above is essential.

Proposition 4 (Finite point-set non-certification)**.** *For any fixed finite $Q \subset \mathbb{T}^2$, there exist smooth gap-2 two-band Hamiltonians H_0, H_1 that agree on a neighborhood of every $q \in Q$ while their occupied bundles have different Chern classes.*

The construction places the localized degree-one texture from Proposition 1 inside an open disk disjoint from Q. Hence finite point agreement alone cannot provide a universal certificate without additional regularity. This statement is specific to finite point evaluations; it does not cover continuous-loop or holonomy observables, adaptive procedures, or restricted function classes with independent regularity guarantees. Section 5 next asks whether particular loop-structured finite-point designs nevertheless offer an empirical advantage at matched query cost.

5 No Stable Loop Advantage in the Tested Finite-Point Benchmark

The finite-point result in Sec. 4.5 shows that sampled agreement is not universally certifying without additional regularity. We therefore test a narrower question: at equal query cost, does loop structure improve failure detection? We compare loop-structured, IID, stratified-IID, and random-path sampling at $B \in \{16, 32, 64, 128\}$ on the same 216 model–setting instances from seven families, including 75 failures. Each protocol uses 16 paired query designs; the score for each instance is the mean of the 16 design-wise maximum occupied-subspace infidelities. We use $B = 64$ and loop-versus-stratified-IID as the primary comparison, with families as the independent units for uncertainty. The numerically best protocol changes with budget: loop-structured sampling leads at $B = 16, 32$, random paths at 64, and IID sampling at 128. At the primary budget, loop-structured and stratified-IID AUROC are 0.8445 and 0.8456, with paired difference -0.0010, family-cluster interval $[-0.039, 0.042]$, and exact family-level permutation $p = 0.5078$. Thus the tested loop-structured schemes show no stable equal-budget advantage.

This result complements, but does not instantiate, the sparse certificate in Sec. 4.4: finite observations become certifying only when combined with coverage and independently justified regularity.

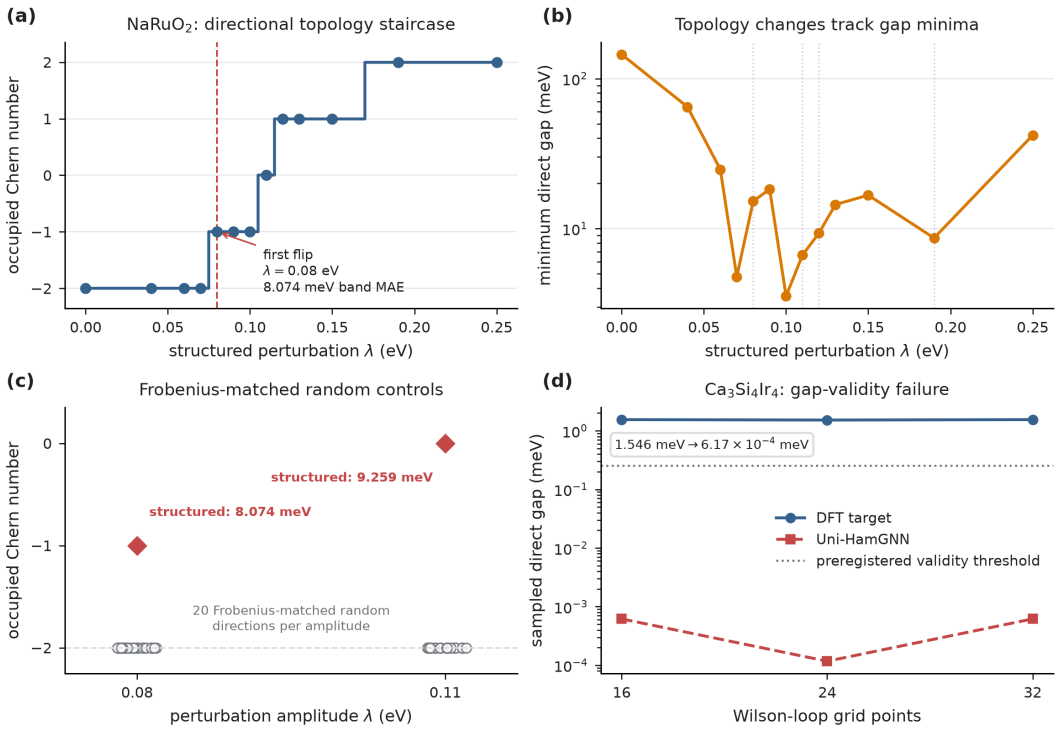


Figure 4: **Structured sensitivity and validity failure in realistic Hamiltonians.** **(a)** $NaRuO_2$ first changes $C = -2 \to -1$ at $\lambda = 0.08$ eV. **(b)** Gap minima accompany successive Chern transitions. **(c)** Frobenius-matched random directions retain $C = -2$ at the tested amplitudes. **(d)** $Ca_3Si_4Ir_4$ undergoes a severe predicted gap collapse, invalidating the raw Z_2 comparison.

The benchmark tests discrete point-query geometry, not continuous holonomy measurements, adaptive procedures, or restricted function classes with additional structure.

6 Realistic Hamiltonians: Sensitivity and Validity

We next test whether the accuracy–topology distinction persists in realistic multiorbital Hamiltonians, considering both structured sensitivity and loss of topological validity.

Structured errors can cross a realistic topological boundary. We study a 32-orbital spin–orbit-coupled $NaRuO_2$ Wannier Hamiltonian (Marzari & Vanderbilt, 1997; Mostofi et al., 2008; Marzari et al., 2012) with occupied rank 17. Perturbing a selected valence–conduction edge pair along $H_\lambda = H_* + \lambda(|v\rangle\langle v| - |c\rangle\langle c|)$ produces the first Chern transition at $\lambda = 0.08$ eV, $C = -2 \to -1$, with 8.074 meV near-Fermi band MAE, 0.1828% relative Hamiltonian RMSE, and a 15.225 meV minimum gap (Fig. 4a,b). The effect is strongly direction dependent in the tested comparison: all 40 Frobenius-matched random perturbations retain $C = -2$, whereas the structured direction changes topology (Fig. 4c). Because the random perturbations have smaller operator norm, this comparison demonstrates enhanced sensitivity to the structured direction without isolating orientation alone.

A learned Hamiltonian can lose topological validity. For the released Uni-HamGNN prediction of $Ca_3Si_4Ir_4$ (Zhong et al., 2026), the representative direct gap collapses from 1.546 meV to 6.17×10^{-4} meV (Fig. 4d), far below the 0.25 meV numerical validity guardrail used for the Wilson-loop analysis. Raw reference–prediction Z_2 labels can therefore differ, but we classify this as a validity failure rather than a valid topological mismatch.

Released-model breadth. All four complete released Uni-HamGNN reference–prediction pairs are included. ZrSiPt and $NaCaTaTiO_6$ show stable trivial agreement; Ta_3AlC_2 preserves all six

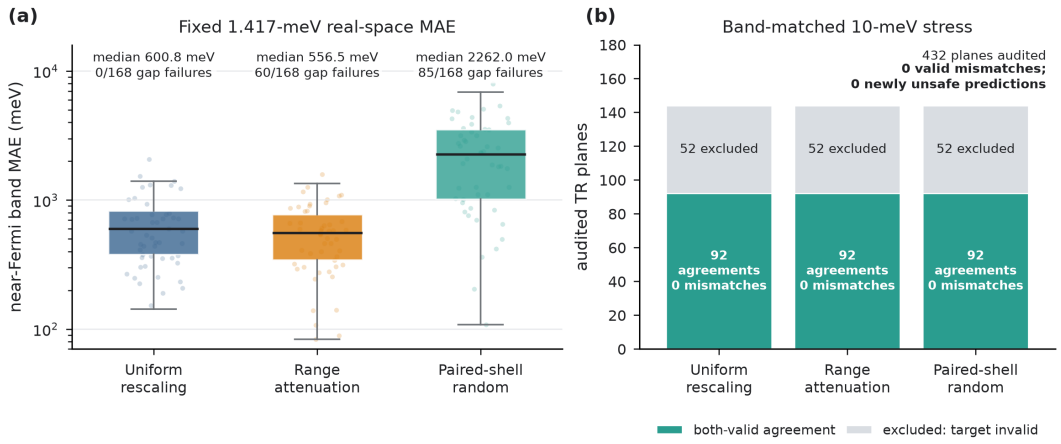


Figure 5: **Possibility does not imply prevalence at small spectral error. (a)** Fixed 1.417 meV matrix MAE can induce hundreds-to-thousands of meV near-Fermi band error, where failures appear. **(b)** At $\sim$ 10 meV band-matched error, all 276 both-valid evaluations preserve topology and none becomes newly unsafe; 156 target-invalid evaluations are excluded.

evaluated plane labels, including two nontrivial planes; and $Ca_3Si_4Ir_4$ exhibits the validity failure above. These examples demonstrate both faithful prediction and loss of validity, but are too few to estimate production failure prevalence. Full results appear in Appendix G.

7 Possibility Is Not Prevalence

The preceding results establish possibility and structured sensitivity, but not prevalence at small physically relevant error. We therefore stress a broader JARVIS material set under both matrix-matched and spectrally matched perturbations.

Small matrix error need not imply small spectral error. We apply uniform rescaling, range attenuation, and paired-shell random perturbations while fixing the real-space Hamiltonian MAE at 1.417 meV. Across 168 materials, the resulting median near-Fermi band MAEs are 600.8, 556.5, and 2262.0 meV, respectively (Fig. 5a). Range attenuation and paired-shell perturbations also produce gap and topology failures, but only under these large spectral distortions. Thus a small entrywise Hamiltonian error need not imply small error in the spectrum governing the occupied bundle.

Band-matched perturbations show no failures in the tested cohort. We next calibrate the same perturbation mechanisms to approximately 10 meV near-Fermi band MAE over 24 materials and six time-reversal-invariant planes, for 432 evaluations. For each mechanism, 92 planes are valid at both endpoints and all preserve their Z_2 label; 52 are already invalid at the reference endpoint and are excluded. Across all mechanisms, we observe zero valid topology mismatches and no target-valid plane becoming newly unsafe (Fig. 5b).

Scope. The exact and realistic examples show that topological failure is possible and can be directionally sensitive, whereas this 10-meV study finds no such failures for the tested generic perturbations. Because these stresses are not residuals sampled from a production learned model, the result bounds the empirical claim rather than estimating real-world failure prevalence.

8 Related Work

Hamiltonian learning and downstream fidelity. Learned electronic Hamiltonians build on first-principles electronic structure (Hohenberg & Kohn, 1964; Kohn & Sham, 1965) and span direct matrix regression, localized-orbital models, and symmetry-aware architectures (Hegde & Bowen, 2017; Schütt et al., 2019; Unke et al., 2021; Zhang et al., 2022; Li et al., 2022), supported by datasets such

as NablaDFT and QH9 (Khrabrov et al., 2022; Yu et al., 2023a). Recent equivariant and transferable models include QHNet, DeepH-E3, HamGNN, Universal HamGNN, DeePTB, DEQHNet, UniversalDeepH, SLEM, SPHNet, TraceGrad, LargeHam, QHNetV2, and MACE-H (Yu et al., 2023b; Gong et al., 2023; Zhong et al., 2023; 2024; Gu et al., 2024; Wang et al., 2024b;a; Zhouyin et al., 2025; Luo et al., 2025; Yin et al., 2025b; Xia et al., 2025; Yu et al., 2025; Qian et al., 2026). Recent work also targets downstream fidelity: WALoss improves eigenspace and energy accuracy (Li et al., 2025), NextHAM combines real- and reciprocal-space supervision with subspace objectives (Yin et al., 2025a), and Uni-HamGNN applies learned Hamiltonians to large-scale topological screening (Zhong et al., 2026). Our question is complementary: rather than improving prediction, we ask when Hamiltonian, band, or subspace accuracy is sufficient evidence that a global topological class is preserved.

Computing and learning topology. Our setting builds on Chern and quantum-Hall topology (Thouless et al., 1982; Haldane, 1988) and time-reversal-protected Z_2 topology (Kane & Mele, 2005; Fu & Kane, 2006; Fu et al., 2007; Fu & Kane, 2007; Moore & Balents, 2007; Hasan & Kane, 2010). Established numerical tools include gauge-invariant lattice Chern calculations (Fukui et al., 2005), Wilson-loop and hybrid-Wannier methods (Yu et al., 2011; Soluyanov & Vanderbilt, 2011; Gresch et al., 2017), and Wannier-based interpolation and software (Marzari & Vanderbilt, 1997; Mostofi et al., 2008; Marzari et al., 2012; Wu et al., 2018). Machine learning has also targeted phases and invariants directly from loop, wavefunction, or Hamiltonian representations (Zhang & Kim, 2017; Deng et al., 2017; Zhang et al., 2018; Scheurer & Slager, 2020), including gauge-equivariant multiband Chern prediction (Huang et al., 2025); learned Hamiltonians have likewise been used for SOC transitions and Z_2 screening (Gong et al., 2023; Zhong et al., 2026). Symmetry-based approaches provide a distinct information source through topological quantum chemistry and symmetry indicators (Bradlyn et al., 2017; Po et al., 2017; Tang et al., 2019), including large-scale catalogues (Zhang et al., 2019; Vergniory et al., 2019). Our focus is narrower: when does ordinary predictive accuracy of an intermediate Hamiltonian itself certify topology?

Perturbative stability and certification. The positive certificates we use rely on classical perturbation theory for spectra and projectors (Kato, 1995; Davis & Kahan, 1970; Stewart & Sun, 1990; Bhatia, 1997) and projection geometry (Avron et al., 1994). Finite samples yield uniform bounds only with additional regularity and coverage assumptions (Wendland, 2005); operator-localizer methods provide alternative finite-volume access to topology (Loring & Hastings, 2010; Cerjan & Loring, 2022). Our material-scale stress tests use JARVIS infrastructure (Choudhary et al., 2020). We do not claim these stability tools as new. Our contribution is to connect them to the finite-p Hamiltonian, band, and occupied-subspace errors used for learned Hamiltonians, separating average accuracy, topological validity, and uniform certification. Appendix Table 1 gives a comparison.

9 Discussion and Conclusion

Our results distinguish *predictive accuracy* from *topological certification*. Integrated Hamiltonian, band, and subspace errors measure prediction quality but do not by themselves control a global topological class: topology-changing error can be arbitrarily localized while remaining finite pointwise. Conversely, topologically distinct valid occupied bundles cannot remain uniformly close in projector space. This motivates sufficient accuracy-based certificates based on uniform projector control or Hamiltonian error relative to a spectral gap.

The scope is limited. The finite-p obstruction concerns unrestricted smooth gapped families, while sparse certification requires independently justified regularity that we do not establish for the learned models here. The sampling benchmark covers only the tested finite-point designs. The $NaRuO_2$ perturbation is deliberately structured, the four released examples cannot estimate production prevalence, and the 10-meV JARVIS study uses generic perturbations rather than production residuals.

Reliable topology claims therefore require *validity* before *fidelity*: the occupied subspace must remain gapped, symmetry-compatible, and numerically resolved before invariants are compared. Within this scope, our results expose a certification gap rather than widespread topological failure. More broadly, evidence should match the downstream claim: average accuracy measures average fidelity, whereas certifying a global discrete property requires ruling out localized failures.

AI USE STATEMENT

We used generative AI tools to assist with literature retrieval and reference organization, manuscript editing, organization of experimental results, and research and mathematical reasoning. All AI-assisted text, citations, code, results, and proofs were reviewed and verified by the authors, who made all final methodological and scientific decisions. We take responsibility for the complete content of this work.

ETHICS STATEMENT

This work studies statistical properties of language-model responses and does not involve human participants or the collection of personal data. The experiments use existing benchmark tasks and pretrained language models. The certificates studied here characterize response distributions under registered representation groups; they should not be interpreted as guarantees of factual correctness, fairness, or safety.

REPRODUCIBILITY STATEMENT

We provide an anonymized supplementary archive containing the implementation, experiment configurations, definitions, random seeds, result artifacts used in the reported analyses, and scripts for reproducing the tables and figures. The supplement also includes environment information and integrity manifests for the experimental artifacts. Experimental settings and evaluation metrics are specified in the main paper and Appendix, and complete proofs of the theoretical results are provided in the Appendix.

REFERENCES

J. Avron, R. Seiler, and B. Simon. The index of a pair of projections. *Journal of Functional Analysis*, 120(1):220–237, 1994. doi: 10.1006/jfan.1994.1031.

Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, New York, 1997. doi: 10.1007/978-1-4612-0653-8.

Barry Bradlyn, L. Elcoro, Jennifer Cano, M. G. Vergniory, Zhijun Wang, C. Felser, M. I. Aroyo, and B. Andrei Bernevig. Topological quantum chemistry. *Nature*, 547:298–305, 2017. doi: 10.1038/nature23268.

Alexander Cerjan and Terry A. Loring. An operator-based approach to topological photonics. *Nanophotonics*, 11(21):4765–4780, 2022. doi: 10.1515/nanoph-2022-0547.

Kamal Choudhary, Kevin F. Garrity, Andrew C. E. Reid, Brian DeCost, Adam J. Biacchi, Angela R. Hight Walker, Zachary Trautt, Jason Hattrick-Simpers, A. Gilad Kusne, Andrea Centrone, Albert Davydov, Jie Jiang, Ruth Pachter, Gowoon Cheon, Evan Reed, Ankit Agrawal, Xiaofeng Qian, Vinit Sharma, Houlong Zhuang, Sergei V. Kalinin, Bobby G. Sumpter, Ghanshyam Pilania, Pinar Acar, Subhasish Mandal, Kristjan Haule, David Vanderbilt, Karin Rabe, and Francesca Tavazza. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Computational Materials*, 6:173, 2020. doi: 10.1038/s41524-020-00440-1.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Machine learning topological states. *Physical Review B*, 96:195145, 2017. doi: 10.1103/PhysRevB.96.195145.

Liang Fu and C. L. Kane. Time reversal polarization and a Z_2 adiabatic spin pump. *Physical Review B*, 74:195312, 2006. doi: 10.1103/PhysRevB.74.195312.

Liang Fu and C. L. Kane. Topological insulators with inversion symmetry. *Physical Review B*, 76: 045302, 2007. doi: 10.1103/PhysRevB.76.045302.

Liang Fu, C. L. Kane, and E. J. Mele. Topological insulators in three dimensions. *Physical Review Letters*, 98:106803, 2007. doi: 10.1103/PhysRevLett.98.106803.

Takahiro Fukui, Yasuhiro Hatsugai, and Hiroshi Suzuki. Chern numbers in discretized brillouin zone: Efficient method of computing (spin) hall conductances. *Journal of the Physical Society of Japan*, 74(6):1674–1677, 2005. doi: 10.1143/JPSJ.74.1674.

Xiaoxun Gong, He Li, Nianlong Zou, Runzhang Xu, Wenhui Duan, and Yong Xu. General framework for E(3)-equivariant neural network representation of density functional theory hamiltonian. *Nature Communications*, 14:2848, 2023. doi: 10.1038/s41467-023-38468-8.

Dominik Gresch, Gabriel Autès, Oleg V. Yazyev, Matthias Troyer, David Vanderbilt, B. Andrei Bernevig, and Alexey A. Soluyanov. Z2Pack: Numerical implementation of hybrid wannier centers for identifying topological materials. *Physical Review B*, 95:075146, 2017. doi: 10.1103/PhysRevB.95.075146.

Qiangqiang Gu, Zhanghao Zhouyin, Shishir Kumar Pandey, Peng Zhang, Linfeng Zhang, and Weinan E. Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with ab initio accuracy. *Nature Communications*, 15:6772, 2024. doi: 10.1038/s41467-024-51006-4.

F. D. M. Haldane. Model for a quantum hall effect without landau levels: Condensed-matter realization of the "parity anomaly". *Physical Review Letters*, 61:2015, 1988. doi: 10.1103/PhysRevLett.61.2015.

M. Z. Hasan and C. L. Kane. Colloquium: Topological insulators. *Reviews of Modern Physics*, 82:3045, 2010. doi: 10.1103/RevModPhys.82.3045.

Ganesh Hegde and R. Chris Bowen. Machine-learned approximations to density functional theory hamiltonians. *Scientific Reports*, 7:42669, 2017. doi: 10.1038/srep42669.

P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136:B864, 1964. doi: 10.1103/PhysRev.136.B864.

Longde Huang, Oleksandr Balabanov, Hampus Linander, Mats Granath, Daniel Persson, and Jan Gerken. Learning chern numbers of multiband topological insulators with gauge equivariant neural networks. In *Advances in Neural Information Processing Systems*, volume 38, pp. 164259–164288, 2025. doi: 10.52202/085713-4948.

C. L. Kane and E. J. Mele. Z_2 topological order and the quantum spin hall effect. *Physical Review Letters*, 95:146802, 2005. doi: 10.1103/PhysRevLett.95.146802.

Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, Berlin, Heidelberg, 2 edition, 1995. doi: 10.1007/978-3-642-66282-9.

Kuzma Khrabrov, Ilya Shenbin, Alexander Ryabov, Artem Tsypin, Alexander Telepov, Anton Alekseev, Alexander Grishin, Pavel Strashnov, Petr Zhilyaev, Sergey Nikolenko, and Artur Kadurin. nablaDFT: Large-scale conformational energy and hamiltonian prediction benchmark and dataset. *Physical Chemistry Chemical Physics*, 24:25853–25863, 2022. doi: 10.1039/D2CP03966D.

W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133, 1965. doi: 10.1103/PhysRev.140.A1133.

He Li, Zun Wang, Nianlong Zou, Meng Ye, Runzhang Xu, Xiaoxun Gong, Wenhui Duan, and Yong Xu. Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nature Computational Science*, 2:367–377, 2022. doi: 10.1038/s43588-022-00265-6.

Yunyang Li, Zaishuo Xia, Lin Huang, Xinran Wei, Samuel Harshe, Han Yang, Erpai Luo, Zun Wang, Jia Zhang, Chang Liu, Bin Shao, and Mark Gerstein. Enhancing the scalability and applicability of kohn–sham hamiltonians for molecular systems. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/035a73893121b4534bb3314e831050b1-Abstract-Conference.html.

T. A. Loring and M. B. Hastings. Disordered topological insulators via C^*-algebras. *EPL (Europhysics Letters)*, 92(6):67004, 2010. doi: 10.1209/0295-5075/92/67004.

Erpai Luo, Xinran Wei, Lin Huang, Yunyang Li, Han Yang, Zaishuo Xia, Zun Wang, Chang Liu, Bin Shao, and Jia Zhang. Efficient and scalable density functional theory hamiltonian prediction through adaptive sparsity. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 41368–41390. PMLR, 2025. URL https://proceedings.mlr.press/v267/luo25l.html.

Nicola Marzari and David Vanderbilt. Maximally localized generalized wannier functions for composite energy bands. *Physical Review B*, 56:12847, 1997. doi: 10.1103/PhysRevB.56.12847.

Nicola Marzari, Arash A. Mostofi, Jonathan R. Yates, Ivo Souza, and David Vanderbilt. Maximally localized wannier functions: Theory and applications. *Reviews of Modern Physics*, 84:1419, 2012. doi: 10.1103/RevModPhys.84.1419.

J. E. Moore and L. Balents. Topological invariants of time-reversal-invariant band structures. *Physical Review B*, 75:121306, 2007. doi: 10.1103/PhysRevB.75.121306.

Arash A. Mostofi, Jonathan R. Yates, Young-Su Lee, Ivo Souza, David Vanderbilt, and Nicola Marzari. wannier90: A tool for obtaining maximally-localised wannier functions. *Computer Physics Communications*, 178(9):685–699, 2008. doi: 10.1016/j.cpc.2007.11.016.

Hoi Chun Po, Ashvin Vishwanath, and Haruki Watanabe. Symmetry-based indicators of band topology in the 230 space groups. *Nature Communications*, 8:50, 2017. doi: 10.1038/s41467-017-00133-2.

Chen Qian, Valdas Vitartas, James R. Kermode, and Reinhard J. Maurer. Equivariant electronic hamiltonian prediction with many-body message passing. *npj Computational Materials*, 12:169, 2026. doi: 10.1038/s41524-026-02020-1.

Mathias S. Scheurer and Robert-Jan Slager. Unsupervised machine learning and band topology. *Physical Review Letters*, 124:226401, 2020. doi: 10.1103/PhysRevLett.124.226401.

K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature Communications*, 10:5024, 2019. doi: 10.1038/s41467-019-12875-2.

Alexey A. Soluyanov and David Vanderbilt. Computing topological invariants without inversion symmetry. *Physical Review B*, 83:235401, 2011. doi: 10.1103/PhysRevB.83.235401.

G. W. Stewart and Ji-guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.

Feng Tang, Hoi Chun Po, Ashvin Vishwanath, and Xiangang Wan. Efficient topological materials discovery using symmetry indicators. *Nature Physics*, 15:470–476, 2019. doi: 10.1038/s41567-019-0418-7.

D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. Quantized hall conductance in a two-dimensional periodic potential. *Physical Review Letters*, 49:405, 1982. doi: 10.1103/PhysRevLett.49.405.

Oliver T. Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess Smidt, and Klaus-Robert Müller. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. In *Advances in Neural Information Processing Systems*, volume 34, pp. 14434–14447, 2021.

M. G. Vergniory, L. Elcoro, Claudia Felser, Nicolas Regnault, B. Andrei Bernevig, and Zhijun Wang. A complete catalogue of high-quality topological materials. *Nature*, 566:480–485, 2019. doi: 10.1038/s41586-019-0954-4.

Yuxiang Wang, Yang Li, Zechen Tang, He Li, Zilong Yuan, Honggeng Tao, Nianlong Zou, Ting Bao, Xinghao Liang, Zezhou Chen, Shanghua Xu, Ce Bian, Zhiming Xu, Chong Wang, Chen Si, Wenhui Duan, and Yong Xu. Universal materials model of deep-learning density functional theory hamiltonian. *Science Bulletin*, 69(16):2514–2521, 2024a. doi: 10.1016/j.scib.2024.06.011.

Zun Wang, Chang Liu, Nianlong Zou, He Zhang, Xinran Wei, Lin Huang, Lijun Wu, and Bin Shao. Infusing self-consistency into density functional theory hamiltonian prediction via deep equilibrium models. In *Advances in Neural Information Processing Systems*, volume 37, 2024b. doi: 10.52202/079017-2846.

Holger Wendland. *Scattered Data Approximation*. Cambridge University Press, Cambridge, 2005. ISBN 978-0-521-84335-5.

QuanSheng Wu, ShengNan Zhang, Hai-Feng Song, Matthias Troyer, and Alexey A. Soluyanov. WannierTools: An open-source software package for novel topological materials. *Computer Physics Communications*, 224:405–416, 2018. doi: 10.1016/j.cpc.2017.09.033.

Chen Hao Xia, Manasa Kaniselvan, Alexandros Nikolaos Ziogas, Marko Mladenović, Rayen Mahjoub, Alexander Maeder, and Mathieu Luisier. Learning the electronic hamiltonian of large atomic structures. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 68236–68257. PMLR, 2025. URL https://proceedings.mlr.press/v267/xia25b.html.

Shi Yin, Zujian Dai, Xinyang Pan, and Lixin He. Advancing universal deep learning for electronic-structure hamiltonian prediction of materials, 2025a.

Shi Yin, Xinyang Pan, Fengyan Wang, and Lixin He. TraceGrad: a framework learning expressive SO(3)-equivariant non-linear representations for electronic-structure hamiltonian prediction. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 72364–72392. PMLR, 2025b. URL https://proceedings.mlr.press/v267/yin25b.html.

Haiyang Yu, Meng Liu, Youzhi Luo, Alex Strasser, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. QH9: A quantum hamiltonian prediction benchmark for QM9 molecules. In *Advances in Neural Information Processing Systems*, volume 36, pp. 40487–40503, 2023a. doi: 10.52202/075280-1761.

Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Efficient and equivariant graph networks for predicting quantum hamiltonian. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 40412–40424. PMLR, 2023b. URL https://proceedings.mlr.press/v202/yu23i.html.

Haiyang Yu, Yuchao Lin, Xuan Zhang, Xiaofeng Qian, and Shuiwang Ji. Efficient prediction of SO(3)-equivariant hamiltonian matrices via SO(2) local frames, 2025.

Rui Yu, Xiao Liang Qi, Andrei Bernevig, Zhong Fang, and Xi Dai. Equivalent expression of Z_2 topological invariant for band insulators using the non-abelian berry connection. *Physical Review B*, 84:075119, 2011. doi: 10.1103/PhysRevB.84.075119.

Liwei Zhang, Berk Onat, Geneviève Dusson, Adam McSloy, Gautam Anand, Reinhard J. Maurer, Christoph Ortner, and James R. Kermode. Equivariant analytical mapping of first principles hamiltonians to accurate and transferable materials models. *npj Computational Materials*, 8:158, 2022. doi: 10.1038/s41524-022-00843-2.

Pengfei Zhang, Huitao Shen, and Hui Zhai. Machine learning topological invariants with neural networks. *Physical Review Letters*, 120:066401, 2018. doi: 10.1103/PhysRevLett.120.066401.

Tiantian Zhang, Yi Jiang, Zhida Song, He Huang, Yuqing He, Zhong Fang, Hongming Weng, and Chen Fang. Catalogue of topological electronic materials. *Nature*, 566:475–479, 2019. doi: 10.1038/s41586-019-0944-6.

Yi Zhang and Eun-Ah Kim. Quantum loop topography for machine learning. *Physical Review Letters*, 118:216401, 2017. doi: 10.1103/PhysRevLett.118.216401.

Yang Zhong, Hongyu Yu, Mao Su, Xingao Gong, and Hongjun Xiang. Transferable equivariant graph neural networks for the hamiltonians of molecules and solids. *npj Computational Materials*, 9:182, 2023. doi: 10.1038/s41524-023-01130-4.

Yang Zhong, Hongyu Yu, Jihui Yang, Xingyu Guo, Hongjun Xiang, and Xingao Gong. Universal machine learning kohn–sham hamiltonian for materials. *Chinese Physics Letters*, 41(7):077103, 2024. doi: 10.1088/0256-307X/41/7/077103.

Table 1: **Positioning of learned-Hamiltonian topology certification.** Each row distinguishes a line of work's primary objective from guarantees that are outside its usual scope; the final column does not assert impossibility.

Line of work	Representative work	What it solves	Typical quantity	Not guaranteed / outside primary scope
Symmetry-equivariant Hamiltonian prediction	DeepH-E3, HamGNN, QHNet, SLEM, Uni-HamGNN, NextHAM (Gong et al., 2023; Zhong et al., 2023; Yu et al., 2023b; Zhouyin et al., 2025; Zhong et al., 2026; Yin et al., 2025a)	Covariant, local, and scalable prediction of Hamiltonian matrix elements	Matrix MAE/RMSE, band error, wavefunction or subspace error	Not guaranteed by sampled losses: preservation of the reference occupied-bundle class between sampled points
Downstream-aware spectral/subspace losses	WALoss and NextHAM (Li et al., 2025; Yin et al., 2025a)	Basis alignment, ghost-state reduction, and improved low-energy spectra or subspaces	Aligned wavefunction, eigenvalue, projector, or subspace loss	Not guaranteed by these objectives alone: a uniform error bound and a gap-preserving path over the full Brillouin zone
Direct invariant classification	Quantum loop topography and topology-sensitive classifiers (Zhang & Kim, 2017; Deng et al., 2017; Zhang et al., 2018; Scheurer & Slager, 2020; Huang et al., 2025)	Prediction of phase labels or invariants from topology-sensitive inputs	Classification accuracy or invariant regression error	Outside the primary task: whether a separately learned Hamiltonian preserves the reference topology and remains valid
Numerical invariant computation	Lattice Chern methods, Wilson loops, and Z2Pack (Fukui et al., 2005; Yu et al., 2011; Soluyanov & Vanderbilt, 2011; Gresch et al., 2017)	Gauge-invariant Chern or Z_2 evaluation for a supplied, sufficiently resolved Hamiltonian	Lattice flux, Wilson phase, Wannier-center flow	Outside the primary scope: prediction validity, surrogate-to-reference agreement, and sparse coverage
Perturbative topological stability	Kato–Nagy, Weyl, and Davis–Kahan theory (Kato, 1995; Davis & Kahan, 1970; Bhatia, 1997)	Sufficient same-phase guarantees under uniform projector or operator perturbations	Operator norm relative to a spectral gap	Outside the theorem's scope: whether conventional average ML metrics estimate the required supremum
Sparse reciprocal-space auditing	IID, stratified, path, and loop point-query designs; fill-distance bounds (Wendland, 2005)	Cost-controlled empirical probing of a fixed error field	Observed maximum, percentile, or mean error at a query budget	Not guaranteed without a regularity or bandwidth bound: the unobserved supremum

Yang Zhong, Rui Wang, Xingao Gong, and Hongjun Xiang. A universal spin–orbit-coupled hamiltonian model for accelerated quantum material discovery. *Nature Machine Intelligence*, 8:403–414, 2026. doi: 10.1038/s42256-026-01196-x.

Zhanghao Zhouyin, Zixi Gan, Shishir K. Pandey, Linfeng Zhang, and Qiangqiang Gu. Learning local equivariant representations for quantum operators. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/7db2ffcbfd0bd361d47b7fa612bd2ba2-Abstract-Conference.html.

A Detailed Literature Positioning

B Mathematical Proofs and Extensions

This appendix provides the assumptions and proofs underlying Sec. 4. The negative result concerns the insufficiency of finite-p integrated errors over unrestricted smooth gapped families. The positive results are sufficient certificates based on uniform projector or Hamiltonian control. We state the class-A results first and then give the fixed-Fermi-level and time-reversal-equivariant extensions.

B.1 Common representation and spectral assumptions

Let $X = \mathbb{T}^d$ denote the compact Brillouin zone, equipped with a fixed periodic metric d and normalized Haar measure μ,

$$\mu(X) = 1.$$

Unless stated otherwise, all matrix norms below are operator norms.

We assume that the reference and predicted Hamiltonians

$$H_*, H_\theta : X \to \mathrm{Herm}(m)$$

are expressed in the same continuous periodic orthonormal orbital trivialization. This common representation is required for a matrix difference such as

$$H_\theta(k) - H_*(k)$$

to have a basis-consistent meaning.

For a generalized eigenvalue problem

$$H(k)u = \varepsilon S(k)u,$$

we assume that the reference and prediction share the same overlap matrix $S(k)$, orbital ordering, and Brillouin-zone parametrization, with $S(k)$ continuous, periodic, and uniformly positive definite:

$$\inf_{k\in X} \lambda_{\min}(S(k)) > 0.$$

All Hamiltonian-norm statements then apply to the common Löwdin representation

$$\widetilde{H}(k) = S(k)^{-1/2} H(k) S(k)^{-1/2}. \tag{2}$$

If S is compatible with an additional symmetry, its positive square root and inverse square root inherit the same equivariance by functional calculus.

For a fixed occupied rank r, we write

$$P_r[H(k)]$$

for the spectral projector onto the lowest r eigenvalues of $H(k)$. Whenever a projector is defined in this way, we assume that the corresponding Hamiltonian has a positive uniform gap,

$$\Delta_j = \inf_{k\in X} [\lambda_{r+1}(H_j(k)) - \lambda_r(H_j(k))] > 0, \qquad j \in \{*, \theta\}. \tag{3}$$

Internal crossings among the lowest r bands are allowed; only separation of the occupied cluster from its complement is required. Standard spectral calculus then implies that

$$P_j(k) = P_r[H_j(k)]$$

is a continuous rank-r orthogonal projector family.

More generally, the projector results below may be applied directly to continuous maps

$$P_0, P_1 : X \to \mathrm{Proj}_r(\mathbb{C}^m)$$

without referring to an underlying Hamiltonian. Their image bundles are

$$\mathcal{E}_j = \{(k, v) : k \in X,\ v \in \mathrm{Ran}\, P_j(k)\}\,.$$

Throughout the theory, a *topological mismatch* means that these well-defined continuous equal-rank bundles are not isomorphic. A gap closure, rank ambiguity, or failure of a numerical invariant to converge is a different type of reliability failure and is not covered by the projector-mismatch statements.

For two-dimensional class-A systems, we use the occupied-projector convention

$$C = \mathrm{Ch}_1(P_{\mathrm{occ}}).$$

For a flattened two-band Hamiltonian

$$H(k) = n(k) \cdot \sigma, \qquad \|n(k)\|_2 = 1,$$

the lower-band projector is

$$P_-(k) = \frac{I - n(k)\cdot\sigma}{2}.$$

With the orientation convention used here,

$$\mathrm{Ch}_1(P_-) = -\deg(n). \tag{4}$$

B.2 PROOF OF FINITE-p NON-CERTIFICATION

We prove Proposition 1 by constructing a smooth degree-one texture whose support can be made arbitrarily small while the spectral gap remains fixed.

Proof of Proposition 1. Fix $1 \le p < \infty$ and $\varepsilon > 0$. Choose a point $k_0 \in \mathbb{T}^2$ and a radius R smaller than the injectivity radius of the torus, so that the disk

$$D_R = \{k : d(k, k_0) < R\}$$

lies inside a single Euclidean coordinate chart. Write the corresponding polar coordinates as (ρ, φ).

Choose a smooth profile

$$\beta : [0, \infty) \to [0, \pi]$$

such that

$$\beta(t) = \pi \quad \text{for } t \le \frac{1}{4}, \qquad \beta(t) = 0 \quad \text{for } t \ge \frac{3}{4}.$$

Define

$$n_R(\rho, \varphi) = (\sin \beta(\rho/R) \cos \varphi,\ \sin \beta(\rho/R) \sin \varphi,\ \cos \beta(\rho/R)) \tag{5}$$

inside D_R, and set

$$n_R(k) = n_0 = (0, 0, 1)$$

outside D_R. Because the profile is constant in neighborhoods of both the center and the boundary of the disk, this definition extends to a C^∞ periodic map on $\mathbb{T}^2$.

The radial texture covers the sphere once. With the above orientation,

$$\deg(n_R) = \frac{1}{4\pi} \int_{\mathbb{T}^2} n_R \cdot \left(\partial_{k_x} n_R \times \partial_{k_y} n_R\right) \, dk = -1.$$

Changing the angular orientation reverses the sign if desired; only the nonzero degree is relevant below.

Define

$$H_R(k) = n_R(k) \cdot \sigma, \qquad H_0(k) = n_0 \cdot \sigma.$$

Since $\|n_R(k)\|_2 = 1$,

$$H_R(k)^2 = I,$$

and hence

$$\operatorname{spec} H_R(k) = \operatorname{spec} H_0(k) = \{-1, +1\}$$

for every k. Both Hamiltonians therefore have a uniform spectral gap equal to 2.

Their lower-band projectors are

$$P_R(k) = \frac{I - H_R(k)}{2}, \qquad P_0(k) = \frac{I - H_0(k)}{2}.$$

By Eq. equation 4,

$$\mathrm{Ch}_1(P_R) = 1, \qquad \mathrm{Ch}_1(P_0) = 0$$

under the stated orientation convention.

Both $H_R - H_0$ and $P_R - P_0$ vanish outside D_R. Moreover,

$$\|H_R(k) - H_0(k)\|_{\mathrm{op}} = \|n_R(k) - n_0\|_2 \le 2,$$

and

$$\|P_R(k) - P_0(k)\|_{\mathrm{op}} = \frac{1}{2} \|H_R(k) - H_0(k)\|_{\mathrm{op}} \le 1.$$

Therefore,

$$\begin{aligned} \|H_R - H_0\|_{L^p}^p &= \int_{D_R} \|H_R(k) - H_0(k)\|_{\mathrm{op}}^p \, d\mu(k) \\ &\le 2^p \, \mu(D_R) = O(R^2), \end{aligned} \tag{6}$$

$$\|P_R - P_0\|_{L^p}^p \le \mu(D_R) = O(R^2). \tag{7}$$

Hence

$$\|H_R - H_0\|_{L^p},\ \|P_R - P_0\|_{L^p} = O(R^{2/p}). \tag{8}$$

Choosing R sufficiently small makes both norms smaller than ε, while the occupied Chern numbers remain different. □

The construction is stated in operator norm because this is the norm used in the positive Hamiltonian certificate below. In finite matrix dimension, all fixed matrix norms are equivalent, so the same non-certification conclusion holds, up to constant factors, for any fixed matrix norm.

The family in Proposition 1 does not obey a uniform derivative or Lipschitz bound as $R \to 0$. This is essential to the scope of the result: additional regularity can prevent a finite-amplitude discrepancy from being concentrated on arbitrarily small support and can therefore relate integrated and uniform errors.

B.3 Principal-angle identities

We next relate the occupied-subspace infidelity used in the main text to the operator-norm distance between orthogonal projectors.

Lemma 1 (Principal-angle identities)**.** *Let $P_0, P_1 \in \mathrm{Proj}_r(\mathbb{C}^m)$ be rank-r orthogonal projectors, and let*

$$0 \le \theta_1 \le \cdots \le \theta_r \le \frac{\pi}{2}$$

be the principal angles between their ranges. Define

$$f = 1 - \frac{1}{r}\,\mathrm{Tr}(P_0 P_1).$$

Then

$$f = \frac{1}{r}\sum_{i=1}^{r} \sin^2\theta_i, \tag{9}$$

and

$$\|P_1 - P_0\|_{\mathrm{op}} = \max_i \sin\theta_i. \tag{10}$$

Consequently,

$$f \le \|P_1 - P_0\|_{\mathrm{op}}^2 \le rf. \tag{11}$$

Proof. Let

$$U_0, U_1 \in \mathbb{C}^{m\times r}$$

be orthonormal frames for $\mathrm{Ran}\, P_0$ and $\mathrm{Ran}\, P_1$, so that

$$P_j = U_j U_j^\dagger.$$

The singular values of

$$U_0^\dagger U_1$$

are $\cos\theta_1, \ldots, \cos\theta_r$. Therefore

$$\begin{aligned}\mathrm{Tr}(P_0P_1) &= \mathrm{Tr}\left(U_0^\dagger U_1 U_1^\dagger U_0\right)\\ &= \sum_{i=1}^{r}\cos^2\theta_i,\end{aligned} \tag{12}$$

which gives Eq. equation 9.

The nonzero singular values of the difference of two equal-rank orthogonal projectors are the corresponding $\sin\theta_i$, yielding Eq. equation 10. Finally,

$$\frac{1}{r}\sum_i \sin^2\theta_i \le \max_i \sin^2\theta_i \le \sum_i \sin^2\theta_i,$$

which proves Eq. equation 11. □

For rank one,

$$f = \|P_1 - P_0\|_{\mathrm{op}}^2.$$

For larger rank, f averages over principal directions. In particular, one completely orthogonal principal direction already gives projector distance 1, while contributing only $1/r$ to the normalized subspace infidelity.

B.4 KATO–NAGY PROJECTOR CERTIFICATE

We now prove Proposition 2. The argument is the standard Kato–Nagy direct-rotation construction (Kato, 1995); related canonical-angle and subspace bounds are reviewed by Stewart & Sun (1990) and originate in the Davis–Kahan theory (Davis & Kahan, 1970).

Proof of Proposition 2. Assume

$$\eta := \sup_{k \in X} \|P_1(k) - P_0(k)\|_{\mathrm{op}} < 1.$$

For each k, define

$$A(k) = P_1(k)P_0(k) + \big(I - P_1(k)\big)\big(I - P_0(k)\big) \tag{13}$$

and

$$B(k) = I - \big(P_1(k) - P_0(k)\big)^2. \tag{14}$$

A direct calculation gives

$$A(k)^\dagger A(k) = B(k).$$

Since

$$\|P_1(k) - P_0(k)\|_{\mathrm{op}} \le \eta < 1,$$

we have

$$B(k) \succeq (1 - \eta^2) I > 0.$$

Thus $B(k)^{-1/2}$ exists and depends continuously on k.

Define the polar unitary

$$U(k) = A(k)B(k)^{-1/2}. \tag{15}$$

Because $A^\dagger A = B$ and A is invertible, $U(k)$ is unitary. Furthermore,

$$A(k)P_0(k) = P_1(k)A(k).$$

The operator $B(k)$ commutes with $P_0(k)$, and hence so does $B(k)^{-1/2}$. Therefore

$$U(k)P_0(k) = P_1(k)U(k),$$

or equivalently,

$$P_1(k) = U(k)P_0(k)U(k)^\dagger. \tag{16}$$

Since P_0 and P_1 are continuous, so is U. Restricting $U(k)$ to $\mathrm{Ran}\, P_0(k)$ therefore gives a continuous fiberwise unitary isomorphism

$$\mathcal{E}_0 \simeq \mathcal{E}_1.$$

For the converse implication used in the main text, suppose that the two bundles are not isomorphic. Then the preceding sufficient condition cannot hold, so

$$\sup_{k \in X} \|P_1(k) - P_0(k)\|_{\mathrm{op}} \ge 1.$$

For orthogonal projectors the pointwise distance is at most 1, hence

$$\sup_{k \in X} \|P_1(k) - P_0(k)\|_{\mathrm{op}} = 1.$$

Because X is compact and the distance is continuous, the supremum is attained at some k_*. Applying Lemma 1,

$$1 = \|P_1(k_*) - P_0(k_*)\|_{\mathrm{op}}^2 \le r f(k_*),$$

and therefore

$$\sup_{k \in X} f(k) \ge \frac{1}{r}.$$

□

The bound $1/r$ is sharp. If exactly one principal angle equals $\pi/2$ and all others vanish, then

$$\|P_1 - P_0\|_{\mathrm{op}} = 1, \qquad f = \frac{1}{r}.$$

Thus increasing occupied rank can dilute a single maximally discrepant principal direction in the normalized infidelity, even though the projector has already reached the boundary of the direct-rotation certificate.

B.5 HAMILTONIAN GAP CERTIFICATE

We next prove Proposition 3. Unlike the projector criterion above, this condition also guarantees that the predicted rank-r spectral subspace remains well defined along the interpolation.

Proof of Proposition 3. Let

$$E(k) = H_\theta(k) - H_*(k), \qquad \delta = \sup_{k \in X} \|E(k)\|_{\mathrm{op}},$$

and consider

$$H_s(k) = H_*(k) + sE(k), \qquad s \in [0, 1].$$

By Weyl's inequality (Bhatia, 1997),

$$\lambda_{r+1}(H_s(k)) \geq \lambda_{r+1}(H_*(k)) - s\|E(k)\|_{\mathrm{op}}, \tag{17}$$
$$\lambda_r(H_s(k)) \leq \lambda_r(H_*(k)) + s\|E(k)\|_{\mathrm{op}}. \tag{18}$$

Subtracting gives

$$\begin{aligned} \lambda_{r+1}(H_s(k)) - \lambda_r(H_s(k)) &\geq \lambda_{r+1}(H_*(k)) - \lambda_r(H_*(k)) - 2s\|E(k)\|_{\mathrm{op}} \\ &\geq \Delta_* - 2s\delta. \end{aligned} \tag{19}$$

If

$$\delta < \frac{\Delta_*}{2},$$

then the right-hand side is strictly positive for every $(k, s) \in X \times [0, 1]$. The lowest r eigenvalues therefore remain spectrally separated from the remaining bands along the entire path.

The corresponding spectral projectors

$$P_s(k) = P_r[H_s(k)]$$

depend continuously on (k, s). Hence

$$\operatorname{Ran} P_s \longrightarrow X \times [0, 1]$$

defines a vector bundle interpolating between the endpoint occupied bundles. The two endpoints are therefore homotopic through gapped class-A Hamiltonians and have isomorphic occupied bundles and identical characteristic classes. □

The condition

$$\delta < \Delta_*/2$$

is sufficient, not necessary. For example, an error of the form

$$E(k) = c(k)I$$

can have arbitrarily large operator norm while leaving every eigenvector, the direct band gap, and the fixed-rank projector unchanged. The Hamiltonian criterion is therefore a conservative certificate rather than a sharp topological decision boundary.

B.6 FIXED-FERMI-LEVEL EXTENSION

Proposition 3 protects the spectral subspace formed by the lowest r bands. If the occupied space is instead defined relative to a fixed chemical potential μ, preservation of the $r/(r+1)$ direct gap alone is insufficient: the entire spectrum may shift across μ without changing the separation between neighboring bands.

Assume that μ lies uniformly inside the reference gap and define

$$g_\mu = \inf_{k \in X} \min\{\mu - \lambda_r(H_*(k)), \lambda_{r+1}(H_*(k)) - \mu\} > 0. \tag{20}$$

If

$$\delta = \sup_k \|H_\theta(k) - H_*(k)\|_{\mathrm{op}} < g_\mu, \tag{21}$$

then along the straight interpolation,

$$\lambda_r(H_s(k)) \le \lambda_r(H_*(k)) + s\delta < \mu, \tag{22}$$

$$\lambda_{r+1}(H_s(k)) \ge \lambda_{r+1}(H_*(k)) - s\delta > \mu. \tag{23}$$

Thus μ remains in the gap for every k and s, and the number of states below μ remains equal to r.

The distinction between the direct-gap and fixed-μ criteria is necessary. For example, consider

$$H_* = \begin{pmatrix} -0.1 & 0 \\ 0 & 100 \end{pmatrix}, \qquad \mu = 0, \qquad E = 0.2I.$$

The reference direct gap is

$$\Delta_* = 100.1,$$

and

$$\|E\|_{\mathrm{op}} = 0.2 < \Delta_*/2.$$

The direct band separation is unchanged, yet both predicted eigenvalues lie above μ, so the occupied rank below the fixed chemical potential changes from one to zero. Condition equation 21, rather than the direct-gap criterion alone, rules out this possibility.

B.7 Time-reversal-equivariant projector certificate

The projector certificate in Proposition 2 is a statement about ordinary complex vector bundles. Symmetry-protected indices require the bundle isomorphism to respect the protecting symmetry. We record the time-reversal-symmetric extension relevant to class AII.

Let the Brillouin-zone involution be

$$k \mapsto -k,$$

and let

$$\Theta_k : \mathcal{H}_k \to \mathcal{H}_{-k}$$

be antiunitary maps satisfying the spinful time-reversal relation

$$\Theta_{-k}\Theta_k = -I. \tag{24}$$

A projector family P is time-reversal equivariant if

$$P(-k)\Theta_k = \Theta_k P(k), \tag{25}$$

or equivalently

$$P(-k) = \Theta_k P(k)\Theta_k^{-1}.$$

Proposition 5 (Time-reversal-equivariant projector certificate)**.** *Let $P_0, P_1 : X \to \mathrm{Proj}_r(\mathbb{C}^m)$ be continuous equal-rank projector families that satisfy Eq. equation 25 with respect to the same time-reversal structure Θ_k. If*

$$\sup_{k\in X} \|P_1(k) - P_0(k)\|_{\mathrm{op}} < 1,$$

then their occupied Quaternionic bundles are equivariantly isomorphic. Consequently, they have the same class-AII Z_2 invariant.

Proof. Use the same operators $A(k)$, $B(k)$, and $U(k)$ as in Eqs. equation 13–equation 15. Since both projectors obey the same time-reversal relation,

$$A(-k) = \Theta_k A(k)\Theta_k^{-1},$$

and

$$B(-k) = \Theta_k B(k)\Theta_k^{-1}.$$

The function $x \mapsto x^{-1/2}$ is real-valued on the positive spectrum of B, so positive functional calculus preserves the antiunitary equivariance:

$$B(-k)^{-1/2} = \Theta_k B(k)^{-1/2}\Theta_k^{-1}.$$

Therefore

$$U(-k) = \Theta_k U(k)\Theta_k^{-1},$$

or

$$U(-k)\Theta_k = \Theta_k U(k). \tag{26}$$

The direct rotation thus maps $\mathrm{Ran}\, P_0(k)$ to $\mathrm{Ran}\, P_1(k)$ while intertwining the time-reversal structures. It is therefore an isomorphism of the corresponding Quaternionic bundles, which preserves their class-AII Z_2 index. □

An analogous Hamiltonian-space statement follows immediately. If H_* and H_θ are equivariant with respect to the same exact time-reversal representation,

$$H_j(-k) = \Theta_k H_j(k) \Theta_k^{-1}, \qquad j \in \{*, \theta\},$$

then the straight interpolation

$$H_s = (1-s)H_* + sH_\theta$$

preserves time-reversal symmetry. If the gap condition of Proposition 3 also holds, the entire path is both gapped and time-reversal symmetric, so the endpoints lie in the same class-AII phase.

For a generalized eigenproblem, this statement additionally requires the common Löwdin transformation in Eq. equation 2 to respect the same time-reversal structure. The exact continuum statement should be distinguished from numerical Z_2 evaluation, which additionally requires gap, discretization, and Kramers-pairing convergence.

B.8 Sparse coverage bounds

We next give the covering argument used in Sec. 4.4. The result is elementary but makes explicit which additional information is needed to convert finite point observations into a continuum bound; the fill-distance formulation is standard in scattered-data approximation (Wendland, 2005).

Let (X, d) be a compact metric space and let

$$\varnothing \neq Q \subset X$$

be finite. Define the fill distance

$$h(Q) = \sup_{k \in X} \min_{q \in Q} d(k, q). \tag{27}$$

Lemma 2 (Sparse coverage bound)**.** *Let $E : X \to \mathrm{Herm}(m)$ satisfy the global Lipschitz bound*

$$\|E(k) - E(k')\|_{\mathrm{op}} \leq L_E d(k, k')$$

for all $k, k' \in X$. Then

$$\|E\|_\infty \leq \max_{q \in Q} \|E(q)\|_{\mathrm{op}} + L_E h(Q). \tag{28}$$

Proof. For arbitrary $k \in X$, choose a nearest query point

$$q(k) \in \arg\min_{q \in Q} d(k, q).$$

Then

$$\begin{aligned} \|E(k)\|_{\mathrm{op}} &\leq \|E(q(k))\|_{\mathrm{op}} + \|E(k) - E(q(k))\|_{\mathrm{op}} \\ &\leq \max_{q \in Q} \|E(q)\|_{\mathrm{op}} + L_E d(k, q(k)) \\ &\leq \max_{q \in Q} \|E(q)\|_{\mathrm{op}} + L_E h(Q). \end{aligned} \tag{29}$$

Taking the supremum over k proves Eq. equation 28. □

Hamiltonian-space consequence. Let

$$E(k) = H_\theta(k) - H_*(k).$$

If

$$\max_{q \in Q} \|E(q)\|_{\mathrm{op}} + L_E h(Q) < \frac{\Delta_*}{2}, \tag{30}$$

then Lemma 2 implies

$$\|E\|_\infty < \Delta_*/2.$$

Proposition 3 therefore guarantees a gapped interpolation and preservation of the fixed-r class-A topology. For fixed-Fermi occupancy, the right-hand side of Eq. equation 30 is replaced by g_μ.

Projector-space consequence. Suppose instead that P_0 and P_1 are already known to be globally well-defined continuous rank-r projector families, and define

$$f(k) = 1 - \frac{1}{r}\,\mathrm{Tr}[P_0(k)P_1(k)].$$

Assume that the scalar field f has a certified global Lipschitz bound

$$|f(k) - f(k')| \le L_f d(k, k').$$

The same nearest-neighbor argument gives

$$\sup_{k\in X} f(k) \le \max_{q\in Q} f(q) + L_f h(Q). \tag{31}$$

If

$$\max_{q\in Q} f(q) + L_f h(Q) < \frac{1}{r}, \tag{32}$$

then

$$\sup_k f(k) < \frac{1}{r}.$$

By Lemma 1,

$$\sup_k \|P_1(k) - P_0(k)\|_{\mathrm{op}}^2 < 1,$$

and Proposition 2 yields bundle isomorphism.

The logical prerequisites of the two sparse routes differ. The Hamiltonian condition can itself protect the occupied/unoccupied spectral separation once the bound falls below the gap margin. The projector condition is conditional on the two projector families already being globally well defined; it does not exclude an unsampled gap closing at which P_1 ceases to define a continuous rank-r bundle.

The quantities L_E and L_f in these statements are global upper bounds relative to the same metric d used in $h(Q)$. Pointwise samples alone do not determine them. In particular, an integrated smoothness penalty does not by itself imply either global Lipschitz bound.

B.9 Fixed finite-point no-go

Finally, we prove Proposition 4. The result formalizes the obstruction faced by finite point evaluations when no quantitative regularity restriction is imposed.

Proof of Proposition 4. Let

$$Q = \{q_1, \dots, q_N\} \subset \mathbb{T}^2$$

be any fixed finite set. Since Q is finite, there exists a point

$$k_0 \in \mathbb{T}^2 \setminus Q$$

and a sufficiently small open disk D_R centered at k_0 whose closure is disjoint from Q.

Let

$$H_0(k) = n_0 \cdot \sigma$$

be the constant trivial Hamiltonian from Sec. B.2. Construct H_1 by inserting the smooth degree-one texture of Eq. equation 5 entirely inside D_R and setting

$$H_1(k) = H_0(k)$$

outside D_R.

Both Hamiltonians have pointwise spectrum

$$\{-1, +1\},$$

so each has gap 2. Their occupied Chern classes differ by one. Because the closure of D_R is disjoint from the finite set Q, every query point has an open neighborhood contained in $\mathbb{T}^2 \setminus D_R$. Hence

$$H_1(k) = H_0(k)$$

on an open neighborhood of every $q \in Q$, even though the two occupied bundles have different topology. □

The proposition concerns a *fixed finite point set*. Its information model does not include continuous access to an entire path, exact holonomy or integral queries, derivative information, or a priori restriction to a function class with known regularity. Likewise, additional arguments would be required to extend the statement to randomized or adaptive query algorithms. The conclusion needed here is narrower: finite point agreement alone, without quantitative control between the sampled points, is not a universal topology certificate.

C Numerical and Topological Robustness

The main results rely on discrete evaluations of continuous Bloch Hamiltonians and their occupied subspaces. We therefore examine whether the reported topological conclusions are stable to momentum-grid resolution, grid origin, local refinement of the largest-error region, and the numerical conditions required by Chern and Wilson-loop calculations. These checks serve two distinct purposes: they test whether the controlled learned-model results can be explained by discretization artifacts, and they separate well-resolved topological disagreement from cases in which the predicted Hamiltonian itself enters a numerically ill-conditioned regime.

C.1 Grid and origin stability

A discrete Chern calculation can in principle depend on both the density and the origin of the momentum mesh, particularly when Berry curvature or projector error is spatially concentrated. We therefore reevaluate the learned two-band models across four grid resolutions,

$$N \in \{48, 64, 96, 128\},$$

and five grid origins at each resolution. This gives 20 discretizations per model. Across 43 learned models, the analysis contains 860 model evaluations; the corresponding target textures contribute an additional 80 evaluations.

The target Chern labels are unchanged in all 80 evaluations. Among localized models that are topologically correct under the reference evaluation, the unregularized models retain their label in 96.7% of grid–origin combinations in aggregate. Their median retention fraction is 1.0, with the least stable model retaining its label in 70% of the tested discretizations. The corresponding Dirichlet-regularized models retain their correct label in every tested grid–origin condition.

Figure 6 summarizes these results. Higher Dirichlet energy is moderately associated with lower label retention (Spearman correlation approximately -0.51), but the effect is not strong enough to make discretization instability the dominant explanation of the learned topological behavior. In particular, most nominally correct models remain correct across changes in both grid spacing and origin.

C.2 Rank-two local-witness refinement

Proposition 2 implies that two topologically different, well-defined rank-r occupied bundles must satisfy

$$\sup_k f(k) \geq \frac{1}{r}.$$

For the learned four-band models in Sec. 3, $r = 2$, so every valid topological mismatch must attain

$$\sup_k f(k) \geq \frac{1}{2}.$$

The original 96^2 evaluation grids place several failed models just below this threshold, raising the natural question of whether finite sampling is underestimating the true supremum.

We therefore refine the evaluation of all 15 IID four-band models with incorrect Chern number. Table 2 reports the distribution of the largest sampled subspace infidelity. On the original 96^2 grid, the minimum, median, and maximum over the 15 models are 0.483804, 0.496735, and 0.499561, respectively. Increasing the resolution to 384^2 raises these values to 0.497468, 0.499561, and 0.499987.

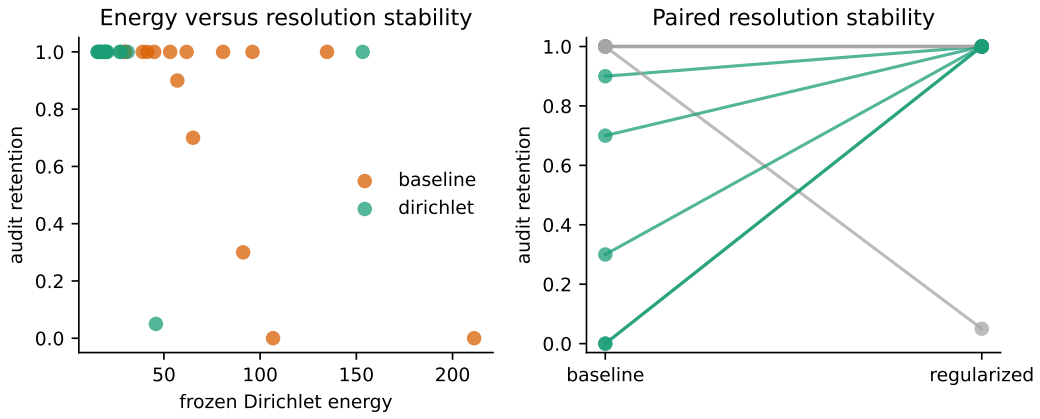


Figure 6: **Grid- and origin-stability analysis of the learned two-band models.** Each model is reevaluated on grids of size $48^2, 64^2, 96^2$, and 128^2, with five grid origins per resolution. Target labels are stable in all 80 target evaluations. Among localized models that are initially topologically correct, unregularized models retain their label in 96.7% of the 20 grid–origin conditions in aggregate, with median retention 1.0, while the corresponding regularized models retain their label in every tested condition. Larger Dirichlet energy is associated with reduced stability, but the high overall retention shows that the primary topology results are not explained by a single discretization or grid origin.

Table 2: **Resolution dependence of the local projector witness in the 15 learned rank-two topology failures.** Entries are the minimum, median, and maximum, across models, of each model's largest sampled occupied-subspace infidelity $f(k)$. The theoretical mismatch threshold for rank $r = 2$ is $1/r = 0.5$.

Evaluation	Min. max-f	Median max-f	Max. max-f
96^2 grid	0.483804	0.496735	0.499561
384^2 grid	0.497468	0.499561	0.499987

For the three cases whose dense-grid maxima remain closest to the threshold, continuous local refinement initialized from the largest-error grid points gives maxima between

$$0.5000007 \quad \text{and} \quad 0.5000088.$$

The 10^{-5}-scale excess above $1/2$ is not in tension with the theoretical result: $1/2$ is a lower bound rather than an upper bound, and a second nonzero principal angle can increase the normalized infidelity slightly beyond the single-direction value.

The topological labels remain unchanged under the same refinement of the global discretization. On both 192^2 and 384^2 Chern grids, all 15 targets have occupied Chern number 1, while all 15 predictions have Chern number 0. The minimum determinant-link singular values of the predicted occupied subspaces are at least 0.573 and 0.787, respectively, on these two resolutions.

These results are a numerical consistency check on the pointwise-witness characterization, not a proof of Proposition 2. They show that the apparent sub-$1/2$ values on a finite mesh are resolved by increasing spatial resolution: the sampled maxima approach the rank-two threshold from below while the topological mismatch remains stable.

C.3 $NaRuO_2$ CONVERGENCE

The $NaRuO_2$ experiment in Sec. 6 involves a substantially larger occupied space and a nonuniform Berry-curvature distribution, making convergence diagnostics particularly important. We therefore evaluate the reference Hamiltonian and the structured perturbation over multiple reciprocal-space grids and monitor the direct gap and link conditioning in addition to the integer Chern result.

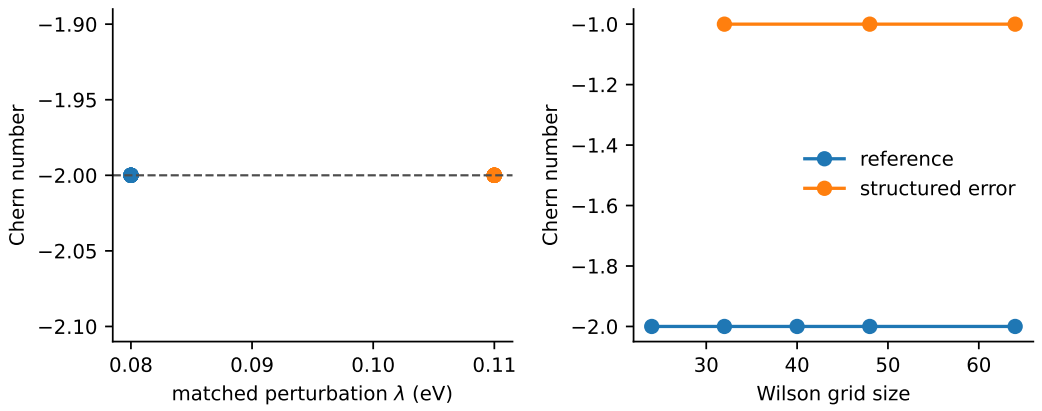


Figure 7: **Numerical convergence and conditioning for the $NaRuO_2$ Wannier Hamiltonian.** The reference occupied Chern number remains $C = -2$ across the tested reciprocal-space discretizations and slices, with minimum direct gap 144.775 meV and minimum occupied-link singular value above 0.939. The structured perturbation is evaluated on 32^2, 48^2, and 64^2 grids; the main text reports the finest-grid result, where the first transition occurs at $\lambda = 0.08$ eV. Gap minima accompany the sequence of Chern changes, supporting their interpretation as resolved topological transitions rather than isolated discretization artifacts.

For the reference 32-orbital spin–orbit-coupled Wannier Hamiltonian, the occupied Chern number is

$$C = -2$$

across all tested grid resolutions and slices. The minimum reference direct gap is 144.775 meV, and the minimum singular value of the occupied-space link matrices remains above 0.939. The latter quantity stays well away from zero, indicating that neighboring occupied subspaces remain numerically well conditioned for the lattice Chern calculation.

The structured perturbation sweep is evaluated at 13 amplitudes on 32^2, 48^2, and 64^2 grids. The main-text result uses the finest resolution. At 64^2, the first observed change occurs at $\lambda = 0.08$ eV,

$$C : -2 \longrightarrow -1,$$

with a minimum direct gap of 15.225 meV. Further increases in λ produce successive Chern sectors, with minima of the direct gap appearing near the changes in the integer invariant. The joint evolution of the gap and Chern number is important: the transitions coincide with the spectral mechanism expected for changes of a well-defined occupied bundle, rather than appearing as isolated jumps of a poorly conditioned numerical estimator.

Figure 7 collects the corresponding convergence and conditioning diagnostics. These checks support the interpretation of the $NaRuO_2$ experiment as a structured sensitivity example. They do not change its scope: the perturbation direction is deliberately aligned with a selected band-edge channel and is not intended to model the frequency of naturally occurring neural-Hamiltonian residuals.

C.4 Production-model gap and Wilson-loop convergence

The released-model analysis involves time-reversal-protected topology, for which an integer or Z_2 label should be interpreted only after checking both spectral validity and numerical convergence. We therefore distinguish stable target–prediction agreement from cases in which the predicted gap becomes too small for a reliable Wilson-loop interpretation.

The four complete released reference–prediction pairs exhibit two qualitatively different regimes. ZrSiPt and $NaCaTaTiO_6$ retain large spectral gaps and stable trivial agreement. Ta_3AlC_2 provides the stronger nontrivial control: all six time-reversal-invariant planes agree between reference and prediction, including two planes with nontrivial target labels. Their representative reference/prediction gaps are $70.807/68.417$ meV. These examples show that the numerical evaluation can preserve both trivial and nontrivial labels when the occupied subspace remains well conditioned.

Table 3: **Released neural-Hamiltonian audit.** Three examples are stable agreements; the small-gap $Ca_3Si_4Ir_4$ prediction fails the validity gate and is not counted as a valid topology mismatch. Material-level band and gap values use each material's frozen primary or released-path setting and are not a common-mesh cross-material benchmark.

Material	Occ. rank	Band MAE (meV)	Gap behavior	Topology result	Interpretation
ZrSiPt	32	30.0	1795 → 1758 meV	$C = 0$ agreement on every evaluated grid and slice	stable trivial agreement
$Ca_3Si_4Ir_4$	106	17.9	1.55 → 0.00062 meV (collapse)	raw Z_2 label is invalid after the predicted gap collapses	gap-collapse / invalid topology regime
$NaCaTaTiO_6$	160	14.7	2308 → 2277 meV	six target and prediction planes agree trivially	stable trivial agreement
Ta_3AlC_2	100	28.4	70.8 → 68.4 meV	all six planes agree; two nontrivial planes preserved	stable nontrivial agreement

$Ca_3Si_4Ir_4$ behaves differently. In the evaluated setting, the reference direct gap is

$$1.546269\ \mathrm{meV},$$

whereas the learned Hamiltonian reduces the representative gap to

$$6.17 \times 10^{-4}\ \mathrm{meV}.$$

Wilson-loop calculations are evaluated at increasing loop resolutions

$$N_{\mathrm{loop}} \in \{16, 24, 32\}.$$

The raw discrete Z_2 output is not resolution-stable in this regime, and the predicted gap remains far below the preregistered 0.25 meV numerical-validity guardrail. The cutoff is an operational, conservative heuristic for the frozen grids, not a physical phase boundary or a universal energy threshold. Across the six evaluated time-reversal-invariant planes, the prediction fails the gap-validity condition rather than yielding a well-resolved alternative topological phase.

This distinction is important for interpreting numerical robustness. A stable invariant difference between two gapped, converged occupied bundles is evidence of a valid topological mismatch. A changing Wilson-loop label accompanied by a nearly collapsed occupied/unoccupied gap is instead evidence that the invariant has entered an ill-conditioned regime. We therefore do not count the $Ca_3Si_4Ir_4$ raw label difference as a topological mismatch.

For comparison, the released-model outcomes are summarized in Table 3, while material-specific Wilson flows, gap-resolution curves, and six-plane results are provided in Appendix G. The combined evidence supports a two-stage interpretation of production topology evaluation:

$$\text{spectral/numerical validity} \quad \longrightarrow \quad \text{topological fidelity}.$$

Only after the first condition is satisfied is comparison of discrete topological labels scientifically meaningful.

D Controlled-Model Ablations

The controlled models isolate how architecture, sampling geometry, and regularization affect the global extension of a Hamiltonian from finite observations. These experiments complement the main-text constructions in two ways. First, they show that the topology failures are not specific to a single architecture or sampling realization. Second, they clarify which interventions improve empirical recovery without providing a mathematical certificate. Throughout the two-band results below, Chern numbers are reported using the occupied-projector convention $C = \mathrm{Ch}_1(P_{\mathrm{occ}})$.

Table 4: **Architecture and sampling ablation for localized two-band Hamiltonians.** Each row aggregates 3 localization radii and 5 initializations per radius. Mean and maximum infidelity are evaluated on a 96^2 grid. Region-centered loop observations use the known texture location and are included as a diagnostic comparison.

Architecture	Sampling	Correct topology	Median mean f	Median max f
Periodic MLP	IID	$1/15$	0.001970	1.000000
Periodic MLP	Region-centered loops	$15/15$	0.000070	0.004652
Fourier features	IID	$3/15$	0.001852	0.994992
Fourier features	Region-centered loops	$8/15$	0.007129	0.935580

D.1 Two-band architecture and sampling results

The learned two-band study compares two periodic function classes under the same finite-observation budget. The first is a periodic MLP using the first harmonic of each momentum coordinate; the second uses Fourier features with harmonics 1–12. Both models have width 64, three hidden layers, and are trained on a 12×12 base grid together with 81 protocol-specific additional observations. We consider localized textures with

$$R \in \{0.22, 0.30, 0.40\},$$

five independent initializations per radius, and evaluate the learned fields on a 96×96 momentum grid.

The two additional-observation protocols use the same budget. IID sampling places the 81 momenta independently, whereas the region-centered protocol places one point at the known texture center together with five discretized rings of 16 points each. The latter therefore has access to the location of the topological texture and should be interpreted as a diagnostic comparison rather than a deployable acquisition strategy.

Table 4 reports the aggregate results. The sampling protocol has a strong effect for the periodic MLP: correct topology increases from $1/15$ under IID observations to $15/15$ under region-centered loop observations. The same intervention is substantially less reliable for the Fourier-feature model, improving recovery from $3/15$ to only $8/15$.

The architecture dependence is most evident within the region-centered setting. The periodic MLP almost eliminates the maximum subspace discrepancy, whereas the Fourier model can fit the observed region while leaving an order-one discrepancy elsewhere. Its topology recovery also degrades as the localized texture grows: the Fourier model is correct in $5/5$ cases at $R = 0.22$, $3/5$ cases at $R = 0.30$, and $0/5$ cases at $R = 0.40$.

The failures are not restricted to reverting to a trivial extension. In the historical field-degree convention, the 15 Fourier loop models produced degrees -1 in eight cases, 0 in four, $+2$ in two, and $+3$ in one. Converting to the occupied-projector convention used in the paper gives

$$C_{\mathrm{occ}} \in \{+1, 0, -2, -3\},$$

with counts $8, 4, 2, 1$, respectively. The -2 and -3 outcomes show that finite supervision can permit *additional* global winding rather than simply missing the target texture.

This result motivates the matched example in Fig. 2: architecture affects how finite constraints are extended through unsampled regions, even when the observations themselves are held fixed.

D.2 QWZ and trivial controls

The localized construction deliberately concentrates its topological texture, so a natural alternative explanation is that the learned models fail simply because nontrivial topology is intrinsically difficult to fit. We test this interpretation using two controls: a QWZ-type target with topological structure distributed over momentum space and a topologically trivial target.

The control outcomes are summarized in Table 5. All six trivial models recover the correct topology. For the QWZ targets, the periodic MLP succeeds in all three cases, whereas the Fourier-feature model succeeds in only one of three.

Table 5: **Two-band topology controls.** The QWZ target distributes its topological structure across the Brillouin zone, whereas the trivial control has no nontrivial winding. Perfect performance on the trivial controls rules out a generic failure of the topology computation; the imperfect Fourier QWZ result shows that localization alone does not explain all learned failures.

Target / architecture	Correct	Total
QWZ / periodic MLP	3	3
QWZ / Fourier features	1	3
Trivial controls	6	6

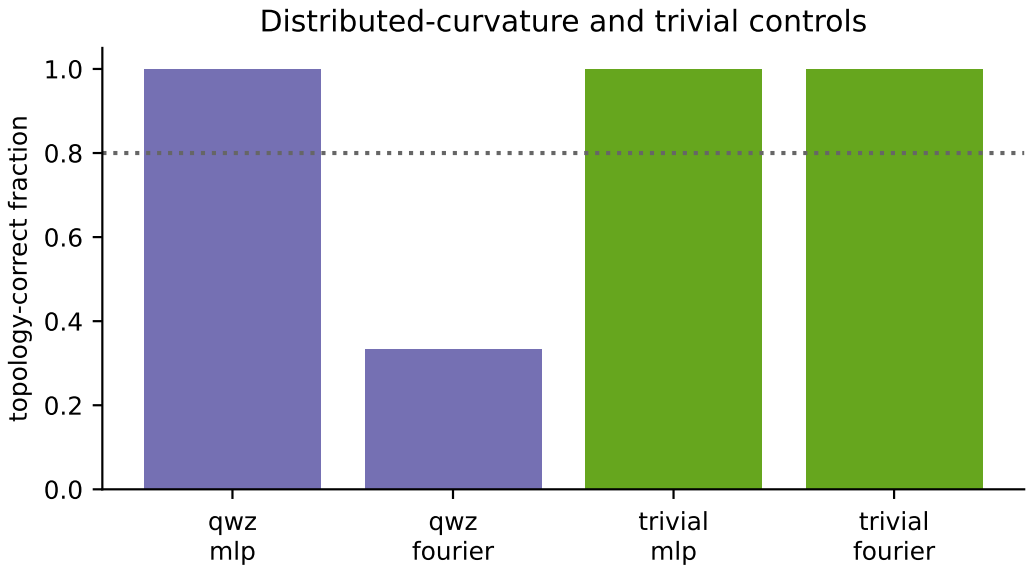


Figure 8: **Distributed-topology and trivial controls for the learned two-band models.** The periodic MLP recovers all three QWZ targets, while the Fourier-feature model recovers one of three; all six trivial controls are correct. The results show that the localized blind spot is not equivalent to a generic inability to represent topology, while also demonstrating that architecture-dependent failures are not unique to sharply localized textures.

These controls constrain the mechanism in both directions. The trivial results show that neither architecture is generically unable to produce a stable Chern label. Conversely, the imperfect Fourier performance on the QWZ targets shows that localized support is not the sole source of difficulty. Architecture, optimization, and interpolation bias can affect the selected global extension even when the target topology is spatially distributed.

For completeness, the corresponding control visualization is shown in Fig. 8.

D.3 Dirichlet regularization

The extra winding exhibited by some Fourier-feature models suggests that an inductive bias toward smoother global extensions might improve empirical topology recovery. We test this hypothesis by adding a Dirichlet-energy penalty on an unlabeled 24×24 collocation grid. For the predicted unit Bloch field $n_\theta(k)$, the additional objective takes the form

$$\mathcal{L}_{\mathrm{Dir}} = \lambda_{\mathrm{D}} \int_X \|\nabla n_\theta(k)\|_2^2 \, dk, \qquad \lambda_{\mathrm{D}} = 10^{-4}. \tag{33}$$

The comparison uses the same Fourier architecture and region-centered observations as above, with five independent initializations at each of the three localization radii.

Table 6: **Dirichlet-regularization ablation for Fourier-feature models with region-centered observations.** Regularization reduces geometric complexity and improves aggregate topology recovery, but one localized case remains incorrect and the discrete label improves in only 20% of paired comparisons.

Variant	Correct topology	Median mean f	Median Dirichlet energy
Unregularized	$12/15$	0.004975	61.81
Dirichlet regularized	$14/15$	0.000504	19.73

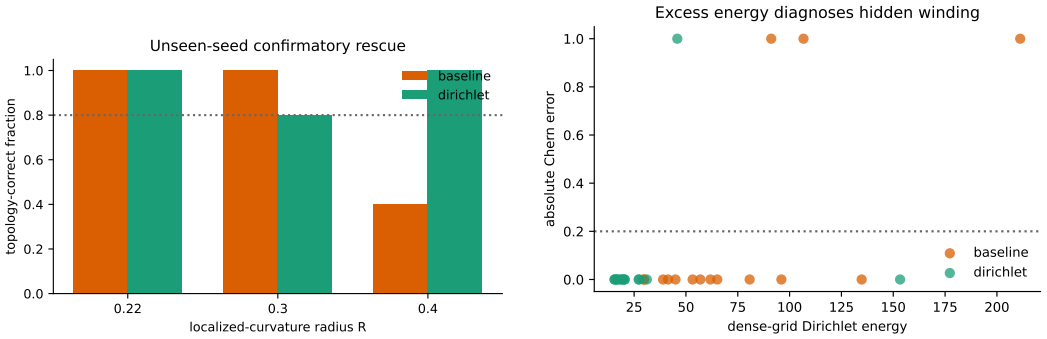


Figure 9: **Effect of Dirichlet regularization on topology recovery and geometric complexity.** Regularization increases localized topology recovery from $12/15$ to $14/15$ and reduces median dense Dirichlet energy from 61.81 to 19.73. The improvement supports smoothness as a useful inductive bias against unnecessary winding in this model class, but the remaining failure and heterogeneous paired response show that the penalty does not select the correct topological sector deterministically.

Regularization improves several aggregate quantities. Correct localized topology increases from $12/15$ to $14/15$, median mean infidelity decreases from 4.975×10^{-3} to 5.04×10^{-4}, and median dense Dirichlet energy decreases from 61.81 to 19.73, or approximately 31.9% of the unregularized value. Table 6 summarizes the comparison.

The improvement is not uniform across paired cases. Only 20% of paired localized fits change from an incorrect to a correct topology label, because the unregularized comparison is already correct in most cases. By radius, the regularized models recover the correct topology in $5/5$, $4/5$, and $5/5$ cases for $R = 0.22, 0.30$, and 0.40, respectively. On additional controls, QWZ recovery improves from $3/5$ to $5/5$, while all three regularized trivial controls remain correct.

Figure 9 shows the corresponding recovery and energy changes.

The distinction between this empirical smoothness effect and the analytical certificate in Sec. 4.4 is important. The implemented penalty controls an integrated squared gradient of the *prediction*. The sparse certificate instead assumes an independently established global Lipschitz bound on either the Hamiltonian error E or the subspace infidelity f. In two dimensions, an L^2 gradient penalty does not provide the required L^∞ derivative bound. Dirichlet regularization is therefore a useful inductive bias in these experiments, not a topology certificate.

D.4 Four-band metric-family comparison

The main text emphasizes the inversion in band MAE between the two learned four-band families. To determine whether that conclusion depends on a single selectively chosen error metric, we compare the same models using Hamiltonian, spectral, and occupied-subspace quantities.

The learned four-band Hamiltonian predicts 16 Hermitian coefficients with a periodic MLP. The occupied rank is $r = 2$, and the localized comparison contains 15 IID-trained and 15 region-centered-loop models over the same three radii and observation budget. The comparison holds fixed the four-layer, width-96 periodic MLP, optimizer, learning rate, 2200 training steps, 12×12 base grid,

Table 7: **Metric-family comparison for learned four-band Hamiltonians.** Values are medians across 15 localized models per sampling protocol. Energy quantities are reported in the native units of the controlled model. For these frozen families, the lower band MAE of the IID cohort gives the opposite ordering from topological correctness, while Hamiltonian and occupied-subspace metrics favor the topologically correct region-centered cohort; no universal ordering beyond these families is asserted.

Sampling	Correct topology	Mean f	Max f	H RMSE	Band MAE	Direct gap
IID	$0/15$	0.001166	0.496735	0.047805	**0.002813**	1.7357
Region-centered loops	$15/15$	0.0000114	0.001122	0.009441	0.006211	1.8489

seeds, and total label count; only the locations of the 81 additional observations differ. Both families remain gapped.

Table 7 reports the median continuous metrics. The result is not that every conventional metric ranks the families incorrectly. Hamiltonian RMSE and both mean and maximum occupied-subspace infidelity strongly favor the region-centered models, consistent with their $15/15$ topology recovery. Band MAE behaves differently: IID models have the smaller median band error, 0.002813 versus 0.006211, while all 15 IID models have the wrong occupied Chern number and all 15 region-centered models are correct.

This ablation makes two points. First, the main band-MAE inversion is not a consequence of gap collapse: both model families retain substantial direct gaps. Second, different continuous metrics constrain different parts of the eigendecomposition. Band MAE measures eigenvalue fidelity and can therefore prefer a model whose occupied eigenspaces form the wrong bundle; projector metrics directly probe those eigenspaces and are substantially more informative in this controlled setting.

At the same time, low *average* projector error is still not a universal certificate. The exact rank-two construction in Sec. 3 has mean subspace infidelity 2.80×10^{-4} while retaining a different Chern class, and Proposition 2 shows that the relevant certificate requires uniform rather than integrated projector control. Thus the metric-family comparison should be read as evidence about empirical model ranking in this controlled family, not as replacing the uniform certificate with a different average loss.

E Alternative Explanations and Independent Checks

The preceding experiments establish several qualitatively different failure modes: localized topological disagreement in controlled models, structured sensitivity in a multiorbital Wannier Hamiltonian, and loss of numerical validity in a released learned Hamiltonian. This section examines alternative explanations that could otherwise weaken those interpretations. We ask whether hidden winding is merely a missed target texture, whether the $NaRuO_2$ transition is reproduced by generic perturbations of comparable matrix norm, whether production-model agreement survives a complete time-reversal-plane analysis, and whether the Wilson-loop implementation agrees with an independent invariant construction.

E.1 Hidden-winding examples across settings

A simple explanation for the two-band failures would be that finite observations merely fail to sample the localized topological region. The learned results show that this explanation is incomplete. Even when the additional observations are explicitly centered on the known texture, different architectures can extend the same finite constraints into different global topological sectors.

The matched example in Fig. 2 makes this distinction particularly clear. The periodic MLP and Fourier-feature model receive the same target, the same localization radius $R = 0.40$, the same initialization seed, and the same 81 region-centered observations. The periodic MLP recovers the target occupied Chern number, whereas the Fourier-feature model develops additional winding away from the observed region and realizes

$$C_{\mathrm{occ}} = -3$$

Table 8: **Global extensions produced by Fourier-feature models under region-centered observations.** All 15 models observe the localized target region. Eight recover the target occupied Chern number $+1$; the remaining seven realize either a trivial extension or additional winding. The counts show that failure is not restricted to omitting the target texture.

Occupied Chern number	Number of models	Interpretation
$+1$	8	target topology
0	4	trivial extension
-2	2	additional winding
-3	1	additional winding

instead of the target value

$$C_{\mathrm{occ}} = +1.$$

Thus the failure cannot be attributed solely to absence of observations near the target texture.

The aggregate results show that this behavior is not confined to the single visualized example. Under region-centered observations, the periodic MLP recovers the target topology in all 15 localized cases, whereas the Fourier-feature model succeeds in only $8/15$. In the historical Bloch-field-degree convention, the 15 Fourier models realize degrees -1, 0, $+2$, and $+3$ with counts 8, 4, 2, and 1, respectively. Using the occupied-projector convention adopted in the paper,

$$C_{\mathrm{occ}} = -\deg(n),$$

the same outcomes are

$$+1, \quad 0, \quad -2, \quad -3.$$

Table 8 gives the corresponding counts.

The same architecture dependence is visible outside this particular localized setting. With IID observations, the periodic MLP and Fourier-feature models recover the correct localized topology in only $1/15$ and $3/15$ cases, respectively, while the QWZ control in Appendix D.2 is recovered by all periodic-MLP models but only one of three Fourier-feature models. These results rule out a single-mechanism explanation in which failures arise only because a compact Berry-curvature feature was missed. Finite observations leave freedom in the global extension, and the realized extension depends on the model and optimization bias.

E.2 $NaRuO_2$ Frobenius-matched random controls

The $NaRuO_2$ experiment uses a deliberately structured perturbation,

$$D = |v\rangle\langle v| - |c\rangle\langle c|,$$

chosen from a selected valence–conduction band-edge pair. The first Chern transition along this direction occurs at $\lambda = 0.08$ eV. A natural alternative explanation is that perturbations of similar global matrix magnitude would generically produce the same transition.

To test this possibility, we compare the structured direction with random traceless Hermitian perturbations at

$$\lambda \in \{0.08, 0.11\}\ \mathrm{eV}.$$

For each amplitude, 20 random directions are normalized to have the same Frobenius norm as D, yielding 40 comparison perturbations in total. The structured and random directions therefore satisfy

$$\|D\|_F = \|D_{\mathrm{rand}}\|_F = \sqrt{2}.$$

All 40 random perturbations retain the reference Chern number

$$C = -2,$$

whereas the structured direction changes the Chern number at both tested amplitudes. The random-control minimum direct gaps remain at least 136.7 meV, substantially above the 15.225 meV minimum reached by the structured perturbation at its first transition. Table 9 reports the norm and outcome comparison.

Table 9: **Structured and Frobenius-matched perturbations of $NaRuO_2$.** Twenty random traceless Hermitian directions are evaluated at each of $\lambda = 0.08$ and 0.11 eV. All 40 retain the reference Chern number, while the structured band-edge perturbation changes topology. The operator norms are not matched, so the comparison does not isolate orientation alone.

Perturbation	$\|D\|_F$	$\|D\|_{\mathrm{op}}$	Evaluations	Chern changes
Structured	$\sqrt{2}$	1.000	2 amplitudes	2/2
Random	$\sqrt{2}$	0.442–0.493	40	0/40

Table 10: **Six-plane topology summary for Ta_3AlC_2.** All six time-reversal-invariant planes agree between the reference and released prediction. Two nontrivial target planes are preserved, providing a positive control in which the learned Hamiltonian retains nontrivial topological structure.

Plane class	Number of planes	Reference–prediction agreement	Valid
Trivial	4	4/4	yes
Nontrivial	2	2/2	yes
All planes	6	6/6	yes

The control is intentionally interpreted at the level it supports. Although the Frobenius norms are identical, the operator norms are not: the random directions have operator norms ranging approximately from 0.442 to 0.493, with median 0.472, whereas the structured direction has operator norm 1. Consequently, the experiment does not identify orientation as the sole cause of the difference. Rather, it establishes that the selected band-edge perturbation is substantially more damaging than the tested Frobenius-matched generic directions, with orientation and operator-norm concentration remaining coupled.

This distinction is also consistent with the conservative certificate in Sec. 4.3, which is formulated in operator norm. Frobenius matching is useful for comparing perturbations of equal total matrix energy, but it is not equivalent to matching the norm that appears in the uniform gap certificate.

E.3 Six-plane Ta_3AlC_2 analysis

The released Ta_3AlC_2 example provides an important positive control because its reference Hamiltonian contains nontrivial time-reversal-invariant planes. It also illustrates why evaluating a single conventional plane can give an incomplete picture of three-dimensional topological structure.

We evaluate all six time-reversal-invariant planes for both the reference and predicted Hamiltonians. The target and prediction agree on every plane. Four planes are trivial in both Hamiltonians, while two planes carry nontrivial target Z_2 labels and both are preserved by the prediction. Thus the production-model agreement is not limited to a large-gap trivial case.

This expanded analysis is scientifically relevant because a single-plane evaluation would not establish preservation of the material's full set of time-reversal-plane labels. In particular, a trivial plane can coexist with nontrivial topology on other symmetry planes. The six-plane result therefore provides a stronger production positive control than a single trivial slice: the released prediction preserves both trivial and nontrivial reference labels in the evaluated planes.

The result should nevertheless be interpreted as agreement for this released example rather than as a general certification of Uni-HamGNN. The complete material-specific numerical results are reported in Appendix G.

E.4 Independent Wilson-Z_2 and Fu–Kane parity comparison

Because several realistic-system conclusions depend on numerical time-reversal invariants, we additionally test whether the Wilson-loop implementation reproduces an invariant obtained by an independent construction. For inversion-symmetric time-reversal controls, the Fu–Kane parity criterion provides such a comparison: it determines the Z_2 index from parity eigenvalues at the time-reversal-invariant momenta rather than from Wilson-loop winding.

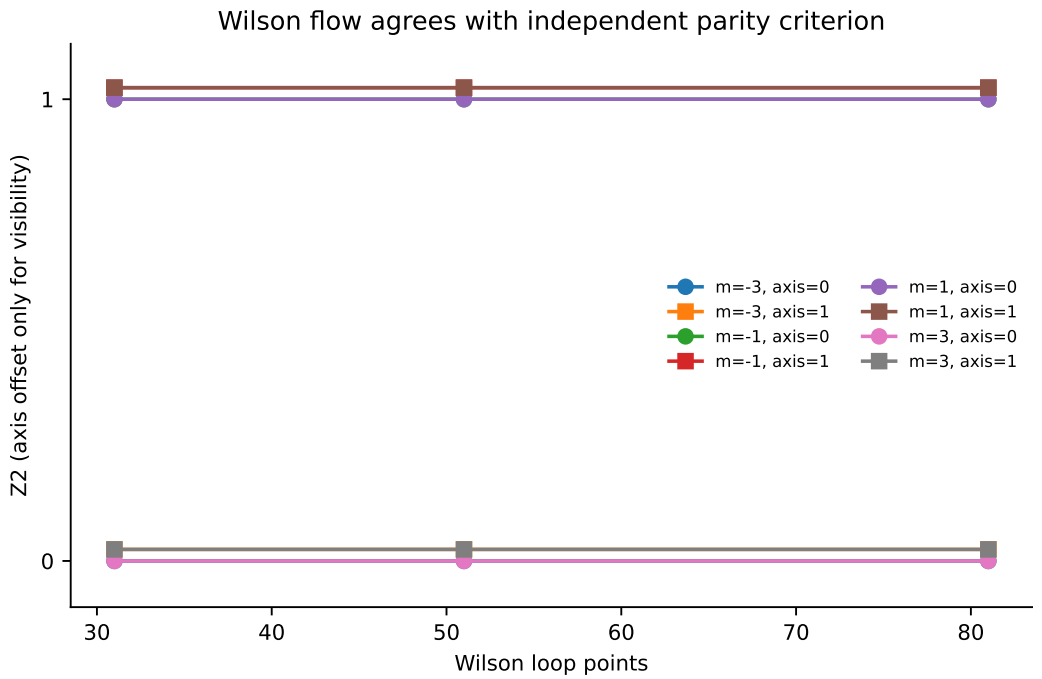


Figure 10: **Independent validation of the Wilson-loop Z_2 implementation.** Across 24 inversion-symmetric time-reversal control settings spanning multiple parameters, plane orientations, and resolutions, Wilson-loop and Fu–Kane parity calculations agree in all 24 cases. Endpoint Kramers error is zero and the minimum link singular value is at least 0.99487. This comparison validates the numerical invariant implementation in the tested regime; it does not by itself establish the correctness of any learned Hamiltonian.

We evaluate 24 control settings spanning multiple model parameters, plane orientations, and numerical resolutions. In every case, the Wilson-loop and Fu–Kane parity calculations return the same Z_2 label:

$$24/24 \quad \text{agreements.}$$

The endpoint Kramers-pairing error is zero in these controls, and the minimum singular value of the Wilson-link overlap matrices is at least

$$0.99487,$$

so the comparison is performed in a numerically well-conditioned regime.

Figure 10 summarizes the independent comparison.

This check addresses a different question from grid convergence. Refining a Wilson calculation can show that one implementation is numerically stable, but it does not exclude a systematic implementation error shared across resolutions. Agreement with the parity construction instead compares two different routes to the same invariant. Their complete agreement on the control suite provides evidence against an implementation-specific explanation for the Z_2 results used elsewhere in the paper.

The cross-check is deliberately limited to settings where the Fu–Kane criterion applies. It is therefore a validation of the topology machinery, not an alternative invariant for every production material or every Hamiltonian considered in this work.

F Sparse-Query Benchmark Details

This appendix expands the equal-budget finite-point comparison from Sec. 5. The benchmark is designed to answer a narrow empirical question: when every protocol observes the same pointwise

occupied-subspace discrepancy and receives the same number of momentum queries, does arranging those queries along loops provide a reproducible failure-detection advantage?

Before outcomes were inspected, the benchmark preregistration fixed $B = 64$ and loop-structured versus stratified-IID sampling as the primary budget and comparator, together with the maximum-score statistic and frozen failure labels. The benchmark contains 216 model–setting instances drawn from seven experimental families, with 75 instances satisfying the common failure criterion. Each sampling protocol is evaluated at four query budgets and over 16 independently generated query designs, yielding 55,296 pointwise query evaluations and 3,456 protocol–instance aggregates. The benchmark records maximum, 95th-percentile, and mean summaries of the observed subspace infidelity. The main-text comparison uses the maximum summary throughout.

F.1 Sampling protocols and query budgets

All protocols query the same pointwise quantity,

$$f(k) = 1 - \frac{1}{r} \operatorname{Tr}\left[P_*(k) P_\theta(k)\right], \tag{34}$$

and differ only in the geometry of the queried momenta. The four protocols are:

1. **Loop-structured sampling.** Query points are placed on discretized closed loops. Each loop contains 16 sampled momenta. Thus budgets $B = 16, 32, 64$, and 128 correspond to $1, 2, 4$, and 8 sampled loops, respectively. For $B > 16$, the union of these loops is not a single globally connected point set; "loop-structured" is therefore the more precise description of the protocol.
2. **IID sampling.** The B momenta are drawn independently over the Brillouin zone.
3. **Stratified IID sampling.** The same budget is distributed across momentum-space strata before sampling, encouraging broader spatial coverage while retaining pointwise, non-path-based observations.
4. **Random paths.** The query points are arranged along randomly generated finite paths. This provides a second spatially correlated sampling baseline without imposing closed-loop geometry.

Every protocol is evaluated at

$$B \in \{16, 32, 64, 128\}.$$

For each protocol and budget, 16 independently generated query designs are applied to the same 216 benchmark instances. This pairing ensures that protocol comparisons are not confounded by different model populations or different failure labels.

For the primary score, each instance receives

$$s_{\max}(Q) = \max_{q \in Q} f(q). \tag{35}$$

For each protocol–budget pair and model–setting instance, we first average $s_{\max}$ over the 16 paired query-design realizations. We then compute one AUROC across the same 216 instances using these seed-averaged scores and the frozen failure labels. Thus query-design seeds quantify acquisition randomness; they are not treated as additional independent model outcomes.

Table 11 reports the resulting maximum-score AUROCs.

Loop-structured sampling has the largest numerical AUROC at budgets 16 and 32, random paths are highest at 64, and IID sampling is highest at 128. The non-monotone ranking is the main reason we interpret the experiment as evidence against a stable geometry-specific advantage rather than as support for a universal best protocol.

F.2 Family-level uncertainty

The 216 benchmark rows are not statistically independent experimental units: they are grouped into seven model families that share architecture, target construction, or training setting. Treating all

Table 11: **Failure-detection AUROC for maximum observed subspace infidelity.** All four protocols receive exactly the same number B of point queries. The numerically best protocol changes with query budget, so no single sampling geometry dominates across the tested range.

Budget	Loop-structured	IID	Stratified IID	Random paths
16	0.7647	0.7351	0.7096	0.7068
32	0.8663	0.7788	0.8398	0.8519
64	0.8445	0.8421	0.8456	0.8778
128	0.8880	0.9008	0.8789	0.8624

rows as independent would therefore understate uncertainty. The seven families are the independent units for primary uncertainty, and we preserve the paired scores within each family when evaluating the primary comparison.

The primary contrast is loop-structured versus stratified IID sampling at $B = 64$. Their aggregate AUROCs are

$$0.844539 \qquad \text{and} \qquad 0.845579,$$

giving

$$\Delta\mathrm{AUROC} = \mathrm{AUROC}_{\mathrm{loop}} - \mathrm{AUROC}_{\mathrm{strat}} = -0.001040. \tag{36}$$

A paired family-cluster bootstrap gives an interval of approximately

$$[-0.038691,\ 0.041667].$$

An exact permutation over all 2^7 family-level protocol-label swaps gives

$$p = 0.5078125.$$

The interval therefore spans effects in both directions and is centered near zero.

Only four of the seven families contain both positive and negative outcomes, which is necessary to define a within-family AUROC. The remaining three single-class families are excluded from the within-family effect display rather than assigned an artificial zero effect. This distinction matters: absence of a definable within-family AUROC is not evidence of equivalence.

Figure 11 shows the family-level comparison.

The limited number of independent families is also why the paper avoids interpreting small row-level differences as precise population effects. The benchmark supports a bounded conclusion about the tested families and sampling procedures rather than a universal ordering of finite-point acquisition strategies.

F.3 Alternative score aggregations

The primary score in Eq. equation 35 emphasizes the largest observed discrepancy because the controlled failure modes can be spatially localized. To separate the choice of query geometry from the choice of within-query aggregation, the benchmark also records

$$s_{95}(Q) = \mathrm{quantile}_{0.95}\{f(q) : q \in Q\}, \tag{37}$$

$$s_{\mathrm{mean}}(Q) = \frac{1}{|Q|}\sum_{q\in Q} f(q). \tag{38}$$

These summaries probe different aspects of the observed error field. The maximum retains a single localized witness; the 95th percentile reduces sensitivity to an isolated extreme query; and the mean measures the average discrepancy over the sampled set. All three are computed from exactly the same queried momenta, so changing the aggregation does not alter the acquisition cost.

The maximum score is used for the main comparison because it is most directly aligned with the localized failure mechanism studied in the paper. This choice should not be confused with a theorem that the maximum must have higher AUROC than a mean or quantile on every model distribution.

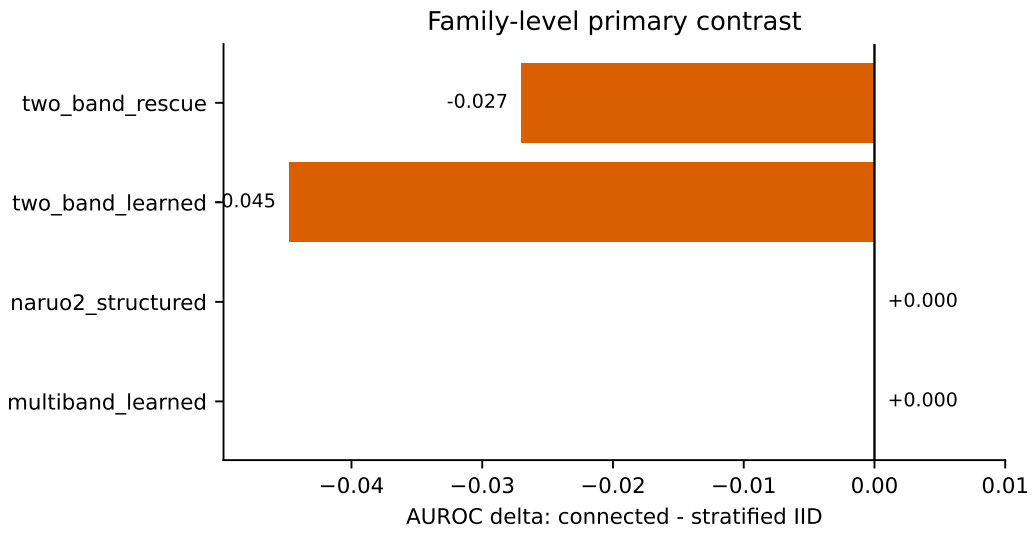


Figure 11: **Family-level effects for the primary equal-budget comparison.** The displayed effects compare loop-structured and stratified-IID sampling at $B = 64$. Four benchmark families contain both failure and non-failure instances and therefore admit within-family AUROC; three single-class families are omitted from the within-family display. The aggregate contrast is -0.00104, with family-cluster interval $[-0.03869, 0.04167]$ and exact family-level permutation $p = 0.5078$. The benchmark therefore provides no stable family-level evidence for a loop-specific advantage at the primary budget.

Proposition 2 concerns a sufficient uniform condition; it does not imply a universal statistical ranking between risk scores.

The alternative summaries are retained in the benchmark outputs to make this distinction explicit. They are useful for assessing whether a protocol advantage depends on a particular aggregation rule, but they are not used to upgrade the main conclusion into a stronger claim. In particular, the equal-budget experiment is interpreted through the instability of the protocol ranking across budgets and the near-zero primary family-level contrast, rather than through selection of whichever score gives the largest separation.

This also distinguishes the present benchmark from the earlier dense-grid analysis. A previously used quantity described as a "connected-query" score was in fact largely the maximum over a dense validation mesh. That quantity can support the descriptive statement that a dense empirical maximum retains information lost by a mean on the controlled suite, but it does not measure the performance of a sparse loop protocol. The equal-budget benchmark reported here supersedes that interpretation.

F.4 Risk–coverage and cost analyses

AUROC evaluates ranking over all benchmark instances, but an auditing system may instead abstain on high-risk cases and report predictions only for a subset of instances. We therefore also examine risk–coverage behavior and average precision. Query cost is controlled directly through B, so every protocol at a fixed budget performs the same number of pointwise subspace evaluations.

At the primary budget $B = 64$, the benchmark summaries are shown in Table 12. Random paths have the lowest numerical area under the risk–coverage curve (AURC) in this particular comparison, while IID sampling attains the largest benchmark-defined usable coverage. The relative ordering is therefore metric-dependent even at a single budget.

Figure 12 shows the corresponding selective prediction curves.

Table 12: **Selective-detection summaries at the primary query budget $B = 64$.** AP denotes average precision and AURC the area under the benchmark risk–coverage curve. "Usable coverage" follows the benchmark's common selective-prediction criterion. These quantities provide complementary views of the same finite-point scores and do not yield a unique dominant sampling geometry.

Protocol	AP	AURC	Usable coverage
Loop-structured	0.627	0.1217	0.2778
IID	0.653	0.1212	0.4583
Stratified IID	0.659	0.1256	0.0833
Random paths	0.702	0.1101	0.4398

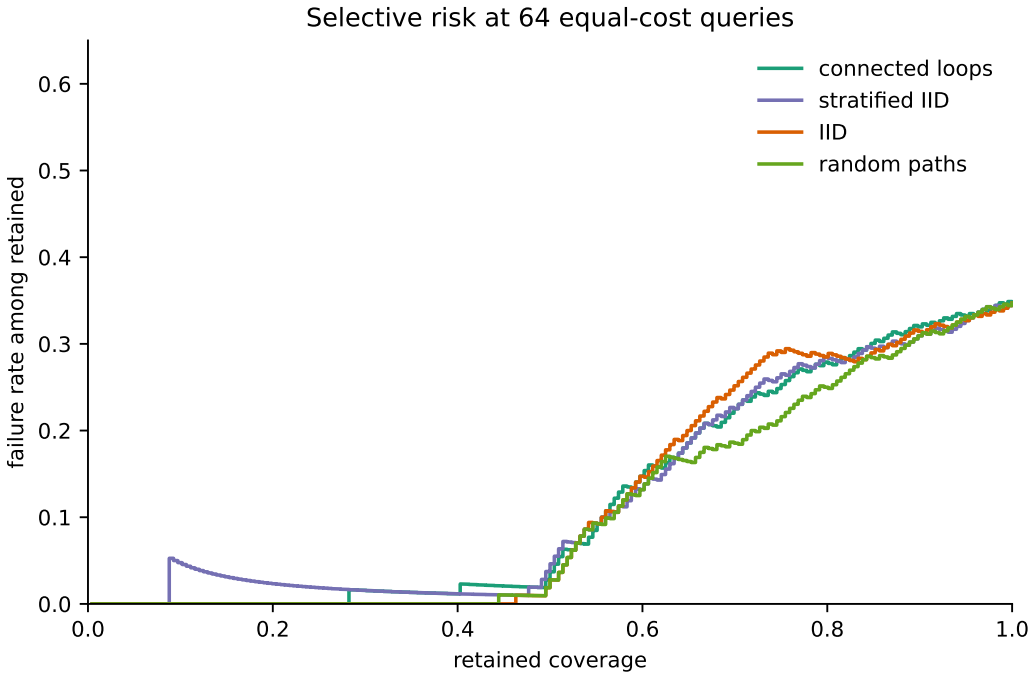


Figure 12: **Risk–coverage tradeoffs for equal-budget sparse-query protocols.** The curves evaluate selective use of the same pointwise occupied-subspace-infidelity scores used in the detection benchmark. At $B = 64$, no protocol dominates simultaneously in AUROC, average precision, AURC, and usable coverage. Random paths give the lowest numerical AURC, whereas IID sampling gives the largest usable coverage under the benchmark criterion. These tradeoffs reinforce the conclusion that loop structure alone does not define a uniformly superior finite-point auditing strategy.

The query budget provides the cleanest cost axis in this experiment. Moving from $B = 16$ to $B = 128$ increases the number of evaluated momenta by a factor of eight for every protocol, while holding the underlying set of model–setting instances fixed. The AUROC ranking nevertheless changes with budget rather than converging monotonically toward one preferred geometry. Thus additional queries improve the information available to all protocols without revealing a persistent loop-specific advantage.

Finally, the finite-query benchmark should not be interpreted as an empirical measurement of the sparse certificates in Appendix B.8. Those certificates involve

$$\max_{q \in Q} f(q) + L_f h(Q),$$

where L_f must be an independently justified global regularity bound. The benchmark observes $f(q)$ but does not certify L_f. Likewise, query geometry alone is not equivalent to the fill-distance term.

For example, approximate fill-distance calculations on the tested query plans do not reproduce the observed AUROC ordering across budgets. The experiment therefore evaluates finite-point failure detection, while the coverage bound remains a sufficient analytical route to certification rather than a calibrated empirical risk score.

G RELEASED NEURAL-HAMILTONIAN AUDITS

The realistic evaluation in Sec. 6 uses the complete reference–prediction Hamiltonian pairs available with the released Uni-HamGNN examples. This appendix gives the material-level results behind Table 3. The four systems span occupied ranks from 32 to 160 and include both trivial and nontrivial time-reversal-plane structure. We report spectral accuracy together with gap and topological validity because agreement of raw discrete invariants is meaningful only when the occupied subspace remains well defined and the numerical invariant is stable.

The reported band and gap quantities follow the evaluation setting associated with each released example; they are not intended as a common-mesh cross-material benchmark. Across the four examples, three predictions remain in valid regimes and preserve the evaluated reference topology, while $Ca_3Si_4Ir_4$ exhibits a severe gap collapse and is classified as a validity failure rather than a valid topological mismatch.

G.1 ZRSIPT

ZrSiPt provides a large-gap reference case for testing whether the released model can preserve a simple topological regime. The spin–orbit-coupled Hamiltonian contains 116 bands in the released representation, with occupied rank

$$r = 32.$$

The prediction has a band MAE of

$$30.033 \text{ meV}.$$

The representative direct gap changes only modestly,

$$1794.72 \text{ meV} \longrightarrow 1758.26 \text{ meV},$$

so the prediction remains far from an occupied–unoccupied gap closing.

The evaluated reference and prediction both remain topologically trivial across the tested discretizations and slices. The mean occupied-subspace infidelity is approximately 2.29×10^{-4}, or 0.0229%. Together with the large retained gap, this provides a production-scale example in which conventional spectral fidelity and topological fidelity agree.

The role of ZrSiPt is therefore primarily diagnostic: it shows that the audit does not systematically turn moderate band error into a topological failure and that the numerical machinery returns stable agreement in a large-gap setting.

G.2 $CA_3SI_4IR_4$

$Ca_3Si_4Ir_4$ probes the opposite regime. The released Hamiltonian contains 396 spin–orbit-coupled bands, with occupied rank

$$r = 106,$$

and the prediction has a band MAE of

$$17.890 \text{ meV}.$$

This error scale is smaller than that of ZrSiPt, but the relevant spectral margin is also dramatically smaller.

In the evaluated reference setting, the representative direct gap is

$$\Delta_{\text{ref}} = 1.546269 \text{ meV}.$$

The learned Hamiltonian reduces this to approximately

$$\Delta_{\text{pred}} = 6.17 \times 10^{-4} \text{ meV},$$

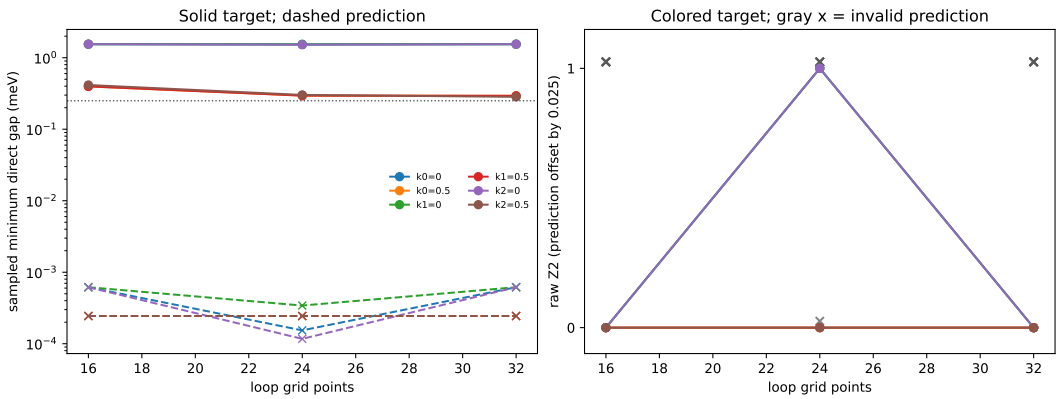


Figure 13: **Gap and Wilson-loop convergence for $Ca_3Si_4Ir_4$.** The reference Hamiltonian has a representative direct gap of 1.546269 meV, whereas the released prediction reduces the gap to approximately 6.17×10^{-4} meV. Wilson-loop calculations at increasing numerical resolution can produce raw target–prediction Z_2 differences, but the predicted gap remains far below the preregistered 0.25 meV numerical-validity guardrail. This cutoff is operational rather than a physical phase boundary. We therefore classify this example as a gap-validity failure rather than a valid topological mismatch.

a reduction of more than three orders of magnitude. This value lies well below the preregistered 0.25 meV numerical-validity guardrail used for the Wilson-loop analysis. This operational cutoff is a conservative heuristic for the frozen discretizations, not a physical phase boundary.

The resulting topology calculation illustrates why numerical validity must precede comparison of discrete labels. Wilson-loop evaluations were performed at increasing loop resolutions

$$N_{\text{loop}} \in \{16, 24, 32\}.$$

Raw target and prediction Z_2 labels can differ, but the prediction fails the gap-validity requirement on the evaluated time-reversal-invariant planes. The discrete disagreement is therefore not interpreted as evidence that two well-defined insulating phases have different topology.

Figure 13 shows the material-specific gap and Wilson-loop convergence behavior.

This example also demonstrates why band MAE cannot be interpreted without the associated spectral margin. A 17.890 meV band error is not large in isolation compared with the other released examples, but it is enormous relative to a 1.5-meV-scale reference gap. The reliability issue is therefore not simply the absolute magnitude of the reported prediction error, but its interaction with the local occupied–unoccupied separation.

G.3 $NaCaTaTiO_6$

$NaCaTaTiO_6$ provides a second large-gap comparison and extends the released audit to a substantially larger occupied subspace. The released spin–orbit-coupled Hamiltonian contains 584 bands and has occupied rank

$$r = 160.$$

The prediction has a band MAE of

$$14.738 \text{ meV},$$

while the representative direct gap changes from

$$2307.56 \text{ meV} \longrightarrow 2277.35 \text{ meV}.$$

All six evaluated time-reversal-invariant planes are trivial for the reference Hamiltonian, and all six remain trivial for the prediction. The large occupied–unoccupied gap is preserved by a wide margin, so the agreement is not close to the numerical-validity boundary encountered for $Ca_3Si_4Ir_4$.

This material is useful because its occupied rank is much larger than that of the controlled two- and four-band examples. The result does not imply that high-rank learned Hamiltonians are automatically topologically reliable; rather, it provides an end-to-end example in which a high-dimensional predicted occupied subspace remains spectrally valid and agrees with the reference topology on every evaluated symmetry plane.

G.4 Ta_3AlC_2

Ta_3AlC_2 is the most informative positive control among the four released examples because the reference topology is not trivial on every evaluated plane. The released Hamiltonian contains 468 spin–orbit-coupled bands, with occupied rank

$$r = 100.$$

The prediction has a band MAE of

$$28.407\ \mathrm{meV}.$$

Its representative direct gap remains open,

$$70.807\ \mathrm{meV} \longrightarrow 68.417\ \mathrm{meV}.$$

We evaluate all six time-reversal-invariant planes. Reference and prediction agree on all six: four planes are trivial in both Hamiltonians, while two reference planes carry nontrivial Z_2 labels and both are preserved by the learned prediction. The complete six-plane summary is given in Table 10.

This result is stronger than agreement on a single conventional slice. A single trivial plane would not establish preservation of the nontrivial structure present elsewhere in the three-dimensional Brillouin zone. The six-plane analysis shows that the released model can preserve nontrivial topological information as well as trivial labels.

Taken together, the four released examples establish a deliberately bounded production-scale conclusion. Stable trivial agreement occurs for ZrSiPt and $NaCaTaTiO_6$; Ta_3AlC_2 preserves nontrivial reference planes; and $Ca_3Si_4Ir_4$ demonstrates that a learned Hamiltonian can instead enter a regime where the topological invariant is no longer numerically interpretable. The set is sufficient to demonstrate the practical distinction between topological fidelity and topological validity, but its size does not support an estimate of failure prevalence across learned Hamiltonians.

H JARVIS Stress Tests and Empirical Scope

The controlled constructions establish that small integrated error can coexist with a different topological class, while the $NaRuO_2$ experiment shows that a realistic Hamiltonian can be sensitive to a deliberately structured perturbation. Neither result determines how frequently such failures should be expected under more generic perturbations of realistic materials. We therefore use a broader set of JARVIS Hamiltonians (Choudhary et al., 2020) to separate two questions: whether a nominally small matrix-space error corresponds to a small physical spectral perturbation, and whether topological failures remain observable once the perturbations are instead calibrated to a genuinely small near-Fermi band error.

The two analyses lead to different conclusions. At fixed real-space Hamiltonian MAE, the tested perturbations can induce very large near-Fermi spectral errors and corresponding topology or validity failures. After recalibration to approximately 10 meV near-Fermi band MAE, however, the same material cohort and perturbation families produce no valid topological mismatch and no newly unsafe prediction. The latter result is an empirical boundary on the scope of the paper rather than evidence against the possibility results established earlier.

H.1 Material selection and screening

The material-scale analysis begins from a frozen stratified draw of 200 JARVIS Hamiltonians. Before any perturbation result, we excluded archives larger than 25 MB, structures with more than 20 atoms, entries with missing or absolute reported magnetic moment above 0.05, and the format-only smoke sample JVASP-816. Sampling without replacement used seed 20260923 and the reported

optB88vdW gap only for pre-result stratification; a recorded pre-result amendment deterministically filled a short lowest-gap stratum. Each material is then screened for a well-defined occupied subspace under the filling convention used in the analysis. Of the 200 systems, 168 have constant even filling across the evaluated Brillouin-zone sampling and form the primary screening cohort.

Each of these 168 materials is evaluated under three generic perturbation families:

1. **uniform rescaling**, which changes the overall Hamiltonian scale;
2. **range attenuation**, which modifies contributions according to their real-space range; and
3. **paired-shell random perturbations**, which introduce structured random changes in matched real-space shells.

These perturbations are intended as generic stress directions rather than as samples from the residual distribution of a particular learned Hamiltonian.

The primary screening stage therefore contains

$$168 \text{ materials} \times 3 \text{ perturbation mechanisms} = 504$$

material–perturbation observations.

A subset of 24 materials is subjected to the more expensive topological analysis on all six time-reversal-invariant planes. It was selected without topology results: 12 eligible systems have the smallest positive target direct gaps, and 12 controls come from the largest target-gap quartile, greedily matched by Wannier dimension and atom count. This gives

$$24 \times 6 = 144$$

plane evaluations for one perturbation mechanism in the fixed-matrix-error analysis, and later

$$24 \times 3 \times 6 = 432$$

plane evaluations in the band-matched analysis.

The distinction between screening and full topology evaluation is important. The 168-material stage establishes how a common matrix-space perturbation translates into spectral and gap distortion. The 24-material six-plane stage then determines whether those distortions correspond to valid topological changes, loss of validity, or stable agreement.

H.2 Fixed matrix-error stress

Let

$$\mathcal{M} = \{(R, a, b) : |H_{*,ab}(R)| > 10^{-12}\}.$$

The implemented real-space error is

$$\mathrm{MAE}_R = \frac{1}{|\mathcal{M}|} \sum_{(R,a,b)\in\mathcal{M}} |\delta H_{ab}(R)|, \tag{39}$$

where the absolute value is the complex modulus. Thus the denominator spans all nonzero target entries across translation blocks and orbital indices. We first hold this quantity fixed at

$$1.417 \text{ meV}$$

for all three perturbation families. If this matrix-space error were a comparable proxy for physically relevant prediction quality across mechanisms, one would expect the induced near-Fermi spectral errors to remain on a similar scale. They do not.

Across the 168 materials, the median near-Fermi band MAEs are

$$600.8 \text{ meV}, \qquad 556.5 \text{ meV}, \qquad 2262.0 \text{ meV}$$

for uniform rescaling, range attenuation, and paired-shell random perturbations, respectively. Thus a matrix-entry MAE of only 1.417 meV can correspond to spectral distortions larger by two to three orders of magnitude. This does not contradict eigenvalue perturbation bounds: the Bloch error is the

Table 13: **JARVIS screening at fixed real-space Hamiltonian MAE.** All perturbation mechanisms are calibrated to the same 1.417 meV matrix-entry MAE, but they induce very different near-Fermi spectral errors and gap-failure rates. A small matrix-space MAE therefore does not imply a comparably small perturbation of the electronic structure relevant to topology.

Perturbation	Median band MAE (meV)	Gap failures / 168
Uniform rescaling	600.8	0
Range attenuation	556.5	60
Paired-shell random	2262.0	85

coherent Fourier sum $\delta H(k) = \sum_R e^{2\pi i k\cdot R}\delta H(R)$, while Eq. equation 39 is an entrywise average rather than an operator-norm bound.

Uniform rescaling acts on $H_*(R) - E_F I\delta_{R0}$, so it preserves Bloch eigenvectors and their topology while changing eigenvalue scale. It is therefore a calibration/negative control at matched matrix MAE, not evidence about generic topological sensitivity.

The gap statistics show the same lack of calibration. Uniform rescaling produces no gap failures in the 168-material screen, whereas range attenuation produces 60 and paired-shell perturbations produce 85 gap failures (Table 13).

We additionally perform a six-plane Z_2 analysis for the 24-material paired-shell subset. Among the resulting 144 time-reversal-invariant planes, 52 are already invalid at the reference endpoint. Of the remaining 92 target-valid planes, 77 remain valid after perturbation and 15 become unsafe. Among the 77 planes valid at both endpoints, three exhibit different Z_2 labels.

Thus the fixed-matrix-error stress does produce valid topological mismatches and validity failures. These outcomes, however, occur in a regime where the associated near-Fermi band distortion is hundreds to thousands of meV. They therefore cannot be interpreted as evidence that topology commonly fails when the *spectral* prediction itself is accurate.

This calibration issue is the main scientific role of the fixed-error experiment. Real-space Hamiltonian MAE and near-Fermi spectral error measure different objects, and a numerically small value of the former does not guarantee that the occupied subspace has undergone a small perturbation.

H.3 Band-matched 10-meV analysis

After the registered fixed-matrix-error results were inspected, but before computing this follow-up, a post-unblinding amendment froze an exploratory band-matched audit. To evaluate the same frozen material cohort at a small spectral-error scale, we recalibrate each perturbation mechanism to approximately

$$10\ \mathrm{meV}$$

near-Fermi band MAE. The achieved median errors are

$$10.0010\ \mathrm{meV}, \qquad 9.9998\ \mathrm{meV}, \qquad 9.9983\ \mathrm{meV}$$

for uniform rescaling, range attenuation, and paired-shell random perturbations, respectively. The three mechanisms are therefore closely matched in the spectral quantity used to define this follow-up. This audit is hypothesis-generating and does not alter the registered fixed-error decision; the perturbations remain controlled directions rather than residuals sampled from a production neural model.

The full evaluation contains

$$24\ \text{materials} \times 3\ \text{mechanisms} \times 6\ \text{planes} = 432$$

material–mechanism–plane evaluations.

For each perturbation mechanism, the same 52 planes are invalid already at the reference endpoint and are therefore not eligible for a topology comparison. The remaining 92 planes are valid at both endpoints. Every one of these 92 planes preserves its reference Z_2 label. Moreover, no target-valid plane becomes newly invalid. Table 14 records the eligible denominators and outcomes for each mechanism.

Table 14: **Post-unblinding exploratory band-matched JARVIS topology analysis.** Each perturbation family is calibrated to approximately 10 meV median near-Fermi band MAE. For every mechanism, all 92 planes valid at both endpoints preserve their Z_2 label; 52 planes are already invalid in the unperturbed reference Hamiltonian and are excluded from topology comparison. Across mechanisms this is 276 both-valid comparisons and 156 target-invalid exclusions among 432 material–mechanism–plane evaluations.

Perturbation	Median MAE	Both valid	Valid mismatch	Newly unsafe
Uniform rescaling	10.0010	92	0	0
Range attenuation	9.9998	92	0	0
Paired-shell random	9.9983	92	0	0

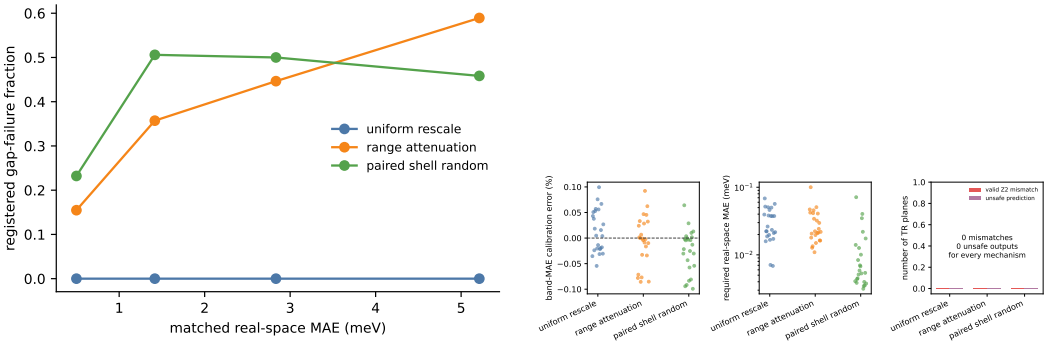


Figure 14: **JARVIS stress calibration separates matrix-space error from small spectral perturbations. Left:** holding the real-space Hamiltonian MAE fixed at 1.417 meV produces median near-Fermi band errors of 600.8, 556.5, and 2262.0 meV for uniform rescaling, range attenuation, and paired-shell random perturbations, respectively. Gap and topology failures in this regime therefore occur under large spectral distortion. **Right:** in the post-unblinding exploratory audit, after calibrating the three mechanisms to approximately 10 meV near-Fermi band MAE, all 92 planes per mechanism that are valid at both endpoints preserve their Z_2 label, and no target-valid plane becomes newly unsafe; 52 target-invalid planes per mechanism are excluded from valid comparison. The null result bounds the empirical claim for this material cohort and these generic perturbation families.

Across all three mechanisms, we therefore observe

$$\boxed{0 \text{ valid topology mismatches}}$$

and

$$\boxed{0 \text{ target-valid planes becoming newly unsafe.}}$$

Figure 14 places the fixed-matrix-error and band-matched analyses side by side.

The contrast between the two stress regimes is essential. The positive failures at fixed matrix MAE do not survive calibration to a small spectral error. Conversely, the absence of failures in the band-matched setting does not contradict the exact non-certification result of Proposition 1: that proposition is existential over unrestricted smooth gapped families, whereas this experiment samples a specific material cohort and three generic perturbation mechanisms.

H.4 Target-invalid planes and exclusion criteria

A Z_2 comparison is meaningful only when the reference plane itself supports a valid invariant. We therefore distinguish three possible outcomes for every target-valid plane:

1. **valid agreement:** both endpoints satisfy the numerical and spectral validity requirements and their Z_2 labels agree;

2. **valid mismatch:** both endpoints are valid but the labels differ; and
3. **newly unsafe:** the reference is valid but the perturbed Hamiltonian fails the validity conditions.

Planes that are already invalid at the reference endpoint are excluded from all topology-agreement denominators. They cannot provide evidence either for or against preservation under perturbation because there is no reliable reference invariant to preserve.

In the 24-material cohort, 52 of the 144 evaluated time-reversal-invariant planes are reference-invalid. This count is a property of the reference material set and is therefore the same for each of the three band-matched perturbation families. The eligible denominator for each mechanism is consequently

$$144 - 52 = 92$$

reference-valid planes.

This exclusion is particularly important when interpreting the zero-failure result. The statement

$$0/92$$

valid mismatches per mechanism means that every plane for which a valid reference–prediction comparison can be made agrees. It does *not* mean that all 144 planes are valid insulating topology calculations. Similarly, the observation of zero newly unsafe predictions means that none of the 92 initially valid planes loses validity at the 10-meV spectral-error scale.

The fixed-matrix-error experiment illustrates why this taxonomy is necessary. For the paired-shell stress, the same 52 reference-invalid planes are excluded, while 15 of the 92 target-valid planes become unsafe and three of the 77 planes valid at both endpoints exhibit a genuine Z_2 mismatch. Without separating these outcomes, gap destruction and topological disagreement would be conflated.

The resulting empirical scope is therefore precise. The fixed-entrywise stress shows that a nominally small Hamiltonian MAE can hide very large spectral distortion and corresponding reliability failures. The band-matched analysis, by contrast, provides no evidence of valid topology failure under the three tested generic perturbations at approximately 10 meV spectral error. Because these perturbations are not residuals drawn from a learned production model, the result characterizes this tested material–perturbation cohort rather than estimating a population-level failure rate for neural Hamiltonians.

I Implementation, Conventions, and Reproducibility

This section records the implementation choices needed to interpret and reproduce the controlled models, topological calculations, and released-model analyses. We emphasize conventions that affect scientific interpretation, including the occupied-projector Chern sign, the treatment of nonorthogonal orbital bases, and the distinction between a valid topological mismatch and a numerically invalid prediction. Detailed numerical results and robustness checks are reported in Appendices C–H.

I.1 Model architectures and optimization

Two-band models. The controlled two-band Hamiltonians have the form

$$H_\theta(k) = n_\theta(k) \cdot \sigma, \qquad \|n_\theta(k)\|_2 = 1,$$

so spectral flattening fixes the two eigenvalues to

$$\{-1, +1\}$$

and maintains a gap of 2 independently of the learned texture. We compare two periodic function classes. The periodic MLP uses first-harmonic periodic coordinate features, while the Fourier-feature model includes harmonics 1–12. Both use width 64, three hidden layers, and 1800 AdamW optimization steps.

The localized training set consists of a 12×12 base momentum grid together with 81 additional observations generated by the sampling protocol. The localized radii are

$$R \in \{0.22, 0.30, 0.40\},$$

with five seeds,

$$13,\ 47,\ 97,\ 131,\ 181.$$

The IID and region-centered protocols use the same number of additional observations. The latter consists of the known texture center and five 16-point rings and is used as a diagnostic oracle comparison rather than as a proposed acquisition procedure. The primary validation grid is 96×96.

Four-band models. The controlled multiband setting predicts a 4×4 Hermitian Hamiltonian through 16 real Hermitian-basis coefficients. The predictor is a periodic MLP with width 96, four hidden layers, and 2200 optimization steps. The occupied rank is $r = 2$. Training uses the same 12×12 base grid plus 81 protocol-specific observations as the two-band setting. The localized study contains three radii, two sampling protocols, and five seeds per condition, yielding 15 localized models per protocol. Additional QWZ and trivial controls are included separately.

Dirichlet-regularized models. The smoothness ablation in Appendix D.3 uses the Fourier two-band architecture and adds

$$\lambda_{\mathrm{D}} \int \|\nabla n_\theta(k)\|_2^2 \, dk, \qquad \lambda_{\mathrm{D}} = 10^{-4},$$

evaluated on an unlabeled 24×24 collocation grid. Optimization uses a 300-step warmup and 1800 total steps, with seeds

$$223,\ 269,\ 307,\ 353,\ 401.$$

This penalty is an integrated smoothness regularizer on the prediction; it is not an implementation of the global Lipschitz constants appearing in Appendix B.8.

Released neural Hamiltonians. The production-scale experiments use the released Uni-HamGNN checkpoint without retraining. The checkpoint contains $52{,}456{,}320$ parameters. The official target structures and released inference inputs are used for the four complete reference–prediction examples analyzed in Appendix G. The production analysis therefore evaluates the released model as supplied rather than fitting a new model to the materials studied here.

I.2 Chern-number convention

All paper-facing Chern numbers refer to the occupied spectral projector:

$$C = \mathrm{Ch}_1(P_{\mathrm{occ}}). \tag{40}$$

This convention requires care for the two-band construction. For

$$H(k) = n(k) \cdot \sigma,$$

the lower-band projector is

$$P_-(k) = \frac{I - n(k) \cdot \sigma}{2}.$$

With the orientation convention used in the numerical implementation,

$$\mathrm{Ch}_1(P_-) = -\deg(n). \tag{41}$$

Some underlying two-band numerical records store the solid-angle degree of the Bloch field n, rather than the occupied-projector Chern number. Consequently, a stored field degree -1 corresponds to occupied Chern number $+1$, and a field degree $+3$ corresponds to occupied Chern number -3.

All figures and claims in the paper are expressed using Eq. equation 40. Absolute Chern discrepancies are unchanged by this sign conversion. The distinction matters only when reporting signed invariants, such as the hidden-winding example in Fig. 2.

I.3 TOPOLOGY COMPUTATION

Class-A Chern number. For multiband occupied subspaces and the realistic class-A calculations, we compute the Chern number from the occupied projectors using a gauge-invariant lattice discretization (Fukui et al., 2005). Neighboring occupied frames are combined through determinant link variables, and the lattice Berry flux is summed over the periodic momentum grid. Internal rotations within the occupied subspace therefore do not affect the result.

The exact two-band visualization also admits a Bloch-vector solid-angle calculation. As discussed above, this quantity evaluates the degree of $n(k)$ and differs by a sign from the lower-band projector Chern number. Paper-facing comparisons use the occupied-projector convention throughout.

Time-reversal Z_2 invariant. For time-reversal-invariant planes, we compute Z_2 indices from Wilson-loop evolution of the occupied subspace (Yu et al., 2011; Soluyanov & Vanderbilt, 2011; Gresch et al., 2017). Production and JARVIS analyses evaluate all relevant time-reversal-invariant planes rather than inferring a three-dimensional conclusion from a single slice. Numerical validity is assessed jointly with the occupied–unoccupied gap, Wilson-loop resolution, endpoint time-reversal consistency, and overlap-link conditioning.

As an independent implementation check, the Wilson-loop calculation is compared with the Fu–Kane parity criterion (Fu & Kane, 2007) on 24 inversion-symmetric time-reversal controls. The two procedures agree in all 24 cases, with zero endpoint Kramers error and minimum link singular value at least 0.99487; see Appendix E.4.

Resolution checks. Chern and Wilson-loop labels used for scientific interpretation are checked for stability under the resolutions appropriate to each experiment. The controlled two-band study additionally varies grid origin, and the rank-two witness study increases the evaluation grid from 96^2 to 384^2. The $NaRuO_2$ structured sweep is evaluated on 32^2, 48^2, and 64^2 grids. For $Ca_3Si_4Ir_4$, Wilson loops are evaluated with 16, 24, and 32 loop points; the instability observed there is interpreted together with the accompanying gap collapse rather than as a resolved alternative topology.

I.4 OVERLAP MATRICES AND LÖWDIN ORTHOGONALIZATION

The controlled Hamiltonians are expressed directly in orthonormal bases. The released electronic-structure examples, however, can be represented as a generalized eigenvalue problem,

$$H(k)u = \varepsilon S(k)u, \tag{42}$$

with overlap matrix $S(k)$.

For each released reference–prediction pair, the target and prediction are evaluated using the same overlap representation. Before applying basis-dependent matrix norms or constructing ordinary orthogonal spectral projectors, we use the symmetric Löwdin transformation

$$\widetilde{H}(k) = S(k)^{-1/2} H(k) S(k)^{-1/2}. \tag{43}$$

Thus reference and prediction are compared in the same orthonormalized orbital frame. This is the representation assumed by the Hamiltonian-space certificate in Sec. 4.3.

The same convention is important for time-reversal comparisons. The equivariant statement in Appendix B.7 assumes a common time-reversal representation after orthogonalization. Numerical spot checks on the four released examples give relative time-reversal residuals of approximately

$$2.7 \times 10^{-13}\text{–}8.5 \times 10^{-13}$$

for the predictions and

$$1.2 \times 10^{-8}\text{–}8.6 \times 10^{-7}$$

for the corresponding reference Hamiltonians at the tested generic momenta. These values support numerical compatibility with the common time-reversal representation used in the calculations; they are not a proof of exact equivariance over the full Brillouin zone.

I.5 Failure and validity definitions

We distinguish topological fidelity from topological validity throughout the paper. For a reference/prediction pair of occupied rank r, an invariant comparison is interpreted only when the corresponding occupied subspaces satisfy the spectral and numerical requirements of the calculation.

We use the following three-way classification.

1. **Valid agreement.** Both endpoints support a well-defined and numerically resolved invariant, and the reference and prediction have the same topological label.
2. **Valid topological mismatch.** Both endpoints remain valid, but their resolved topological labels differ.
3. **Invalid or unsafe prediction.** The reference is valid, but the prediction fails a prerequisite for interpreting its invariant, such as loss of the occupied–unoccupied gap or failure of the numerical convergence criteria.

A plane that is already invalid at the reference endpoint is excluded from topological-agreement denominators. This convention is used explicitly in the JARVIS analysis of Appendix H.4.

The validity criteria depend on the numerical setting rather than on a single universal energy threshold. For the $Ca_3Si_4Ir_4$ production analysis, the preregistered gap-validity guardrail is

$$0.25\ \mathrm{meV}.$$

The frozen plane-level rule also requires endpoint Kramers error below 0.01, minimum Wilson-link singular value above 10^{-4}, and agreement of the two finest Wilson-loop resolutions. A reported plane invariant is treated as valid only when all these conditions hold at the relevant endpoints. Its predicted representative gap,

$$6.17 \times 10^{-4}\ \mathrm{meV},$$

falls far below this threshold, and the raw Wilson-loop label is therefore classified as invalid rather than counted as a Z_2 mismatch. The exact 0.25-meV cutoff is an operational, conservative heuristic for the frozen grids. It is neither a universal validity threshold nor a physical phase boundary; the diagnosis is supported jointly by gap collapse, convergence behavior, Kramers/link checks, and resolution stability.

This taxonomy is distinct from the composite failure label used in the finite-query detection benchmark. In particular, the projector witness theorem applies only to two continuous, equal-rank, well-defined bundles with different topology; it does not imply

$$\text{generic benchmark failure} \implies \sup_k f(k) \geq 1/r$$

when the failure itself is a gap or numerical-validity failure.

I.6 Experimental provenance and preregistered endpoints

The empirical claims in the paper combine confirmatory comparisons, diagnostic follow-ups, and exploratory boundary analyses. We preserve these roles in the interpretation rather than treating every reported result as an equivalent confirmatory endpoint.

Table 15 summarizes the principal experimental stages and the scientific status of their primary conclusions.

Several distinctions are important for reproducibility of the interpretation. The rank-two local-witness refinement in Appendix C.2 is a numerical consistency check on already trained models rather than an independent test of the theorem. The 10-meV JARVIS study is a scope analysis performed after the large fixed-matrix-error stress revealed that matrix MAE was poorly calibrated to near-Fermi spectral error. Its null result is therefore used to constrain the empirical claim rather than to estimate a prospective population failure rate.

For the released-model analyses, the available provenance is not identical for all four materials. ZrSiPt and $Ca_3Si_4Ir_4$ have material-specific analysis plans, whereas the $NaCaTaTiO_6$ and Ta_3AlC_2 evaluations use the common settings established for the subsequent breadth analysis. We do not treat this difference as evidence about model quality; it is recorded to make clear which material-level endpoints were specified independently.

Table 15: **Experimental provenance and interpretation of the principal endpoints.** The table records whether the planned comparison supported its intended scientific claim; negative and null outcomes are retained because they bound the conclusions of the paper. "Exploratory" denotes an analysis introduced to characterize scope rather than a preregistered primary test.

Study	Primary question	Observed endpoint	Paper interpretation
Two-band controlled models	Can finite-observation learning realize the localized topology blind spot?	IID recovery is poor; region-centered recovery is architecture dependent.	Controlled evidence for underdetermined global extension; no universal loop remedy.
Dirichlet regularization	Does smoothness reliably remove hidden winding?	Recovery improves to $14/15$, but only 20% of paired labels improve; the planned paired-effect criterion is not met.	Smoothness is a useful inductive bias, not a certificate.
Grid/origin stability	Are topology labels primarily discretization artifacts?	Targets are fully stable; correct baseline models retain 96.7% of labels and regularized models 100%.	Numerical aliasing does not explain the main controlled results.
Four-band transfer	Does the mismatch survive a rank-two occupied space?	IID $0/15$ versus region-centered $15/15$, with lower band MAE for the incorrect IID family.	Primary learned multiband evidence.
$NaRuO_2$	Can a realistic multiorbital Hamiltonian cross a topology boundary at modest spectral error along a vulnerable direction?	First transition at 8.074 meV near-Fermi band MAE.	Evidence of structured sensitivity, not prevalence.
JARVIS fixed-error stress	What does a small real-space matrix MAE imply for spectral accuracy?	1.417 meV matrix MAE produces hundreds to thousands of meV band error.	Demonstrates metric calibration failure.
JARVIS 10-meV analysis	Do the same generic mechanisms produce failures after matching small spectral error?	$276/432$ evaluations are valid at both endpoints; among them, 0 valid mismatches and 0 newly unsafe predictions. The other 156 are target-invalid exclusions.	Post-unblinding exploratory empirical boundary; not a production residual estimate.
Equal-budget query benchmark	Do tested loop-structured point queries have a stable equal-budget advantage?	At $B = 64$, loop minus stratified-IID AUROC is -0.00104, with interval spanning zero and $p = 0.5078$.	No supported loop-specific advantage in the tested benchmark.

I.7 Released-data availability and scope

The production-scale analysis is restricted by the contents of the official release rather than by a selection of only unfavorable examples. The released examples archive contains four complete material cases for which the information required for a paired reference–prediction Hamiltonian analysis is available:

$$\mathrm{ZrSiPt}, \quad \mathrm{Ca_3Si_4Ir_4}, \quad \mathrm{NaCaTaTiO_6}, \quad \mathrm{Ta_3AlC_2}.$$

All four are included in the paper.

A separate released topology archive contains 138 crystal structures, but these files do not provide the paired graph/model inputs and reference Hamiltonians required to perform the same target–prediction audit. We therefore do not treat those structures as additional paired production examples.

This availability boundary explains the role of the four-material production study. It demonstrates that the evaluation pipeline can be applied to a released large neural Hamiltonian, and it supplies

both stable agreements and a gap-validity failure. It does not estimate the prevalence of topological errors across the broader material population.

The broader JARVIS perturbation study in Appendix H addresses a different question using many more reference Hamiltonians, but its perturbations are generic synthetic stress directions rather than residuals drawn from Uni-HamGNN or another production predictor. The two evidence sources are therefore complementary and should not be combined into a single failure-rate estimate.

For reproducibility, the principal analyses are backed by saved model checkpoints, resolved configurations, tabular result files, and material-specific topology summaries. The controlled study contains 72 saved two-band checkpoints, 43 smoothness-ablation checkpoints, and 38 four-band checkpoints. The equal-budget query analysis retains the pointwise query observations and their protocol-level aggregates, while the released-model and JARVIS analyses retain the corresponding gap, spectral, and plane-wise topology summaries. These artifacts are sufficient to reproduce the quantitative tables and figures reported in the paper without retraining the learned controlled models.