COMPLEX: A CLOSED-FORM CERTIFIED EMBEDDING OF MULTIPARAMETER PERSISTENCE MODULES

Anonymous authors
Paper under double-blind review

ABSTRACT

Every multiparameter persistence vectorization we know of carries a one-sided Lipschitz *upper* bound and nothing below it: without a lower gauge there is no sense in which the features are faithful, and no per-prediction guarantee can be built on them. This paper supplies the missing side. **COMPLEX** is a closed-form, training-free embedding of multiparameter modules—slice the module along a fixed near-diagonal net, embed each slice barcode by the certified PLACE/PALACE landmark map, concatenate. Under a checkable *witnessing-slice coherence* condition, holding on 100% of audited pairs on Orbit5k, a single slice carries a closed-form lower gauge: separated modules stay separated in the embedding. With the standard upper bound this gives, to our knowledge, the first two-sided distortion bound for a multiparameter feature map, making faithfulness *measurable*. Measuring it, we find the floor tight within a small factor of realized distances yet operationally *local*: an RBF-SVM reaches 91% where 1-NN reaches 78% on the same features. Local per-prediction certification therefore fails for a structural reason common to every landmark embedding whose lower gauge is witnessed by one coordinate. With no learned embedding and no held-out calibration—only a cross-validated SVM head—COMPLEX matches the best published accuracy on Orbit5k (91.95%) and exceeds it on Orbit100k (92.98%), level with Euler-characteristic surfaces and above transformers and graphcode. On graphs it exceeds GRIL on all four shared molecular benchmarks with one fixed configuration, including the only multiparameter method to clear COX2's majority baseline by more than three points. Closed-form selection—of the landmark radius, the kernel (certificate-preserving), and the bifiltration set—buys further accuracy; gradient-shaped adaptation buys no reliable gain.

1 INTRODUCTION

Multiparameter persistence assigns a dataset not a single barcode but a persistence module over $\mathbb{R}^d$, arising for instance from a density–Rips or a sublevel bifiltration. A growing literature turns such modules into features for learning: multiparameter persistence landscapes (Vipond, 2020) and images (Carrière & Blumberg, 2020), signed measures and sliced barcodes (Botnan et al., 2022; Loiseaux et al., 2023b), the GRIL landscape and its differentiable successor (Xin et al., 2023; Mukherjee et al., 2026), and graphcodes (Kerber & Russold, 2024). Two further families compete on the same benchmarks without being persistence methods: Euler-characteristic curves and surfaces (Hacquard & Lebovici, 2024) read the same bifiltration but bypass homology altogether, and transformer architectures (Reinauer et al., 2021; Kim, 2024) read one-parameter diagrams directly. These methods split along two orthogonal axes (Table 3, Appendix A), and on the *guarantee* axis they are unanimous: each carries a one-sided Lipschitz upper bound but never a lower bound with explicit constants. To our knowledge the certified column is empty: no multiparameter-persistence feature map supplies a computable lower distortion bound, and hence none yields a split-free per-prediction certificate derived from the feature geometry itself.

In one parameter, PLACE (Majhi et al., 2026a) and its data-adaptive successor PALACE (Majhi et al., 2026b) already occupy the certified, training-free cell: a closed-form landmark embedding of persistence diagrams with a lower distortion certificate ρ_- derived from labels alone, no tuning. But they certify only *one* parameter. This paper carries the certificate into multiparameter modules, then spends its second half measuring what the certificate is worth; a differentiable variant (Section 5) shows it

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

Table 1: Point-cloud accuracy (%, mean±std over folds). Bold = best in family; * literature numbers under each method's own protocol; the COMPLEX RBF row uses $\alpha = 4$ (and PALACE's three-bifiltration recipe on Orbit5k), the nearest-centroid row the $\alpha = 1.75$ default. †Nyström feature map at 4,000 components (3 seeds × 3 folds), calibrated against the exact head on Orbit5k (Appendix E).

Method	Orbit5k	Orbit100k
One-parameter persistence		
SW kernel*	83.6 ± 0.9	–
Fisher kernel*	85.9 ± 0.8	–
PersLay*	87.7 ± 1.0	89.2 ± 0.3
xPerT (transformer)*	87.0	91.1
Persformer (transformer)*	91.2 ± 0.8	$\mathbf{92.0 \pm 0.4}$
PALACE*	$\mathbf{91.3 \pm 1.0}$	–
Euler characteristic (homology-free)		
ECS+XGB*	$\mathbf{91.8 \pm 0.4}$	–
ECS+XGB, our re-run (matched bifiltration)	89.5 ± 0.9	–
Multiparameter persistence		
Graphcode (GNN)* (Kerber & Russold, 2024)	88.5 ± 1.1	92.3 ± 0.3
COMPLEX, nearest-centroid (certified)	73.4 ± 2.6	71.6 ± 0.5
COMPLEX, RBF-SVM (uncertified)	$\mathbf{91.95 \pm 1.3}$	$\mathbf{92.98 \pm 0.15}^{\dagger}$

survives gradient adaptation as well. The mechanism is the *slice-stack*: embed a module by stacking the per-slice diagram embeddings over a finite slice net, and certify the *sliced matching distance* d_{mat}^{S}. We call the construction **COMPLEX**—*Certified One-slice Multiparameter Persistence Landmark EXtension*—"one-slice" because a single witnessing slice carries the certificate to the module level (Theorem 2.3). This paper develops it self-containedly on top of PLACE/PALACE (Majhi et al., 2026a;b), whose population-level statistics—covariance estimation, two-sample tests and confidence sets for the mean embedding—are developed in Bagchi et al. (2026).

Contribution. *(i) A two-sided bound.* The slice-stack Ψ (Definition 2.1) is 1-Lipschitz up to $\sqrt{S}\, N_{\max}$ for d_{mat}^{S} (Proposition 2.2) and, under a checkable *witnessing-slice coherence* condition (Definition 2.4), obeys the closed-form lower gauge $\|\Psi(M) - \Psi(M')\| \geq \rho_{-}(\tau) = \frac{1}{4}\tau\, w_{\min}$ (Theorem 2.3); coherence needs no multiparameter matching, has a matching-free sufficient condition (Proposition B.7), and holds on 100% of audited cross-class pairs at the headline configuration. *(ii) The net.* The certificate's value is angle-independent but its price is not, which is what makes the fixed near-diagonal fan the default; a certified data-adaptive net generalizes it (Proposition 2.5). *(iii) Predictions.* A certified nearest-centroid rule inherits PALACE's split-free per-prediction guarantee (Theorem 3.1), which holds for any bounded embedding; the rule that reads ρ_{-} is nearest neighbor (Corollary 3.2), with a stability radius in the module metric, and B/ρ_{-} controls a—numerically vacuous—generalization bound (Proposition D.1). *(iv) Adaptation keeps the certificate.* The floor constrains the output embedding, so the bifiltration can be selected in closed form or learned by gradient (D-COMPLEX, Proposition E.1) and the kernel chosen among Gaussian, χ^2 and Hellinger (Lemma D.3) without losing ρ_{-}, once its hypotheses are re-audited on the adapted embedding. *(v) Faithfulness, measured.* The floor is tight within a small factor of realized distances, yet every per-prediction rule built on it accepts under 1.5% of test modules at the budgets that maximize accuracy: an RBF-SVM reaches 91% where 1-NN reaches 78% on the same embedding—the accuracy is *global*, carried by the kernel, while ρ_{-} certifies a *local* separation in classes interleaved at neighbor scale; nothing in the argument is specific to persistence, though it is measured on Orbit5k, not proved in general (Section 4). *(vi) Accuracy.* With no learned embedding and no held-out calibration, COMPLEX matches the best published accuracy on Orbit5k (91.95%) and exceeds it on Orbit100k (92.98%), level with Euler-characteristic surfaces and above transformers and the GNN-trained graphcode, and exceeds GRIL on all four shared molecular graph sets with one fixed configuration; closed-form selection of radius, kernel and bifiltration set buys further accuracy, gradient-shaped adaptation no reliable gain (related work: Appendix A).

2 COMPLEX: THE SLICE-STACK AND ITS TWO-SIDED BOUND

2.1 MODULES, SLICES, AND THE SLICED MATCHING DISTANCE

We recall finitely presented multiparameter modules, slicing, and the sliced matching distance; all are standard (see (Carlsson & Zomorodian, 2009; Lesnick, 2015)). Throughout, k indexes landmarks, s slices and q homological degree; slice angles are ϑ_s, the FPS radius factor is α, the RBF bandwidth h and the kernel κ; Table 4 in Appendix B collects the notation. Every result in this paper is proved; proofs not given in the text are in Appendix C.

A *d-parameter persistence module* $M\colon \mathbb{R}^d_{\geq 0} \to \mathrm{Vec}_{\Bbbk}$ is taken finitely presented (which also implies it is pointwise finite-dimensional). Every filtration in this paper produces non-negative grades (Appendix B). A *slice* is a line $\ell \subset \mathbb{R}^d$ of positive slope; the restriction $M|_\ell$ is a one-parameter module with barcode $\mathrm{dgm}(M|_\ell) \in \mathcal{D}_N$, and $\ell \mapsto \mathrm{dgm}(M|_\ell)$ is the *fibered barcode*, equivalent to the rank invariant (Cerri et al., 2013; Lesnick & Wright, 2015). We write $\mathcal{R}_N$ for the modules with at most N bars in every slice.

Slices, parametrized. A line of positive slope enters $\mathbb{R}^d_{\geq 0}$ at a single point β with $\min_i \beta_i = 0$; writing it as the ray from its entry point, $\ell_{\beta,u} = \{\beta + tu : t \geq 0\}$ with $u \in \mathbb{S}^{d-1}_{>0} := \{u \in \mathbb{R}^d_{>0} : \|u\|_2 = 1\}$ (1), names every slice exactly once, with t the Euclidean *arclength* from β, and a generator at grade γ enters $M|_{\ell_{\beta,u}}$ at

$$s_{\beta,u}(\gamma) \;=\; \min\{t \geq 0 :\; \beta + tu \geq \gamma\} \;=\; \max_{1\leq i\leq d} (\gamma_i - \beta_i)/u_i \;\geq\; 0, \tag{2}$$

so slice barcodes land in $\mathcal{D}_N$ with non-negative births. An ε-interleaving shifts grades diagonally by ε, which moves (2) by at most $\varepsilon/\min_i u_i$ whatever the offset, so the weight making a slice a lower gauge for the interleaving distance is $\omega(u) = \min_{1\leq i\leq d} u_i \in (0, 1/\sqrt{d}]$ (3), maximal on the diagonal and $\to 0$ toward the coordinate hyperplanes (Appendix B relates this to the classical ℓ^∞ convention). For a finite *slice net* $\mathcal{L}_S = \{\ell_{\beta_1,u_1}, \ldots, \ell_{\beta_S,u_S}\}$ with weights $\omega_s = \omega(u_s)$, the *sliced matching distance* is

$$d^S_{\mathrm{mat}}(M, M') = \max_{1\leq s\leq S} \omega_s\, d_{\mathcal{B}}\big(M|_{\ell_s}, M'|_{\ell_s}\big), \tag{4}$$

a finite-net surrogate for the continuous matching distance $d_{\mathrm{mat}} = \sup_\ell \omega(\ell)\, d_{\mathcal{B}}(M|_\ell, M'|_\ell)$ (Cerri et al., 2013; Landi, 2018), exactly computable only in impractically high-degree polynomial time (Kerber et al., 2019). Here $d_{\mathcal{B}}$ is the with-diagonal bottleneck distance on the diagram space of PALACE (Majhi et al., 2026b): the *space of one-point diagrams* is $\mathcal{D}_1 = T \cup \{\Delta\}$, $T = \{(b,d) : d > b \geq 0\}$ and Δ an abstract point standing for the diagonal, with $d_{\mathcal{B}}((b,d),\Delta) = \frac{1}{2}(d-b)$ and $d_{\mathcal{B}}(x,y) = \min\{\|x-y\|_\infty, \max\{d_{\mathcal{B}}(x,\Delta), d_{\mathcal{B}}(y,\Delta)\}\}$; the *space of diagrams on N points* is $\mathcal{D}_N = \mathcal{D}_1^N/S_N$ under the max metric,

$$d_{\mathcal{B}}(A,B) \;=\; \min_{\psi\in S_N} \max_{1\leq i\leq N} d_{\mathcal{B}}\big(x_i, y_{\psi(i)}\big), \tag{5}$$

a diagram with fewer points being padded with copies of Δ (Appendix B states the one-point metric in full as (14)). The per-slice ingredient we invoke is the closed-form certified landmark embedding of $(\mathcal{D}_N, d_{\mathcal{B}})$ established for PLACE/PALACE (Majhi et al., 2026a;b).

2.2 THE EMBEDDING

Per-slice configuration. Following PALACE (Majhi et al., 2026b), a *landmark configuration* for $\mathcal{D}_N$ is $\nu = \{(p_k, r_k, w_k)\}_{k=1}^K$ with landmark *points* $p_k \in T \subset \mathcal{D}_1$, radii

$$0 < r_k \;\leq\; d_{\mathcal{B}}(p_k, \Delta) \;=\; \tfrac{1}{2}\big(d(p_k) - b(p_k)\big), \tag{6}$$

and weights $w_k > 0$ with $\sum_k w_k^2 = 1$. The *coordinate function* on $\mathcal{D}_1$, and its sum-pool to N-point diagrams, are

$$\varphi_{p,r}(x) = \max\{r - d_{\mathcal{B}}(p,x), 0\}, \qquad \varphi_{p,r}(A) = \sum_{a\in A} \varphi_{p,r}(a), \tag{7}$$

and the *summation landmark embedding* is $\Phi_k(A;\nu) = w_k\, \varphi_{p_k,r_k}(A)$, so $\Phi(\cdot;\nu)\colon \mathcal{D}_N \to \mathbb{R}^K$. Landmarks are placed by (class-aware) farthest-point sampling and weights set to $w_k = K^{-1/2}$—deterministic and closed-form (Majhi et al., 2026b).

Definition 2.1 (Slice-stack embedding)**.** Fix a slice net $\mathcal{L}_S$ and a per-slice configuration ν_s on each slice. The *COMPLEX configuration* is $\mathcal{C} = \{(\ell_s, \omega_s, \nu_s)\}_{s=1}^{S}$, and the embedding of $M \in \mathcal{R}_N$ is the slice-stack

$$\Psi(M; \mathcal{C}) = \big(\, \omega_s \, \Phi(M|_{\ell_s}; \nu_s) \,\big)_{s=1}^{S} \;\in\; (\mathbb{R}^K)^S \cong \mathbb{R}^{SK}. \tag{8}$$

Slicing and embedding are applied degree by degree and the blocks concatenated over the degrees used; pooling the degrees into one multiset before embedding preserves the upper bound but not the hypothesis of the certificate (Appendix B).

Proposition 2.2 (Upper distortion bound)**.** *The slice-stack is* 1*-Lipschitz up to* $\sqrt{S}$*:* $\|\Psi(M) - \Psi(M')\|_{\ell^2} \le \sqrt{S}\, N_{\max}\, d_{\mathrm{mat}}^{S}(M, M')$*, so* $(\sqrt{S}\, N_{\max})^{-1}\Psi$ *is* 1*-Lipschitz for* d_{mat}^{S}*.*

Proof. Deferred to Appendix C.

Three conventions of the released pipeline are discussed in Appendix B: it does not enforce the cap (6), so we run an *uncapped* variant carrying the accuracy headline and a *capped* one to which Proposition 2.2 and Corollary 3.2(ii) attach; the classifier reads the *unweighted* stack Ψ°, on which both gauges survive; and it keeps the 50 most persistent bars per slice, so every result holds on $\mathcal{R}_{50}$.

2.3 The witnessing-slice certificate

A subtlety the single-parameter theory does not see: separation at scale τ in d_{mat}^{S} is separation at the *realized* scale $t_s = \tau/\omega_s$ in the bottleneck distance on slice s, by (4). A hat never exceeds its radius, so we require of a witnessing landmark that its radius reach the realized scale; the landmarks that can witness at all form the *active set* $\mathcal{A}_s(\tau) = \{k : r_{s,k} \ge \tau/(4\omega_s)\}$ (9), and when it is nonempty $w_{\min}(\tau; \nu_s) := \min\{w_{s,k} : k \in \mathcal{A}_s(\tau)\}$ is well defined and computable in time linear in the number of landmarks.

Theorem 2.3 (Witnessing-slice certificate)**.** *Let* $\mathcal{C}$ *be a net whose active sets* $\mathcal{A}_s(\tau)$ *are all nonempty, let* (M, M') *be* $\mathcal{C}$*-coherent (Definition 2.4) with* $d_{\mathrm{mat}}^{S}(M, M') \ge \tau$*, and let* $s^\star = \arg\max_s \omega_s\, d_{\mathcal{B}}(M|_{\ell_s}, M'|_{\ell_s})$ *be the witnessing slice. Then*

$$\|\Psi(M) - \Psi(M')\|_{\ell^2} \;\ge\; \tfrac{1}{4}\,\tau\, w_{\min}(\tau; \nu_{s^\star}) \;\ge\; \rho_-(\tau; \mathcal{C}) \;:=\; \tfrac{1}{4}\,\tau \min_{1 \le s \le S} w_{\min}(\tau; \nu_s), \tag{10}$$

and, under equal weights $w_{s,k} = K^{-1/2}$*,* $\rho_-(\tau) = \tau/(4\sqrt{K})$*.*

Proof. Deferred to Appendix C.

The proof is three lines: the ℓ^2 norm dominates the $s^\star$ block; coherence supplies one landmark $k^\star \in \mathcal{A}_{s^\star}(\tau)$ of that block whose coordinate moves by at least $w_{k^\star} t^\star/4$ at the realized scale $t^\star = \tau/\omega_{s^\star}$; and the block weight $\omega_{s^\star}$ in (8) multiplies this back to $\frac{1}{4}\tau\, w_{k^\star}$, so the reparametrization cancels, and the minimum over slices makes ρ_- a property of the configuration alone, computable before any pair is examined. The *value* in (10) is thus angle-independent, the block weight $\omega_{s^\star}$ cancelling the realized scale, but the *cost* of earning it is not: by (9) a slice with small ω_s needs radii larger by $1/\omega_s$ to have a nonempty active set at all (Remark B.4), which is what makes the near-diagonal fan the cheapest certifiable net at a given budget.

2.4 Witnessing-slice coherence

Coherence asks, without any multiparameter matching, that a single witnessing slice genuinely separate the two modules—the multiparameter analogue of PALACE's witnessing-landmark condition.

Definition 2.4 (Witnessing-slice ν-coherence)**.** A pair (M, M') with $d_{\mathrm{mat}}^{S}(M, M') \ge \tau$ is $\mathcal{C}$*-coherent* if its witnessing slice $s^\star$ has restricted pair $(M|_{\ell_{s^\star}}, M'|_{\ell_{s^\star}})$ single-parameter $\nu_{s^\star}$-coherent at the realized scale $t^\star := \tau/\omega_{s^\star}$: some landmark $k^\star$ active at that scale, $k^\star \in \mathcal{A}_{s^\star}(\tau)$ as in (9), realizes the single-coordinate floor

$$|\Phi_{k^\star}(M|_{\ell_{s^\star}}) - \Phi_{k^\star}(M'|_{\ell_{s^\star}})| \;\ge\; w_{k^\star}\, t^\star/4 \;=\; w_{k^\star}\, \tau/(4\omega_{s^\star}).$$

This is PALACE's ν-coherence (Majhi et al., 2026b) on the witnessing slice, with the floor set by the scale $t^\star$; the audit of Section 4 checks the stronger realized-distance form.

Admissibility, at the realized scale. Following PALACE, a configuration ν covering the support $\mathcal{X} \subset T$ of a slice's diagram points has *functional Lebesgue number* $\lambda_0(\nu) = \inf_{x\in\mathcal{X}} \max_k \varphi_{p_k,r_k}(x)$ (Appendix B) and is *t-admissible* if $\lambda_0(\nu) \geq t/4$ and $\max_k r_k \leq (t + \lambda_0)/2$; a net $\mathcal{C}$ is *τ-admissible* if each ν_s is (τ/ω_s)-admissible, at the realized scale. Its real content is that it derives coherence: Proposition B.7 (Appendix B) shows that if $\mathcal{C}$ is τ-admissible, D bounds d_{mat}^S over the data set, and along every slice the generator and relation grades of M enter more than $4D/\omega_{\min}$ apart while every bar of $M|_{\ell_s}$ and $M'|_{\ell_s}$ is longer than $2D/\omega_{\min}$, then (M, M') is $\mathcal{C}$-coherent—both hypotheses read off the two barcodes separately, with no cross-pair matching. We state plainly that the configurations run in Section 4 are *not* τ-admissible, at any τ (Remark B.5); Theorem 2.3 is unaffected, consuming nonempty active sets and coherence, both measured directly, but Corollary 3.2 and Proposition D.1 have their hypothesis verified by audit rather than by construction. Coherence is a hypothesis because sum-pooling is many-to-one (Remark B.6); it fails on 2.5% of separated cross-class pairs on Orbit5k at the $\alpha = 1.75$ default and on none at the headline $\alpha = 4$, while the certificate's conclusion holds on 100% of pairs in both cases.

2.5 The slice net, fixed or selected

By (9) a slice of weight ω_s can be certified at a fixed budget only if its radii reach $\tau/(4\omega_s)$, unbounded as the slice approaches an axis, so the closed-form default is the uniform near-diagonal fan (for $d = 2$, the case of every experiment)

$$\mathcal{L}_S = \{\ell_{0,u_s}\}_{s=1}^S, \qquad \vartheta_s \text{ equally spaced in } [\tfrac{\pi}{4} - a, \tfrac{\pi}{4} + a], \quad u_s = (\cos\vartheta_s, \sin\vartheta_s), \tag{11}$$

of half-width a; no slice is optimized. *Certified slice selection.* Since the certificate value does not degrade off the diagonal, the net may also be selected from the training labels and frozen: rank the directions of a dense candidate fan by a closed-form score, keep those whose FPS configuration ν_u is admissible at its realized scale and clears a floor $\rho_-(\nu_u) \geq \rho_0$, diversify by farthest-point sampling in angle, and freeze (Appendix B).

Proposition 2.5 (Certified selection)**.** *For* any *net produced by this rule and then frozen, the slice-stack satisfies the witnessing-slice certificate (Theorem 2.3) on every $\mathcal{C}$-coherent pair, with floor $\rho_-(\tau) \geq \rho_0$; and the farthest-point step is the standard 2-approximation of the k-center objective (the min–max covering radius) among the certifiable directions.*

Proof. Deferred to Appendix C. Empirically the adaptive net falls at or slightly below the fixed fan (Appendix E).

3 Certified classification

The certificate is a statement about *pairs* of modules; this section turns it into statements about *predictions*.

3.1 Nearest centroid: a guarantee that does not use ρ_-

Every embedded module lies in a ball of explicit radius, $\|\Psi(M;\mathcal{C})\|_{\ell^2} \leq B := \sqrt{S}\, N_{\max} \max_{s,k} r_{s,k}$ (Lemma C.1, Appendix C), which is what the concentration argument needs.

Theorem 3.1 (Certified prediction)**.** *Let $(M_i, y_i)_{i=1}^m$ be i.i.d. from a distribution on $\mathcal{R}_N \times [C]$ with m_c samples in class c, let $\widehat{\mu}_c, \mu_c \in \mathbb{R}^{SK}$ be the empirical and population means of Ψ on class c, $V_c = \mathbb{E}_c \|\Psi(M) - \mu_c\|^2$, and $b_\delta = \log(2C/\delta)$. Put*

$$r_m := \max_{1\leq c\leq C} \left(\sqrt{\tfrac{2b_\delta V_c}{m_c}} + \tfrac{4Bb_\delta}{3m_c} \right), \tag{12}$$

with B as above. Then $\Pr[\max_c \|\widehat{\mu}_c - \mu_c\| \leq r_m] \geq 1 - \delta$, and on that event the empirical and population nearest-centroid rules assign the same label to every M whose empirical decision gap

$$\mathrm{gap}(M) := \min_{c\neq\widehat{c}(M)} \|\Psi(M) - \widehat{\mu}_c\| - \left\|\Psi(M) - \widehat{\mu}_{\widehat{c}(M)}\right\|$$

exceeds $2r_m$. If moreover $r_m < \frac{1}{2}\widehat{\Delta}$, $\widehat{\Delta} := \min_{c\neq c'} \|\widehat{\mu}_c - \widehat{\mu}_{c'}\|$, the population means are distinct and the population rule is well posed.

Proof. Deferred to Appendix C.

The training-time inequality $r_m < \frac{1}{2}\widehat{\Delta}$ plus the per-test-point check $\mathrm{gap}(M) > 2r_m$ certifies predictions with no held-out calibration split; r_m becomes computable on substituting $V_c \le B^2$. One condition of the sampling model deserves to be stated rather than assumed: the configuration $\mathcal{C}$—landmarks, radii, τ—must be fixed independently of the sample for the $\Psi(M_i)$ to be i.i.d., and the same is true of every probabilistic statement in this section (the deterministic margin and stability radius of Corollary 3.2 need no independence). The protocol of Section 4 places the landmarks by class-aware farthest-point sampling on the training fold, so the configuration is selected from the very sample being certified; conditioning on such a selection does not in general preserve the i.i.d. structure, so the probability guarantees are not established for that implementation. Sample splitting—landmarks from one half of the training fold, means and radii from the other—restores the hypothesis exactly at the cost of half the data; we did not run it, and we report the certified fractions with this caveat in place. What the theorem does *not* do is use the certificate: ρ_- appears nowhere in it, and it holds for any bounded embedding, because a lower bound on *pairwise* cross-class distances says nothing about *class means*—two classes can be pairwise separated at every scale and share a centroid (the population version of this obstruction is worked out in Bagchi et al. (2026)).

3.2 Nearest neighbor: where the certificate does the work

Corollary 3.2 (Certified 1-NN prediction, with a stability radius in d^S_{mat})**.** *Let $\mathcal{C}$ be τ-admissible and $T = \{(M_i, y_i)\}_{i=1}^m$ a training set. Let M have nearest training neighbor $M_{i^\star}$ at embedded distance $\varepsilon = \|\Psi(M) - \Psi(M_{i^\star})\|$, and suppose every pair $(M_{i^\star}, M_j)$ with $y_j \neq y_{i^\star}$ is $\mathcal{C}$-coherent with $d^S_{\mathrm{mat}} \ge \tau$. If $\varepsilon < \frac{1}{2}\rho_-(\tau; \mathcal{C})$ then*

(i) every differently labeled training module lies at embedded distance at least $\rho_-(\tau) - \varepsilon$ from $\Psi(M)$, so the 1-nearest-neighbor label $y_{i^\star}$ is attained with margin $\rho_-(\tau) - 2\varepsilon > 0$ (the conclusion is the margin: that $M_{i^\star}$ is the nearest neighbor is its definition); and

(ii) that label is unchanged for every module $M' \in \mathcal{R}_N$ with

$$d^S_{\mathrm{mat}}(M, M') < \frac{\rho_-(\tau; \mathcal{C}) - 2\varepsilon}{2\sqrt{S}\, N_{\max}}.$$

Proof. Deferred to Appendix C.

The two-sided bound thus yields a two-sided guarantee—a label determined with a positive margin that survives a quantified perturbation of the input, though not certified correct—evaluated at inference time; the separation half of the hypothesis costs nothing on taking $\tau := \min\{d^S_{\mathrm{mat}}(M_i, M_j) : y_i \neq y_j\}$ (13), the cross-class separation of the training set itself, the τ we use below.

A generalization bound. The certificate also controls the *population* risk of the 1-NN rule, with ρ_- appearing essentially, through the algorithmic-robustness route of Xu & Mannor (2012): with τ fixed before the sample is drawn, a training set that is separated and coherent at τ makes the rule $(\mathcal{K}, 0)$-robust with $\mathcal{K} \le C(1 + 8B/\rho_-(\tau))^{SK}$, so the sample complexity is governed by the two-sided distortion B/ρ_-, which this paper computes and no prior multiparameter vectorization can (Proposition D.1, Appendix D). As a number the bound is vacuous, $\mathcal{K}$ being exponential in $SK = 1{,}600$, and the vacuity is a matter of scale rather than dimension: test modules sit 6–12 times ρ_- from their nearest training module, so no two training modules share a cell at the certificate's scale.

Kernel heads. The same floor certifies the higher-accuracy Gaussian-SVM head: with $\kappa(M, M') = \exp(-\|\Psi(M) - \Psi(M')\|^2 / 2h^2)$ and feature map ι into its RKHS $\mathcal{H}$, monotonicity and (10) give, on every coherent cross-class pair with $d^S_{\mathrm{mat}} \ge \tau$, $\kappa(M, M') \le \kappa_-(\tau) := \exp(-\rho_-(\tau)^2/2h^2) < 1$ and $\|\iota(M) - \iota(M')\|_{\mathcal{H}} \ge \sqrt{2(1 - \kappa_-(\tau))} > 0$ (Lemma D.2). This certifies the separation the kernel sees, not the margin of the soft-margin SVM actually fitted (a pairwise floor does not bound the distance between convex hulls), and closing that gap is what Section 4 measures. The floor carries to the χ^2 and Hellinger kernels with explicit constants (Lemma D.3, Appendix D), so selecting the kernel by inner cross-validation never loses it.

4 EXPERIMENTS

We test one question: does the closed-form slice-stack, with no learned embedding and no held-out calibration, match trained multiparameter baselines? All numbers are produced by the released pipeline (bifiltration $\to$ fixed near-diagonal net $\to$ per-slice PALACE embedding $\to$ concatenation $\to$ classifier), averaged over three seeds, on Orbit5k (5,000 point clouds, 5 classes), Orbit100k, and the TU graph benchmarks COX2, DHFR, MUTAG, NCI1, NCI109, PTC_MR, PROTEINS, DD, IMDB-BINARY and IMDB-MULTI; point clouds carry the alpha–DTM bifiltration, graphs a sublevel bifiltration (heat-kernel signature $\times$ degree, and $\times$ node label for the labeled molecular sets).

Protocol (no feature tuning). The net (11) is the fixed fan ($S = 8$, $a = 0.6$); each slice is PALACE-embedded with $K = 200$ class-aware FPS landmarks, equal weights, radius factor $\alpha = 4$ (PALACE's published Orbit value, adopted unchanged), the top $N = 50$ bars, essential classes capped at the largest finite death, and τ at the median half-persistence (the *audit* scale below is the median cross-class d^S_{mat}); on Orbit5k the headline additionally uses PALACE's three DTM bifiltrations at $K = 200$ per block. We classify Ψ° by (a) nearest centroid, with the certificate of Theorem 3.1, and (b) an RBF-SVM. The *embedding* has no hyperparameter tuned on our data, but the head is not tuning-free: following PALACE, the bandwidth and C are chosen by inner 3-fold cross-validation on the training fold; no held-out data is used, and evaluation is 10-fold stratified cross-validation (3-fold at Orbit100k) with the embedding frozen from the training fold. Baselines, and which numbers are literature figures under their own protocol (*) against re-runs on our clouds and folds, are in Appendix E.

Point clouds. Table 1: on Orbit5k COMPLEX reaches 91.95 ± 0.11 over seeds (± 1.3 over folds); the single-bifiltration stack reads 91.74 ± 0.10 at the same α and 91.2 at the $\alpha = 1.75$ default. That places the closed-form embedding ahead of every diagram-based kernel and vectorization, of the transformers xPerT and Persformer (91.2 ± 0.8; a reproducibility caveat attaches, Appendix E), of its predecessor PALACE (91.3 ± 1.0) and, for the first time in this line of work, of the Euler-characteristic surface (91.8 ± 0.4), previously the ceiling among training-free descriptors. The ECS lead is 0.15 points, within one fold's deviation, so *level* is the defensible reading; against a *matched* re-run of the single-parameter pipeline (PALACE-1p, Table 5) COMPLEX is ahead by 3.7. On Orbit100k COMPLEX reads 92.98 ± 0.02 over three seeds, ahead of xPerT (91.1), Persformer (92.0) and the GNN-trained graphcode (92.3 ± 0.3), the strongest previously published number at this scale; the certified capped variant reads 92.78 ± 0.20 there (91.31 against 91.74 on Orbit5k), itself ahead of every published number. Under outlier contamination (Table 5) the margin over PALACE-1p grows from $+3.7$ clean to $+12.4$ at contamination rate 0.3, but against ECS the clean lead becomes a 0.5-point deficit: what COMPLEX buys over the Euler family is the certificate and applicability to bifiltered data, not robustness.

Graphs. Table 2 reports the TU benchmarks. The geometry-only bifiltration (HKS $\times$ degree) is already competitive with GRIL; adding the node-label axis lifts every molecular set and takes COMPLEX past GRIL on MUTAG, DHFR, and PROTEINS—using the same node information the trained baselines see, but with no training. Concatenating three bifiltrations with the homology degrees kept separate—the "multi" column, and the per-degree form the certificate requires—is stronger still: *one fixed configuration*, applied unchanged to every data set, exceeds GRIL on all four shared benchmarks ($+1.9$ COX2, $+2.5$ DHFR, $+0.7$ MUTAG, $+1.3$ PROTEINS) and clears the majority baseline everywhere. Enforcing the radius cap costs at most 0.8 points and changes no ordering. Trained topological pipelines outside the family report stronger numbers on several sets (TopoGCL (Chen et al., 2024)) and we do not claim otherwise. COX2 is degenerate—all seven multiparameter methods tabulated by D-GRIL (Mukherjee et al., 2026), itself included, report exactly its majority rate 78.16 ± 0.41—and the entries clearing it by a real margin are all variants of this construction; D-GRIL itself is comparable only on IMDB-BINARY, without node labels, where COMPLEX reads 70.0 against its 62.60. At OGB scale the training-free embedding loses: on `ogbg-molhiv` it reads 72.9 test ROC-AUC against 75.6–77.1 for GIN and GCN, so the parity demonstrated here is parity on point clouds and the TU benchmarks (Appendix E). Selecting the kernel among Gaussian, χ^2 and Hellinger by the same inner cross-validation is worth $+2.9$ on NCI1, $+5.4$ on PTC_MR and $+1.4$ on COX2, neutral on PROTEINS and IMDB-B, and costs about a point on DHFR and MUTAG (Table 6); by Lemma D.3 none of it touches the certificate.

Table 2: Graph accuracy (%, 10-fold CV; COMPLEX columns mean±std over 3 seeds). "geom." = HKS×degree; "+label" adds the node-label axis; "multi" concatenates three bifiltrations with the homology degrees kept separate, *one fixed configuration* for every data set; "sel." is SELECT-MULTI (Section 5), chosen per fold on the training fold only; "majority" is the constant-classifier rate. Bold = best; * literature, single bifiltration, own protocol.

		COMPLEX						
Dataset	majority	geom.	+label	multi	sel.	GRIL*	mp-i*	mp-l*
MUTAG	66.5	88.1 ± 1.1	$\mathbf{89.7 \pm 0.7}$	88.5 ± 0.2	88.8 ± 0.4	87.8	85.6	85.7
DHFR	61.0	77.7 ± 0.8	80.6 ± 0.4	80.1 ± 0.5	$\mathbf{81.8 \pm 1.9}$	77.6	80.2	79.5
COX2	78.2	78.0 ± 0.1	78.1 ± 0.2	$\mathbf{81.7 \pm 0.4}$	81.5 ± 0.4	79.8	77.9	79.0
PROTEINS	59.6	69.5 ± 0.4	71.3 ± 0.5	$\mathbf{72.2 \pm 0.5}$	70.9 ± 0.7	70.9	67.3	65.8
NCI1	50.1	69.5 ± 0.1	73.1 ± 0.2	74.4 ± 0.1	$\mathbf{78.1 \pm 0.2}$	–	–	–
PTC_MR	55.8	–	–	57.0 ± 1.7	54.1 ± 1.4	–	–	–
IMDB-B	50.0	$\mathbf{70.0 \pm 0.4}$	–	–	–	–	–	–

Coherence audit, and the certificate instantiated. On Orbit5k we sample 2,000 cross-class pairs, keep the 1,000 whose sliced matching distance clears the median scale τ, and test Definition 2.4 on each pair's witnessing slice over active landmarks, in the realized-distance form $|\varphi_k(A) - \varphi_k(B)| \geq d_{\mathcal{B}}(A, B)/4$. Coherence holds on 97.5% of pairs at the $\alpha = 1.75$ default while the certificate's *conclusion* holds on 100%; at the adopted $\alpha = 4$ the hypothesis itself saturates at 100.0%, capped and uncapped alike—larger radii enlarge every active set, so the hypothesis is *easier* to satisfy at the configuration the accuracy prefers. The floor itself, on the certified capped variant at $\alpha = 4$ ($S = 8$, $K = 200$, 20,000 sampled cross-class pairs of which the ≈10,000 with $d^S_{\mathrm{mat}} \geq \tau$ enter the distance rows; `certificate_instantiation.py`), reads: $\tau = 1.220$, $\rho_-(\tau) = \tau/(4\sqrt{K}) = 0.0216$, minimum embedded distance over kept pairs 0.0950 ($4.4 \times \rho_-$; 2.5× at the $\alpha = 1.75$ default), median 0.2271 (10.5×), minimum centroid gap $\widehat{\Delta} = 0.0584$ (2.7×), largest class radius 0.577. The floor is thus *tight to within a small factor*; and the naive route from the certificate to the centroid guarantee is quantitatively hopeless, the within-class radius any pairwise-to-centroid bound must subtract being 27 times the floor, although the centroids *are* separated.

What the certified 1-NN rule delivers, and why it is not more. Corollary 3.2 accepts a test module when $\varepsilon < \rho_-(\tau)/2$. Measured on Orbit5k over $K \in \{8, \ldots, 200\}$ and five choices of τ (`cert_1nn.py`, 500 train / 400 test), **coverage never exceeds** 1.5% **and is** 0 **in most cells**, at the $\alpha = 1.75$ default and at the headline configuration alike. It is *not* that the floor is loose: granting a perfectly tight certificate leaves coverage below 1.5% as well. The binding constraint is that the median nearest-training-neighbor distance (0.056 at $K = 200$) is several times the smallest cross-class distance. Four further levers—the landmark budget, sampling density, the ℓ^∞ reading of the floor, and the set-valued relaxation—fail at the same wall (Appendix E, Figure 1). They share a cause: at $K = 200$ the median ratio $d_{(2)}/d_{(1)}$ of the distances to the nearest training module of the runner-up and of the nearest class is 1.10—the classes are interleaved at nearest-neighbor scale.

Local certificates, global accuracy. The conclusion we draw is not that the embedding is weak. On the *same* embedding a q-tuned RBF-SVM reaches 91.2% where 1-NN reaches 78%: the discriminative information is there, but it is *global*, carried by the kernel across all training points, and not concentrated in local neighborhoods. Every acceptance rule we have tested is local—ε to one neighbor, a ball, a per-class nearest distance—and each inherits the same weak local geometry, which is why five independent levers fail at the same wall. This reframes the certified/accurate gap as something other than a deficiency of the nearest-centroid rule: ρ_- certifies a *local* separation, and the accuracy lives in *global* structure. Attaching the guarantee to the kernel head is possible—each one-vs-one decision function is Lipschitz in Ψ with constant $L_{ab} = e^{-1/2}\sigma^{-1}\sum_i |\alpha_i^{ab}|$, and Proposition 2.2 converts a margin into a radius in d^S_{mat}—and vacuous: measured (`cert_svm.py`, $\alpha = 4$, capped), the median radius is 1.3×10^{-6}, three orders below ρ_-, because L_{ab} runs from 3×10^2 to 1.5×10^4; even $L = 1$ would leave it near 7×10^{-3}, below one outlier's displacement. A guarantee for the accurate head, if one exists, is not a Lipschitz one; nothing in the argument is particular to persistence, and we report it as measured on Orbit5k, not proved in general. Two further measurements are in Appendix E: ranking test points by the empirical centroid margin gives a

discriminative selective classifier (Figure 2), though only 1.5% certify under the strict threshold of Theorem 3.1; and reading that theorem per prediction (Proposition D.5, Table 8) moves the guarantee from operationally empty to covering most predictions on three of eleven benchmarks (74.0% of NCI1, 71.4% of DD, 68.7% of NCI109), still for the nearest-centroid head, which trails the kernel head by 18 points.

5 Adapting the filtration: a certified spectrum

Everything so far fixes the bifiltration; we now relax that along a spectrum—*fixed* $\rightarrow$ *selected* $\rightarrow$ *learned*—every step of which keeps the certificate, since the floor of Theorem 2.3 constrains the *output* embedding however the bifiltration is produced. **SELECT-COMPLEX** ranks candidate bifiltration pairs from a bank of node functions on the training fold by a shrinkage-regularized Fisher–Mahalanobis margin (20) and freezes the best pair (Appendix E); it pays only where the fixed pair is genuinely suboptimal (Table 9). **SELECT-MULTI** selects the *set*, keeping the top $k = 5$ of the 21 pairs by inner nearest-centroid accuracy at the fixed total budget; against the fixed three-pair bank ("sel." in Table 2) it reads $+3.7$ on NCI1 and $+1.7$ on DHFR at a small cost on the smaller sets—closed-form selection is unreliable at the single-pair level and useful at the set level. **D-COMPLEX** learns the bifiltration: a network emits two positive grades per node, $g_\theta(v) = \mathrm{softplus}(f_\theta(X_v)) \in \mathbb{R}^2_{>0}$, in place of the hand-picked scale and codensity, the net (11) staying frozen since learning the slice directions overfits; along slice s a node appears at $u_v^{(s)} = \max(g_\theta(v)_1/\cos\vartheta_s, g_\theta(v)_2/\sin\vartheta_s)$ and a simplex at $s_\sigma^{(s)} = \max_{v\in\sigma} u_v^{(s)}$, a valid filtration for every g_θ.

Proposition E.1 (Appendix E) gives the exact gather gradient: wherever the persistence pairing is locally constant and the maximizing vertex of every birth and death simplex is unique, each bar endpoint is differentiable in g_θ with derivative $1/\cos\vartheta_s$ or $1/\sin\vartheta_s$ on the attaining grade of that vertex, so the slice-stack is differentiable in θ; the uniqueness condition is not generic—on labeled molecular graphs ties are structural—and there the gather rule returns a Clarke subgradient. The certificate's *hypotheses*—admissibility and coherence—are functions of g_θ, and a cold start destroys admissibility outright (Appendix E), so the certificate is re-audited after training (`dcomplex_audit.py`: MUTAG, DHFR, PROTEINS, same pairs, frozen landmarks and τ): coherence and the ρ_- conclusion hold on 100% of audited pairs for the learned filtration on all three sets. D-COMPLEX adds a hinge margin term to the cross-entropy, a differentiable surrogate for the embedded separation that a purely accuracy-trained model such as D-GRIL cannot express; it needs the PLACE/PALACE head and a warm start at the fixed filtration or fails outright (Appendix E), and with both the learned filtration helps on three of five sets and hurts on two (Table 10). Gradient-shaped adaptation buys no reliable gain over closed-form selection; the differentiable route earns its place as the way to *certify a trained model*—the cell D-GRIL leaves empty—not as a path to accuracy.

6 Discussion

COMPLEX lifts the closed-form, certified single-parameter kernel of PLACE/PALACE to multi-parameter modules without leaving closed form or conceding accuracy, and the certificate survives adapting the bifiltration; a finite net certifies the sliced surrogate $d_{\mathrm{mat}}^S \leq d_{\mathrm{mat}}$, not the matching distance itself, and the constants depend on the bar budget N, intrinsically (Carrière & Bauer, 2019; Zava, 2025). The certified/accurate gap—73.4 against 91.2 from the same run—remains the honest limitation: a local certificate against a global head, closed for every Lipschitz route.

AI use statement

In this work, we used generative AI tools for two tasks: writing and debugging experiment and analysis code, and editing prose (grammar, wording, and condensing the full-length manuscript to the conference page limit). We have not used generative AI tools to generate research ideas, to produce or verify mathematical statements or proofs, to design experiments, or to produce, analyse or interpret experimental results. We have reviewed all AI-assisted work: AI-written code was checked and tested by the authors, and every AI-edited passage was read against the original for meaning, with all claims and numbers verified against the experiment logs. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Reproducibility statement

All proofs deferred from the main text are in Appendix C. The protocol, baselines, the priced conventions, the certificate-1NN sweep, the selection criteria, the per-prediction certificate and the D-COMPLEX training conditions are in Appendices B–E; every measurement names the script in the released pipeline that produces it. A full-length version of this paper, with every result and discussion in the main text, is available as a preprint.

References

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101. Originally arXiv:2107.07511.

Pramita Bagchi, Sushovan Majhi, Atish Mitra, and Žiga Virk. Statistical inference for persistence diagrams via landmark embeddings: Minimax theory and finite approximation. arXiv:2609.07691, 2026.

Magnus Bakke Botnan, Steffen Oppermann, and Steve Oudot. Signed barcodes for multi-parameter persistence via rank decompositions. In *38th International Symposium on Computational Geometry (SoCG 2022)*, volume 224 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 19:1–19:18. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2022. doi: 10.4230/LIPIcs.SoCG.2022.19.

Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009. doi: 10.1007/s00454-009-9176-0.

Mathieu Carrière and Ulrich Bauer. On the metric distortion of embedding persistence diagrams into separable Hilbert spaces. In *Proceedings of the 35th Annual Symposium on Computational Geometry (SoCG)*, pp. 21:1–21:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, June 2019.

Mathieu Carrière and Andrew J. Blumberg. Multiparameter persistence image for topological machine learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020.

Mathieu Carrière, Marco Cuturi, and Steve Oudot. Sliced Wasserstein kernel for persistence diagrams. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pp. 664–673. PMLR, 2017.

Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda. PersLay: A neural network layer for persistence diagrams and new graph topological signatures. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108, pp. 2786–2796, 2020.

Mathieu Carrière, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hariprasad Kannan, and Yuhei Umeda. Optimizing persistent homology based functions. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pp. 1294–1303, 2021.

Andrea Cerri, Barbara Di Fabio, Massimo Ferri, Patrizio Frosini, and Claudia Landi. Betti numbers in multidimensional persistent homology are stable functions. *Mathematical Methods in the Applied Sciences*, 36(12):1543–1557, 2013. doi: 10.1002/mma.2704.

Yuzhou Chen, Jose Frias, and Yulia R. Gel. TopoGCL: Topological graph contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 11453–11461, 2024.

Olympio Hacquard and Vadim Lebovici. Euler characteristic tools for topological data analysis. *Journal of Machine Learning Research*, 25(240):1–39, 2024.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 22118–22133, 2020.

Michael Kerber and Florian Russold. Graphcode: Learning from multiparameter persistent homology using graph neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, 2024.

Michael Kerber, Michael Lesnick, and Steve Oudot. Exact computation of the matching distance on 2-parameter persistence modules. In *35th International Symposium on Computational Geometry (SoCG 2019)*, volume 129 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 46:1–46:15. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2019. doi: 10.4230/LIPIcs.SoCG.2019.46.

Sehun Kim. xPerT: Extended Persistence Transformer, 2024. URL https://arxiv.org/abs/2410.14193.

Claudia Landi. The rank invariant stability via interleavings. In *Research in Computational Topology*, volume 13 of *Association for Women in Mathematics Series*, pp. 1–10. Springer, 2018. doi: 10.1007/978-3-319-89593-2_1.

Tam Le and Makoto Yamada. Persistence Fisher kernel: A Riemannian manifold kernel for persistence diagrams. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, pp. 10028–10039, 2018.

Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015. doi: 10.1007/s10208-015-9255-y.

Michael Lesnick and Matthew Wright. Interactive visualization of 2-d persistence modules, 2015. RIVET software.

Jacob Leygonie, Steve Oudot, and Ulrike Tillmann. A framework for differential calculus on persistence barcodes. *Foundations of Computational Mathematics*, 22(4):1069–1131, 2022.

David Loiseaux, Mathieu Carrière, and Andrew J. Blumberg. A framework for fast and stable representations of multiparameter persistent homology decompositions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023a.

David Loiseaux, Luis Scoccola, Mathieu Carrière, Magnus Bakke Botnan, and Steve Oudot. Stable vectorization of multiparameter persistent homology using signed barcodes as measures. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023b.

Sushovan Majhi, Atish Mitra, Žiga Virk, and Pramita Bagchi. A closed-form persistence-landmark pipeline for certified point-cloud and graph classification. *Transactions on Machine Learning Research*, 2026a. ISSN 2835-8856. URL https://openreview.net/forum?id=4kZxNlE5Ve.

Sushovan Majhi, Atish Mitra, Žiga Virk, and Pramita Bagchi. A closed-form adaptive-landmark kernel for certified point-cloud and graph classification, 2026b. URL https://arxiv.org/abs/2605.04046.

Soham Mukherjee, Shreyas N. Samaga, Cheng Xin, Steve Oudot, and Tamal K. Dey. D-GRIL: End-to-end topological learning with 2-parameter persistence. In *42nd International Symposium on Computational Geometry (SoCG)*, volume 367 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 79:1–79:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi: 10.4230/LIPIcs.SoCG.2026.79. arXiv:2406.07100.

Iosif Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Annals of Probability*, 22(4):1679–1706, 1994.

Raphael Reinauer, Matteo Caorsi, and Nicolas Berkouk. Persformer: A transformer architecture for topological machine learning, 2021. URL https://arxiv.org/abs/2112.15210.

Oliver Vipond. Multiparameter persistence landscapes. *Journal of Machine Learning Research*, 21 (61):1–38, 2020.

Table 3: The landscape of topological features computed from a multiparameter filtration (ECS reads the filtration but computes no persistence). Every prior method is uncertified; this paper supplies the certified version of *both* regimes.

	uncertified	**certified (this paper)**
training-free	mp-landscape/image, signed measures, ECS	**COMPLEX**
trained	GRIL, D-GRIL, graphcode, Persformer	**D-COMPLEX**

Cheng Xin, Soham Mukherjee, Shreyas N. Samaga, and Tamal K. Dey. GRIL: A 2-parameter persistence based vectorization for machine learning. In *Proceedings of the 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*, volume 221 of *Proceedings of Machine Learning Research (PMLR)*, pp. 313–333, 2023.

Huan Xu and Shie Mannor. Robustness and generalization. *Machine Learning*, 86(3):391–423, 2012.

Nicolò Zava. Coarse and bi-Lipschitz embeddability of subspaces of the Gromov–Hausdorff space into Hilbert spaces. *Algebraic & Geometric Topology*, 25(8):5153–5174, 2025. doi: 10.2140/agt.2025.25.5153.

A RELATED WORK

Multiparameter vectorizations. Most feature maps for multiparameter modules read the fibered barcode and vectorize it: multiparameter persistence landscapes (Vipond, 2020) and images (Carrière & Blumberg, 2020) sample a piecewise-linear or convolved summary over a line net; signed barcodes and Hilbert/Euler signed measures (Botnan et al., 2022; Loiseaux et al., 2023b;a) compare modules through sliced-Wasserstein kernels. All are 1-Lipschitz *upper* bounds—small module distance implies small feature distance—but none supplies the converse lower bound that a separation guarantee needs. COMPLEX reuses the same fibered-barcode input and the same near-diagonal line net, but embeds each slice by the closed-form PLACE/PALACE landmark map (Majhi et al., 2026a;b), which *does* carry a two-sided certificate, and lifts it to the module by a single witnessing slice.

Learned multiparameter features. GRIL (Xin et al., 2023) generalizes the persistence landscape to a 2-parameter rank-invariant landscape, and D-GRIL (Mukherjee et al., 2026) makes the underlying bifiltration *learned*, differentiating the landscape through a graph neural network. Euler-characteristic surfaces with gradient-boosted trees (Hacquard & Lebovici, 2024) bypass homology for a large speed gain, transformer architectures (Reinauer et al., 2021; Kim, 2024) read the diagrams directly, and graphcodes (Kerber & Russold, 2024) feed a two-parameter hierarchy of barcodes to a graph neural network, previously the strongest reproducible accuracy at the Orbit100k scale (92.3; Section 4). These are the trained or tuned methods COMPLEX aims to match with no learned embedding. On graphs its natural comparison within the multiparameter family is GRIL, which it exceeds on all four shared molecular sets (Section 4) while additionally certifying its predictions—something a learned pipeline cannot express. Beyond that family, trained topological pipelines that couple persistence with graph neural networks or attention—TopoGCL (Chen et al., 2024) among them—report higher accuracy on several TU benchmarks. Section 4 names them where they lead; the comparison this paper makes is with no learned embedding and no held-out calibration.

Certified single-parameter embeddings. PLACE (Majhi et al., 2026a) and its data-adaptive successor PALACE (Majhi et al., 2026b) give a closed-form landmark embedding of one-parameter diagrams with a label-derived lower distortion certificate and a per-prediction nearest-centroid guarantee. Both are deterministic and split-free, and they certify the geometry of a prediction—its stability radius and its agreement with the population rule—rather than the marginal coverage of the true label that distribution-free conformal prediction (Angelopoulos & Bates, 2023) guarantees under exchangeability from a calibration split. The two are complementary rather than substitutes: any certificate margin is itself a valid nonconformity score, so a conformal wrapper can be placed on top of the certified head without touching the certificate. COMPLEX is their multiparameter lift: the per-slice certificate is PALACE's, and the new content is the slice-stack that carries it to a module through a single witnessing slice, certifying the sliced matching distance d_{mat}^{S}.

Table 4: Notation.

$\ell_{\beta,u},\ s_{\beta,u},\ \omega$	slice, entry parameter, weight	$\nu = \{(p_k, r_k, w_k)\}$	landmarks: points, radii, weights
$\mathcal{L}_S,\ \vartheta_s,\ \omega_s$	net, its angles and weights	$\varphi_{p,r},\ \Phi,\ \Psi$	hat coordinate, slice map, stack
$\mathcal{D}_N,\ d_{\mathcal{B}},\ d_{\mathrm{mat}}^S$	diagrams, bottleneck, sliced matching	$K,\ S,\ N_{\max},\ \alpha$	landmarks/slice, slices, bar budget, radius factor
$\tau,\ t^\star = \tau/\omega_{s^\star}$	scale, its realized value on $s^\star$	$\lambda_0,\ \mathcal{A}_s(\tau),\ w_{\min}$	func. Lebesgue number, active set, weight
$\rho_-(\tau) = \frac{1}{4}\tau w_{\min}$	certified floor	$B,\ B_\infty$	norm and coordinate bounds
$\widehat{\Delta},\ r_m,\ \mathrm{gap}(M)$	centroid gap, radius, decision gap	$\varepsilon,\ d_{(1)},\ d_{(2)}$	nearest-neighbor distances
$\kappa,\ h,\ \kappa_-,\ c_D$	kernel, bandwidth, ceiling, constant	$g_\theta,\ \theta$	learned bifiltration, its parameters

B Remarks on the construction

The one-point metric, offsets, and conventions. The space of one-point diagrams $\mathcal{D}_1 = T \cup \{\Delta\}$ is metrized by

$$d_{\mathcal{B}}\big((b,d),\Delta\big) = \tfrac{1}{2}(d-b), \qquad d_{\mathcal{B}}(x,y) = \min\big\{\|x-y\|_\infty,\ \max\{d_{\mathcal{B}}(x,\Delta), d_{\mathcal{B}}(y,\Delta)\}\big\}, \quad (14)$$

so two points are compared directly or by sending both to the diagonal, whichever is cheaper; the padding of (5) embeds $\mathcal{D}_k \hookrightarrow \mathcal{D}_N$ isometrically for $k \le N$. Both parameters of a slice (1) earn their place. A comparable pair $\gamma \le \gamma'$ lies on a common ray through the origin only when γ' is a positive multiple of γ, so the directions alone read a one-parameter subfamily of the slices, and the offsets are what let the fibered barcode see the rest. (A slice through both γ and γ' exists exactly when $\gamma'_i > \gamma_i$ in every coordinate; the equivalence with the full rank invariant extends to the remaining comparable pairs by constructibility (Cerri et al., 2013; Lesnick & Wright, 2015).) They also make the construction translation-covariant: translating a module and every base point of the net by the same $c \in \mathbb{R}^d$ leaves each slice barcode unchanged. Two consequences, worth separating. The continuous matching distance d_{mat} is a supremum over *all* slices, and translation permutes that family, so d_{mat} does not depend on where a module sits in the orthant—which is what makes it the right target for the finite-net surrogate (4) to approximate. A fixed finite net inherits less: the nets of Section 2.5 set $\beta_s = 0$ for every module, so d_{mat}^S is unchanged by a translation applied to the data set together with the net, but not by translations applied to modules individually. We apply neither, since the grades above are already non-negative. Normalizing u in ℓ^∞ instead rescales every slice barcode by $\max_i u_i$ and replaces (3) by $\min_i u_i / \max_i u_i$; the product $\omega\, d_{\mathcal{B}}$—and hence everything below—is unchanged, but the two conventions must not be mixed. That second form is the classical one: in $d = 2$ a line of slope m has ℓ^∞-normalized direction $(1,m)/\max\{1,m\}$ and weight $\min\{m, 1/m\}$, the weight of Cerri et al. (2013). Our d_{mat} is therefore their matching distance, written in the arclength parametrization; the ℓ^2 convention is a reparametrization, not a different distance.

Admissibility, read literally. The functional Lebesgue number of a configuration ν covering the support $\mathcal{X}$ is

$$\lambda_0(\nu) \;=\; \inf_{x \in \mathcal{X}} \ \max_k \ \varphi_{p_k,r_k}(x). \quad (15)$$

The qualifier is meant literally. This is not the Lebesgue number of the cover $\{B(p_k, r_k)\}$ in the covering sense—nothing here asserts that a ball of some radius about any point fits inside a single cover element—but its functional analogue: the height of the lowest point of the upper envelope of the hats over the support. The two meet at one identity, since $\varphi_{p,r}(x) \ge \rho$ exactly when $x \in \overline{B}(p, r-\rho)$: $\lambda_0(\nu) \ge \rho$ holds precisely when the uniformly shrunk cover $\{\overline{B}(p_k, r_k - \rho)\}_k$ still contains $\mathcal{X}$, so λ_0 is how far every radius can be pulled in before coverage fails. Reading it through the functions rather than the cover is what makes it computable from the embedding and what lets it enter the floor $\frac{1}{4}\tau w_{\min}$ as a number rather than a hypothesis.

Homological degree. Slicing and embedding are applied degree by degree: M above denotes the degree-q module, d_{mat}^S the degree-q sliced matching distance, and the descriptor of a data set is the concatenation of the blocks $\Psi(M^{(q)})$ over the degrees used. Everything below is stated for one degree and extends to the concatenation with d_{mat}^S replaced by $\max_q d_{\mathrm{mat}}^S$. We flag this because pooling the degrees into a single multiset before embedding—the single-parameter habit, and what our point-cloud implementation inherits—is *not* harmless in the multiparameter setting. Pooling grants the optimal matching extra freedom, so $d_{\mathcal{B}}(A_0 \cup A_1,\ B_0 \cup B_1) \le \max_q d_{\mathcal{B}}(A_q, B_q)$. The upper bound (Proposition 2.2) therefore survives pooling, but the hypothesis $d_{\mathrm{mat}}^S \ge \tau$ of Theorem 2.3 does

not transfer to the pooled barcode, and the certificate is lost in exactly the direction our contribution lives in. In one parameter this costs nothing: the degree-q barcode is a *complete* invariant of the degree-q module, so pooling is a feature-map choice made on top of complete information. For $d \geq 2$ no complete discrete invariant exists (Carlsson & Zomorodian, 2009) and the fibered barcode is already lossy, so pooling compounds an avoidable loss on an unavoidable one, with no completeness to appeal to when a referee asks what the pooled object represents. We therefore take the per-degree stack as the definition, and price the pooled shortcut empirically in Section 4.

Remark B.1 (The radius cap). Proposition 2.2 needs (6), and for a reason prior to Lipschitzness: a hat whose radius exceeds $d_{\mathcal{B}}(p_k, \Delta)$ does not vanish at Δ, so $\varphi_{p_k,r_k}(A)$ depends on how many copies of Δ are used to present A as a point of $\mathcal{D}_N$, and the coordinate is not well defined independently of N. Under the cap the padded and unpadded sums agree, which is what the proof uses. Lipschitzness fails with it: for $p = (0, 10)$ the cap is 5, and taking $r = 6$, $A = \{(4.9, 5.1)\}$, $B = \varnothing$ gives $d_{\mathcal{B}}(A, B) = 0.1$ while the coordinate moves by 1.1. The released implementation, following PALACE (Majhi et al., 2026b), sets $r_k = \alpha \cdot d_{\mathrm{NN}}(p_k)$ (a multiple of the nearest-landmark distance) clipped to $[\tau/2, 4\tau]$ and does not enforce the cap: on Orbit5k ($S = 8$, $K = 200$) 10% of landmarks violate it at the default $\alpha = 1.75$ and 26% at the headline $\alpha = 4$. We therefore run two variants: the *uncapped* embedding, which carries the accuracy headline, and the *capped* one, to which Proposition 2.2 and through it the stability radius of Corollary 3.2(ii) attach. The gap between them—under half a point on Orbit5k at the headline $\alpha = 4$, a full point at the $\alpha = 1.75$ default, and 0.2 on Orbit100k—is the embedding-level price of the certificate, the same trade this paper measures at the head level; Section 4 prices it.

Remark B.2 (The block weights, and what the classifier reads). Definition 2.1 weights block s by ω_s. The released classification pipeline stacks the blocks unweighted, so every accuracy number in this paper is measured on $\Psi^\circ = \big(\Phi(M|_{\ell_s}; \nu_s)\big)_{s=1}^{S}$, the image of Ψ under the fixed invertible diagonal map $\mathrm{diag}(\omega_s^{-1})$; the audits of Section 4 build the weighted stack of Definition 2.1. Both gauges survive the rescaling, with different constants, so Ψ° is certified on the same hypotheses. The lower gauge does not weaken. Coherence bounds a single coordinate of the witnessing block below by $w_{k^\star} t^\star/4 = w_{k^\star}\tau/(4\omega_{s^\star})$, and it is the prefactor $\omega_{s^\star}$ in (8) that cancels it down to $\frac{1}{4}\tau w_{k^\star}$; without that prefactor the bound survives intact, giving $\|\Psi^\circ(M) - \Psi^\circ(M')\| \geq \rho_-(\tau; \mathcal{C})/\max_s \omega_s$. The upper gauge weakens by $1/\omega_{\min}$, each block now carrying $d_{\mathcal{B}}$ in place of $\omega_s d_{\mathcal{B}}$: $\|\Psi^\circ(M) - \Psi^\circ(M')\| \leq \sqrt{S}\, N_{\max}\, d_{\mathrm{mat}}^S(M, M')/\omega_{\min}$. The two-sided distortion is therefore worse by the fan's weight spread $\max_s \omega_s/\omega_{\min}$, which is $0.644/0.184 = 3.5$ at the default net—the price of the unweighted stack, paid in the constant of Proposition D.1 and nowhere else. Which stack to classify is then an accuracy question, and Section 4 measures it: a wash.

Remark B.3 (What the guarantees are about). $N_{\max}$ is the per-slice bar budget defining $\mathcal{R}_N$; it enters both gauges because $\varphi_{p,r}(A)$ sums over the points of A. The implementation keeps the 50 most persistent bars per slice, and on Orbit5k that cap is active on 99.5% of slices. The budget is deliberate rather than incidental: it is what keeps the computation manageable, bounding both the cost of a coordinate—$O(K|A|)$ on a slice—and that of the pairwise audits, which are quadratic in the bar count. Truncation $T_{50}\colon \mathcal{D}_N \to \mathcal{D}_{50}$ is not 1-Lipschitz for $d_{\mathcal{B}}$: two diagrams within ε can order their 50th and 51st bars differently and truncate to diagrams far apart. Every result in this paper therefore holds verbatim on $\mathcal{R}_{50}$, i.e. for the truncated fibered barcode—the invariant the pipeline computes, and the one on which d_{mat}^S, τ and the audit of Section 4 are evaluated; we claim no bound relating it to the untruncated module, which would need a gap between the 50th and 51st persistences. A soft, genuinely 1-Lipschitz alternative to top-n is the clean fix and we leave it open. The other bar-level convention—capping essential classes at the largest finite death (Section 4)—is likewise measured rather than assumed: replacing it by extended persistence moves the six TU sets by -0.46 points on average, helping three and hurting three with swings to ± 3.8 on a single seed, so the cap costs nothing measurable.

Remark B.4 (What is and is not angle-independent). The *value* in (10) does not depend on the witnessing angle: $\omega_{s^\star}$ cancels, and under equal weights $\rho_-(\tau) = \tau/(4\sqrt{K})$ whatever the net. The *cost* of earning it does. By (9) a slice with small ω_s has a nonempty active set only if its landmarks have radii larger by the factor $1/\omega_s$, and is τ-admissible (Section 2.4) only if its cover reaches that inflated scale—at a fixed budget K, covering a coarser scale. The near-diagonal fan of Section 2.5 is thus not merely a heuristic default: it is the cheapest net that is τ-admissible at a given landmark budget.

Remark B.5 (Admissibility is not satisfied by the configurations we run). We state plainly that the configurations used in every experiment of this paper are *not* τ-admissible, at any τ, and separate

what this costs from what it does not (`admissibility_check.py`). On the certified capped variant at the audit scale $\tau = 1.220$ the functional Lebesgue number is $\lambda_0 = 0$ on all eight slices, because the cap deliberately excludes the near-diagonal mass and 1.8–2.2% of the support then lies under no hat at all; without the cap λ_0 is positive (0.105–0.724) but still an order below the required $t_s/4$, and on the outer slices the largest radius exceeds $(t_s + \lambda_0)/2$ as well (5.341 against 3.736 on slice 7). Theorem 2.3 is unaffected: it consumes nonempty active sets and coherence, both of which we measure directly and both of which hold at the headline configuration (100% of audited pairs, Section 4). What admissibility is a hypothesis of is the *derivation* of coherence—Proposition B.7—and the two statements that inherit it, Corollary 3.2 and Proposition D.1; for those, the hypothesis is currently verified by audit rather than by construction. The gap is closable in principle and we priced it: a greedy maximal $(t_s/4)$-separated net with uniform radii $t_s/2$, admissible by construction on the support it is built from, needs 11–39 landmarks per slice against 200 and reads $90.3 \pm 1.1\%$ against the radius rule's $91.2 \pm 0.9\%$ with the same head on Orbit5k (3 folds, 218 embedding dimensions against 1,600; nearest centroid is 0.7 points *better* at 74.0%). Even that net leaves up to 0.03% of the support below $t_s/4$, so what is achievable cheaply is admissibility on the covered support with the uncovered fraction reported, not on the full support; we regard closing this as the first thing the construction owes, and we prefer to report the number than to let the hypothesis stand unexamined.

Remark B.6 (Why coherence is a hypothesis, and how it fails). It is fair to ask why this is assumed rather than derived from τ-admissibility. The obstruction is sum-pooling. Each coordinate reads $\varphi_{p_k,r_k}(A) = \sum_{a\in A} \varphi_{p_k,r_k}(a)$, a many-to-one map, so a single coordinate can be blind to a rearrangement of mass within its cone: one point at ℓ^∞-distance $r - t$ from p_k, contributing t, and two points at distance $r - t/2$, contributing $t/2$ each, give the same value, however far apart the two diagrams are in $d_{\mathcal{B}}$. Admissibility guarantees *coverage*—some landmark sees the region where the two diagrams differ—but coverage does not by itself guarantee that the landmark which sees the difference also *reports* it. Coherence asks that not all K coordinates be simultaneously blind, which is a conjunction of K inequalities and so fails rarely but genuinely: on Orbit5k it fails on 2.5% of separated cross-class pairs at the $\alpha = 1.75$ default, and on none at the headline $\alpha = 4$ (Section 4). The failures are informative rather than fatal, because Theorem 2.3's *conclusion* holds on 100% of the pairs we audit—coherence is sufficient, not necessary—but we prefer a checkable hypothesis with a measured failure rate to an unchecked one. Nor is the hypothesis circular. Coherence is a statement about *one* coordinate clearing a fixed floor, verified independently of the ℓ^2 conclusion it implies, and strictly stronger than it: the pairs on which coherence fails while the conclusion holds are exactly those where several coordinates each fall short of the floor and their sum does not—the compensation regime PALACE identifies (Majhi et al., 2026b).

The sufficient condition.

Proposition B.7 (Within-module sufficient condition)**.** *Write* $\omega_{\min} = \min_{1\le s\le S} \omega_s$ *and let* D *be any a-priori upper bound on* d_{mat}^S *over the data set. Let* (M, M') *be a cross-class pair with* $d_{\mathrm{mat}}^S(M, M') \ge \tau$ *and* $\mathcal{C}$ τ*-admissible, and suppose that,* along every slice of the net*,*

$$\textit{(i)}\quad \min_{i\neq j} \min_{1\le s\le S} \left|s_{\beta_s,u_s}(\gamma_i) - s_{\beta_s,u_s}(\gamma_j)\right| \;>\; \frac{4D}{\omega_{\min}} \quad \textit{for the generator and relation grades } \gamma_1, \gamma_2, \dots \textit{ of } M, \tag{16}$$

(ii) every bar of $M|_{\ell_s}$ *and of* $M'|_{\ell_s}$ *is longer than* $2D/\omega_{\min}$ *for every* s*,*

where $s_{\beta,u}(\cdot)$ *is the entry parameter* (2) *on the slice* $\ell_{\beta,u}$*. Then* (M, M') *is* $\mathcal{C}$*-coherent. Both hypotheses are read off the two barcodes separately—(i) from* M*, (ii) from each of* M *and* M'*—with no cross-pair matching.*

Proof. Appendix C.

Remark B.8 (Two-level audit). Like PLACE's disjoint-activation condition (Majhi et al., 2026a), (16) is a strong geometric condition; on chemical graph data it may hold on few pairs. Coherence (Definition 2.4) is far weaker, and we report its empirical rate directly: the fraction of cross-class pairs with $d_{\mathrm{mat}}^S \ge \tau$ that are $\mathcal{C}$-coherent, and the fraction for which the embedded distance clears ρ_- (Section 4). The classifier of Section 3 operates at the class-margin level, independent of pairwise coherence, so accuracy does not hinge on the audit.

The slice-selection rule. The default ignores the labels. But the discriminating structure of a module may live along an off-diagonal direction—a feature visible only at a particular slope, with

no one-parameter analogue—which the near-diagonal fan misses. Since the certificate value does not degrade off the diagonal, we may chase such directions and still certify them: select the net data-adaptively subject to a certifiability constraint, exactly as PALACE selects landmarks (FPS, frozen) one level down. Adaptive landmarks ask *where to look within a slice*; adaptive slices ask *which one-parameter projection of the module to expose*.

A single embedding needs one net for all modules, so the net is selected once from the training labels and *frozen* (like FPS landmarks); the distance is already pair-adaptive through the witnessing slice, and only the net is data-adaptive. Given a dense candidate fan $\mathcal{U}$ and a per-slice discrimination score $g(u)$ computed in closed form from the training embeddings (e.g. the class-margin $\widehat{\Delta}/\sqrt{K}$ of (Majhi et al., 2026b), $\widehat{\Delta}$ being the minimum inter-centroid gap of Section 3):

1. *Certifiability filter.* Build ν_u by FPS on the training diagrams of slice $\ell_{0,u}$, and keep $u \in \mathcal{U}$ if ν_u is admissible at its realized scale $\tau/\omega(u)$ and clears a design floor, $\rho_-(\nu_u) \geq \rho_0 > 0$. (Under equal weights $\rho_-(\nu_u) = \tau/(4\sqrt{K})$ for every u, so the floor test has teeth only for non-uniform weights and admissibility is the operative filter.)
2. *Score.* Rank the certifiable directions by $g(u)$.
3. *Diversify.* From the top-ranked pool, pick S directions by farthest-point sampling in angle (so the slices are not redundant), seeded by the highest score.
4. *Freeze.* Fix the selected net across all modules.

Everything is closed-form: g has an explicit formula, the selection is greedy, and nothing is tuned by gradient or held-out search. We state the rule over directions, which is what we ran; it applies verbatim to candidate pairs (β, u), since every step reads only $\omega(u)$ and the configuration built on the slice.

Remark B.9 (Selection does not break the guarantee). The net is supervised, so it can overfit a separating direction. Three things contain it: selection runs on the training fold and is frozen, so cross-validation stays honest; the certifiability filter shrinks the admissible slice space, ruling out arbitrary separating lines; and the diversity step blocks redundant directions. The certified-adaptive net is the strict generalization of the fixed fan (11), recovered when g is constant. Empirically the adaptive net falls at or slightly below the fixed fan (Section 4), so we present slice selection as certificate-preserving generality, not an accuracy lever. The adaptivity that *does* pay comes from selecting the bifiltration *set* rather than the net (Section 5).

C Deferred proofs

Proof of Theorem 2.3

Proof. The ℓ^2 norm dominates its $s^\star$ block. Coherence supplies a landmark $k^\star \in \mathcal{A}_{s^\star}(\tau)$ realizing the single-coordinate floor on the witnessing slice at the realized scale $t^\star = \tau/\omega_{s^\star}$, $|\Phi_{k^\star}(M|_{\ell_{s^\star}}) - \Phi_{k^\star}(M'|_{\ell_{s^\star}})| \geq w_{k^\star}\, t^\star/4$. The $s^\star$ block of Ψ carries the weight $\omega_{s^\star}$ (8), so its contribution is at least $\omega_{s^\star} w_{k^\star} t^\star/4 = \frac{1}{4}\tau\, w_{k^\star} \geq \frac{1}{4}\tau\, w_{\min}(\tau; \nu_{s^\star})$; the reparametrization weight cancels. Taking the minimum over s removes the dependence on which slice witnesses the pair, which is what makes ρ_- a property of the configuration alone—computable once, before any pair is examined. □

Proof of Proposition 2.2

Proof. Fix a slice, write $A = M|_{\ell_s}$, $B = M'|_{\ell_s}$ and $\delta = d_{\mathcal{B}}(A, B)$, represent both as Δ-padded N-tuples $[x_1, \ldots, x_N]$, $[y_1, \ldots, y_N]$, and let $\psi \in S_N$ realize δ in (5). For a landmark (p, r) obeying (6), $\varphi_{p,r}$ is 1-Lipschitz for $d_{\mathcal{B}}$ on $\mathcal{D}_1$ and vanishes at Δ, so

$$|\varphi_{p,r}(A) - \varphi_{p,r}(B)| \;\leq\; \sum_{i=1}^{N} \big|\varphi_{p,r}(x_i) - \varphi_{p,r}(y_{\psi(i)})\big| \;\leq\; \sum_{i=1}^{N} d_{\mathcal{B}}\big(x_i, y_{\psi(i)}\big) \;\leq\; N_{\max}\, \delta,$$

the padding entries included (the Δ-terms contribute $|\varphi_{p,r}(x_i)| \leq d_{\mathcal{B}}(x_i, \Delta) \leq \delta$). Squaring, summing over landmarks and using $\sum_k w_k^2 = 1$ gives $\|\Phi(A; \nu_s) - \Phi(B; \nu_s)\| \leq N_{\max}\, d_{\mathcal{B}}(A, B)$. Finally, by (8), $\|\Psi(M) - \Psi(M')\|^2 = \sum_s \omega_s^2\, \|\Phi(A_s) - \Phi(B_s)\|^2 \leq N_{\max}^2 \sum_s \big(\omega_s d_{\mathcal{B}}(A_s, B_s)\big)^2 \leq S\, N_{\max}^2\, d_{\mathrm{mat}}^S(M, M')^2$. □

Proof of Proposition B.7

Proof. Let $s^\star$ be the witnessing slice, $A = M|_{\ell_{s^\star}}$, $B = M'|_{\ell_{s^\star}}$, $\delta = d_{\mathcal{B}}(A, B)$ and $t^\star = \tau/\omega_{s^\star}$. Since $\omega_{s^\star}\delta = d_{\mathrm{mat}}^S(M, M') \in [\tau, D]$ and $\omega_{s^\star} \geq \omega_{\min}$, we have $t^\star \leq \delta \leq D/\omega_{\min}$. Two distinct points of A are bars that differ in a birth or a death, i.e. in the entry parameter of a generator or a relation, so by (i) they are more than $4D/\omega_{\min} \geq 4\delta$ apart in $\|\cdot\|_\infty$. By (ii) every point of $A \cup B$ has $d_{\mathcal{B}}(\cdot, \Delta) > D/\omega_{\min} \geq \delta$, so a matching ψ realizing δ in (5) pairs no point with a padding Δ (that would cost more than δ) and compares every matched pair directly, the diagonal route in (14) costing more than δ: $|A| = |B|$, every $a \in A$ is paired with $\psi(a) \in B$ at $\|a - \psi(a)\|_\infty \leq \delta$, and some pair $(a^\star, b^\star = \psi(a^\star))$ attains $\|a^\star - b^\star\|_\infty = \delta$.

Admissibility of $\nu_{s^\star}$ at the realized scale gives $\lambda_0 := \lambda_0(\nu_{s^\star}) \geq t^\star/4$ and $\max_k r_k \leq (t^\star + \lambda_0)/2$; since $\lambda_0 \leq \max_k r_k$ always, this forces $\max_k r_k \leq t^\star$. By the definition (15) of the functional Lebesgue number there is $k^\star$ with $\varphi_{k^\star}(a^\star) \geq \lambda_0$, i.e. $\|p - a^\star\|_\infty \leq r - \lambda_0$ for $p = p_{k^\star}$, $r = r_{k^\star}$; as $r \geq \varphi_{k^\star}(a^\star) \geq \lambda_0 \geq t^\star/4$, $k^\star \in \mathcal{A}_{s^\star}(\tau)$. Three facts locate the mass in the cone of $k^\star$. (1) For $a' \in A \setminus \{a^\star\}$: $\|a' - p\|_\infty \geq \|a' - a^\star\|_\infty - \|a^\star - p\|_\infty > 4\delta - r \geq 3\delta \geq r$, so $\varphi_{k^\star}(a') = 0$. (2) $\|b^\star - p\|_\infty \geq \|a^\star - b^\star\|_\infty - \|a^\star - p\|_\infty \geq \delta - r + \lambda_0 \geq t^\star - \frac{1}{2}(t^\star + \lambda_0) + \lambda_0 = \frac{1}{2}(t^\star + \lambda_0) \geq r$, so $\varphi_{k^\star}(b^\star) = 0$. (3) For $b \in B \setminus \{b^\star\}$, paired with some $a = \psi^{-1}(b) \neq a^\star$: if $\|b - p\|_\infty < r$ then $\|a - a^\star\|_\infty \leq \|a - b\|_\infty + \|b - p\|_\infty + \|p - a^\star\|_\infty < \delta + r + r \leq 3\delta$, contradicting the separation in (1); so $\varphi_{k^\star}(b) = 0$. Hence $\varphi_{k^\star}(A) = \varphi_{k^\star}(a^\star) \geq \lambda_0 \geq t^\star/4$ and $\varphi_{k^\star}(B) = 0$, so $|\Phi_{k^\star}(A) - \Phi_{k^\star}(B)| = w_{k^\star}\varphi_{k^\star}(a^\star) \geq w_{k^\star}t^\star/4$, which is Definition 2.4. Stating (i) along the slices is essential: an ℓ^∞ separation of the grades in $\mathbb{R}^d$ does *not* imply separation after restriction, since a slice contracts grade space—at $\vartheta = \pi/4$ the grades $(0, t)$ and $(t, 0)$ are ℓ^∞-far but enter at the same parameter. Hypothesis (ii) is what rules out short bars of M' inside the witnessing cone, which would otherwise be free to cancel the witness; without it the conclusion can fail. □

Proof of Proposition 2.5

Proof. Theorem 2.3 needs nonempty active sets on the net, which τ-admissibility supplies (Section 2.4), and coherence of the pair. The certifiability filter admits a direction u only if ν_u is admissible at its realized scale $\tau/\omega(u)$ and $\rho_-(\nu_u) \geq \rho_0$; a net assembled from admitted directions is therefore τ-admissible by construction, and the floor of (10), a minimum over the selected slices of $\frac{1}{4}\tau w_{\min}(\tau; \nu_u)$, is at least ρ_0. Coherence is a property of the pair on its witnessing slice, which the selection rule does not see, so the theorem applies verbatim once the net is frozen—the only role of freezing is that the net does not depend on the module being embedded. For the second claim, farthest-point sampling on the metric space of admitted angles with the arc metric is the greedy k-center heuristic, whose 2-approximation of the optimal covering radius is the guarantee PALACE proves for its landmark placement (Majhi et al., 2026b) and holds on any metric space; here it is applied to angles rather than diagram points. □

Lemma C.1 (boundedness of the slice-stack) and its proof

Lemma C.1 (Boundedness of the slice-stack)**.** *For every* $M \in \mathcal{R}_N$, $\|\Psi(M; \mathcal{C})\|_{\ell^2} \leq B := \sqrt{S}\, N_{\max} \max_{s,k} r_{s,k}$.

Proof. $0 \leq \varphi_{p,r} \leq r$ pointwise and a slice barcode carries at most $N_{\max}$ points, so $0 \leq \Phi_k(A; \nu_s) \leq w_{s,k} N_{\max} \max_k r_{s,k}$; squaring, summing over k and using $\sum_k w_{s,k}^2 = 1$ gives $\|\Phi(A; \nu_s)\| \leq N_{\max} \max_k r_{s,k}$. Summing the S blocks with $\omega_s \leq 1$ gives the claim. □

Proof of Theorem 3.1

Proof. Condition on the class labels, so that the class sizes m_c are fixed and the m_c embeddings $\Psi(M_i)$ with $y_i = c$ are i.i.d. from the class-conditional law; the bound proved under this conditioning holds unconditionally (classes with $m_c = 0$ have no empirical mean and are excluded from the maximum). $\Psi(M) - \mu_c$ is a centered $\mathbb{R}^{SK}$-valued variable with $\|\Psi(M) - \mu_c\| \leq 2B$ a.s. (Lemma C.1), so Pinelis's Bernstein inequality in a 2-smooth Banach space (Pinelis, 1994), applied in the Hilbert space $\mathbb{R}^{SK}$, gives $\Pr[\|\widehat{\mu}_c - \mu_c\| > r] \leq 2\exp\big(-\frac{m_c r^2}{2(V_c + 2Br/3)}\big)$. Setting the

right side to δ/C and solving the resulting quadratic, then using $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, yields the sufficient radius (12); a union bound over the C classes gives the coverage event. On that event $\big|\|x - \widehat{\mu}_c\| - \|x - \mu_c\|\big| \le \|\widehat{\mu}_c - \mu_c\| \le r_m$ for every x and c. Hence if $\mathrm{gap}(M) > 2r_m$, writing $\widehat{c} = \widehat{c}(M)$ and $x = \Psi(M)$, for every $c \neq \widehat{c}$

$$\|x - \mu_{\widehat{c}}\| \;\le\; \|x - \widehat{\mu}_{\widehat{c}}\| + r_m \;<\; \|x - \widehat{\mu}_c\| - 2r_m + r_m \;\le\; \|x - \mu_c\|\,,$$

so the population rule also selects $\widehat{c}$. Finally $\|\mu_c - \mu_{c'}\| \ge \widehat{\Delta} - 2r_m > 0$ under $r_m < \frac{1}{2}\widehat{\Delta}$. □

Proof of Corollary 3.2

Proof. (i) For $y_j \neq y_{i^\star}$, Theorem 2.3 gives $\|\Psi(M_{i^\star}) - \Psi(M_j)\| \ge \rho_-(\tau)$, so $\|\Psi(M) - \Psi(M_j)\| \ge \rho_-(\tau) - \varepsilon > \varepsilon = \|\Psi(M) - \Psi(M_{i^\star})\|$, the last step by $\varepsilon < \rho_-/2$; the gap is $(\rho_- - \varepsilon) - \varepsilon$. (ii) By Proposition 2.2 a d^S_{mat}-perturbation of size η moves $\Psi(M)$ by at most $\sqrt{S}\, N_{\max}\, \eta$, changing each of the two distances in (i) by at most that amount; the margin $\rho_- - 2\varepsilon$ survives whenever $2\sqrt{S}\, N_{\max}\, \eta < \rho_- - 2\varepsilon$. □

Proof of Proposition D.1

Proof. Write $x_i = \Psi(M_i)$ for the training points, $\rho = \rho_-(\tau)$, and $x = \Psi(M)$ for an arbitrary module $M \in \mathcal{R}_N$; by Lemma C.1 all of these lie in the ball $B_{\mathbb{R}^{SK}}(B)$. The partition constructed below depends on τ, $\mathcal{C}$ and B only, not on T, as Theorem 3 of Xu & Mannor (2012) requires; the training set enters only through $\varepsilon(T)$, which that theorem allows to depend on it. If T is not separated at τ the robustness condition holds trivially with $\varepsilon(T) = 1$, the loss being bounded by 1, so assume it is. The argument then has three steps: the classes are separated, so the 1-NN rule is locally constant, so a fine enough partition has zero within-cell loss variation.

Separation. By Theorem 2.3, $\|x_i - x_j\| \ge \rho$ whenever $y_i \neq y_j$; in particular no two training modules with different labels share an embedding, so $h_T(x_i) = y_i$ for every i and $\widehat{R}(h_T) = 0$.

The 1-NN rule is constant near each training point. Fix a training point x_i and any x with $\|x - x_i\| < \rho/2$. The nearest training neighbor of x—nearest among all m training points, with no reference to any cell—is some x_k with $\|x - x_k\| \le \|x - x_i\| < \rho/2$, hence

$$\|x_k - x_i\| \;\le\; \|x_k - x\| + \|x - x_i\| \;<\; \rho,$$

and separation forces $y_k = y_i$. Every training point tied with x_k at the minimal distance satisfies the same inequality, so the conclusion is independent of how ties are broken: $h_T(x) = y_i$ on the open ball of radius $\rho/2$ about x_i.

The partition, and why its cells have diameter $\rho/2$ rather than ρ. Cover $B_{\mathbb{R}^{SK}}(B)$ by $\mathcal{N}(\rho/4; B_{\mathbb{R}^{SK}}(B))$ open balls of radius $\rho/4$ and turn the cover into a partition by assigning each point to the first ball that contains it; every cell then has diameter less than $\rho/2$. Robustness in Xu & Mannor (2012) is a property of a partition of the sample space, here $\Psi(\mathcal{R}_N) \times \{1, \dots, C\}$, and the $0/1$ loss reads the label, so the cells of that partition are the products of these $\mathcal{N}(\rho/4; B_{\mathbb{R}^{SK}}(B))$ cells with a single label: $\mathcal{K} = C \cdot \mathcal{N}(\rho/4; B_{\mathbb{R}^{SK}}(B))$ cells in all. The volumetric bound $\mathcal{N}(r; B_{\mathbb{R}^n}(B)) \le (1 + 2B/r)^n$ with $n = SK$ and $r = \rho/4$ gives the inequality in (17). Cells of diameter below ρ, which separation alone suggests, would not do: a test point x and a training point x_i in such a cell have $\|x - x_i\| < \rho$, which bounds the distance from x_i to the nearest neighbor of x only by 2ρ, and a training point of the other label can sit there. Halving the diameter is exactly what closes the gap.

Zero within-cell variation. Let a training sample (x_i, y_i) and a point (x, y) of the sample space lie in the same cell. Then $y = y_i$ and $\|x - x_i\| < \rho/2$, so by the second step $h_T(x) = y_i = h_T(x_i)$, and both $0/1$ losses are 0. This is $(\mathcal{K}, \varepsilon(T))$-robustness with $\varepsilon(T) = 0$ on separated training sets, and Theorem 3 of Xu & Mannor (2012), with the loss bounded by 1, gives the displayed bound. The expression for B/ρ_- substitutes Lemma C.1 and $\rho_- = \tau/(4\sqrt{K})$. □

Proof of Proposition E.1

Proof. Fix θ_0 satisfying (a) and (b) and a neighborhood U of θ_0 on which the pairing is constant. On U every finite bar i has a fixed birth simplex σ_b and death simplex σ_d, so $b_i(\theta) = s^{(s)}_{\sigma_b}(\theta)$ and

$d_i(\theta) = s^{(s)}_{\sigma_d}(\theta)$ with $s^{(s)}_{\sigma} = \max_{v\in\sigma} u^{(s)}_v$. By (b) the maximizing vertex $v_b = \arg\max_{v\in\sigma_b} u^{(s)}_v$ is unique at θ_0; the finitely many maps $\theta \mapsto u^{(s)}_v(\theta)$ are continuous, so v_b remains the unique maximizer on a smaller neighborhood, on which $s^{(s)}_{\sigma_b} = u^{(s)}_{v_b}$ and hence $\partial b_i/\partial g_\theta(v) = \mathbf{1}[v = v_b]\,\partial u^{(s)}_{v_b}/\partial g_\theta(v_b)$. The inner map $u^{(s)}_v = \max\big(g_\theta(v)_1/\cos\vartheta_s,\ g_\theta(v)_2/\sin\vartheta_s\big)$ is the maximum of two affine functions of $g_\theta(v)$; wherever the attaining grade is unique it coincides locally with that branch, with derivative $1/\cos\vartheta_s$ in the first grade and 0 in the second, or $1/\sin\vartheta_s$ and 0 respectively, which is the displayed inner derivative; at a tie between the two grades the map is locally Lipschitz and the gather rule returns one element of its Clarke subdifferential (Remark E.2). The same argument applies to d_i. Finally, a vectorization differentiable in the bar coordinates composes with (b_i, d_i) by the chain rule; the hat coordinates (7) are piecewise linear in the bar coordinates and differentiable off the finitely many hyperplanes where $\|p_k - (b_i, d_i)\|_\infty = r_k$ or where the two coordinates of that ℓ^∞ norm tie, so $\Psi_\theta(M)$ is differentiable in θ at every θ_0 satisfying (a), (b) and lying off those hyperplanes. □

D CERTIFIED CLASSIFICATION: FURTHER DISCUSSION

THE GENERALIZATION BOUND (SECTION 3)

The certificate also controls the *population* risk of the rule, with ρ_- appearing essentially, through the algorithmic-robustness route of Xu & Mannor (2012); τ is now fixed before the sample is drawn—the strict scale (13) is a function of T, and a partition chosen after seeing T is not what a robustness bound tolerates—and the training set enters only through whether it clears that scale.

Proposition D.1 (Robustness generalization of the certified rule)**.** *Fix $\tau > 0$ and a τ-admissible $\mathcal{C}$ in advance, let $T = \{(M_i, y_i)\}_{i=1}^m$ be i.i.d., let h_T be the 1-NN rule on Ψ and $\ell \in [0,1]$ the 0/1 loss, and say that T is* separated at τ *if its cross-class separation (13) is at least τ and every cross-class pair of T is $\mathcal{C}$-coherent. Then h_T is $(\mathcal{K}, \varepsilon(T))$-robust in the sense of Xu & Mannor (2012), with $\varepsilon(T) = 0$ whenever T is separated at τ and $\varepsilon(T) = 1$ otherwise, where*

$$\mathcal{K} \;=\; C{\cdot}\mathcal{N}\big(\rho_-(\tau)/4;\, B_{\mathbb{R}^{SK}}(B)\big) \;\le\; C\Big(1+\tfrac{8B}{\rho_-(\tau)}\Big)^{SK}, \qquad \frac{B}{\rho_-(\tau)} \;=\; \frac{4\sqrt{SK}\,N_{\max}\max_{s,k} r_{s,k}}{\tau}, \tag{17}$$

C is the number of classes and $\mathcal{N}(r; B_{\mathbb{R}^{SK}}(B))$ is the number of open balls of radius r needed to cover the ball of radius B; consequently, with probability at least $1-\delta$ over T,

$$R(h_T) \;\le\; \widehat{R}(h_T) \;+\; \varepsilon(T) \;+\; \sqrt{\frac{2\mathcal{K}\log 2 + 2\log(1/\delta)}{m}}\,,$$

where $\widehat{R}(h_T) = 0$, every training point being its own nearest neighbor, and $\varepsilon(T) = 0$ on every training set that is separated at τ.

Proof. Deferred to Appendix C.

The content is the middle term of (17): the sample complexity is governed by the two-sided distortion B/ρ_-, which this paper computes and no prior multiparameter vectorization can. As a number the bound is vacuous—$\mathcal{K}$ is exponential in $SK = 1{,}600$—and the vacuity is a matter of scale rather than dimension (Appendix D): test modules sit 6–12 times ρ_- from their nearest training module, so no two training modules share a cell at the certificate's scale.

On the generalization bound. The content is the middle term of (17): the sample complexity is governed by B/ρ_-, the ratio of the upper gauge to the lower one—that is, by the two-sided distortion of the embedding, which is precisely what this paper computes and no prior multiparameter vectorization can. Improving either gauge improves the bound, and a method with only an upper bound cannot state it at all.

We state its numerical status bluntly, because it would be easy to oversell. $\mathcal{K}$ is exponential in SK (here 1,600), so the right-hand side is astronomically larger than 1 and the bound is *vacuous as a number*: it is a structural statement about which quantity controls generalization, not a usable risk estimate. The exponent is the ambient embedding dimension; a non-vacuous version needs the covering number of the *image* $\Psi(\mathcal{R}_N)$, whose intrinsic dimension we do not control. This is the

same N- and K-dependence the field already concedes as intrinsic—no quantitative embedding of the space of modules into a fixed-dimensional Euclidean space exists—and we prefer to state it than to bury it.

Nor is the vacuity an artifact of the ambient dimension. Transporting a cover of the module space through the upper gauge—a cover of $\mathcal{R}_N$ in d_{mat}^S at scale $\rho_-/(4\sqrt{S}\,N_{\max})$ pushes forward, by Proposition 2.2, to a cover of the image at scale $\rho_-/4$—replaces the exponent SK in (17) by $2N_{\max}S$, the number of bar coordinates, with K demoted to the base; that is the one inequality in which both gauges act at once, and it is still astronomically large. The cause is the scale, not the dimension, and it is measurable. A robustness bound is informative only when the data concentrate on $\mathcal{K} \ll m$ cells of diameter ρ_-, and the certificate-1NN sweep of Section 4 finds the median distance from a test module to its nearest training module to be 6–12 times ρ_- at every sample size and at both configurations—and training modules, drawn from the same distribution, sit at the same scale from one another. No two training modules share a cell: the empirical covering number of the training set at the certificate's scale is m itself, and the complexity term is at least $\sqrt{2\log 2} > 1$ whatever the sample size. This is the local/global mechanism of Section 4 in learning-theoretic form—the certificate lives below the sampling resolution of the data—and the same fact that empties the certified rule's coverage empties its risk bound.

The kernel floor. The same floor certifies the higher-accuracy Gaussian-SVM head. Let $\kappa(M, M') = \exp(-\|\Psi(M) - \Psi(M')\|^2/2h^2)$ with feature map ι into its RKHS $\mathcal{H}$, so $\|\iota(M) - \iota(M')\|_{\mathcal{H}}^2 = 2(1 - \kappa(M, M'))$. (The bandwidth is h and the kernel κ, a Greek kappa matching its ceiling κ_- and distinct from the landmark index k; σ stays reserved for simplices.)

Lemma D.2 (Kernel separation floor)**.** *Fix $h > 0$ and let (M, M') be a $\mathcal{C}$-coherent cross-class pair with $d_{\mathrm{mat}}^S(M, M') \geq \tau$, as in Theorem 2.3. Then*

$$\kappa(M, M') \leq \kappa_-(\tau) := \exp\Big(-\tfrac{\rho_-(\tau)^2}{2h^2}\Big) < 1, \qquad \|\iota(M) - \iota(M')\|_{\mathcal{H}} \geq \sqrt{2\,(1 - \kappa_-(\tau))} > 0.$$

Proof. $x \mapsto e^{-x^2/2h^2}$ is decreasing, so (10) gives the bound on κ; substitute into $\|\iota(M) - \iota(M')\|_{\mathcal{H}}^2 = 2(1 - \kappa)$. As $\rho_-(\tau) > 0$ we have $\kappa_-(\tau) < 1$. □

Monotonicity alone carries the linear floor ρ_- through the kernel: coherent cross-class pairs separated at scale τ stay separated in the RKHS, by an explicit amount. We are deliberately careful about what this does *not* say. A uniform lower bound on *pairwise* cross-class distances does not lower-bound the margin of a max-margin separator, which is half the distance between the two *convex hulls*; interleaved configurations have large pairwise distances and arbitrarily small margin. Lemma D.2 therefore certifies the separation the kernel sees—the quantity that governs nearest-centroid and nearest-neighbor decisions in $\mathcal{H}$—not the margin of the soft-margin SVM actually fitted, whose bandwidth h is moreover selected per fold. Closing that gap, so that the guarantee attaches to the head that attains the reported accuracy, is what Section 4 measures—and finds no Lipschitz route can do. The bound relaxes as $h \to \infty$ and tightens as h shrinks.

Nothing in Lemma D.2 is specific to the Gaussian kernel beyond monotonicity, and the coordinates of Ψ are sums of hat functions—non-negative and histogram-like, the regime where χ^2 and Hellinger kernels are classically competitive. The floor carries to that whole family, with constants supplied by the same boundedness that powers Lemma C.1.

Lemma D.3 (The floor survives non-Euclidean kernels)**.** *Write $B_\infty := \max_{s,k} \omega_s w_{s,k} N_{\max} r_{s,k}$ for the per-coordinate bound of the stack (from the proof of Lemma C.1; $B_\infty \leq N_{\max} \max_{s,k} r_{s,k}/\sqrt{K}$ under equal weights), and for a divergence $D \geq 0$ with $D(x, x) = 0$ let $\kappa_D(M, M') = \exp\big(-\gamma\, D(\Psi(M), \Psi(M'))\big)$, $\gamma > 0$. If D dominates the Euclidean geometry on the embedding's range,*

$$D(x, y) \geq c_D\, \|x - y\|_2^2 \qquad \text{for all } x, y \in [0, B_\infty]^{SK}, \tag{18}$$

then for every $\mathcal{C}$-coherent cross-class pair with $d_{\mathrm{mat}}^S \geq \tau$,

$$\kappa_D(M, M') \leq \exp\big(-\gamma\, c_D\, \rho_-(\tau)^2\big) < 1, \qquad \|\iota_D(M) - \iota_D(M')\|_{\mathcal{H}_D} \geq \sqrt{2\big(1 - e^{-\gamma c_D \rho_-(\tau)^2}\big)} > 0,$$

whenever κ_D is positive definite. The three heads at issue satisfy (18) with explicit constants: the Gaussian kernel ($D = \|\cdot\|_2^2$, $c_D = 1$, $\gamma = 1/2h^2$, recovering Lemma D.2); the exponential χ^2

kernel, $D(x,y) = \sum_i (x_i - y_i)^2/(x_i + y_i)$, with $c_D = 1/(2B_\infty)$, since every coordinate obeys $x_i + y_i \le 2B_\infty$; and the Hellinger kernel, $D(x,y) = \left\|\sqrt{x} - \sqrt{y}\right\|_2^2$, with $c_D = 1/(4B_\infty)$, since $|\sqrt{x_i} - \sqrt{y_i}| = |x_i - y_i|/(\sqrt{x_i} + \sqrt{y_i}) \ge |x_i - y_i|/(2\sqrt{B_\infty})$. (Coordinates with $x_i = y_i = 0$ contribute zero to both sides, so the degenerate case is harmless; positive definiteness holds for all three—the χ^2 distance is conditionally negative definite on the non-negative orthant, so $\exp(-\gamma\,\chi^2)$ is positive definite for every $\gamma > 0$ by Schoenberg's theorem, and the Hellinger kernel is a Gaussian precomposed with the coordinatewise square root.)

Proof. Theorem 2.3 gives $\|\Psi(M) - \Psi(M')\|_2 \ge \rho_-(\tau)$ on coherent pairs; (18) and the monotonicity of $t \mapsto e^{-\gamma t}$ give the kernel ceiling, and $D(x,x) = 0$ normalizes the diagonal, so $\|\iota_D(M) - \iota_D(M')\|_{\mathcal{H}_D}^2 = 2\big(1 - \kappa_D(M, M')\big)$ as before. □

The same holds for the χ^2 and Hellinger kernels with explicit constants, so choosing the kernel by cross-validation never loses the certificate.

Remark D.4 (Certified kernel selection). Lemma D.3 makes the *choice* of kernel a certified degree of freedom: selecting among the Gaussian, χ^2 and Hellinger heads by inner cross-validation on the training fold—exactly as the bandwidth and the SVM constant are already selected—changes only the constant c_D in the floor, never its existence. Whether the non-Euclidean heads pay in accuracy is an empirical question; what the lemma settles is that no choice in this family leaves the certificate behind.

THE PER-PREDICTION CERTIFICATE

Reading the same theorem per prediction. The gap test of Theorem 3.1 is already a per-test-point check, but the radius it compares against is not: r_m bounds $\max_c \|\widehat{\mu}_c - \mu_c\|$ in *norm*, so it carries B (hence $N_{\max}$ and the radii) and the second moments V_c, and it must hold uniformly over test points. For a single test module drawn independently of the training sample, less is needed. Writing $\delta_c = \widehat{\mu}_c - \mu_c$ and $x = \Psi(M)$, the identity

$$\|x - \widehat{\mu}_c\|^2 - \|x - \mu_c\|^2 \;=\; -2\langle x - \mu_c,\, \delta_c\rangle + \|\delta_c\|^2 \tag{19}$$

is exact, so each pairwise margin of the nearest-centroid rule moves by a *scalar* sample mean—of the variables $\langle x - \mu_c, \Psi(M_i) - \mu_c\rangle$, with variance $(x - \mu_c)^\top \Sigma_c (x - \mu_c)/m_c$—plus a term of order $1/m_c$. Neither the embedding dimension nor a trace appears. Proposition D.5 in Appendix D turns this into a certificate: if for every competing class the empirical squared-distance margin exceeds $2z\big(\sigma_{\widehat{c}}(x)/\sqrt{m_{\widehat{c}}} + \sigma_{c'}(x)/\sqrt{m_{c'}}\big) + r^2$, with z a Gaussian quantile and r any radius holding $\max_c \|\delta_c\| \le r$, then the population and empirical rules agree at M, at a confidence charged by scalar Berry–Esseen; replacing the Gaussian quantile by a Bernstein bound gives a fully non-asymptotic variant. Section 4 measures both, and the gain over the norm-based radius is large on the benchmarks where the rule-level condition fails. Three caveats travel with the statement, and we keep them attached: the variance is plugged in, the Berry–Esseen charge is not negligible at these class sizes (0.09–0.37), and the guarantee is per test point rather than simultaneous over a test set. It also certifies what Theorem 3.1 certifies—agreement with the population nearest-centroid rule, not correctness—and that head remains the weaker one.

Throughout, $x = \Psi(M)$ for a test module M drawn independently of the training sample, $\delta_c = \widehat{\mu}_c - \mu_c$, $\sigma_c^2(x) = (x - \mu_c)^\top \Sigma_c (x - \mu_c)$ with Σ_c the class covariance of Ψ, and $\kappa_c(x) = \mathbb{E}|\langle x - \mu_c, \Psi(M') - \mu_c\rangle|^3/\sigma_c(x)^3$ for M' drawn from class c.

Proposition D.5 (Per-prediction agreement)**.** *Let r satisfy $\Pr[\max_c \|\delta_c\| \le r] \ge 1 - \delta_2$—for instance $r = r_m$ of (12) at level δ_2—and let $z = \Phi^{-1}(1 - \delta_1/(2C))$. Write $\widehat{c} = \widehat{c}(M)$ and $G_{c'}(x) = \|x - \widehat{\mu}_{c'}\|^2 - \|x - \widehat{\mu}_{\widehat{c}}\|^2$. If*

$$G_{c'}(x) \;>\; 2z\Big(\tfrac{\sigma_{\widehat{c}}(x)}{\sqrt{m_{\widehat{c}}}} + \tfrac{\sigma_{c'}(x)}{\sqrt{m_{c'}}}\Big) \;+\; r^2 \qquad \textit{for every } c' \neq \widehat{c},$$

then the population nearest-centroid rule assigns M the label $\widehat{c}$, with probability at least $1 - \delta_1 - \delta_2 - 0.9496 \sum_c \kappa_c(x)/\sqrt{m_c}$. Replacing $z\sigma_c(x)/\sqrt{m_c}$ by $\sqrt{2b\,\sigma_c^2(x)/m_c + 2\beta_c(x)b/(3m_c)}$, where $b = \log(4C/\delta_1)$ and $\beta_c(x)$ bounds $|\langle x - \mu_c, \Psi(M') - \mu_c\rangle|$ almost surely, gives the same conclusion with no Berry–Esseen term.

Proof. By (19) the difference between the empirical and population margins at x is $-2\ell_{c'} + 2\ell_{\widehat{c}} + \|\delta_{c'}\|^2 - \|\delta_{\widehat{c}}\|^2$ with $\ell_c = \langle x - \mu_c, \delta_c \rangle$. Each ℓ_c is a mean of m_c i.i.d. centred scalars with variance $\sigma_c^2(x)$, so the Berry–Esseen theorem with Shevtsova's constant 0.4748, applied two-sided to each class, gives $|\ell_c| \le z\sigma_c(x)/\sqrt{m_c}$ for all c simultaneously with probability at least $1 - \delta_1 - 0.9496 \sum_c \kappa_c(x)/\sqrt{m_c}$. On that event intersected with $\{\max_c \|\delta_c\| \le r\}$, the margin moves by at most the displayed bound, so every population margin stays positive. The Bernstein variant replaces the first step by the scalar Bernstein inequality at level $\delta_1/(2C)$. □

Three limitations are worth stating exactly. The variances $\sigma_c(x)$ and the direction $x - \mu_c$ are population quantities and are estimated in the measurements of Section 4 by $\widehat{\Sigma}_c$ and $x - \widehat{\mu}_c$, the same plug-in convention PALACE's variance-aware radius uses; a self-contained statement needs an effective-rank bound on $\left\|\widehat{\Sigma}_c - \Sigma_c\right\|_{\mathrm{op}}$ along that direction, which we do not prove here; the Hilbert-space covariance estimators and confidence balls of Bagchi et al. (2026) are the natural source for it. The Berry–Esseen charge is additive and, at the class sizes of these benchmarks, not small: the plug-in $\widehat{\kappa}$ gives 0.09 on NCI109 and 0.33 on MUTAG, so the Gaussian form's nominal $1 - \delta$ is reduced accordingly, and a self-normalized moderate-deviation bound—whose error is relative to the tail rather than additive—is the natural repair. And the statement is per test point: certifying a whole test set of size T simultaneously costs a $\log T$ in z.

E Experimental details

Cost (the price of "training-free"). On 16 CPU cores—no GPU anywhere in the pipeline—the full Orbit5k experiment (diagrams for all $5{,}000$ clouds, per-fold landmarks, embedding and the tuned head) runs in 6.8 minutes end to end: 1.2 minutes for the first fold including all diagram computation, then embedding and head per fold (`timing.py`). One Orbit100k fold costs 166 minutes, 149 of them in the Nyström head at $4{,}000$ components, so the full 3-fold protocol is ≈8 hours on the same 16 cores. For scale, D-GRIL reports 5–30 minutes *per fold* of GPU training on the far smaller TU sets (their Table 4, on an A10), and the transformer baselines train for hours on accelerators before their first prediction.

Baselines. For point clouds: the sliced-Wasserstein (Carrière et al., 2017) and Fisher (Le & Yamada, 2018) kernels, PERSLAY (Carrière et al., 2020) (whose table is also the source of the two kernel rows), the Euler-characteristic surface with gradient-boosted trees (ECS+XGB, tuned to its best resolution), the single-parameter certified kernel PALACE, and the transformers Persformer and xPerT. For graphs: multiparameter persistence landscapes (mp-l) and images (mp-i), and the GRIL landscape—the trained or hyperparameter-tuned methods COMPLEX aims to match with no learned embedding.

Baseline protocol. Because the tables mix literature numbers with re-runs, we state per baseline which is which. *Literature, own protocol* (* in the tables): the sliced-Wasserstein and Fisher kernels, PERSLAY ($70/30$ splits, 100 runs), xPerT and Persformer (their own splits and trained heads), PALACE (10-fold), graphcode (Orbit: $70/30$, 20 and 10 runs; graphs: $80/20$, 20 runs), GRIL together with the mp-i and mp-l rows it re-tabulates (5 stratified train/test splits, XGBoost head), TopoGCL (10-fold), and the ECS headline of Hacquard & Lebovici (2024). *Re-run on our clouds and folds*: ECS (matched bifiltration, best resolution, XGB), and in the contamination sweep ECS, ECC, HT and PALACE-1p, each scored under both a gradient-boosted and an RBF-SVM head and reported at the better, so that no baseline is handicapped by a head it was not designed for. End-to-end architectures are reported natively, since forcing an SVM head would handicap them below their design. Where a re-run and a literature number differ (ECS: 89.5 against 91.8) both are shown and the comparison is drawn against the literature figure.

The headline configuration, in detail. On Orbit5k COMPLEX reaches 91.95 ± 0.11 over seeds (± 1.3 over folds), using PALACE's published Orbit recipe lifted unchanged to the slice-stack—radius factor $\alpha = 4$ and three DTM bifiltrations ($k \in \{10, 15, 20\}$) concatenated at $K = 200$ per block, so the embedding is three times wider ($4{,}800$ against $1{,}600$); the single-bifiltration stack at the same α reads 91.74 ± 0.10, and 91.2 at the $\alpha = 1.75$ default (the radius sweep, monotone in α,

and the certified capped variant are priced below). Multiplicity pays only with the extra width: the same three bifiltrations at a *matched* total budget ($K = 60$ per block, 1,440 dimensions) read 91.53 ± 0.01, below the single bifiltration—so the $+0.2$ is bought by dimension, exactly as in PALACE's published recipe, and not by the second and third filtrations per se. That places the closed-form embedding ahead of every diagram-based kernel and vectorization (PERSLAY 87.7, Fisher 85.9, sliced Wasserstein 83.6), of xPerT, and of the strongest one-parameter methods, Persformer (91.2 ± 0.8)[1] and its own predecessor PALACE (91.3 ± 1.0)—and, for the first time in this line of work, of the Euler-characteristic surface (91.8 ± 0.4), previously the accuracy ceiling among training-free descriptors. Two comparisons still deserve care. The ECS lead we overturn is 0.15 points, within one fold's deviation, so *level* is the defensible reading and *ahead* the nominal one; and ECS is not a persistence method—it bypasses homology entirely, which is exactly its advantage in cost and its limitation in what it can certify. Within its own family COMPLEX has one published competitor: graphcode (Kerber & Russold, 2024), a GNN trained on a two-parameter barcode hierarchy, reads 88.5 ± 1.1 on Orbit5k under its 70/30 protocol—3.5 below COMPLEX—and 92.3 ± 0.3 at the Orbit100k scale, previously the strongest reproducible number there (see below). Two readings of the PALACE comparison are possible and the paper owes both. Against PALACE's *published* 91.3, obtained under its own protocol, the multiparameter construction is ahead (91.95 against 91.3)—using PALACE's own radius and filtration multiplicity, so the gap isolates the slice-stack. Against a *matched* re-run of the same single-parameter pipeline on the same clouds and the same folds—"PALACE-1p" in Table 5, where COMPLEX reads 90.9 and PALACE-1p 87.2 under one protocol—it is ahead by 3.7 points. The second is the like-for-like comparison and the first is not, so we report the published-protocol gap as the conservative claim and the matched gap as the informative one; comparing our re-run to another paper's headline in either direction is the asymmetry we are at pains to avoid elsewhere. A slice-level ablation at the $\alpha = 1.75$ default locates that gap: a *single* diagonal slice of the bifiltration already reaches 89.1 at the same total embedding dimension, and stacking to $S = 8$ adds a further $+2.3$, saturating there ($S = 16$ on a wider fan gains $+0.3$ over $S = 4$; `slice_value.py`, 5-fold, one seed, so its $S = 8$ figure of 91.4 is not the 10-fold 91.2 of Table 7). So the second parameter and the stack each contribute, and neither is redundant. What the construction additionally buys is a module-level certificate and applicability to genuinely bifiltered data, with no learned embedding. At the transformer scale the standing sharpens: on Orbit100k COMPLEX reads 92.98 ± 0.02 over three seeds (± 0.15 over folds; 3-fold, as the protocol states; single bifiltration, $\alpha = 4$, Nyström at 4,000 components), ahead of xPerT (91.1), of Persformer (92.0, with the reproduction caveat above), and of the GNN-trained graphcode (92.3 ± 0.3), the strongest previously published number at this scale—still with no learned embedding, subject to the approximation caveat in the caption. The certified capped variant reads 92.78 ± 0.20 here—the certificate's construction is itself ahead of every published number at this scale (Remark B.1). At the $\alpha = 1.75$ default with 1,000 Nyström components the same pipeline reads 92.3; the gain decomposes cleanly, $+0.4$ from provisioning the head at its calibrated optimum and $+0.2$ from the radius factor, both fixed on Orbit5k before touching Orbit100k.

The Nyström head. Orbit100k uses a Nyström feature map in place of the exact kernel (a 90k $\times$ 90k kernel is $\sim$65 GB), provisioned at 4,000 components. Calibrated against the exact kernel on Orbit5k at 5 folds and the headline $\alpha = 4$ (`nystrom_calib.py`; exact head 91.58 ± 1.44), the approximation reads $-0.48/-0.24/+0.04$ points at 250/500/1,000 components and $+0.30/+0.36$ at 2,000/4,000, where it acts as a regularizer rather than an approximation (at $\alpha = 1.75$ the ladder is $-0.74/-0.04/+0.34/+0.52/+0.44$; the shape is the same). The row is 3 seeds $\times$ 3 folds, the protocol stated in the text; at the 10-fold protocol used elsewhere in this paper the same configuration reads 93.10 ± 0.28 (seed 0). At the $\alpha = 1.75$ default with 1,000 components the pipeline reads 92.28 ± 0.02 at 3 folds and 91.94 ± 0.35 at 10.

The radius cap, priced. Remark B.1 leaves two variants; here is what the certified one costs. Enforcing (6) after construction reads $91.25 \to 90.24$ at $\alpha = 1.75$ and $91.74 \to 91.31$ at $\alpha = 4$ on Orbit5k (single bifiltration, 10-fold RBF-SVM), and $92.98 \to 92.78 \pm 0.20$ on Orbit100k (3-

[1] A reproducibility caveat attaches to the Persformer rows: the same table reports 99.1 when the model reads the *raw point clouds*, alone or alongside the diagrams—an input that is no longer persistence (its saliency-filtering experiment reads 91.1)—while xPerT's independent re-run of the diagram-input model failed to train (28.2 ± 7.3; "the model was not trainable in both ORBIT5K and ORBIT100K" (Kim, 2024)). We quote the commonly cited raw accuracy and flag that the transformer rows carry more reproduction risk than their variances suggest.

fold, seed 0). The radius sweep is monotone in α for both variants ($91.25/91.69/91.74$ uncapped at $\alpha = 1.75/2.5/4$), so certification and the radius factor point the same way, and the certified construction itself stays above every previously published number at the 100k scale, graphcode's 92.3 included. The lost accuracy localizes where the unconstrained radii earn their accuracy: coordinates that read persistence mass *near the diagonal*, a small-bar count signal the certified construction deliberately excludes.

The block weights, priced. Remark B.2 leaves the choice of stack to accuracy, so we ran both at the head that carries it. On the single bifiltration the weighted stack of Definition 2.1 reads 91.79 ± 0.12 against the unweighted 91.74 ± 0.10; on the three-bifiltration headline 91.99 ± 0.07 against 91.95 ± 0.11; on Orbit100k 92.99 against 92.96 (3-fold, seed 0, $\alpha = 4$, Nyström at 4,000). All three are ties within seed noise, in the weighted stack's favour by a hair, so the reported numbers would not move if the pipeline weighted its blocks (`omega_sweep.py`). The certified nearest-centroid head is the one that notices, preferring the unweighted stack by 1.5 points (73.87 ± 0.38 against 72.39 ± 0.18 at $\alpha = 4$): it decides by Euclidean distances in the stack directly, and the weights shrink the extreme slices—the most class-discriminative ones—by up to $3.5\times$ relative to the diagonal, while the kernel head reabsorbs the same rescaling into its per-fold bandwidth.

Pooled degrees, priced. The Orbit numbers pool H_0 and H_1 into one multiset per slice before the top-50 cap, the shortcut Section 2.2 warns against; the certified statements attach to the per-degree stack. On Orbit5k the per-degree stack reads within 0.1 of the pooled one at matched embedding dimension (91.54 against 91.65, $K = 100$ per degree-block, 5-fold, three seeds, $\alpha = 1.75$) and 0.3 below it at matched per-block K and twice the dimension, so the shortcut buys no accuracy and the headline is not resting on it (`dimsep_pointcloud.py`). Under a shared cap the two regimes see different bar counts (399 against 747 per cloud), which is the only respect in which they differ.

The slice weights, priced. Definition 2.1 multiplies block s by ω_s; the classifier stacks the blocks unweighted, so the accuracy tables describe the plain concatenation while the certificate statements—and every audit in this section—describe the weighted stack. The gap is measured and small. On the headline configuration the weighted stack reads 91.99 ± 0.07 against 91.95 ± 0.11 on Orbit5k and 92.99 against 92.96 on Orbit100k at seed 0 (`omega_sweep.py`), inside seed noise both times; the certified nearest-centroid head is the one that notices, losing 1.5 points at $\alpha = 4$ (72.39 ± 0.18 against 73.87 ± 0.38; Table 1's centroid row is at the $\alpha = 1.75$ default), since a centroid rule reads the raw Euclidean geometry that the weights rescale by up to $3.5\times$ across the fan. Nor does the theory turn on it: dropping the weights multiplies the upper gauge of Proposition 2.2 by $1/\omega_{\min}$ (5.4 for the default fan) and *improves* the floor of Theorem 2.3 by $1/\max_s \omega_s$ ($1.6\times$), because the witnessing block is no longer shrunk—the two-sided bound survives with both constants explicit. We report the unweighted numbers because every table in this paper is computed that way.

Robustness to outlier contamination. A codensity axis exists to suppress outliers, so the sharpest test of the second parameter is a dirty-data sweep against the descriptors that do not have one. Table 5 replaces each point of each Orbit5k cloud, independently with probability p, by a uniform draw from the ambient box, and scores every method on the same clouds and the same folds. The Euler descriptors are given their best resolution (40×40; at the 20×20 default ECS loses 3.2 points and the comparison is not a fair one), and each is reported at the better of gradient-boosted trees and an RBF-SVM head.

Two readings, and they differ. Against a *matched single-parameter* pipeline the second parameter does exactly what it is designed to do: the margin over PALACE-1p grows from $+3.7$ on clean data to $+12.4$ at $p = 0.3$, and PALACE-1p has the worst degradation in the table (-37.8). The same holds against the Euler characteristic curve (ECC, -36.7) and the hybrid transform (HT, -30.1). Against *ECS*, it does not. COMPLEX leads by 1.8 points on clean data, the lead is gone by $p = 0.1$, and at $p = 0.2$ and $p = 0.3$ ECS is ahead by 0.2 and 0.5 points; ECS also degrades less overall (-26.7 versus -29.1). We had predicted the opposite—the Euler characteristic is a signed alternating sum, so outlier simplices should cancel unpredictably rather than be suppressed—and the prediction is not borne out. The honest statement is that on contaminated Orbit5k COMPLEX and ECS are empirically equivalent, with COMPLEX slightly ahead when clean and slightly behind when heavily contaminated.

Table 5: Outlier contamination on Orbit5k (accuracy %, mean±std over 3 seeds, 5-fold CV, 1,000 clouds per class). Each point is replaced by a uniform draw with probability p. Euler descriptors use a 40×40 surface and the better of XGB and RBF-SVM. DROP is the accuracy lost from $p = 0$ to $p = 0.3$; smaller magnitude is more robust. Bold = best in column.

Method	p=0	0.05	0.1	0.2	0.3	DROP
COMPLEX	$\mathbf{90.9 \pm 0.0}$	$\mathbf{83.8 \pm 0.4}$	$\mathbf{80.7 \pm 0.2}$	72.5 ± 0.3	61.9 ± 0.2	-29.1
ECS	89.1 ± 0.2	83.4 ± 0.2	80.6 ± 0.3	$\mathbf{72.7 \pm 0.2}$	$\mathbf{62.4 \pm 0.2}$	$\mathbf{-26.7}$
HT	86.5 ± 0.1	76.8 ± 0.0	73.7 ± 0.2	65.0 ± 0.3	56.4 ± 0.2	-30.1
ECC	83.6 ± 0.1	72.3 ± 0.1	65.7 ± 0.3	55.6 ± 0.1	46.9 ± 0.3	-36.7
PALACE-1p	87.2 ± 0.1	78.1 ± 0.1	71.3 ± 0.3	61.4 ± 0.1	49.4 ± 0.3	-37.8

This is the same concession Table 1 already makes on clean accuracy, and it has the same answer. ECS matches COMPLEX here without computing homology at all, which is what makes it cheap—and is also why it can certify nothing: there is no module, no interleaving, and no distortion bound to instantiate. What COMPLEX buys over the Euler family on this benchmark is not accuracy and not robustness, but the per-prediction guarantee of Section 3 and applicability to genuinely bifiltered data.

Comparison with D-GRIL. D-GRIL (Mukherjee et al., 2026) is the closest point of comparison—a *learned* two-parameter bifiltration—and reports MUTAG 85.09 ± 5.99, PROTEINS 69.45 ± 4.11, DHFR 61.24 ± 4.37 and IMDB-BINARY 62.60 ± 6.56. Most of that table is not comparable with ours: they standardize every multiparameter signature onto a 3-layer MLP head and say so, their protocol is 5-fold to our 10, and on the labeled molecular sets their bifiltration collapses to the constant classifier—their DHFR row sits at ≈ 61 for *all seven* methods, and DHFR's majority class is 60.98%, so three of those entries reproduce it to two decimals. The original papers report ≈ 81 there, and our +label column reaches 80.6; that gap is what a node-label axis opens. Comparing our 80.6 to their 61.2 there would measure their feature set, not our embedding.

The comparison that does hold is IMDB-BINARY, which carries no node labels, so neither method can differ in what it reads from the graph. COMPLEX reaches 70.0 against D-GRIL's 62.60, with GRIL, mp-l and the persistence-image baseline all at chance (50.0) and mp-i at 56.60. A closed-form, training-free embedding is 7.4 points ahead of a gradient-trained bifiltration on the one benchmark where the two are measured on equal footing—and IMDB-BINARY is also where capping essential classes helped most ($66.9 \to 70.0$), since in a social graph the unfilled cycles are the signal.

At OGB scale, the training-free embedding loses. The TU sets are small, and a fair reading of them has to ask whether the result holds where the data is larger and the split harder. We therefore ran the graph pipeline unchanged on `ogbg-molhiv` (Hu et al., 2020)—41,127 molecules, the official scaffold split, ROC-AUC on a 3.5% positive rate—with the landmarks fit on the training split alone and two changes the benchmark forces: a Nyström head with a linear SVM in place of the exact kernel, and ROC-AUC in place of accuracy (`graph_ogb.py`, three seeds). COMPLEX reads 72.92 ± 0.41 test ROC-AUC with the HKS×degree bifiltration and 72.85 ± 0.19 with HKS×label, against 76.1 for a plain GCN, 75.6 for GIN and 77.1 for GIN with virtual nodes on the OGB leaderboard. That is a loss of three to four points, and we report it as one. Two honest readings are available and we decline to choose between them: the scaffold split is a deliberate distribution shift, which a fixed closed-form embedding has no mechanism to adapt to, and message passing on molecular graphs has a decade of architecture search behind it that a landmark embedding of a bifiltration does not. Neither excuses the gap. What the result bounds is the scope of the accuracy claim in this paper—training-free parity with trained multiparameter baselines is demonstrated on point clouds and on the TU benchmarks, not at OGB scale—while leaving the certificate, which no baseline in either regime supplies, untouched.

Kernel head, selected per data set. Remark D.4 makes the kernel a certified degree of freedom; here is what it is worth. Table 6 scores the same five-bifiltration embedding under a fixed Gaussian head and under *select*—the kernel chosen among Gaussian, χ^2 and Hellinger inside the same train-only inner cross-validation that already tunes the bandwidth and C. The coordinates of Ψ are non-negative sums of hat functions, and the histogram kernels behave exactly as that suggests: selection is worth $+2.9$ on NCI1 and $+5.4$ on PTC_MR (where χ^2 is chosen in over half the folds),

Table 6: Kernel selection on the five-bifiltration stack (accuracy %, 10-fold CV, mean±std over 3 seeds). "Gaussian" fixes the RBF head; "select" chooses among Gaussian, χ^2 and Hellinger by inner CV on the training fold (12 bandwidth $\times$ C candidates per kernel), which Remark D.4 shows is certificate-preserving. "picks" counts the inner-CV winner over the 30 test folds. This pipeline differs from Table 2 (multi-big bank at $K = 40$ per block, no per-feature standardization), so only within-row comparisons are meaningful.

Dataset	Gaussian	select	picks (of 30)
MUTAG	$\mathbf{86.9 \pm 0.2}$	86.2 ± 0.9	rbf 15, Hellinger 9, χ^2 6
COX2	77.2 ± 0.6	$\mathbf{78.6 \pm 0.6}$	Hellinger 19, χ^2 9
DHFR	$\mathbf{82.2 \pm 0.6}$	81.1 ± 0.5	Hellinger 15, rbf 10
PROTEINS	73.3 ± 0.2	73.2 ± 0.5	rbf 19, Hellinger 8
NCI1	76.3 ± 0.3	$\mathbf{79.2 \pm 0.2}$	χ^2 17, Hellinger 13
PTC_MR	52.0 ± 2.3	$\mathbf{57.4 \pm 0.2}$	χ^2 16, Hellinger 9
IMDB-B	66.0 ± 0.1	66.0 ± 0.2	χ^2 20, rbf 8

$+1.4$ on COX2 (Hellinger), is neutral on PROTEINS and IMDB-B, and costs about a point on DHFR and MUTAG, where the Gaussian was already the winner and the inner CV occasionally guesses wrong. On the per-degree stack the effect repeats ($77.7 \to 79.6$ on COX2, $54.1 \to 58.3$ on PTC_MR). The lesson mirrors slice selection: adaptivity pays where the default is wrong, costs a little where it is right—and, by Lemma D.3, none of it touches the certificate.

Slice-net ablation. Table 7 varies the only design constant, the net. Two findings. First, the near-diagonal fan decisively beats the axis-aligned percentile slices (91.2 vs 85.7), confirming Section 2.5. Second, and central to the closed-form thesis: making the net *data-adaptive* does not help. The certified-adaptive net, which selects slices by a discrimination score, and even a gradient-trained net (angles optimized end-to-end through the persistence layer) both fall at or below the fixed fan and tend to overfit the selection. (In a separate small-scale probe—150 training clouds, 6 angles, persistence statistics in place of the landmark head—gradient-optimized angles scored 75.6 on held-out clouds against the fixed fan's 77.2; the absolute level is not comparable to the table, only the ordering.) No net we tried improves on the fixed fan by more than 0.7 points—inside the 1.5-point fold standard deviation—so we read this as a null result rather than an optimality claim: the closed-form default costs no accuracy, and net design is not where the headroom lies.

Offset slices. Table 7 varies the fan's angles; its base points are all at the origin, while (1) allows any point of the orthant boundary. Here is what that generality is worth (`offset_sweep.py`, $\alpha = 4$, 10-fold, three seeds, offsets in calibrated grade units, where the axis calibration puts both grades mostly in $[0, 1]$). The origin fan (8 angles, $K = 200$) reads 91.74 ± 0.10, reproducing the single-bifiltration figure of the headline paragraph seed for seed under a harness that shares no code path with it. Trading half the angles for offsets at the same block count and width (4 angles $\times \{-0.25, +0.25\}$) reads 91.59 ± 0.04; giving all eight angles a ± 0.25 pair reads 91.51 ± 0.03 at matched dimension (24 blocks, $K = 65$, 1,560 dimensions) and 91.65 ± 0.02 at matched per-block budget (24 blocks, 4,800 dimensions, three times the width); five offsets per angle read 91.07 ± 0.08. Every offset net sits at or below the origin fan and the loss grows with the number of offsets, and since even the three-times-wider variant does not close it, this is not a capacity effect. The mechanism is visible in (2): below its base point an offset slice reads one axis no further—for $\beta = (c, 0)$ a simplex with scale under c enters at its codensity alone—so at a fixed budget the offsets buy a coarser view of the bifiltration rather than a new one. We keep them in the theory, where they are what makes the fibered barcode the full rank invariant and d_{mat}^{S} independent of the translation into the orthant, and report the origin fan as the default because it is what the data prefers—the same verdict this section returns for adaptive angles.

Coherence audit. Theorem 2.3 is proved under witnessing-slice coherence, and Remark B.8 promised its empirical rate; we report it here. On Orbit5k we sample 2,000 cross-class pairs, keep the 1,000 whose sliced matching distance clears the median scale τ, and test Definition 2.4 on each pair's witnessing slice—over active landmarks only ($k \in \mathcal{A}_{s^\star}(\tau)$), where the weights cancel and the check reduces to $|\varphi_k(A) - \varphi_k(B)| \geq d_{\mathcal{B}}(A, B)/4$—the realized-distance form, stronger than

Table 7: Slice-net ablation on Orbit5k (10-fold CV accuracy %, mean over seeds). Only the net changes; embedding and classifier are fixed at the $\alpha = 1.75$ single-bifiltration default, so the absolute level sits below the headline configuration and only the ordering matters here. ‡The adaptive rows use a non-degeneracy proxy for the certifiability filter of Proposition 2.5, not the full $\rho_- \geq \rho_0$ admissibility test.

Net	(S, a)	nearest-centroid	RBF-SVM
percentile (axis-aligned)	$(4, -)$	44.0	85.7
near-diagonal fan	$(4, 0.6)$	72.6	90.9
near-diagonal fan	$(8, 0.6)$	**73.4**	**91.2**
near-diagonal fan	$(8, 0.3)$	68.6	90.7
certified-adaptive‡	$(4, 0.7)$	67.9	89.0
certified-adaptive	$(8, 0.7)$	68.8	90.4

Definition 2.4's scale form, so the rates below lower-bound the rate of the definition itself. Coherence holds on 97.5% of pairs for the $S = 8$ fan and 98.0% for $S = 4$ at the $\alpha = 1.75$ default, while the certificate's *conclusion*—embedded distance at least $\rho_-(\tau)$—holds on 100% of pairs for both. The residual 2.5% at $\alpha = 1.75$ is the expected gap, and the reason Remark B.8 separates the two levels: coherence is sufficient, not necessary, so the separation it guarantees survives on the pairs where the hypothesis itself cannot be verified. At the adopted radius factor $\alpha = 4$ the hypothesis itself saturates: 100.0% of 1,000 audited pairs are coherent ($\rho_- = 0.0220$ at $\tau = 1.244$, the median of this 2,000-pair sample; the 20,000-pair estimate of Remark B.5 is 1.220), capped and uncapped alike—larger radii enlarge every active set, so the certificate's hypothesis is *easier* to satisfy at exactly the configuration the accuracy prefers.

Numeric instantiation of the certificate. The theory promises a finite, reportable floor; here it is, on Orbit5k with the $S = 8$ fan and $K = 200$ (Ψ into $\mathbb{R}^{1600}$), for the *certified capped* variant at $\alpha = 4$ (Remark B.1)—the construction the certified statements attach to. All quantities are computed from the embeddings over 20,000 sampled cross-class pairs, of which the $\approx$10,000 with $d^S_{\mathrm{mat}} \geq \tau$—the theorem's hypothesis—enter the distance rows (`certificate_instantiation.py`).

quantity	value	
τ (median d^S_{mat} over cross-class pairs)	1.220	the audit scale
$\rho_-(\tau) = \tau/(4\sqrt{K})$	0.0216	the certified floor
$\min \lVert\Psi(M) - \Psi(M')\rVert$, over kept pairs	0.0950	$4.4 \times \rho_-$
median $\lVert\Psi(M) - \Psi(M')\rVert$, over kept pairs	0.2271	$10.5 \times \rho_-$
$\widehat{\Delta}$ (min centroid gap)	0.0584	$2.7 \times \rho_-$
$\max_c$ class radius in $\mathbb{R}^{1600}$	0.577	

(The uncapped variant reads 6.3× and 13.9× with class radius 3.48; at the $\alpha = 1.75$ default the ratios were 2.5× and 6.2×. Larger radii move every realized distance up while the equal-weight floor stays $\tau/(4\sqrt{K})$, so the ratio measures the pairing of floor and configuration, not the theory alone.)

Two readings, and the second is the one we did not expect. The floor is *tight to within a small factor*: the smallest embedded distance attained by any pair satisfying the hypothesis is 0.0950 against a guaranteed 0.0216—a factor 4.4 here, and 2.5 at the $\alpha = 1.75$ default, where the realized distances sit lower. For a bound routed through a functional Lebesgue number, a quarter-scale constant and an equal-weight $K^{-1/2}$, that is close. (A minimum over sampled pairs can only fall as the sample grows; at the $\alpha = 1.75$ default it moves from 2.7× at 1,000 pairs to 2.5× at 20,000, so the estimate is stable. The median, which is not sensitive to the sample size in this way, is the more robust statistic and gives 10.5×.) And the naive route from the certificate to the centroid guarantee is not merely false in principle (Section 3) but quantitatively hopeless: the within-class radius that any pairwise-to-centroid bound must subtract is 0.577, a factor 27 above the floor, so the triangle bound $\widehat{\Delta} \geq d_{\min} - r_c - r_{c'}$ evaluates to -1.06. Notably $\widehat{\Delta} = 0.0584$ is itself $2.7\rho_-$—the centroids *are* separated—but the certificate cannot be what proves it.

What the certified 1-NN rule delivers, and why it is not more. Corollary 3.2 accepts a test module when $\varepsilon < \rho_-(\tau)/2$. We measured its coverage on Orbit5k over $K \in \{8, \dots, 200\}$ and five choices of τ (`cert_1nn.py`, 500 train / 400 test), first at the $\alpha = 1.75$ default, where the mechanism is cleanest, and then—at the end of this section—at the headline configuration. **Coverage never exceeds 1.5% and is 0 in most cells.** We report this plainly because it bounds what the certificate currently buys, and because the reason is diagnostic rather than incidental.

It is *not* that the floor is loose—we just measured it at $2.5\times$. Replacing ρ_- by the realized minimum cross-class distance, i.e. granting a perfectly tight certificate, leaves coverage below 1.5% as well. The binding constraint is that the median nearest-training-neighbor distance (0.056 at $K = 200$) is several times the smallest cross-class distance, so a test module is typically no closer to its own class than the classes are to each other. That is a statement about sampling density relative to class separation, and it also explains why 1-NN on this embedding attains only 66.7% at $n = 500$ (rising to 78.0% at $n = 4{,}000$): the head the certificate reaches is not the head that is accurate. Three further levers—the landmark budget, the sampling density and the metric—are swept in Appendix E; none moves the shortfall by more than a bounded factor, for the reasons summarized next.

The obstruction, stated. Four routes, four mechanisms, all measured on the same data (Figure 1 shows two): the floor cannot be tightened (it is within $2.5\times$ of realized, and a perfect floor leaves coverage below 1.5%), the landmark budget is invariant, sampling density is stationary, and the metric returns a bounded factor. A fifth closes analytically. The natural set-valued relaxation—accept the labels of every training module inside the certified ball—is sound, but that ball is nonempty exactly when $\varepsilon < \rho_-/2$, so it fires on the same event and certifies a set in place of a label. Exclusion runs the other way: ruling a class *out* needs the Lipschitz *upper* bound of Proposition 2.2, which every multiparameter vectorization already has, and only inclusion reads ρ_-.

They share a cause, and it is visible in one measurement. For a test module let $d_{(1)}$ and $d_{(2)}$ be the distances to the nearest training module of the nearest and of the runner-up class. At $K = 200$ the median ratio $d_{(2)}/d_{(1)}$ is 1.10: the second class is ten per cent further away than the first. Only 2.8% of test modules have a runner-up even $1.5\times$ behind, and covering the true class at a $1.25\times$ radius takes 2.7 of the 5 labels. The classes are interleaved at nearest-neighbor scale, and more landmarks make it worse, not better ($d_{(2)}/d_{(1)}$ falls from 1.19 at $K = 16$ to 1.10 at $K = 200$).

That single fact accounts for everything above. Corollary 3.2 accepts when $\varepsilon < \rho_-(\tau)/2$, and ρ_- is at best a constant factor below the smallest embedded cross-class distance; with the classes 10% apart at that scale, the acceptance ball contains a handful of near-duplicates and nothing else. The same geometry caps 1-NN at 78%, empties the set-valued rule, and would force conformal sets to more than half the label space.

Re-run at the headline configuration ($\alpha = 4$, capped, three seeds), every number above survives—coverage $\leq 0.5\%$ at $K = 200$, $d_{(2)}/d_{(1)}$ at 1.12–1.13—with one change: under the radius cap the shortfall grows with the budget ($9\times$ at $K = 8$ to $25\times$ at $K = 200$) instead of staying invariant, so the budget is a lever *against* the certificate there (Appendix E).

Local certificates, global accuracy. The conclusion we draw is not that the embedding is weak. On the *same* embedding a q-tuned RBF-SVM reaches 91.2% where 1-NN reaches 78%—the discriminative information is there, but it is *global*, carried by the kernel across all training points, and not concentrated in local neighborhoods. Every acceptance rule we have tested is local: ε to one neighbor, a ball, a per-class nearest distance. Each inherits the same weak local geometry, which is why five independent levers fail at the same wall.

This also reframes the certified/accurate gap of Section 3 as something other than a deficiency of the nearest-centroid rule. ρ_- certifies a *local* separation; the accuracy lives in *global* structure. The natural response is to attach the guarantee to the kernel head itself, and the two-sided bound makes that possible: each one-vs-one RBF decision function is Lipschitz in Ψ with constant $L_{ab} = e^{-1/2}\sigma^{-1}\sum_i |\alpha_i^{ab}|$, exact from the duals, and Proposition 2.2 converts a margin in the embedding to a radius in d_{mat}^S: if the predicted class wins every pairwise contest by m_{co}, the prediction is unchanged under any module perturbation smaller than $\min_o m_{co}/(L_{co}\sqrt{S}N_{\max})$. We measured it (`cert_svm.py`, $\alpha = 4$, capped, 10 folds, three seeds). The head reaches 91.6% and wins every contest on 99.8% of test modules, so the certificate applies almost everywhere—and its median radius

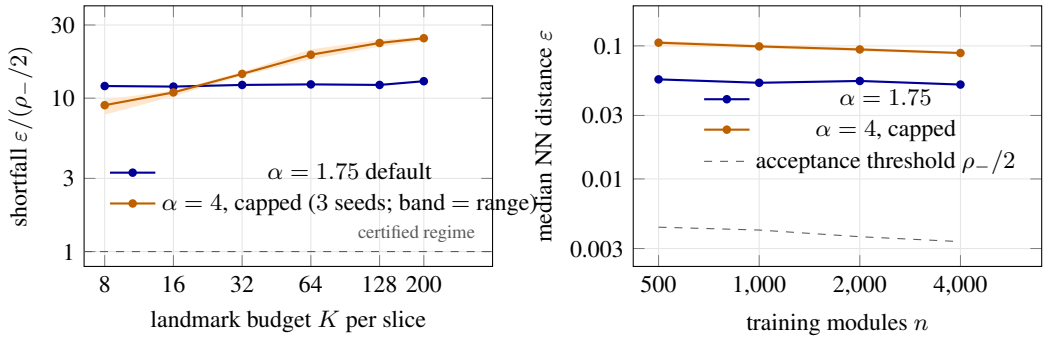


Figure 1: The two swept levers behind "The obstruction, stated" (Orbit5k, $K = 200$ unless swept, `cert_1nn.py`). *Left:* the shortfall $\varepsilon/(\rho_-/2)$ of Corollary 3.2's acceptance test against the landmark budget at $n = 500$—flat near $12\times$ at the $\alpha = 1.75$ default, where ε and ρ_- both scale as $K^{-1/2}$, and rising from $9\times$ to $25\times$ at the capped $\alpha = 4$ headline; the dashed line is the certified regime. *Right:* the median nearest-training-neighbor distance ε against training-set size, stationary over an eightfold range, with the acceptance threshold $\rho_-/2$ (equal for the two α to within 1%) 12–$25\times$ below it. The floor's tightness and the ℓ^∞ reading are reported in the text.

is 1.3×10^{-6}, three orders below ρ_- and five below a single outlier's displacement; coverage is 0 at every radius above 10^{-4}. The constants are the reason: L_{ab} runs from 3×10^2 to 1.5×10^4, because the fitted head is a *high-gain* function of the embedding—thousands of support vectors at $C = 10$—while its margins are $O(1)$. Even an impossibly tight $L = 1$ would leave the radius near 7×10^{-3}, still below one outlier. So the sixth lever closes with the others, and for the same reason from the other side: the kernel reads global structure through a sharp decision function, which is exactly what a Lipschitz certificate cannot cover. A guarantee for the accurate head, if one exists, is not a Lipschitz one. Nothing in the argument is particular to persistence: it applies to any landmark embedding whose lower gauge is witnessed by a single coordinate, and says that per-prediction certification by local acceptance requires local class separation that a globally-discriminative embedding need not provide.

We report this as measured on Orbit5k, not proved in general: a regime of lower intrinsic dimension or wider class separation could behave differently, and we would expect it to. But within the geometry this paper works in, the shortfall is structural—neither the landmark budget nor the sample size moves it—and the operational content of the certificate lies at the level of the *embedding*, where its conclusion holds on 100% of the pairs audited above, rather than at the level of individual predictions. The certified/accurate gap therefore has two distinct causes, not one—the nearest-centroid rule the certificate *cannot* reach (the convex-hull obstruction, quantified above), and the nearest-neighbor rule it *can* reach but which is weak here.

One design consequence is worth recording, since it is free. $\rho_- = \frac{1}{4}\tau w_{\min}$ is maximized by concentrating the weight budget on the active set $\mathcal{A}_s(\tau)$: landmarks below the radius threshold cannot witness at scale τ regardless, so weight on them is wasted from the certificate's point of view. With uniform weights on the K_{act} active landmarks the floor improves by $\sqrt{K/K_{\mathrm{act}}}$. Measured at $K = 200$: $K_{\mathrm{act}} = 98$ at $\tau = 0.226$ ($1.4\times$) and $K_{\mathrm{act}} = 29$ at $\tau = 0.491$ ($2.6\times$). The uniform choice $w_k = K^{-1/2}$ that the pipeline uses is thus the *worst* case for the certificate, and a certificate-first configuration would spend its weight differently—at some cost in accuracy, since the inactive landmarks are not useless to the classifier.

Certified selective classification. Figure 2 plots selective accuracy of the certified nearest-centroid head against coverage, ranking test points by their certificate margin. The margin is discriminative: accuracy climbs monotonically from 73.3% at full coverage (this curve is a single seed; the 3-seed mean is 73.4) to 91.6% on the most-certain tenth, and the certified minority reaches the *uncertified* RBF-SVM accuracy of the same run (90.8%; 91.2% over three seeds) at about 24% coverage—i.e., the top-ranked subset is as accurate as the best uncertified classifier. Two caveats keep this honest. The ranking statistic is the empirical nearest-centroid margin, a proxy for the ρ_- floor rather than the floor itself, so this is selective classification, not certification; under the strict threshold of Theorem 3.1

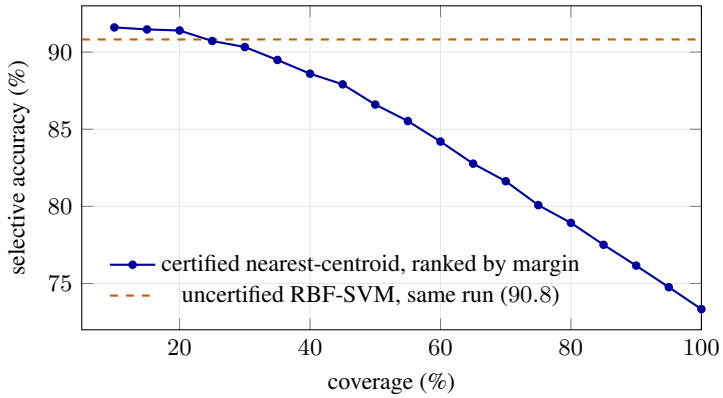


Figure 2: Certified selective classification (Orbit5k, near-diagonal fan $S = 8$). Selective accuracy of the certified nearest-centroid head vs. coverage; dashed line is the uncertified RBF-SVM.

only 1.5% of points certify (at 92.2% accuracy), and the strict certificate is not operational at this sample size. And Lemma D.2 bounds the kernel separation of coherent pairs, not the fitted SVM's margin, so the 91.2% head is not itself certified: the per-prediction guarantee of Theorem 3.1 attaches to the nearest-centroid head, at 73.4%. Narrowing that gap is the main open problem this paper leaves; the paragraph below reports how far a per-prediction reading of the same theorem goes toward it.

The certificate, read per prediction. Section 3 noted that the radius of Theorem 3.1 is a norm bound while the quantity a single prediction needs is a scalar; Table 8 measures what that distinction is worth, on the eleven benchmarks of PALACE's firing study (`pointwise_cert.py`, five seeds $\times$ ten folds). Read per point, the paper's own radius (12) certifies 8.3% of Orbit5k test modules at 84.7% accuracy against 63.3% overall, and between 32% and 54% on the four graph sets where it fires at all. The non-asymptotic per-point form is stronger wherever the classes are well separated—74.0% of NCI1, 71.4% of DD, 68.7% of NCI109—and it is on Orbit5k, where it triples coverage (26.9% against 8.3%) at the same accuracy, that the mechanism is clearest: the norm-based radius pays for the ambient dimension, and a single prediction does not have to. The Gaussian form goes further still, at a confidence reduced by the Berry–Esseen charge in the last column. The pattern where nothing certifies is the one a certificate should have: on COX2, DHFR, PTC_MR and IMDB-MULTI the nearest-centroid head is at or below the constant classifier, and neither theorem-backed form certifies a single point there, the Gaussian form reaching only 7.8% on DHFR and 0.1% on IMDB-MULTI. IMDB-BINARY is the one omitted set where the head is well above its baseline (62.6 against 50.0) and the two theorem-backed forms are still silent, while the Gaussian form certifies 45.8% at 73.6%—the clearest case that the Berry–Esseen charge, not the geometry, is what stands between the measured coverage and a guarantee. This does not close the gap of the preceding paragraph—the head being certified is still the nearest-centroid one, which trails the kernel head by 18 points at the headline configuration—but it moves the per-prediction guarantee from "operationally empty" to "covers most predictions on three of eleven benchmarks", which is a different sentence about the same theorem.

THE CERTIFICATE-1NN SWEEP: BUDGET, SAMPLING DENSITY, METRIC, AND THE HEADLINE CONFIGURATION

The paragraphs below continue the study of Section 4 at the $\alpha = 1.75$ default and then repeat it at the headline configuration; the numbers they establish are the ones summarized under "The obstruction, stated."

Nor is the landmark budget a lever, in either direction. Writing the shortfall as the ratio $\varepsilon/(\rho_-/2)$, both quantities scale as $K^{-1/2}$—ρ_- by construction and ε because $w_{\min} = K^{-1/2}$ makes the embedded norm an average over landmarks—so the ratio is invariant. Measured across $K \in \{8, 16, 32, 64, 128, 200\}$ it moves from 12.0 to 12.9 at $\tau = 0.491$, and from 26.1 to 28.1 at $\tau = 0.226$: flat to within the sampling noise over a $25\times$ range of budgets. There is therefore no small-K regime

Table 8: The nearest-centroid certificate read per prediction (five seeds × ten folds; `pointwise_cert.py`). Cells give the percentage of test points certified and, in parentheses, nearest-centroid accuracy on them; "NC" is that head's accuracy over all test points. "Radius (12)" is the gap test of Theorem 3.1 evaluated per point, which is valid on the concentration event whether or not the rule-level condition $r_m < \frac{1}{2}\widehat{\Delta}$ holds; "Bernstein" and "Gaussian" are the two forms of Proposition D.5; "BE" is the Berry–Esseen charge subtracted from the Gaussian form's confidence. Orbit5k is the capped single-bifiltration configuration at $\alpha = 4$; the graph rows are the multi configuration of Table 2. The five omitted benchmarks certify nothing under the radius and Bernstein columns.

Data set	NC	Radius (12)	Bernstein	Gaussian	BE
Orbit5k	63.3	8.3 (84.7)	26.9 (85.9)	67.0 (73.2)	0.28
MUTAG	87.0	0	0	66.1 (91.8)	0.33
PROTEINS	58.5	31.6 (66.6)	9.5 (90.3)	84.1 (61.0)	0.26
NCI1	63.6	54.0 (67.8)	**74.0** (66.4)	88.7 (65.2)	0.10
NCI109	63.6	47.5 (68.2)	**68.7** (67.1)	85.8 (65.3)	0.09
DD	74.8	45.1 (80.0)	**71.4** (79.2)	92.0 (76.4)	0.13

in which the certificate becomes operational, contrary to what the $K^{-1/2}$ dependence of ρ_- alone would suggest.

Sampling density barely moves it either. Over an eightfold range of training set size (500 to 4,000 modules) the median nearest-neighbor distance ε is stationary—0.056, 0.053, 0.055, 0.051—because in a 1,600-dimensional embedding those distances concentrate, so more data does not bring a test module much closer to its own class. The scale τ meanwhile *falls* (0.491, 0.468, 0.417, 0.383), being a quantile of the training cross-class separation, which denser sampling can only tighten; at that outermost τ the shortfall therefore widens rather than closes, from 12.9 to 15.1 (from 4.9 to 8.6 measured against the active-set floor). Coverage is exactly 0 up to $n = 2{,}000$ and reaches a single accepted test module in 400 at $n = 4{,}000$ (0.3%, correctly classified). We record that as the first nonzero cell rather than as a trend: it is one point. What does improve with density is the head the certificate can reach—1-NN accuracy rises from 66.7% to 78.0% across the same range—so the obstruction is specific to the acceptance radius, not to the representation. Larger n was not testable here: the audit materializes an $n \times n \times 1000$ array, which needs 128 GB at $n = 4{,}000$ and grows quadratically.

The metric is the last free choice, and it comes closest. Coherence asserts only that *some* landmark witnesses, so the ℓ^2 bound already rests on a single coordinate and pays $w_{\min} = K^{-1/2}$ for the privilege. Read the same conclusion in ℓ^∞ over the ω-scaled coordinates and the floor is $\tau/4$ rather than $\tau/(4\sqrt{K})$, larger by $\sqrt{K}$; what it costs is that ε becomes a maximum rather than a root-mean-square, so the shortfall improves by $\sqrt{KS}/(\max/\mathrm{RMS})$. Measured at $K = 200$, $\max/\mathrm{RMS} = 9.4$, an improvement of 4.3× against the 12× required—the median test module sits 32× above the ℓ^2 threshold and 7.6× above the ℓ^∞ one. Nor does the improvement grow with the budget: $\max/\mathrm{RMS}$ itself rises roughly as $\sqrt{K}$ (2.7, 4.3, 5.7, 6.9, 8.7, 9.4 for $K = 8, \ldots, 200$), because $\Delta\varphi$ spreads over more landmarks as the budget grows, so the net gain returns only the $\sqrt{S}$ and is flat in K. Not one additional test module crosses the threshold.

The same sweep at the headline configuration. Everything above was measured at the $\alpha = 1.75$ default. Re-running the full sweep on the certified capped construction at $\alpha = 4$ (3 seeds, $K \in \{8, \ldots, 200\}$, $n \in \{500, \ldots, 4{,}000\}$) confirms the obstruction and sharpens it in one respect. Coverage never exceeds 0.5% at $K = 200$ in any cell; the only cells above 1% are at $K = 8$ (up to 2.5%), where the accepted handful is classified well below the head's accuracy—acceptance there is noise, not certification. The median nearest-neighbor distance is larger, 0.106 at $n = 500$ falling only to 0.088 at $n = 4{,}000$, so the shortfall at the outermost τ is 25× rather than 12×, and flat in n (24.6, 24.1, 25.4, 26.0). The one qualitative change is that the shortfall is no longer invariant in K: it grows from 9.0× at $K = 8$ to 24.6× at $K = 200$ (from 19× to 53× at the median τ), because under the radius cap ε falls more slowly than $K^{-1/2}$ while ρ_- still does. That does not open a small-K regime—the ratio never approaches 1, and the $K = 8$ coverage is the noise just described—but it means the landmark budget is a lever *against* the certificate at the headline configuration, not a neutral

one. The nearest-neighbor geometry is unchanged: $d_{(2)}/d_{(1)}$ is 1.12–1.13 at every n, 1-NN accuracy 74–78%, the ℓ^∞ reading returns $\max/\mathrm{RMS} = 7.8$ at $K = 200$ (rising $2.8 \to 7.8$ across the budget, as before), and 5–7% of test modules have a runner-up class even $1.5\times$ behind. The conclusion is the one already drawn, now at the configuration the paper reports.

Why the simpler selection criterion fails

The score itself deserves a word, because a natural choice fails. Ranking pairs by the margin-like ratio (minimum inter-class centroid gap over *mean* within-class spread) divides by a scalar, so it cannot tell a low-variance discriminative direction from a high-variance nuisance one—and in fact prefers the latter: on a synthetic check, moving the same centroid gap onto a high-variance axis *raises* the ratio from 0.27 to 0.37. On MUTAG this is not hypothetical; the ratio selects degree×clustering, which separates the training subsample and transfers worse (-3.7). Following PALACE, we instead rank by the shrinkage-regularized Fisher–Mahalanobis margin (20), which normalizes *per direction* and collapses to 0.08 on the same synthetic check. We use a diagonal $\widehat{S}_W$—the full covariance is $K \times K$ with few samples per class and singular in practice—and $\lambda = \overline{\mathrm{var}}$. The regularization is not incidental: at $\lambda = 0.1\,\overline{\mathrm{var}}$ the criterion is under-regularized, $1/(\mathrm{var} + \lambda)$ hands large weight to coordinates whose training variance is near zero by accident, and it walks into the MUTAG trap on every fold (-13.9).

SELECT-COMPLEX in detail

From a bank of candidate node functions (heat-kernel signatures at several diffusion times, degree, closeness and eigenvector centrality, clustering), SELECT-COMPLEX scores each candidate pair on the training fold, keeps the certifiable ones, and freezes the best pair. There is no gradient and no held-out split: selection is train-only, so the certificate and the fold protocol are untouched.

Following PALACE, we rank candidate pairs by a shrinkage-regularized Fisher–Mahalanobis margin under the pooled within-class covariance,

$$J(\nu) \;=\; \min_{c \neq c'} (\widehat{\mu}_c - \widehat{\mu}_{c'})^\top \big(\widehat{S}_W + \lambda I\big)^{-1} (\widehat{\mu}_c - \widehat{\mu}_{c'}), \tag{20}$$

with a diagonal $\widehat{S}_W$ (the full $K \times K$ covariance is singular at these sample sizes) and $\lambda = \overline{\mathrm{var}}$. Appendix E records why the simpler ratio of centroid gap to mean within-class spread fails—it prefers high-variance nuisance directions and, on MUTAG, selects a pair that transfers 3.7 points worse—and why the shrinkage level matters (-13.9 at $\lambda/10$).

Table 9 reports selected against fixed at equal net and pipeline, under both the closed-form criterion and an inner-cross-validated one. Selection pays only where the fixed pair is genuinely suboptimal—$+1.9$ on COX2 (closed-form; $+2.8$ under inner-CV), against -2.1 on MUTAG and -1.0 on the label-heavy DHFR—and the ranking of the two criteria is the same on all three sets. Inner-CV is 0.7 points better on average, and we nonetheless report the closed-form criterion as the default: it keeps the selector genuinely closed-form, as Section 2.5 claims of the whole construction, and it is far more stable, agreeing with itself on $5/5$ folds on COX2 where inner-CV splits over three pairs and, on DHFR, chooses five different pairs in five folds. We report both so the cost of that choice is visible. Once essential classes are capped, the fixed HKS×degree baseline is itself much stronger on MUTAG ($85.6 \to 88.8$), and the headroom selection used to occupy is gone. On DHFR the selection is also unstable, choosing a different pair in every one of the five folds and inflating the fold standard deviation to ± 7.4. The value is therefore not an accuracy jump—on two of three sets there is none—but a *training-free, certified* route to a data-adaptive filtration: no gradient, no warm start, and more stable than the differentiable version below.

SELECT-MULTI: selecting the set, not the pair. Single-pair selection asks which *one* bifiltration to keep; the stronger question is which k to *concatenate*. SELECT-MULTI scores all $\binom{7}{2} = 21$ candidate pairs from the node-function bank on the training fold (3-fold inner nearest-centroid accuracy on a cheap $K = 40$ stack), keeps the top k, concatenates them at the same total budget as the fixed multi configuration, and freezes—per fold, train-only, so the protocol and the certificate are untouched. The gains are monotone in k on four of six sets, and at $k = 5$ (the "sel." column of Table 2) the comparison against the fixed three-pair bank reads $+3.7$ on NCI1, $+1.7$ on DHFR,

Table 9: SELECT-COMPLEX: selected vs. fixed bifiltration (5-fold CV accuracy %, RBF-SVM), under two selection criteria. Both are train-only, so the certificate and the fold protocol hold for either. The *closed-form* criterion is the shrinkage-regularized Fisher/Mahalanobis margin (20); *inner-CV* is a 3-fold cross-validated nearest-centroid accuracy on the training fold. Inner-CV is 0.7 points better on average but is a held-out search, so we report the closed-form criterion as the default and inner-CV as the variant. "pair (folds)" gives the modal choice and how many of the five folds agree on it—the closed-form criterion is markedly more stable. This experiment uses a lighter configuration than Table 2 (5 folds, $K = 120$ landmarks, $S \leq 6$ slices), so only the within-row comparison is meaningful.

		closed-form (Mahalanobis)		inner-CV	
Dataset	fixed (HKS×deg)	selected	pair (folds)	selected	pair (folds)
COX2	78.2 ± 1.6	$\mathbf{80.1 \pm 3.0}$	HKS_1×clus. (5/5)	81.0 ± 3.8	HKS_1×clos. (3/5)
MUTAG	$\mathbf{88.8 \pm 1.0}$	86.7 ± 4.0	HKS_{10}×deg. (4/5)	88.3 ± 2.7	HKS_{10}×clos. (2/5)
DHFR	$\mathbf{76.9 \pm 1.3}$	75.9 ± 3.2	HKS_1×clus. (2/5)	75.6 ± 7.4	(five pairs in five folds)

Table 10: Adapting the filtration: gain over the fixed HKS×degree bifiltration (accuracy points, mean over 3 seeds, fixed net). Neither route is reliable: closed-form selection helps on COX2 and hurts on MUTAG and DHFR; gradient learning helps on three of five sets and hurts on two. The selector column is the fold-matched difference of Table 9 under the closed-form criterion and is reported only where that experiment was run.

Dataset	selected − fixed	learned − fixed
COX2	+1.9	+0.8
MUTAG	−2.1	−2.0
DHFR	−1.0	−2.0
NCI1	–	+3.4
PROTEINS	–	+1.0

+0.3 on MUTAG, −0.2 on COX2, −1.3 on PROTEINS and −2.9 on PTC_MR. The pattern is the one selection should produce: it pays precisely where the hand-picked bank omits the discriminating axes—on NCI1 the selector concentrates on an eigenvector-centrality family (eigen×label plus HKS×eigen, stable across folds and seeds) that the fixed bank does not contain—and it costs a little where the fixed bank was already right. This revises the verdict of Table 9: closed-form selection is unreliable at the single-pair level, and useful at the set level. (On IMDB-B the bank harness lacks the essential-class cap that matters most there, and every configuration including selection reads 64–66 against the capped single bifiltration's 70.0; we report the capped number in Table 2 and flag the harness gap rather than mix the two.)

D-COMPLEX: DIFFERENTIABILITY AT TIES, AND THE IMPLEMENTATION

Proposition E.1 (Gather gradient)**.** *Fix θ_0 such that (a) the persistence pairing is constant on a neighborhood of θ_0 and (b) for every simplex σ appearing as a birth or death simplex, the maximizing vertex $\arg\max_{v \in \sigma} u_v^{(s)}$ is unique. Then on that neighborhood each finite bar has $b_i = s_{\sigma_b}^{(s)}$, $d_i = s_{\sigma_d}^{(s)}$, and*

$$\frac{\partial b_i}{\partial g_\theta(v)} = \mathbf{1}\big[v = \arg\max_{u \in \sigma_b} u_u^{(s)}\big] \, \frac{\partial u_v^{(s)}}{\partial g_\theta(v)},$$

the inner derivative being $1/\cos\vartheta_s$ or $1/\sin\vartheta_s$ on whichever grade attains the max in $u_v^{(s)}$ (and likewise for d_i). Consequently any barcode vectorization differentiable in the bar coordinates—in particular the PLACE/PALACE coordinates (7)—is differentiable in g_θ at θ_0.

Proof. Deferred to Appendix C.

Remark E.2 (Where this is and is not differentiable in Proposition E.1). Condition (a) is the standard one: the pairing is locally constant off a closed set of measure zero in the space of filtration values, and persistence is differentiable there in the sense of Leygonie et al. (2022), with the practical

gather/scatter implementation of Carrière et al. (2021). Condition (b) is stronger and is *not* generic; we state it separately for that reason. A max-of-vertices filtration ties whenever two vertices of a simplex carry the same value, and on labeled molecular graphs such ties are structural, not accidental—atoms of the same type receive identical node features, so entire orbits of vertices tie exactly, on a set of positive measure in θ. What is generically unique is not the value but the *argmax vertex*; at a tie $\max_v u_v$ remains locally Lipschitz but is not differentiable, and the gather rule returns one element of the Clarke subdifferential $\partial^\circ \max = \mathrm{conv}\{\partial u_v : v \in \arg\max\}$ rather than a gradient. This is the usual situation for subgradient descent on a locally Lipschitz objective, and it is what the implementation does; we call D-COMPLEX differentiable in that sense—almost everywhere, with a Clarke subgradient at the ties—and not in the sense of a globally C^1 map.

In practice the pairing is computed once per forward pass, and the barcode endpoints are re-expressed as gathers of the differentiable tensor $s^{(s)}$; the backward pass flows through the gathers and the max-reductions to f_θ, recovering the slice-stack $\Psi_\theta(M)$ as a differentiable function of the network.

D-COMPLEX TRAINING CONDITIONS

Gradient learning of the filtration needs two conditions or it fails outright. The head must be the PLACE/PALACE embedding—with a lightweight top-k vectorization the learned filtration does not separate from the fixed one (on DHFR both sit near 66%, far below the 81% the same data supports), the head rather than the filtration being the bound. And f_θ must be warm-started to reproduce the fixed filtration—otherwise its random bars miss the frozen landmarks, the embedding vanishes, and the model collapses to the majority class.