Binary Measure Sketches for Persistent Homology: All-Scale Interleavings and Memory Lower Bounds

Anonymous authors
Paper under double-blind review

Abstract

Persistent homology summarizes geometry across scales, but computing it from a compressed representation requires knowing which geometric information must survive compression. For distribution-sensitive filtrations, this question concerns both distances and the masses they organize, rather than the number of distinct distance values alone. We introduce binary measure sketches that aggregate a sample into weighted cells and replace representative coordinates by calibrated binary geometry. For finite distance-to-a-measure weighted Vietoris–Rips persistence, we show that mass-preserving aggregation and uniform distance distortion compose into an all-scale interleaving with error $2r_{\mathrm{part}}+\eta$, separating partition loss from coding loss without an additional penalty for the scale axis or homological degree. A complementary packing of degree-zero persistence diagrams gives a topology-specific memory lower bound, showing that the information required by the topological output cannot in general be eliminated by quantizing distances. Experiments examine the resulting error decomposition and the utility of the retained representation. Together, these results connect the geometry a sketch preserves to the persistence it can recover, providing a principled basis for allocating representative and coding budgets.

1 Introduction

Persistent homology converts a metric dataset into a multiscale record of connected components, loops, and higher-dimensional voids (Carlsson, 2009; Edelsbrunner & Harer, 2010). Its robustness is well understood, but using it directly on a large sample remains expensive: the Vietoris–Rips complex may contain exponentially many simplices even when its edge lengths take only a few distinct values. This distinction matters for binary hashing. A b-bit Hamming code has only $b+1$ possible pairwise distances, yet it can still induce a dense complex on arbitrarily many vertices. Quantizing *scale values* is not the same as bounding *state size*.

The useful compression target is therefore the weighted geometry that defines the filtration. Partitioning a sample into M cells reduces the number of retained vertices, but storing their representatives can remain expensive in high dimension. Binary encoding reduces this second cost. Unlike an unweighted subsample, however, a distributional sketch must also retain the mass assigned to each cell: those masses determine when vertices enter a DTM-weighted filtration. This raises our central question: *can weighted representative geometry be replaced by binary codes while certifying persistence across the entire scale axis?*

Our answer separates two approximations (Figure 1). Mass-preserving aggregation incurs a geometric loss governed by the cell radius r_{part}; replacing representative distances by calibrated Hamming distances incurs a distortion η. These losses compose to $2r_{\mathrm{part}}+\eta$, without a further penalty for the number of scales, simplices, or homological degrees. Angular and unit-sphere geometries admit a pure code-and-mass state; unnormalized Euclidean geometry additionally requires radial information. In either case, original-space representative coordinates need not be retained.

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

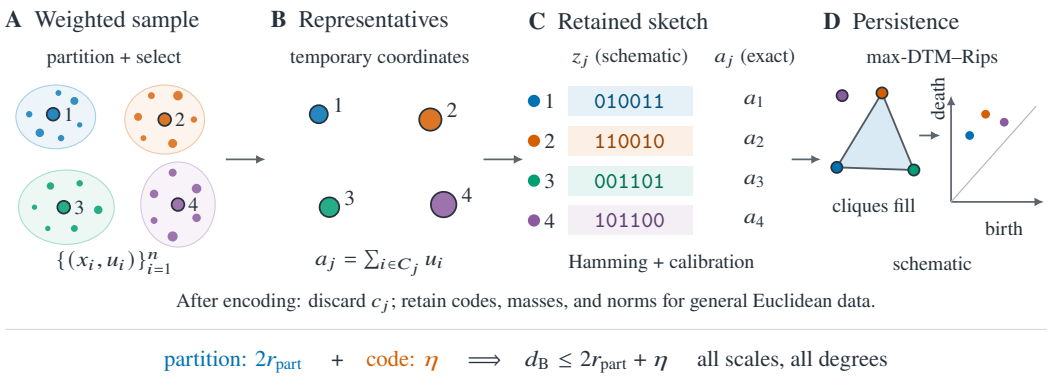


Figure 1: From a sample to a binary measure sketch. Aggregation preserves cell masses; encoding replaces representative coordinates by calibrated binary geometry. The two approximation steps contribute $2r_{\text{part}}$ and η to one all-scale persistence bound. The pure code-and-mass state applies to angular or unit-sphere geometry; unnormalized Euclidean geometry also retains norms. Point positions, codes, and the diagram are schematic, not experimental measurements.

Our contributions are as follows:

1. We prove that, for finite max-DTM–Rips persistence, uniform additive dissimilarity error η gives an η-interleaving on a fixed weighted vertex set. A single random-code distortion event therefore certifies all scales and degrees simultaneously.
2. We prove that a mass-preserving fibre quotient is a scale-natural homotopy equivalence for the pullback geometry. Combining it with code stability yields the sample-to-sketch bound $2r_{\text{part}} + \eta$ and separates the roles of representative count and code length.
3. We construct a packing of explicit H_0 diagrams that gives a memory lower bound for arbitrary decoders, including offline compression. The obstruction concerns persistence itself, not the stronger requirement of reconstructing an input metric.
4. We evaluate the two error terms and retained-state scaling in paired point-cloud experiments; independent graph evaluation tests predictive utility relative to a fixed Gaussian encoder.

The construction builds on established DTM stability and binary embeddings. The new connection is between a labelled, mass-preserving quotient, its code-only all-scale certificate, and an output-level information bound. Appendix G distinguishes these statements from the imported stability results; Appendix B.3 extends the certificate to multiplicative and log scales.

2 Related work

The distance to a measure was introduced as a robust surrogate for distance to a support (Chazal et al., 2011). Its statistical and geometric properties were developed by Chazal et al. (2018; 2016b). For probability measures μ, ν, the classical order-two bound $\| \mathrm{DTM}_{\mu,\beta} - \mathrm{DTM}_{\nu,\beta} \|_\infty \le \beta^{-1/2} W_2(\mu, \nu)$ controls persistence of ambient DTM sublevel sets (Buchet et al., 2016). Weighted Čech/Rips and DTM-based filtrations make these ideas finite and computable (Buchet et al., 2016; Anai et al., 2019). Our fixed-measure theorem is different: it perturbs the metric while keeping the atomic weights fixed, so its sharp constant has no $\beta^{-1/2}$ loss. Kang et al. (2024) provide careful DTM sensitivity and lower-bound arguments in a privacy model; our memory lower bound uses a different diagram packing model.

Function stability and the algebraic stability/isometry framework connect sup-norm or interleaving bounds to bottleneck distance (Cohen-Steiner et al., 2007; Chazal et al., 2016a; Lesnick, 2015). Rips persistence is stable under metric-space perturbations (Chazal et al., 2014). Sparse Rips constructions reduce complex size under geometric assumptions (Sheehy, 2013); subsampling and diagram sketching reduce different parts of the pipeline (Chazal et al., 2015; Sheehy & Sheth, 2021). Recent multicover sparsification and hardness results sharpen the boundary for measure bifiltrations (Alonso,

2025; Kolbe & Mayr, 2026; McCabe, 2026). Our state stores weighted codes *before* persistence is formed.

Random-hyperplane hashing estimates angular distance (Charikar, 2002; Plan & Vershynin, 2014). Uniform finite-set binary embeddings use $b = O(\eta^{-2} \log M)$ bits per code in standard settings, with corresponding lower bounds (Yi et al., 2015; Oymak & Recht, 2015). These results provide one way to satisfy our deterministic distortion premise. Sign codes capture directional information; unnormalized Euclidean reconstruction additionally uses norms. We state the main theorem for an arbitrary calibrated symmetric dissimilarity and specify the target geometry in each corollary.

Persistence landscapes and diagram layers turn topological summaries into finite-dimensional predictors (Bubenik, 2015; Carriere et al., 2020). Recent graph-learning systems instead learn edge filtrations or combine persistent homology with graph convolutions and tensor representations (Shin et al., 2024; Wen et al., 2024). Those methods optimize predictive representations; our question is complementary: after a filtration target has been fixed, how much of its persistence can a low-state binary measure sketch retain? Accordingly, our graph experiments separate structure-only, topology-specific, and node-attribute baselines and never interpret their cross-filtration scores as bottleneck comparisons to max-DTM–Rips.

3 Binary measure filtrations

3.1 Finite DTM weights

Let $\mathcal{X} = (V, d, a)$ be a finite weighted metric space: $V = \{1, \dots, N\}$, $a_x > 0$, and $\sum_x a_x = 1$. Fix a DTM mass parameter $\beta \in (0, 1]$. At $x \in V$, the radial mass distribution, generalized inverse, and order-two DTM weight are

$$F_{x,d}(r) = \sum_{y:\, d(x,y) \le r} a_y, \qquad Q_{x,d}(u) = \inf\{r \ge 0 : F_{x,d}(r) \ge u\}, \tag{1}$$

$$w_{d,a,\beta}(x) = \left(\frac{1}{\beta} \int_0^\beta Q_{x,d}(u)^2 \, du \right)^{1/2}. \tag{2}$$

The integral has an exact finite implementation. Order the vertices so that $r_i = d(x, y_i)$ is nondecreasing and set $A_i = \sum_{\ell \le i} a_{y_\ell}$, with $A_0 = 0$. Then

$$w_{d,a,\beta}(x)^2 = \frac{1}{\beta} \sum_{i=1}^{N} r_i^2 \big(\min\{\beta, A_i\} - \min\{\beta, A_{i-1}\} \big).$$

Thus accumulation stops at β, allowing a fractional final atom. Distance ties can be grouped. Mass at distance zero, including the query atom itself, contributes toward β; if it reaches β, the DTM weight is zero. After aggregation, an unweighted nearest-neighbor average would generally compute a different quantity: a representative stands for its entire cell mass, not one original observation. Appendix A.2 proves the formula and its endpoint conventions.

Definition 1 (Max-DTM–Rips filtration)**.** For nonempty $\sigma \subseteq V$, define its entry time and the sublevel complex by

$$\rho_{d,a,\beta}(\sigma) = \max\left\{ \max_{x \in \sigma} w_{d,a,\beta}(x), \ \max_{x,y \in \sigma} d(x, y) \right\}, \tag{3}$$

$$\mathcal{K}_{d,a,\beta}(r) = \{\varnothing\} \cup \{\varnothing \ne \sigma \subseteq V : \rho_{d,a,\beta}(\sigma) \le r\}.$$

Thus vertices with DTM weight at most r enter first, and the flag complex uses the edge convention $d(x, y) \le r$.

The maximum in (3) makes face monotonicity immediate and yields a direct metric-perturbation constant. This finite filtration is the target of our guarantees; Appendix G describes its relation to ambient DTM sublevel sets and power-weighted Rips constructions.

3.2 State and code dissimilarity

Construction first fixes the representatives and their cell masses, then encodes these same M representatives with b bits each. The representative set does not change between bits. Once the calibration is

fixed, Hamming distances replace the original pairwise distances in the finite DTM and simplex-time calculations.

A binary measure sketch retains M labelled states $\mathsf{S} = \{(z_j, a_j) : j \in [M]\}$, with $z_j \in \{0,1\}^b$, $a_j > 0$, and $\sum_j a_j = 1$. A calibration $g : \{0, 1/b, \dots, 1\} \to \mathbb{R}_+$ induces

$$\widehat{d}(j,k) = g\Big(b^{-1} d_{\mathrm{H}}(z_j, z_k)\Big), \qquad \widehat{d}(j,j) = 0. \tag{4}$$

The downstream procedure computes (1)–(3) using $(\widehat{d}, a)$ only. For an arbitrary g, $\widehat{d}$ is a symmetric zero-diagonal dissimilarity and need not satisfy the triangle inequality. If $g(0) = 0$ and g extends to a nondecreasing subadditive transform, then $\widehat{d}$ is a pseudometric; the stability theorem below does not need this extra condition. Duplicate codes remain distinct labelled vertices. If they are collapsed, their masses must be aggregated and the quotient lemma below must be applied explicitly. For angular or unit-sphere data this is a pure code-and-mass state. The Euclidean extension retains quantized norms as well and uses a radial decoder rather than the one-argument calibration in (4); Appendix F states its distortion and storage costs.

4 All-scale guarantees

4.1 Pure code-only interleaving

Theorem 2 (Fixed-measure dissimilarity stability)**.** *Let d and $\widehat{d}$ be nonnegative symmetric functions with zero diagonal on the same finite labelled set V. Use the same weights a and mass parameter β on both sides. If $\|d - \widehat{d}\|_\infty = \max_{x,y\in V} |d(x,y) - \widehat{d}(x,y)| \le \eta$, the identity on V induces inclusions for every $r \in \mathbb{R}$:*

$$\mathcal{K}_{d,a,\beta}(r) \hookrightarrow \mathcal{K}_{\widehat{d},a,\beta}(r+\eta), \qquad \mathcal{K}_{\widehat{d},a,\beta}(r) \hookrightarrow \mathcal{K}_{d,a,\beta}(r+\eta).$$

They form an η-interleaving of filtered complexes. Consequently, over any field and for every $q \ge 0$,

$$d_{\mathrm{I}}\Big(H_q\mathcal{K}_{d,a,\beta}, H_q\mathcal{K}_{\widehat{d},a,\beta}\Big) \le \eta, \qquad d_{\mathrm{B}}\Big(\mathrm{Dgm}_q\,\mathcal{K}_{d,a,\beta}, \mathrm{Dgm}_q\,\mathcal{K}_{\widehat{d},a,\beta}\Big) \le \eta.$$

The proof does not discretize the scale axis. Uniform dissimilarity distortion first gives ball inclusions, then $|Q_{x,d} - Q_{x,\widehat{d}}| \le \eta$. The reverse triangle inequality in $L^2(0,\beta)$ gives

$$|w_{d,a,\beta}(x) - w_{\widehat{d},a,\beta}(x)| \le \beta^{-1/2}\|Q_{x,d} - Q_{x,\widehat{d}}\|_2 \le \eta. \tag{5}$$

Every entry in the simplex maximum therefore moves by at most η. Finiteness gives q-tameness, so the isometry theorem applies. The full proof, including commuting diagrams and ties, is in Appendix B.

The coefficient of η cannot be reduced in general. For two equally weighted points and $\beta \le 1/2$, both DTM weights vanish. Changing their distance from $R > 2\eta$ to $R + \eta$ moves the unique finite H_0 bar by exactly η: matching it to the diagonal would cost more. Appendix K gives this construction and distinguishes it from sharpness of the partition term at the metric level.

Corollary 3 (A single random-code certificate)**.** *Suppose a random code and calibration obey*

$$\Pr\left[\max_{j,k} |d(c_j, c_k) - \widehat{d}(j,k)| \le \eta\right] \ge 1 - \delta.$$

With probability at least $1 - \delta$, the conclusion of Theorem 2 holds simultaneously for all scales and all homological degrees. No additional union bound over scales, simplices, or degrees is needed.

4.2 Mass aggregation and the end-to-end theorem

Let (V, d, u) be the raw finite weighted sample. A surjection $\pi : V \to [M]$ assigns every point to a cell; choose $c_j \in V$ in cell j and aggregate

$$a_j = \sum_{x:\pi(x)=j} u_x, \qquad r_{\mathrm{part}} = \max_{x\in V} d(x, c_{\pi(x)}).$$

Write $d_C(j,k) = d(c_j, c_k)$.

Theorem 4 (Fibre quotient and binary end-to-end bound)**.** *In the setting above, let $\widehat{d}$ be any code dissimilarity on $[M]$ satisfying $\|d_C - \widehat{d}\|_\infty \le \eta$. Then, for every $q \ge 0$,*

$$d_{\mathrm{B}}\Big(\mathrm{Dgm}_q\, \mathcal{K}_{d,u,\beta}, \mathrm{Dgm}_q\, \mathcal{K}_{\widehat{d},a,\beta}\Big) \le 2r_{\mathrm{part}} + \eta. \tag{6}$$

More strongly, the persistence modules are $(2r_{\mathrm{part}} + \eta)$-interleaved.

To compare different vertex sets, define the pullback pseudometric $\bar{d}(x, y) = d_C(\pi(x), \pi(y))$ on the *original labelled vertices*. Then $\|d - \bar{d}\|_\infty \le 2r_{\mathrm{part}}$, so Theorem 2 compares the raw and pullback filtrations. This step uses the triangle inequality of the original metric, whereas the code perturbation theorem itself does not. For $j = \pi(x)$, grouping each fibre's mass gives the exact identity

$$F_{x,\bar{d},u}(r) = \sum_{k:\, d_C(j,k) \le r} \; \sum_{y:\, \pi(y)=k} u_y = \sum_{k:\, d_C(j,k) \le r} a_k = F_{j,d_C,a}(r).$$

Consequently, radial quantiles and DTM weights agree, and every simplex has the same entry time as its image under π. Retaining the exact masses is essential for this equality; equal weighting of representatives does not in general preserve it.

Choose one fixed section $s(j) = c_j$. Projection and section induce simplicial maps at every scale, with $\pi \circ s$ the identity. The other composition need not fix labels, but it is contiguous to the identity: $\sigma \cup s\pi(\sigma)$ has the same projected vertex set as σ and therefore the same entry time. The induced homology maps are inverses. Because the vertex maps do not depend on scale, these inverses commute with all structure maps, giving an isomorphism of persistence modules rather than only equality of Betti numbers at selected thresholds. Applying the code perturbation adds η, proving (6). Appendix C supplies the full commutative-diagram argument.

The decomposition provides two complementary design levers. Increasing the representative budget can reduce r_{part}, whereas increasing code length can reduce η. The remaining term indicates where further allocation is useful: improving coverage brings coding precision into focus, while refining the code brings partition quality into focus. The certificate supports this allocation rule, with individual realized diagram errors also depending on the geometry and random code.

The certificate also identifies which features must survive compression. Writing $E = 2r_{\mathrm{part}} + \eta$, a finite bar of lifetime greater than $2E$ cannot be matched to the diagonal in a bottleneck matching of cost at most E. It therefore has a counterpart whose birth and death each differ by at most E. Shorter bars may disappear without violating the bound. This does not imply identical Betti numbers at an arbitrary fixed threshold, since a nearby birth or death can cross that threshold. The guarantee concerns multiscale features, rather than equality of complexes at each scale.

Remark 5 (What is independent of n). The input must still be scanned, so construction costs at least $\Omega(n)$ time. The number of retained weighted vertices and the downstream complex in (6) depend on M, b, β, not on the discarded sample size n. Exact empirical counts nevertheless cost $M\lceil \log_2(n+1) \rceil$ bits; fixed-precision masses remove this residual dependence but change the target measure unless a mass-error theorem is added. If M grows with n, the downstream cost grows too.

5 A topology-specific memory lower bound

Metric-reconstruction lower bounds do not automatically imply topological lower bounds: distinct metric measures may have identical diagrams.

Theorem 6 (Topology-specific H_0 bit lower bound)**.** *Fix $0 < \epsilon \le 1/16$, $0 < \beta \le 1$, and an atom budget n. Let*

$$q = \left\lfloor \min\left\{ \frac{n}{2}, \frac{1}{2\beta}, \frac{1}{16\epsilon} \right\} \right\rfloor,$$

and assume $q \ge 1$. There are 2^q uniformly weighted finite metric spaces of diameter at most one, each with $2q \le n$ atoms, whose degree-zero max-DTM–Rips diagrams are pairwise at bottleneck distance at least 4ϵ. Hence:

1. *every deterministic state that permits worst-case bottleneck error at most ϵ uses at least q bits;*
2. *if a randomized B-bit state succeeds on every input with probability at least $1 - \delta$, then*

$$B \ge q + \log_2(1 - \delta).$$

Thus $B = \Omega(\min\{n, \beta^{-1}, \epsilon^{-1}\})$ *deterministically, and also for every fixed* $\delta < 1$*, whenever the three terms are nonvacuous up to constants.*

For $\theta \in \{0,1\}^q$, the construction has q pairs. The i-th within-pair distance is $\ell_i^\theta = 8\epsilon i + 4\epsilon\theta_i$; every between-pair distance is one. Each atom has mass $1/(2q) \geq \beta$, so its DTM weight is zero. The H_0 diagram has one finite bar dying at each ℓ_i^θ, $q-1$ common bars dying at one, and one essential bar. Slot separation, distance to the diagonal, and distance to the common bars are all at least 4ϵ. Two ϵ-balls around target diagrams are disjoint, so counting the 2^B states proves the claim. Appendix D handles every bottleneck matching case and the randomized argument.

The dependence on β^{-1} above comes from forcing zero DTM vertex weights in the same max filtration used by the upper bound. For completeness, Appendix E gives a stronger companion construction for the standard $p = 2$ DTM-power weighted Rips filtration: six-point regular-graph gadgets yield $\Omega(\min\{n, \epsilon^{-1}\})$ bits for *every fixed* $\beta \in (0,1)$, and are realizable as Euclidean point clouds.

6 Memory, computation, and usable instantiations

Table 1: Worst-case costs after partitioning separate the retained geometric budget from raw sample size. w is the machine-word size and k the largest requested homology degree.

object or operation	storage / time	raw-n dependence
binary codes	Mb bits	none
exact count masses	$M\lceil \log_2(n+1) \rceil$ bits	logarithmic
fixed-precision masses	Mb_a bits	none
all code distances	$O(M^2\lceil b/w \rceil)$ word operations	none
all discrete DTM weights	$O(M^2 \log M)$ naive time	none
simplices through degree $k+1$	at most $\sum_{j=1}^{k+2} \binom{M}{j}$	none
input assignment	at least $\Omega(n)$ time	linear

Hamming distances take only $b+1$ values and support histogram-based DTM calculation. The M vertex weights contribute additional filtration values, while simplex counts depend on the retained vertex budget and requested homology degree. Downstream computation therefore operates on *a fixed geometric budget in* M, b*, with at most logarithmic* n*-dependence for exact count masses*. This separates compression of the input from the subsequent complexity of persistence reduction.

For calibrations with $g(0) = 0$, the code-only computation forms, at each vertex j, the histogram $H_{j,\ell} = \sum_{k:\, d_{\mathrm{H}}(z_j, z_k) = \ell} a_k$, $\ell = 0, \ldots, b$. Sweeping these masses in increasing order of $g(\ell/b)$ evaluates the finite DTM formula exactly; a nonmonotone calibration requires sorting the calibrated levels first. This groups equal-distance contributions without discarding mass or expanding cells back into raw points. Vertices then enter at their DTM weights, and edges at $\max\{w_j, w_k, \widehat{d}(j,k)\}$; higher simplices enter when all their faces are present. The radial Euclidean decoder must also use stored norms, so Hamming bins alone do not suffice for that extension.

For unit vectors fixed independently of the projection, $z_j = \mathrm{sign}(Ac_j)$ with independent Gaussian rows estimates normalized angular distance. Hoeffding's inequality and a union bound over the $\binom{M}{2}$ pairs give, for normalized Hamming distance,

$$\Pr\left[\max_{j<k} \left| b^{-1} d_{\mathrm{H}}(z_j, z_k) - d_{\mathbb{S}}(c_j, c_k) \right| > \eta \right] \leq M(M-1)e^{-2b\eta^2}.$$

For $M \geq 2$, $b \geq [2\eta^2]^{-1} \log(M(M-1)/\delta)$ is a sufficient condition for Corollary 3; the single-vertex case has no pairwise error. The target here is angular topology. For unit-sphere chordal distance, the calibration $g(s) = 2\sin(\pi s/2)$ is π-Lipschitz and changes the error constant to $\pi\eta$. Unnormalized Euclidean data require stored norm information or a different sketch.

Measuring r_{part} in the same angular metric, a target diagram error $\tau > 2r_{\mathrm{part}}$ gives a direct budget rule: use $\eta = \tau - 2r_{\mathrm{part}}$ in the bit bound and round b up to an integer. If $2r_{\mathrm{part}} \geq \tau$, the partition term leaves no positive allowance for coding error. Increasing M can improve coverage, but also increases code

storage, pairwise work, and the logarithmic factor in the uniform guarantee. Selecting a useful allocation therefore requires a measured or modelled coverage curve, not only a bit concentration bound. The theorem separates these costs without claiming that any one representative budget is universally optimal. The prospective bound assumes a partition independent of code randomness; a measured uniform distortion instead certifies any realized partition and code through the deterministic theorem.

The mass parameter β plays a different role: it sets how much radial mass enters each DTM average. Increasing it does not decrease the vertex weights and leaves the stored code state unchanged. On a fixed weighted metric space of diameter at most D, replacing β by $\beta' \in [\beta, 1]$ changes the persistence diagrams in bottleneck distance by at most $D\sqrt{(\beta' - \beta)/\beta'}$ (Appendix K). At each vertex, splitting the quantile integral at β gives $w^2_{\beta'} \leq (\beta/\beta')w^2_{\beta} + (1 - \beta/\beta')D^2$, where the metric and masses are fixed. Monotonicity then bounds the weight change; pairwise distances are unchanged, so the same bound controls every simplex entry time. This separates a change of filtration target from an approximation to that target: same-target fidelity comparisons must fix β, while sensitivity studies may vary it explicitly.

The same argument also extends to multiplicative distortion. If $(1-\epsilon)d \leq \widehat{d} \leq (1+\epsilon)d$, radial quantiles, DTM weights, and all simplex times obey the same multiplicative sandwich. The filtrations are $(1-\epsilon)^{-1}$-interleaved multiplicatively, or $-\log(1-\epsilon)$-interleaved after positive log reparameterization. Zero times require a common-zero-complex convention; Appendix B.3 states three rigorous options.

7 Empirical evaluation

The evaluation follows the theorem's two-stage structure. We first isolate partition and coding error, then ask whether discarding coordinates yields a useful fidelity–memory trade-off, and finally test whether the compressed persistence retains predictive utility. Point-cloud results use five paired repetitions; graph classification uses the repeated nested protocol in Section 7.1. Same-target comparisons use bottleneck distance to exact max-DTM–Rips persistence. VR, empirical DTM–Rips, and Sparse–Rips have different filtration targets and are compared only through feature recovery and computational cost. Appendix L specifies datasets, baseline definitions, parameter ranges, and uncertainty conventions.

On a contaminated circle, the dense parameter trends in Figure 2A–D separate the effects predicted by Theorem 4. All 335 certificate comparisons satisfy the partition, coding, and end-to-end inequalities within numerical tolerance. Increasing M improves coverage, but at fixed $b = 100$ the mean BMS[1] error changes little from $M = 128$ to 640 (0.085 to 0.082), while FPS-EXACT reaches below 0.001. The plateau is thus attributable to coding, not to inadequate coverage. Increasing b at fixed $M = 128$ instead reduces both distance distortion and coding-only diagram error. The descriptive log–log distortion slope is -0.481, consistent with the $b^{-1/2}$ concentration dependence. These observations support joint allocation of M and b, rather than treating more representatives as a substitute for code precision.

The unmodified optical and range point clouds distributed with JavaPlex (Adams et al., 2014) contain 15,000 unit-sphere observations in dimensions 8 and 25. Their geometry permits a pure code-and-mass state. The full-data comparison evaluates coding fidelity against exact persistence on the same 128 farthest-point representatives; the subset comparison uses the exact filtration on 128 sampled rows. At $b = 100$, BMS-code retains 2,624 bytes including masses and reduces error by 1.9–3.6 times relative to word-rounded JL-bits, using 15% less state (Figure 2E). JL-dim occupies a higher-memory operating point, using 39.4 times as much state for a 1.5–1.9-fold improvement in accuracy. Together, these operating points locate BMS's advantage in the low-memory regime. Replacing original-space representatives by codes is central to this saving and connects the practical comparison to the code-only guarantee.

At fixed M, b, increasing raw sample size from 128 to 32,768 increases partition time while encoding, persistence time, and the 2,624-byte retained state remain approximately flat (Figure 2F). This matches the separation between input processing and post-partition computation in Section 6. The contaminated-circle and real-subset comparisons further distinguish same-target fidelity from cross-

[1]The source code is available in `https://anonymous.4open.science/r/BMS/`.

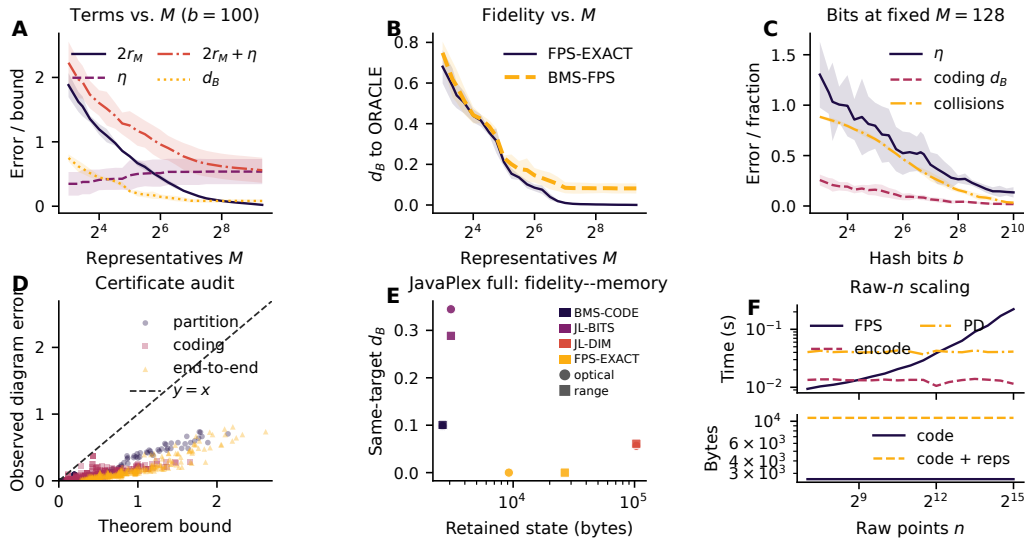


Figure 2: Dense theory-linked validation over five paired repetitions. A–C show mean curves with shaded 95% Student-t confidence intervals: certificate terms and fidelity versus M, then distortion, coding-only d_B, and sign-code collisions versus b. D checks every M, b, and β triple against its corresponding bound; all points lie below $y = x$. E isolates the full-mode JavaPlex comparison: color identifies BMS-CODE, matched-budget JL-BITS, high-memory JL-DIM, or FPS-EXACT, while circles and squares identify optical and range. BMS-CODE occupies the low-memory region and has lower error than JL-BITS; FPS-EXACT stores the original-space representatives. F varies raw n at fixed $M = 128, b = 100$; its mean curves are shown without uncertainty shading, and post-partition time and retained state are flat. Every curve joins measured values only.

filtration feature recovery (Figures 4 and 5). Sparse–Rips emphasizes smaller complexes and faster reduction, while BMS emphasizes a compact weighted representation with a persistence certificate; the two approaches address complementary computational goals.

7.1 Graph classification under a memory budget

Preserving a diagram has a direct interpretation for a fixed landscape representation and affine score: their perturbations can be controlled by the diagram bound (Appendix M). Whether this fidelity also preserves accuracy after training is an empirical question.

We test whether code-only compression preserves useful graph-classification features under a common memory cap. The evaluation uses 792 IMDB-BINARY graphs and 4,032 COLLAB graphs held out from development (Morris et al., 2020). Graph resistance embeddings are converted to DTM–Rips persistence landscapes and classified by logistic regression. At each nominal byte cap, BMS, Gaussian JL, and FPS-EXACT use the same representatives, vertex labels, and masses. Encoder settings are selected on separate development data and then fixed throughout evaluation, so the comparison characterizes these encoders across the common nominal budgets.

All methods use the same 10×10 stratified grouped outer cross-validation and five-fold inner validation. Isomorphic graphs remain in the same group throughout development and evaluation; feature scaling, landscape grids, and classifier selection use only training data. The retained representation includes codes or coordinates, masses, and any required norm information, but excludes the shared classifier and working memory. Appendix M.3 gives the datasets, encoder settings, budget ranges, and additional task-level baselines.

Figure 3 summarizes the graph geometry, persistence diagrams, budget–accuracy curves, and paired effects. The comparison asks whether a compact encoding retains useful topological features under the same nominal memory constraints. Detailed paired inference, repeated-cross-validation corrections, and the definitions of the interval markers are given in Appendix M.3.

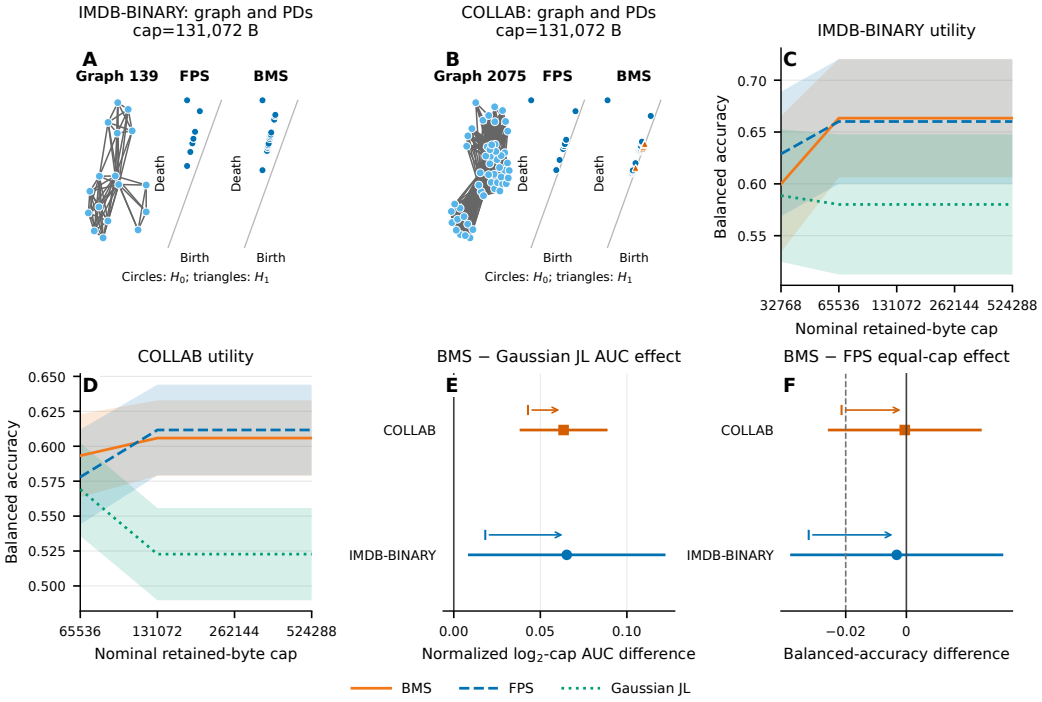


Figure 3: Compression and predictive utility on IMDB-BINARY and COLLAB. A/B show fixed representative graphs and FPS/BMS persistence diagrams on a shared birth/death scale within each panel; circles denote H_0, triangles denote H_1, and essential classes are displayed at the upper edge. C/D show mean balanced accuracy against the $\log_2$ nominal per-object byte cap, with shaded uncertainty bands. E/F summarize paired BMS–JL and BMS–FPS effects. Estimates use 100 outer folds per dataset; the bands, interval markers, and non-inferiority reference are defined in Appendix M.3.

BMS improves the normalized log-budget accuracy area over Gaussian JL by 0.065 on IMDB-BINARY and 0.063 on COLLAB. Its maximum representation size is 2,344 bytes on both datasets, approximately $1/15$ and $1/54$ of the corresponding FPS-EXACT maxima. Together, these results highlight the value of BMS for memory-constrained graph classification, with an improved accuracy–memory trade-off relative to the fixed Gaussian encoder on the evaluated datasets. Comparisons with exact geometry further characterize the balance between representation size and predictive accuracy, with detailed paired analyses provided in Appendix M.3.

8 Discussion and conclusion

Binary measure sketches certify finite DTM-weighted persistence from compressed weighted geometry. Mass-preserving aggregation and uniform dissimilarity control contribute $2r_{\mathrm{part}}$ and η to one all-scale guarantee. Degree-zero diagram packing complements this certificate with an output-level memory obstruction governed by atom and mass-parameter budgets. Together, the bounds link retained geometry to recoverable persistence.

The experiments support jointly allocating representative count and code precision, and demonstrate a low-memory fidelity advantage. The graph study shows useful predictive features relative to a fixed Gaussian encoder. These observations complement the deterministic certificate; generalization and retraining stability additionally depend on the learning pipeline.

The guarantees concern fixed atomic masses and finite max-DTM–Rips persistence, with radial information used for Euclidean reconstruction. The retained vertex budget controls the downstream problem size; simplex counts and reduction costs remain relevant. Appendix G connects this target to ambient DTM theory.

AI USE STATEMENT

Generative AI assisted with language editing, literature and reference checks, and the checking of theoretical statements and proofs. This introduced no new experimental measurements. The authors retain responsibility for the final manuscript, including verification of all AI-assisted content.

ETHICS STATEMENT

This work concerns TDA. No new human-subject data were collected. TDA can amplify representation bias, so deployment in consequential settings requires dataset-specific fairness, privacy, and error analysis.

REFERENCES

Henry Adams, Andrew Tausz, and Mikael Vejdemo-Johansson. javaPlex: A research software package for persistent (Co)Homology. In *Mathematical Software – ICMS 2014*, volume 8592 of *Lecture Notes in Computer Science*, pp. 129–136. Springer Berlin Heidelberg, 2014. doi: 10.1007/978-3-662-44199-2_23. URL https://doi.org/10.1007/978-3-662-44199-2_23.

Ángel Javier Alonso. A sparse multicover bifiltration of linear size. In *41st International Symposium on Computational Geometry (SoCG 2025)*, volume 332 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 6:1–6:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi: 10.4230/LIPIcs.SoCG.2025.6. URL https://doi.org/10.4230/LIPIcs.SoCG.2025.6.

Hirokazu Anai, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hiroya Inakoshi, Raphaël Tinarrage, and Yuhei Umeda. DTM-based filtrations. In *35th International Symposium on Computational Geometry (SoCG 2019)*, volume 129 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 58:1–58:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi: 10.4230/LIPIcs.SoCG.2019.58. URL https://doi.org/10.4230/LIPIcs.SoCG.2019.58.

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16(3):77–102, 2015. URL https://jmlr.org/papers/v16/bubenik15a.html.

Mickaël Buchet, Frédéric Chazal, Steve Y. Oudot, and Donald R. Sheehy. Efficient and robust persistent homology for measures. *Computational Geometry*, 58:70–96, 2016. doi: 10.1016/j.comgeo.2016.07.001. URL https://doi.org/10.1016/j.comgeo.2016.07.001.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009. doi: 10.1090/S0273-0979-09-01249-X. URL https://doi.org/10.1090/S0273-0979-09-01249-X.

Mathieu Carriere, Frederic Chazal, Yuichi Ike, Theo Lacombe, Martin Royer, and Yuhei Umeda. PersLay: A neural network layer for persistence diagrams and new graph topological signatures. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2786–2796. PMLR, 2020. URL https://proceedings.mlr.press/v108/carriere20a.html.

Moses S. Charikar. Similarity estimation techniques from rounding algorithms. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pp. 380–388. ACM, 2002. doi: 10.1145/509907.509965. URL https://doi.org/10.1145/509907.509965.

Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011. doi: 10.1007/s10208-011-9098-0. URL https://doi.org/10.1007/s10208-011-9098-0.

Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014. doi: 10.1007/s10711-013-9937-z. URL https://doi.org/10.1007/s10711-013-9937-z.

Frederic Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Subsampling methods for persistent homology. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2143–2151. PMLR, 2015. URL https://proceedings.mlr.press/v37/chazal15.html.

Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer, Cham, 2016a. doi: 10.1007/978-3-319-42545-0. URL https://doi.org/10.1007/978-3-319-42545-0.

Frédéric Chazal, Pascal Massart, and Bertrand Michel. Rates of convergence for robust geometric inference. *Electronic Journal of Statistics*, 10(2):2243–2286, 2016b. doi: 10.1214/16-EJS1161. URL https://doi.org/10.1214/16-EJS1161.

Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance to a measure and kernel distance. *Journal of Machine Learning Research*, 18(159):1–40, 2018. URL https://jmlr.org/papers/v18/15-484.html.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007. doi: 10.1007/s00454-006-1276-5. URL https://doi.org/10.1007/s00454-006-1276-5.

William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(5):1550066, 2015. doi: 10.1142/S0219498815500668. URL https://doi.org/10.1142/S0219498815500668.

Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010. ISBN 978-0-8218-4925-5. URL https://bookstore.ams.org/mbk-69.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. URL https://www.jstor.org/stable/4615733.

Taegyu Kang, Sehwan Kim, Jinwon Sohn, and Jordan Awan. Differentially private topological data analysis. *Journal of Machine Learning Research*, 25(189):1–42, 2024. URL https://jmlr.org/papers/v25/23-0585.html.

Benedikt Kolbe and Tim Mayr. Persistence meets resistance: Doubling down on hardness. In *53rd International Colloquium on Automata, Languages, and Programming (ICALP 2026)*, volume 374 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 131:1–131:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi: 10.4230/LIPIcs.ICALP.2026.131. URL https://doi.org/10.4230/LIPIcs.ICALP.2026.131.

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. URL https://doi.org/10.1007/s10208-015-9255-y.

Kenneth McCabe. Lower bounds for approximating the Vietoris–Rips filtration. *arXiv preprint arXiv:2607.06524*, 2026. URL https://arxiv.org/abs/2607.06524.

Christopher Morris, Nils M. Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. TUDataset: A collection of benchmark datasets for learning with graphs. In *ICML 2020 Workshop on Graph Representation Learning and Beyond*, 2020. URL https://arxiv.org/abs/2007.08663.

Claude Nadeau and Yoshua Bengio. Inference for the generalization error. *Machine Learning*, 52(3):239–281, 2003. doi: 10.1023/A:1024068626366. URL https://doi.org/10.1023/A:1024068626366.

Samet Oymak and Benjamin Recht. Near-optimal bounds for binary embeddings of arbitrary sets. *arXiv preprint arXiv:1512.04433*, 2015. URL https://arxiv.org/abs/1512.04433.

Yaniv Plan and Roman Vershynin. Dimension reduction by random hyperplane tessellations. *Discrete & Computational Geometry*, 51(2):438–461, 2014. doi: 10.1007/s00454-013-9561-6. URL https://doi.org/10.1007/s00454-013-9561-6.

Donald R. Sheehy. Linear-size approximations to the Vietoris–Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013. doi: 10.1007/s00454-013-9513-1. URL https://doi.org/10.1007/s00454-013-9513-1.

Donald R. Sheehy and Siddharth Sheth. Sketching persistence diagrams. In *37th International Symposium on Computational Geometry (SoCG 2021)*, volume 189 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 57:1–57:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi: 10.4230/LIPIcs.SoCG.2021.57. URL https://doi.org/10.4230/LIPIcs.SoCG.2021.57.

Jaesun Shin, Eunjoo Jeon, Taewon Cho, Namkyeong Cho, and Youngjune Gwon. Line graph Vietoris–Rips persistence diagram for topological graph representation learning. *Journal of Machine Learning Research*, 25(362):1–36, 2024. URL https://jmlr.org/papers/v25/23-1610.html.

Michael E. Van Huffel, Olympio Hacquard, Vadim Lebovici, and Matteo Palo. Discrete transforms of quantized persistence diagrams. In *2025 Proceedings of the Symposium on Algorithm Engineering and Experiments (ALENEX)*, pp. 68–80. Society for Industrial and Applied Mathematics, 2025. doi: 10.1137/1.9781611978339.6. URL https://doi.org/10.1137/1.9781611978339.6.

Oliver Vipond, Joshua A. Bull, Philip S. Macklin, Ulrike Tillmann, Christopher W. Pugh, Helen M. Byrne, and Heather A. Harrington. Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors. *Proceedings of the National Academy of Sciences*, 118(41): e2102166118, 2021. doi: 10.1073/pnas.2102166118. URL https://doi.org/10.1073/pnas.2102166118.

Tao Wen, Elynn Chen, and Yuzhou Chen. Tensor-view topological graph neural network. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 4330–4338. PMLR, 2024. URL https://proceedings.mlr.press/v238/wen24a.html.

Xinyang Yi, Constantine Caramanis, and Eric Price. Binary embedding: Fundamental limits and fast algorithm. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2162–2170. PMLR, 2015. URL https://proceedings.mlr.press/v37/yi15.html.

Appendices

A Foundations, conventions, and finite DTM calculus

This appendix specifies the conventions and assumptions used throughout the proofs, including the generalized inverse, fixed atomic masses, and the relation between edge-threshold and ball-radius parameterizations.

A.1 Notation at a glance

Table 2: Symbols used throughout the paper and appendices.

symbol	meaning
V	finite labelled vertex set
$d, \widehat{d}$	reference and reconstructed dissimilarities
$a = (a_x)_{x\in V}$	positive atomic masses summing to one
$\beta \in (0, 1]$	DTM mass parameter
$F_{x,d}, Q_{x,d}$	radial mass CDF and its generalized inverse
$w_{d,a,\beta}$	order-two discrete DTM vertex weight
$\rho_{d,a,\beta}$	max-DTM–Rips simplex filtration value
$\mathcal{K}_{d,a,\beta}(r)$	subcomplex at edge-threshold scale r
$\pi : V \to [M]$	partition map
c_j, a_j	representative and aggregated mass of cell j
r_{part}	maximum point-to-representative assignment distance
$z_j \in \{0, 1\}^b$	b-bit representative code
o	chosen Euclidean translation center
r_j, R_o	centered representative radius and maximum radius
b_r	radial quantization precision
$b_a = 64$	float64 mass precision in the Euclidean experiments
h_{jk}	normalized Hamming distance
ϵ_h, Δ_{b_r}	angular-proxy and radial reconstruction errors
$d_{\mathrm{B}}, d_{\mathrm{I}}$	bottleneck and interleaving distances

All homology is taken over a fixed field $\Bbbk$. A "diagram" always includes multiplicity. Essential intervals can either be represented by a point with death $+\infty$ or treated as the common essential part of a barcode. In every lower-bound family the number and birth of essential bars are constant, so either convention gives the same finite bottleneck lower bound.

A.2 Radial mass quantiles

Let (V, d, a) be a finite weighted pseudometric space. The pseudometric qualification allows different labelled vertices to have distance zero. Nothing in this subsection uses the triangle inequality.

Definition (Generalized inverse)**.** For $x \in V$ and $u \in (0, 1]$, define

$$Q_{x,d}(u) = \inf\{r \in [0, \infty) : F_{x,d}(r) \geq u\}, \qquad F_{x,d}(r) = \sum_{y:\, d(x,y)\leq r} a_y.$$

We set $F_{x,d}(r) = 0$ for $r < 0$.

Because V is finite, the infimum is attained and belongs to the finite set $\{d(x, y) : y \in V\}$. The value of Q at a jump endpoint depends on a left/right convention only on finitely many u's, a Lebesgue-null set; the integral in (2) is therefore convention-independent.

Lemma A.1 (Finite DTM calculus)**.** *(a) Fix x. Order the vertices so that $r_1 = d(x, y_1) \leq r_2 = d(x, y_2) \leq \cdots \leq r_N = d(x, y_N)$, and let $A_j = \sum_{\ell=1}^{j} a_{y_\ell}$, $A_0 = 0$. Then*

$$w_{d,a,\beta}(x)^2 = \frac{1}{\beta} \sum_{j=1}^{N} r_j^2 \left(\min\{\beta, A_j\} - \min\{\beta, A_{j-1}\}\right). \tag{7}$$

The value is independent of how equal-distance vertices are ordered.

(b) If $Q_{x,d}(u) \le Q_{x,\widehat{d}}(u)$ for almost every $u \in (0,\beta]$, then $w_{d,a,\beta}(x) \le w_{\widehat{d},a,\beta}(x)$. If $Q_{x,\widehat{d}} \le cQ_{x,d}$ for $c \ge 0$, then $w_{\widehat{d},a,\beta}(x) \le c\, w_{d,a,\beta}(x)$.

(c) For any two nonnegative measurable functions $f, g \in L^2(0,\beta)$,

$$\left|\beta^{-1/2}\|f\|_2 - \beta^{-1/2}\|g\|_2\right| \le \beta^{-1/2}\|f-g\|_2.$$

In particular, if $\|f-g\|_\infty \le \eta$, then the left-hand side is at most η.

Proof. (a) For $u \in (A_{j-1}, A_j]$, the generalized inverse equals r_j, except that multiple equal values can split the same constant interval into adjacent pieces. Intersecting each interval with $(0,\beta]$ gives length

$$\lambda_j = \left|(A_{j-1}, A_j] \cap (0,\beta]\right| = \min\{\beta, A_j\} - \min\{\beta, A_{j-1}\}.$$

Integrating $Q_{x,d}(u)^2$ yields (7). Reordering tied terms changes individual λ_j's but their sum over the tied block is fixed and is multiplied by the same squared radius.

(b) Both claims follow by squaring the pointwise inequalities between nonnegative quantiles, integrating, dividing by β, and taking nonnegative square roots.

(c) The first claim is the reverse triangle inequality in the normed space $L^2(0,\beta)$. For the second,

$$\|f-g\|_2^2 = \int_0^\beta |f(u)-g(u)|^2\, du \le \int_0^\beta \eta^2\, du = \beta\eta^2.$$

□

Remark (Self mass and zero weights). If the mass at pseudometric distance zero from x is at least β, then $Q_{x,d}(u) = 0$ for almost every $u \in (0,\beta]$, and hence $w_{d,a,\beta}(x) = 0$. In a genuine metric this mass is at least a_x; in a pseudometric it is the total mass of the zero-distance equivalence class.

The cancellation of $\beta^{-1/2}$ in Lemma A.1(c) is central. It occurs because the same atoms and masses are compared pointwise in radial quantile space. When two *measures* change, an optimal-transport argument controls an integral coupling rather than a uniform quantile difference and produces the familiar $\beta^{-1/2}W_2$ term (Chazal et al., 2011; Buchet et al., 2016; Chazal et al., 2018).

A.3 The max-DTM–Rips filtration is a filtration

Lemma A.2 (Finite maxima and filtration properties)**.** *(a) If $\tau \subseteq \sigma \subseteq V$ are nonempty, then $\rho_{d,a,\beta}(\tau) \le \rho_{d,a,\beta}(\sigma)$. Consequently, $\mathcal{K}_{d,a,\beta}(r)$ is an abstract simplicial complex and $\mathcal{K}_{d,a,\beta}(r) \subseteq \mathcal{K}_{d,a,\beta}(s)$ whenever $r \le s$.*

(b) For two vectors $u, v \in \mathbb{R}^L$,

$$\left|\max_{\ell\le L} u_\ell - \max_{\ell\le L} v_\ell\right| \le \max_{\ell\le L} |u_\ell - v_\ell|.$$

Proof. (a) The set of vertex-weight terms indexed by τ is a subset of those indexed by σ, and likewise for pairwise-distance terms. A maximum cannot increase when terms are removed. Both conclusions follow directly from the sublevel definition.

(b) Let i maximize u. Then

$$\max_\ell u_\ell - \max_\ell v_\ell \le u_i - v_i \le \max_\ell |u_\ell - v_\ell|.$$

Exchange u and v for the reverse difference. □

Remark (Rips scale convention). We place an edge at scale r when $d(x,y) \le r$. Some literature defines Rips complexes using intersecting radius-r balls, resulting in edge threshold $2r$. If only the edge rule is changed, edge perturbations are halved but the DTM vertex-weight perturbation is not, so the overall maximum need not improve. Every filtration time and every additive constant is halved only under the global reparameterization $\rho_{\text{ball}} = \rho/2$, which rescales both vertex and edge times. We never switch conventions inside a theorem.

A.4 Persistence modules, shifts, and interleavings

We recall only the definitions needed to verify the maps in later proofs; see Chazal et al. (2016a); Lesnick (2015) for a full treatment.

Definition (Filtered complex and shift)**.** A real-indexed filtered complex K is a functor from $(\mathbb{R}, \leq)$ to abstract simplicial complexes. We write $\iota^K_{r,s} : K(r) \hookrightarrow K(s)$ for its structure map when $r \leq s$. Its additive η-shift is $(T_\eta K)(r) = K(r + \eta)$.

Definition (Additive interleaving)**.** An η-interleaving between filtered complexes K, L consists of natural families of simplicial maps $\varphi_r : K(r) \to L(r + \eta)$ and $\psi_r : L(r) \to K(r + \eta)$ such that

$$\psi_{r+\eta} \circ \varphi_r = \iota^K_{r,r+2\eta}, \qquad \varphi_{r+\eta} \circ \psi_r = \iota^L_{r,r+2\eta}$$

for every r. Replacing equality by equality after homology gives an interleaving of persistence modules.

For same-vertex sublevel filtrations, inclusions induced by the identity are especially strong: all naturality squares commute strictly and both interleaving equations are literally the standard inclusion.

Definition (q-tameness)**.** A persistence module U is q-tame if $\operatorname{rank}(U(r) \to U(s)) < \infty$ for all $r < s$.

Proposition A.3 (Finite persistence and diagram stability)**.** *(a) For finite V, every homology persistence module $H_q(\mathcal{K}_{d,a,\beta}; \Bbbk)$ is pointwise finite-dimensional and q-tame. It changes at only finitely many parameter values.*

(b) If two finite filtered complexes are additively η-interleaved, then for every q

$$d_{\mathrm{I}}(H_q K, H_q L) \leq \eta, \qquad d_{\mathrm{B}}(\mathrm{Dgm}_q K, \mathrm{Dgm}_q L) \leq \eta.$$

Proof. (a) There are at most $2^{|V|} - 1$ nonempty simplices. Each simplex has one finite entry time $\rho_{d,a,\beta}(\sigma)$, so the filtration changes only at the finite set of those values. Each chain group, cycle group, boundary group, and homology group is finite-dimensional. Every linear map between finite dimensional spaces has finite rank.

(b) Homology is functorial, so an interleaving of filtered complexes induces one of persistence modules with the same shift. Part A.3(a) supplies q-tameness. Algebraic stability, or equivalently the isometry theorem for the resulting finite interval-decomposable modules, gives the bottleneck bound (Chazal et al., 2016a; Crawley-Boevey, 2015). □

A.5 Two sanity checks

Example (Only $b+1$ edge lengths is not constant complexity)**.** Take N vertices with distinct codes in $\{0, 1\}^b$ and choose a threshold at which their Hamming graph becomes complete. The Rips complex then contains $2^N - 1$ nonempty simplices although all edge times lie in a set of size $b + 1$. Restricting to a k-skeleton still permits $\sum_{j=1}^{k+1} \binom{N}{j}$ simplices. Quantized distances may reduce sorting or event scheduling, but do not bound matrix size.

Example (Why masses must be part of the state)**.** Consider two labelled vertices at distance one and let $\beta = 1/2$. If a vertex's own mass changes from $1/2$ to $1/2 - \xi$, however small $\xi > 0$ is, its radial quantile on a set of positive u-measure changes from zero to one. Thus a mass threshold can be crossed discontinuously with respect to coordinatewise mass error. A theorem for approximate masses needs a different assumption, such as Wasserstein control, a margin at every relevant cumulative mass, or an explicit diameter term.

B Pure code-only all-scale interleavings

We now prove Theorem 2 in full. The argument is deterministic. Randomness enters only when a particular coding scheme is used to establish its hypothesis.

B.1 Additive metric perturbations

Lemma B.1 (Additive perturbation from radial balls to simplex times)**.** *Assume* $\max_{x,y\in V}|d(x,y)-\widehat{d}(x,y)|\le\eta$. *With exactly the same masses a and parameter β, the following hold.*

(a) For every $x\in V$ *and* $r\in\mathbb{R}$,

$$\{y : d(x,y)\le r\}\subseteq\{y : \widehat{d}(x,y)\le r+\eta\},\qquad \{y : \widehat{d}(x,y)\le r\}\subseteq\{y : d(x,y)\le r+\eta\}.$$

Consequently,

$$F_{x,\widehat{d}}(r+\eta)\ge F_{x,d}(r),\qquad F_{x,d}(r+\eta)\ge F_{x,\widehat{d}}(r). \tag{8}$$

(b) For every $x\in V$ *and* $u\in(0,1]$, $|Q_{x,d}(u)-Q_{x,\widehat{d}}(u)|\le\eta$.

(c) The DTM vertex weights satisfy

$$\max_{x\in V}|w_{d,a,\beta}(x)-w_{\widehat{d},a,\beta}(x)|\le\eta.$$

(d) For every nonempty simplex $\sigma\subseteq V$, $|\rho_{d,a,\beta}(\sigma)-\rho_{\widehat{d},a,\beta}(\sigma)|\le\eta$.

Proof. (a) If $d(x,y)\le r$, then $\widehat{d}(x,y)\le d(x,y)+\eta\le r+\eta$, proving the first set inclusion. The second follows after exchanging d and $\widehat{d}$. Atomic masses are identical and nonnegative, so summing over the included sets gives (8).

(b) Fix x,u, and set $q=Q_{x,d}(u)$. Finiteness of V means the infimum in the generalized inverse is attained; hence $F_{x,d}(q)\ge u$. The sandwich (8) and the definition of the infimum give

$$F_{x,\widehat{d}}(q+\eta)\ge F_{x,d}(q)\ge u,\qquad Q_{x,\widehat{d}}(u)\le q+\eta=Q_{x,d}(u)+\eta.$$

Repeating the argument with d and $\widehat{d}$ exchanged yields $Q_{x,d}(u)\le Q_{x,\widehat{d}}(u)+\eta$. Combining the two one-sided bounds proves the claim.

(c) Fix x. Apply Lemma A.1(c) with $f=Q_{x,d}|_{(0,\beta)}$ and $g=Q_{x,\widehat{d}}|_{(0,\beta)}$. Lemma B.1(b) gives $\|f-g\|_\infty\le\eta$. Therefore

$$\begin{aligned}|w_{d,a,\beta}(x)-w_{\widehat{d},a,\beta}(x)| &= \frac{1}{\sqrt{\beta}}\left|\|Q_{x,d}\|_{L^2(0,\beta)}-\|Q_{x,\widehat{d}}\|_{L^2(0,\beta)}\right|\\ &\le \frac{1}{\sqrt{\beta}}\|Q_{x,d}-Q_{x,\widehat{d}}\|_{L^2(0,\beta)}\le\frac{1}{\sqrt{\beta}}\sqrt{\beta\eta^2}=\eta.\end{aligned}$$

The bound is uniform in x.

(d) Index a finite vector by all vertices and unordered pairs in σ. For d, fill its vertex coordinates with $w_{d,a,\beta}(x)$ and its pair coordinates with $d(x,y)$; use the corresponding hatted quantities for $\widehat{d}$. Every coordinate changes by at most η, by Lemma B.1(c) and the metric-distortion premise. The filtration values are the maxima of these vectors. Apply Lemma A.2(b). □

Proof of Theorem 2. If $\sigma\in\mathcal{K}_{d,a,\beta}(r)$, then

$$\rho_{\widehat{d},a,\beta}(\sigma)\le\rho_{d,a,\beta}(\sigma)+\eta\le r+\eta.$$

Thus the identity on labelled vertices and the reverse bound in Lemma B.1(d) induce inclusions

$$\varphi_r:\mathcal{K}_{d,a,\beta}(r)\hookrightarrow\mathcal{K}_{\widehat{d},a,\beta}(r+\eta),\qquad \psi_r:\mathcal{K}_{\widehat{d},a,\beta}(r)\hookrightarrow\mathcal{K}_{d,a,\beta}(r+\eta).$$

Both maps act as the identity on every vertex and every simplex in their domain. Hence for $r\le s$, the square

$$\begin{array}{ccc}\mathcal{K}_d(r) & \xrightarrow{\varphi_r} & \mathcal{K}_{\widehat{d}}(r+\eta)\\ \downarrow & & \downarrow\\ \mathcal{K}_d(s) & \xrightarrow[\varphi_s]{} & \mathcal{K}_{\widehat{d}}(s+\eta)\end{array}$$

commutes strictly; the analogous square for ψ does as well. Moreover,

$$\psi_{r+\eta} \circ \varphi_r = \iota^{\mathcal{K}_d}_{r,r+2\eta}, \qquad \varphi_{r+\eta} \circ \psi_r = \iota^{\mathcal{K}_{\widehat{d}}}_{r,r+2\eta},$$

again as literal simplex inclusions. This is an η-interleaving of filtered complexes.

Proposition A.3(a) establishes q-tameness in every homological degree, and Proposition A.3(b) yields both the interleaving and bottleneck bounds. □

B.2 Why the result is genuinely all-scale

The proof never enumerates the distinct values of ρ, nor does it map a finite scale grid to another grid. For every real r, the same identity map is valid. Consequently:

- simultaneous simplex births and distance ties require no generic position assumption;
- the code dissimilarity may assign distance zero to distinct labels;
- the result applies to every homological degree that exists on the finite complex, without a dimension-specific probability event;
- the number of possible Hamming distances is irrelevant to the logical interleaving statement.

Proof of Corollary 3. Let $\mathcal{E} = \{\|d_C - \widehat{d}\|_\infty \le \eta\}$. Conditioned on $\mathcal{E}$, Theorem 2 is deterministic and simultaneously asserts all of its inclusions and diagram bounds. Therefore

$$\Pr[\text{all conclusions of Theorem 2 hold}] \ge \Pr(\mathcal{E}) \ge 1 - \delta.$$

No countable or uncountable union over scales is taken. □

B.3 Multiplicative perturbations

Multiplicative control is often more natural for embeddings of positive distances. It should not be converted silently to an additive bound without a diameter or scale restriction.

Theorem B.2 (Multiplicative code-only interleaving)**.** *Let $0 \le \epsilon < 1$, and suppose*

$$(1-\epsilon)d(x,y) \le \widehat{d}(x,y) \le (1+\epsilon)d(x,y) \quad \text{for all } x, y \in V.$$

With identical a and β, the following hold:

$$(1-\epsilon)Q_{x,d}(u) \le Q_{x,\widehat{d}}(u) \le (1+\epsilon)Q_{x,d}(u), \tag{9}$$

$$(1-\epsilon)w_{d,a,\beta}(x) \le w_{\widehat{d},a,\beta}(x) \le (1+\epsilon)w_{d,a,\beta}(x), \tag{10}$$

$$(1-\epsilon)\rho_{d,a,\beta}(\sigma) \le \rho_{\widehat{d},a,\beta}(\sigma) \le (1+\epsilon)\rho_{d,a,\beta}(\sigma). \tag{11}$$

For every $r \ge 0$,

$$\mathcal{K}_{d,a,\beta}(r) \subseteq \mathcal{K}_{\widehat{d},a,\beta}((1+\epsilon)r), \tag{12}$$

$$\mathcal{K}_{\widehat{d},a,\beta}(r) \subseteq \mathcal{K}_{d,a,\beta}\Big(\frac{r}{1-\epsilon}\Big). \tag{13}$$

Thus the filtrations are multiplicatively $C = (1-\epsilon)^{-1}$-interleaved. On a positive log scale, they are additively $\gamma = -\log(1-\epsilon)$-interleaved.

Proof. The metric sandwich gives ball inclusions

$$B_d(x,r) \subseteq B_{\widehat{d}}(x,(1+\epsilon)r), \qquad B_{\widehat{d}}(x,r) \subseteq B_d\Big(x, \frac{r}{1-\epsilon}\Big).$$

Repeat the achieved-infimum argument from Lemma B.1(b), with a multiplicative radius change, to obtain (9). Since all quantiles are nonnegative, taking normalized L^2 norms preserves the pointwise multiplicative sandwich and proves (10). All terms in the simplex maximum are nonnegative and obey the same sandwich, which gives (11). The two filtration inclusions follow by taking sublevel sets.

Both directions fit the common multiplicative factor $C = \max\{1+\epsilon, 1/(1-\epsilon)\} = 1/(1-\epsilon)$, because $(1+\epsilon)(1-\epsilon) \le 1$. On setting $r = e^s$, a multiplicative factor C becomes an additive shift $\log C = -\log(1-\epsilon)$. □

The multiplicative inclusions are valid on $[0, \infty)$, including $r = 0$. However, $\log 0$ is undefined. A log-diagram statement therefore uses one of the following three rigorous, but generally different, constructions:

- Fix $r_{\min} > 0$ and truncate the filtrations to $[r_{\min}, \infty)$.
- Note from (11) that $\rho_d(\sigma) = 0$ if and only if $\rho_{\widehat{d}}(\sigma) = 0$. The zero-scale complex is common, so study relative persistence $H_q(\mathcal{K}(r), \mathcal{K}(0))$ for $r > 0$.
- Assign the common zero-scale simplices time $-\infty$ on an extended parameter line and allow intervals with an endpoint at $-\infty$.

Truncation moves events below $r_{\min}$ to the boundary, while relative persistence quotients the common zero-scale complex. The extended-parameter convention retains ordinary zero-born finite bars directly. Specifying the construction therefore determines how zero-scale features are represented in a logarithmic diagram.

B.4 Variants that change the constant

If a filtration uses $d(x, y)^2$ or $w(x)^2$, then additive distance error η no longer gives error η. Under $0 \le d, \widehat{d} \le D$,

$$|d^2 - \widehat{d}^2| = |d - \widehat{d}|(d + \widehat{d}) \le 2D\eta,$$

if D bounds both metrics, or at most $2D\eta + \eta^2$ if only $d \le D$ is known. Squared DTM weights obey the same conversion.

The quantile proof compares the same atomic set with the same mass assigned to every radial ball. If $a \ne \widehat{a}$, even $\|a - \widehat{a}\|_1 \ll 1$ does not prevent a jump at β. An extension can use a cumulative-mass margin around every relevant quantile or compare two measures by the classical $\beta^{-1/2} W_2$ theorem. These assumptions account for the additional mass perturbation separately from fixed-measure metric distortion.

An expected error $\mathbb{E}|\widehat{d}(x, y) - d(x, y)| \le \eta$, or a separate high-probability event for each pair, does not imply the uniform event used above. One must either derive a uniform bound (for example by concentration plus a union bound) or prove a different topology-aware certificate.

C Mass aggregation, zero-distance fibres, and the quotient theorem

This appendix proves Theorem 4. The essential point is that aggregating a fibre is not merely deleting vertices. We first pull the representative metric back to the full labelled set, where the fixed-vertex interleaving theorem applies, and only then quotient the resulting zero-distance fibres by natural homotopy equivalences.

C.1 Partition and pullback pseudometric

Let (V, d, u) be a finite weighted metric space, with $\sum_{x \in V} u_x = 1$. Let $\pi : V \to [M]$ be surjective, choose $c_j \in \pi^{-1}(j)$, and define

$$a_j = \sum_{x \in \pi^{-1}(j)} u_x, \qquad d_C(j, k) = d(c_j, c_k), \qquad r_{\text{part}} = \max_{x \in V} d(x, c_{\pi(x)}).$$

Here r_{part} is the maximum assignment radius. Define a pseudometric on the original labels by

$$\bar{d}(x, y) = d_C(\pi(x), \pi(y)) = d(c_{\pi(x)}, c_{\pi(y)}). \tag{14}$$

Distinct vertices in the same fibre have $\bar{d}(x, y) = 0$, which is why $\bar{d}$ need not be a metric.

Lemma C.1 (Pullback distortion)**.**

$$\|d - \bar{d}\|_\infty \le 2r_{\text{part}}.$$

Proof. For arbitrary $x, y \in V$, the triangle inequality for d gives

$$d(x, y) \le d(x, c_{\pi(x)}) + d(c_{\pi(x)}, c_{\pi(y)}) + d(c_{\pi(y)}, y),$$

and hence $d(x,y) - \bar{d}(x,y) \le 2r_{\text{part}}$. Applying the triangle inequality to the representative pair through x, y gives

$$\bar{d}(x,y) \le d(c_{\pi(x)}, x) + d(x,y) + d(y, c_{\pi(y)}),$$

So $\bar{d}(x,y) - d(x,y) \le 2r_{\text{part}}$. Take the maximum. □

Theorem 2, with $\eta = 2r_{\text{part}}$, immediately gives a $2r_{\text{part}}$-interleaving between $\mathcal{K}_{d,u,\beta}$ and $\mathcal{K}_{\bar{d},u,\beta}$. It remains to compare the latter with the M-vertex representative filtration.

C.2 Radial spectra commute with mass aggregation

Lemma C.2 (Exact mass-aggregation identities)**.** *(a) For every $x \in V$, writing $j = \pi(x)$, and for every $r \in \mathbb{R}$, $F_{x,\bar{d},u}(r) = F_{j,d_C,a}(r)$. Consequently, for every $s \in (0,1]$, $Q_{x,\bar{d},u}(s) = Q_{j,d_C,a}(s)$ and*

$$w_{\bar{d},u,\beta}(x) = w_{d_C,a,\beta}(j). \tag{15}$$

(b) For every nonempty $\sigma \subseteq V$,

$$\rho_{\bar{d},u,\beta}(\sigma) = \rho_{d_C,a,\beta}(\pi(\sigma)), \tag{16}$$

where $\pi(\sigma) = \{\pi(x) : x \in \sigma\} \subseteq [M]$.

Proof. (a) Group the full sum by fibres:

$$F_{x,\bar{d},u}(r) = \sum_{y:\, \bar{d}(x,y) \le r} u_y = \sum_{k:\, d_C(j,k) \le r} \; \sum_{y:\, \pi(y) = k} u_y = \sum_{k:\, d_C(j,k) \le r} a_k = F_{j,d_C,a}(r).$$

The CDFs are identical as functions, so their generalized inverses are identical. Integrating the squared quantiles over $(0, \beta]$ proves (15).

(b) By Lemma C.2(a), the multiset of vertex-weight values appearing on the left is obtained by repeating values indexed by $\pi(\sigma)$. Repeating entries does not change their maximum. Moreover,

$$\max_{x,y \in \sigma} \bar{d}(x,y) = \max_{j,k \in \pi(\sigma)} d_C(j,k).$$

Taking the maximum of the vertex and pair terms yields (16). □

C.3 Projection and section

For each scale r, define a vertex map

$$p_r : \mathcal{K}_{\bar{d},u,\beta}(r) \to \mathcal{K}_{d_C,a,\beta}(r), \qquad p_r(x) = \pi(x).$$

An abstract simplicial map is allowed to identify vertices, so its image on σ is the set $\pi(\sigma)$. Lemma C.2(b) shows that p_r is simplicial.

Choose a section $s : [M] \to V$, for example $s(j) = c_j$, and define

$$i_r : \mathcal{K}_{d_C,a,\beta}(r) \to \mathcal{K}_{\bar{d},u,\beta}(r), \qquad i_r(j) = s(j).$$

Because $\pi \circ s = \mathrm{id}_{[M]}$, another application of Lemma C.2(b) shows that i_r is simplicial and

$$p_r \circ i_r = \mathrm{id}_{\mathcal{K}_{d_C,a,\beta}(r)} \tag{17}$$

as a simplicial map.

Proposition C.3 (Fibre contraction and natural quotient isomorphism)**.** *(a) For every scale r, the maps $\mathrm{id}_{\mathcal{K}_{\bar{d},u,\beta}(r)}$ and $i_r \circ p_r$ are contiguous. Therefore they induce the same map on homology.*

(b) For every $q \ge 0$, the maps $H_q(p_r)$ form a natural isomorphism of persistence modules

$$H_q(\mathcal{K}_{\bar{d},u,\beta}) \cong H_q(\mathcal{K}_{d_C,a,\beta}),$$

with inverse $H_q(i_r)$.

Proof. *(a)* Two simplicial maps $f, g : K \to L$ are contiguous if for every simplex $\sigma \in K$, the union $f(\sigma) \cup g(\sigma)$ is a simplex of L. Here, for $\sigma \in \mathcal{K}_{\bar{d},u,\beta}(r)$, consider $\tau = \sigma \cup s(\pi(\sigma))$. Every vertex in τ lies in a fibre already represented in σ, so $\pi(\tau) = \pi(\sigma)$. Lemma C.2(b) gives

$$\rho_{\bar{d},u,\beta}(\tau) = \rho_{d_C,a,\beta}(\pi(\tau)) = \rho_{d_C,a,\beta}(\pi(\sigma)) = \rho_{\bar{d},u,\beta}(\sigma) \le r.$$

Thus τ is a simplex at the same scale. This proves contiguity. Contiguous maps have chain-homotopic simplicial chain maps and hence identical maps on homology.

(b) The identities (17) and part C.3(a), respectively, give

$$H_q(p_r)H_q(i_r) = \mathrm{id}, \qquad H_q(i_r)H_q(p_r) = \mathrm{id}\,.$$

Thus the maps are inverse isomorphisms at every scale.

It remains to verify persistence naturality. The vertex maps $p_r = \pi$ do not depend on r, and the maps $i_r = s$ do not depend on r. Therefore, for $r \le t$, both squares

$$\begin{array}{ccc} \mathcal{K}_{\bar{d}}(r) & \xrightarrow{p_r} & \mathcal{K}_{d_C}(r) \\ \downarrow & & \downarrow \\ \mathcal{K}_{\bar{d}}(t) & \xrightarrow[p_t]{} & \mathcal{K}_{d_C}(t) \end{array} \qquad \begin{array}{ccc} \mathcal{K}_{d_C}(r) & \xrightarrow{i_r} & \mathcal{K}_{\bar{d}}(r) \\ \downarrow & & \downarrow \\ \mathcal{K}_{d_C}(t) & \xrightarrow[i_t]{} & \mathcal{K}_{\bar{d}}(t) \end{array}$$

commute strictly before homology. Hence the scale-wise isomorphisms are natural. □

C.4 Composing the two errors

Proof of Theorem 4. Introduce three persistence modules $U_q = H_q(\mathcal{K}_{d,u,\beta})$, $\bar{U}_q = H_q(\mathcal{K}_{\bar{d},u,\beta})$, and $C_q = H_q(\mathcal{K}_{d_C,a,\beta})$, and the code module $\widehat{C}_q = H_q(\mathcal{K}_{\widehat{d},a,\beta})$. Lemma C.1 and Theorem 2, Proposition C.3(b), and the code premise with Theorem 2, respectively, give

$$d_{\mathrm{I}}(U_q, \bar{U}_q) \le 2r_{\mathrm{part}}, \qquad d_{\mathrm{I}}(\bar{U}_q, C_q) = 0, \qquad d_{\mathrm{I}}(C_q, \widehat{C}_q) \le \eta.$$

The triangle inequality for interleaving distance yields the first bound below; since all modules arise from finite filtrations and are q-tame, the isometry theorem gives the second:

$$d_{\mathrm{I}}(U_q, \widehat{C}_q) \le 2r_{\mathrm{part}} + \eta, \qquad d_{\mathrm{B}}(\mathrm{Dgm}\, U_q, \mathrm{Dgm}\, \widehat{C}_q) \le 2r_{\mathrm{part}} + \eta.$$

□

For completeness, one can avoid invoking the abstract triangle inequality: compose the raw-to-pullback shifted identity, the quotient projection, and the decoded-to-code shifted identity. In the reverse direction compose the code identity, the quotient section, and the pullback-to-raw identity. The two compositions equal the required structure maps after homology because $i_r p_r$ is contiguous to the identity. Their total shift is exactly $2r_{\mathrm{part}} + \eta$.

Corollary C.4 (Randomized end-to-end certificate)**.** *If the partition is fixed and a random code satisfies* $\Pr[\|d_C - \widehat{d}\|_\infty \le \eta] \ge 1 - \delta$*, then with probability at least* $1 - \delta$*, for every degree* q*,*

$$d_{\mathrm{B}}\Big(\mathrm{Dgm}_q\, \mathcal{K}_{d,u,\beta}, \mathrm{Dgm}_q\, \mathcal{K}_{\widehat{d},a,\beta}\Big) \le 2r_{\mathrm{part}} + \eta.$$

Proof. The partition part of the proof is deterministic. Condition on the single uniform code event and apply Theorem 4. □

C.5 Variants of the representative construction

The proof extends when c_j belongs to a common ambient metric space rather than the raw sample V. The section $s(j) \in \pi^{-1}(j)$ no longer equals the representative, but the pullback pseudometric and radial CDF identity are unchanged. To prove contiguity, any selected raw section suffices because all vertices in a fibre have identical pullback profiles. The distortion bound still uses $d(x, c_{\pi(x)}) \le r_{\text{part}}$.

Empty cells carry zero mass and are omitted. Surjectivity of π is thus without loss of generality. Including a zero-mass vertex would violate the positive-mass convention and could introduce topology not represented by the measure.

The code dissimilarity may put two different cells at distance zero. Theorem 2 still holds because the labelled cells are retained. If an implementation merges equal-code cells, it creates a second quotient. This quotient is exact when $g(0) = 0$: equal codes then have zero mutual distance and identical code-distance profiles, and their masses are summed. For $g(0) > 0$, equal codes need not have zero mutual dissimilarity, so merging them is not covered by the exact fibre argument. Under the zero-distance condition, the same projection/section proof applies to the *code* filtration, although comparison with decoded geometry still depends on the uniform distortion event.

If only a certificate $\widetilde{r} \ge \max_x d(x, c_{\pi(x)})$ is available, replace r_{part} by $\widetilde{r}$. Lemma C.1 uses an upper bound on the maximum assignment distance, matching the worst-case nature of the interleaving certificate; average quantization loss provides a complementary summary of partition quality.

D Topology-specific bit and memory lower bounds

This appendix proves Theorem 6. We first state a general diagram-packing principle, then construct a packing inside the exact max-DTM–Rips model used by the upper bound.

D.1 A general output-packing lemma

Let $\mathfrak{X}$ be an input class and $\Phi : \mathfrak{X} \to (\mathfrak{D}, d_{\mathrm{B}})$ a target persistence-diagram map. For $r > 0$, define the strict packing number

$$\mathcal{M}_\Phi(r) = \sup\left\{L : \exists X_1, \dots, X_L \in \mathfrak{X}, \quad d_{\mathrm{B}}(\Phi(X_i), \Phi(X_j)) > r \text{ for } i \ne j\right\}.$$

Using a strict $>$ avoids irrelevant equality issues when decoder error is stated with $\le$. A construction with separation at least 4ϵ is, for example, a strict 2ϵ-packing.

Theorem D.1 (Diagram packing implies memory)**.** *Suppose an encoder maps each input to one of at most 2^B states and a decoder maps the retained state to a diagram.*

- *If the scheme is deterministic and* $\sup_{X\in\mathfrak{X}} d_{\mathrm{B}}(\widehat{D}(X), \Phi(X)) \le \epsilon$*, then*

$$B \ge \log_2 \mathcal{M}_\Phi(2\epsilon).$$

- *If the scheme is randomized and, for every input,*

$$\Pr[d_{\mathrm{B}}(\widehat{D}(X), \Phi(X)) \le \epsilon] \ge 1 - \delta, \qquad 0 \le \delta < 1,$$

  then for every finite strict 2ϵ-packing of size L,

$$B \ge \log_2((1-\delta)L). \tag{18}$$

Proof. For the deterministic claim, assume two packing inputs X_i, X_j produce the same state. Their decoded diagrams are identical, so the triangle inequality gives

$$d_{\mathrm{B}}(\Phi(X_i), \Phi(X_j)) \le d_{\mathrm{B}}(\Phi(X_i), \widehat{D}) + d_{\mathrm{B}}(\widehat{D}, \Phi(X_j)) \le 2\epsilon,$$

contradicting strict separation. Hence the encoder is injective on the packing, and $2^B \ge L$.

For the randomized claim, place the uniform distribution on the L packing inputs and regard every random decision made by the encoder and decoder as part of one master random string R, independent of the input. The assumed per-input success probability implies an average success probability at least

$1-\delta$ over the joint draw of input and R. Therefore some fixed realization $R = r$ succeeds on at least $(1-\delta)L$ inputs. With this random string fixed, the scheme is deterministic. The first argument shows that all successful inputs must use different states, so $2^B \geq (1-\delta)L$. Taking base-two logarithms proves (18). □

Remark (Fano form). If one prefers an information-theoretic statement, the decoded ϵ-neighbourhood identifies the packing index on success. Fano's inequality and data processing, applied to the actual error probability $P_e \leq \delta$, give

$$B \geq \log_2 L - h_2(P_e) - P_e \log_2(L-1) \geq (1-\delta)\log_2 L - 1,$$

where h_2 is binary entropy and the last step uses $h_2(P_e) \leq 1$. The fixed-randomness counting bound in (18) is sharper for the worst-case success model used here.

Remark (The model is deliberately broad). The decoder may perform unlimited computation and output an arbitrary diagram; it need not reconstruct distances or use our algorithm. The encoder may inspect the entire input offline. Therefore the lower bound automatically applies to one-pass, multi-pass, mergeable, and streaming sketches whose retained state has B bits.

D.2 The paired metric family

Fix $q \geq 1$, $0 < \epsilon \leq 1/16$, and assume

$$q \leq \frac{1}{16\epsilon}. \tag{19}$$

The common vertex set is $V_q = \{u_1, v_1, \ldots, u_q, v_q\}$, with uniform masses $a_x = 1/(2q)$. For every bit vector $\theta = (\theta_1, \ldots, \theta_q) \in \{0,1\}^q$, define $\ell_i^\theta = 8\epsilon i + 4\epsilon\theta_i$ and set

$$d_\theta(u_i, v_i) = \ell_i^\theta, \qquad d_\theta(x, y) = 1 \quad \text{if } x, y \text{ belong to different pairs.} \tag{20}$$

The diagonal is zero and the definition is symmetric.

Lemma D.2 (Geometry, weights, and barcode of the paired family)**.** *(a) Every d_θ is a metric of diameter at most one. If $q \geq 2$, its diameter is exactly one. Moreover, $8\epsilon \leq \ell_i^\theta \leq 3/4$.*

(b) Every paired metric (V_q, d_θ) embeds isometrically into $\mathbb{R}^{2q}$.

(c) If $\beta \leq 1/(2q)$, then for every θ and $x \in V_q$, $w_{d_\theta,a,\beta}(x) = 0$.

(d) If $\beta \leq 1/(2q)$, then $\mathrm{Dgm}_0 \mathcal{K}_{d_\theta,a,\beta}$ consists of:

- *one point $(0, \ell_i^\theta)$ for every $i \in [q]$;*
- *$q-1$ copies of $(0,1)$;*
- *one essential interval born at zero.*

Proof. (a) The lower bound is immediate. Using (19) and $\epsilon \leq 1/16$,

$$\ell_i^\theta \leq 8\epsilon q + 4\epsilon \leq \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

Thus every within-pair distance is in $(0,1)$, and every between-pair distance equals one.

For a triangle, either all three vertices lie in at least two different pairs, or two share a pair and the third does not. In the latter case the side lengths are $\ell_i^\theta, 1, 1$, and $\ell_i^\theta \leq 2$ while $1 \leq 1 + \ell_i^\theta$. In the former case all relevant inequalities are immediate from side lengths at most one and at least one on the longest side. Hence the triangle inequality holds. If $q \geq 2$, a between-pair distance attains one; if $q = 1$, the sole nonzero distance is $\ell_1^\theta \leq 3/4$.

(b) Let $e_1, \ldots, e_q, f_1, \ldots, f_q$ be an orthonormal basis of $\mathbb{R}^{2q}$. Put $r_i = \ell_i^\theta/2$ and $s_i = \sqrt{1/2 - r_i^2}$. Lemma D.2(a) gives $r_i \leq 3/8$, so s_i is real. Define

$$\varphi(u_i) = s_i e_i + r_i f_i, \qquad \varphi(v_i) = s_i e_i - r_i f_i.$$

Within pair i, $\|\varphi(u_i) - \varphi(v_i)\| = 2r_i = \ell_i^\theta$. For points from different pairs $i \neq j$, all four relevant basis directions are mutually orthogonal. Independently of the two signs,

$$\|\varphi(x_i) - \varphi(x_j)\|^2 = s_i^2 + r_i^2 + s_j^2 + r_j^2 = \frac{1}{2} + \frac{1}{2} = 1.$$

Thus every prescribed distance in (20) is realized exactly.

(c) The atom at x alone has mass $a_x = 1/(2q) \geq \beta$ and distance zero from x. Hence $F_{x,d_\theta}(0) \geq \beta$ and $Q_{x,d_\theta}(u) = 0$ for almost every $u \in (0, \beta]$. The definition (2) gives a zero weight.

(d) Thus the max-DTM–Rips filtration reduces to the ordinary edge-threshold Rips filtration on this family. Degree-zero persistence can be read from the minimum spanning forest, but we compute it directly.

At scale zero, all $2q$ vertices are present and no distinct pair is joined, so there are $2q$ connected components. Before scale one, the only possible edges are the q disjoint within-pair edges. At $r = \ell_i^\theta$, the edge $\{u_i, v_i\}$ joins two previously distinct components and kills one H_0 class. These events contribute the q listed points, with simultaneous events allowed.

After all within-pair edges have appeared, exactly q components remain, one for each pair. At scale one every between-pair edge appears, so the graph becomes complete. Joining q components kills $q - 1$ further classes at one. The final component persists forever. The total number of finite bars is $q + (q - 1) = 2q - 1$, as required for a filtration beginning with $2q$ isolated vertices and ending connected. □

D.3 Every bottleneck matching case

Lemma D.3 (Slot and exact diagram separation)**.** *(a) For any $i, j \in [q]$ and bits $s, t \in \{0, 1\}$,*

$$(i, s) \neq (j, t) \quad \Longrightarrow \quad |8\epsilon i + 4\epsilon s - (8\epsilon j + 4\epsilon t)| \geq 4\epsilon.$$

(b) If $\beta \leq 1/(2q)$ and $\theta \neq \theta'$, then

$$d_{\mathrm{B}}\big(\mathrm{Dgm}_0 \mathcal{K}_{d_\theta, a, \beta}, \mathrm{Dgm}_0 \mathcal{K}_{d_{\theta'}, a, \beta}\big) = 4\epsilon.$$

Proof. (a) If $i = j$, then $s \neq t$ and the difference is exactly 4ϵ. If $j \geq i + 1$, the smallest possible positive difference occurs for $(s, t) = (1, 0)$ and $j = i + 1$: $8\epsilon(i + 1) - [8\epsilon i + 4\epsilon] = 4\epsilon$. The case $i \geq j + 1$ is symmetric.

(b) To prove the upper bound, match the essential bars and all $q - 1$ copies of $(0, 1)$ identically. For each slot i, match $(0, \ell_i^\theta)$ to $(0, \ell_i^{\theta'})$. This costs either zero or 4ϵ. Hence a bottleneck matching of cost 4ϵ exists.

For the lower bound, choose an index i with $\theta_i \neq \theta_i'$, and consider the point $p_i = (0, \ell_i^\theta)$. We enumerate every possible destination under a bottleneck matching of finite points augmented by the diagonal.

The point in the same slot in the other diagram has death differing by exactly 4ϵ.

For a point in a different internal slot, Lemma D.3(a) shows that its death differs by at least 4ϵ. Since all births are zero, the L^∞ point cost is at least 4ϵ.

For a common bar dying at one, Lemma D.2(a) gives

$$1 - \ell_i^\theta \geq \frac{1}{4} \geq 4\epsilon,$$

where the last inequality uses $\epsilon \leq 1/16$.

The L^∞ distance from $(0, \ell_i^\theta)$ to the diagonal is half its persistence:

$$\frac{\ell_i^\theta}{2} \geq \frac{8\epsilon}{2} = 4\epsilon.$$

A finite-death point cannot match an infinite-death point at finite cost.

Thus p_i has no possible partner at cost strictly below 4ϵ. Every bottleneck matching costs at least 4ϵ. □

The lower-bound proof explicitly checks the diagonal because deleting a short bar is often cheaper than matching it to a perturbed copy. The initial offset 8ϵ rules out that escape.

D.4 Proof of the headline lower bound

Proof of Theorem 6. The definition of q ensures

$$2q \le n, \qquad \beta \le \frac{1}{2q}, \qquad q \le \frac{1}{16\epsilon}.$$

The paired-family lemma, parts D.2(a)–D.2(d), and Lemma D.3(b) therefore construct 2^q admissible inputs whose target diagrams have mutual distance exactly 4ϵ, and hence form a strict 2ϵ-packing.

The deterministic part of Theorem D.1 and, for the randomized part, $L = 2^q$ in (18) give, respectively,

$$B \ge \log_2(2^q) = q, \qquad B \ge \log_2((1-\delta)2^q) = q + \log_2(1-\delta).$$

□

Corollary D.4 (Word memory and error as a function of bits)**.** *(a) If one RAM word stores w bits, the same family requires at least $\lceil q/w \rceil$ words deterministically, and at least $\lceil (q + \log_2(1-\delta))/w \rceil$ words under the randomized guarantee, whenever the numerator is positive.*

(b) If $n \ge 2(B+1)$ and $\beta \le 1/(2(B+1))$, then every deterministic B-bit sketch for the max-DTM–Rips H_0 diagram has worst-case bottleneck error strictly greater than $1/(16(B+1))$ on inputs of diameter at most one. In particular, the construction gives an $\Omega(B^{-1})$ error lower bound when $B = O(\min\{n, \beta^{-1}\})$.

Proof. (a) Divide the deterministic and randomized bit lower bounds by w and round up to an integer number of words.

(b) Set $q = B+1$ and use the hard family with construction parameter $\epsilon_0 = 1/(16q)$. The assumptions ensure $2q \le n$ and $\beta \le 1/(2q)$. If worst-case error were at most ϵ_0, Theorem 6 would require at least $q = B + 1$ bits, a contradiction. □

D.5 Why an input-space packing would not suffice

The classical DTM stability inequality has the direction

$$d_{\mathrm{B}}(\mathrm{Dgm}(\mathrm{DTM}_{\mu,\beta}), \mathrm{Dgm}(\mathrm{DTM}_{\nu,\beta})) \le \beta^{-1/2} W_2(\mu, \nu).$$

It does not lower-bound diagram distance by transport distance. Consequently, a large W_2-packing can collapse to a single diagram.

Example (Positive transport, identical topology)**.** On the two-point metric space $\{0, 1\}$, let $\mu_p = p\delta_0 + (1-p)\delta_1$. Choose $p \ne p'$ with $\beta < \min\{p, 1-p, p', 1-p'\}$. For both measures, each support point has at least β self mass, so all discrete DTM weights are zero. The max-DTM–Rips H_0 diagram is one finite bar dying at distance one plus an essential bar, independent of the chosen mass split. Yet $W_2(\mu_p, \mu_{p'}) > 0$. The explicit construction above avoids this logical gap by packing diagrams themselves.

E A mass-uniform lower bound for DTM-power weighted Rips

The lower bound in Appendix D uses the max-DTM–Rips filtration and forces DTM vertex weights to zero, which introduces the β^{-1} term. This appendix gives a stronger companion result for the standard DTM-power weighted Rips filtration (Buchet et al., 2016; Anai et al., 2019). A regular-graph gadget makes every point's radial distance multiset independent of the encoded bit. The DTM weights are therefore identical across the entire hard family for any fixed $\beta \in (0, 1)$.

E.1 The power-weighted convention

Let a finite metric space have a nonnegative vertex weight $w(x)$. For $t \geq w(x)$, define the order-two weighted radius $r_x(t) = \sqrt{t^2 - w(x)^2}$; for $t < w(x)$, the vertex is absent. At time t, an edge $\{x, y\}$ appears when both vertices are present and $r_x(t) + r_y(t) \geq d(x, y)$. The weighted Rips complex is the flag completion of this graph. We use $w(x) = w_{d,a,\beta}(x)$. This appendix concerns this power-weighted filtration only; constants are not transferred to Definition 1.

If $w(x) = w(y) = w$ and $d(x, y) = r$, the edge time has the closed form

$$T_w(r) = \sqrt{w^2 + \frac{r^2}{4}}. \tag{21}$$

E.2 Six-point gadgets at separated scales

Theorem E.1 (Mass-uniform Euclidean diagram packing)**.** *Fix any $\beta \in (0, 1)$ and integer $K \geq 1$. There are 2^K metric-measure spaces of diameter at most one, $\{\mathcal{X}_\theta : \theta \in \{0, 1\}^K\}$, each with $6K$ uniform atoms, such that:*

- *every $\mathcal{X}_\theta$ is isometrically realizable in Euclidean dimension at most $6K$;*
- *the order-two DTM weight of every labelled atom is independent of θ;*
- *their degree-zero power-weighted Rips diagrams satisfy*

$$d_{\mathrm{B}}(\mathrm{Dgm}_0 \mathcal{X}_\theta, \mathrm{Dgm}_0 \mathcal{X}_{\theta'}) \geq \Delta_K := \frac{1}{16\sqrt{65}(2K + 9)} \quad \text{whenever } \theta \neq \theta'.$$

Consequently, worst-case error below $\Delta_K/2$ requires at least K bits deterministically and at least $K + \log_2(1 - \delta)$ bits for per-input success probability $1 - \delta$.

Before diameter normalization, set $X = [K] \times [6]$ and $\mu = (6K)^{-1} \sum_{x \in X} \delta_x$. For gadget i, define

$$a_i^2 = 2i + 8, \qquad b_i^2 = a_i^2 + 1 = 2i + 9, \qquad b_{\max} = \sqrt{2K + 9}, \qquad L = 4b_{\max}.$$

If $\theta_i = 0$, let the short-distance graph G_i be the six-cycle C_6. If $\theta_i = 1$, let $G_i = C_3 \sqcup C_3$, the disjoint union of two triangles. For two distinct points in the same gadget, set

$$d_\theta(x, y) = \begin{cases} a_i, & \{x, y\} \in E(G_i), \\ b_i, & \{x, y\} \notin E(G_i). \end{cases} \tag{22}$$

For points in different gadgets, set $d_\theta(x, y) = L$. Denote this unnormalized metric-measure space by $\widetilde{\mathcal{X}}_\theta = (X, d_\theta, \mu)$.

E.3 Geometry, bit-independent weights, and the internal barcode

Both C_6 and $C_3 \sqcup C_3$ are two-regular on six vertices. Thus every point in gadget i, for every value of θ_i, sees the same distance multiset:

$$\left\{0,\ a_i, a_i,\ b_i, b_i, b_i,\ \underbrace{L, \ldots, L}_{6K-6 \text{ entries}}\right\}. \tag{23}$$

Let $\alpha = 1/(6K)$, the mass of one atom, and define

$$u = \frac{(\min\{\beta, 3\alpha\} - \alpha)_+}{\beta}, \tag{24}$$

$$v = \frac{(\min\{\beta, 6\alpha\} - 3\alpha)_+}{\beta}, \tag{25}$$

$$c = \frac{(\beta - 6\alpha)_+}{\beta}. \tag{26}$$

The first α units of radial mass lie at zero, the next 2α units at a_i, the next 3α at b_i, and the remaining mass at L.

Lemma E.2 (The gadget geometry and its encoded internal barcode)**.** *(a) For every θ, d_θ is a metric of diameter at most L. If $K \geq 2$, its diameter is exactly L.*

(b) All six vertices of gadget i have the common weight w_i, independent of θ, where

$$w_i^2 = ua_i^2 + vb_i^2 + cL^2 = (u+v)a_i^2 + v + cL^2. \tag{27}$$

Writing $\gamma = u + v \in [0,1]$ and $C = v + cL^2$, we have

$$w_i^2 = \gamma a_i^2 + C. \tag{28}$$

(c) For gadget i, let

$$A_i = T_{w_i}(a_i) = \sqrt{w_i^2 + \frac{a_i^2}{4}}, \qquad B_i = T_{w_i}(b_i) = \sqrt{w_i^2 + \frac{b_i^2}{4}}. \tag{29}$$

The identities (28) and (29) give

$$A_i^2 = C + \left(\gamma + \frac{1}{4}\right) a_i^2, \tag{30}$$

$$B_i^2 = A_i^2 + \frac{1}{4}. \tag{31}$$

Then

$$A_1 < B_1 < A_2 < B_2 < \cdots < A_K < B_K.$$

Moreover, the squares of any two distinct values in this list differ by at least $1/4$.

(d) Ignoring the later, bit-independent cross-gadget mergers, gadget i contributes to its degree-zero persistence diagram

$$\begin{aligned} \theta_i = 0: &\quad 5 \text{ copies of } (w_i, A_i), \\ \theta_i = 1: &\quad 4 \text{ copies of } (w_i, A_i) \sqcup 1 \text{ copy of } (w_i, B_i). \end{aligned}$$

Proof. (a) Within one gadget, every nonzero distance is either a_i or b_i. Because $b_i^2 = a_i^2 + 1 < 4a_i^2$, we have $b_i < 2a_i$. Hence the longest possible side of an intragadget triangle is shorter than the sum of the other two.

For a triangle with vertices in at least two gadgets, every cross-gadget side has length L. If exactly two vertices share a gadget, the side lengths are r, L, L with $0 < r < L$, so all triangle inequalities hold. If all vertices belong to different gadgets, all sides are L. Positivity and symmetry are immediate. If $K \geq 2$, cross-gadget pairs attain L; if $K = 1$, all nonzero distances are at most $b_1 < L$.

(b) Apply the finite-sum formula in Lemma A.1(a) to (23). The normalized amount of $(0, \beta]$ assigned to radius a_i, radius b_i, and radius L is respectively u, v, c. This proves the first identity. Since $b_i^2 = a_i^2 + 1$, the second identity and (28) follow. The coefficients depend on β, K, but not on i or θ. Their nonnegativity and the fact that their sum is at most one give $\gamma \in [0,1]$.

(c) The identity (31) gives $A_i < B_i$ and a squared gap $1/4$. Because $a_{i+1}^2 - a_i^2 = 2$,

$$A_{i+1}^2 - B_i^2 = \left(\gamma + \frac{1}{4}\right)(a_{i+1}^2 - a_i^2) - \frac{1}{4} = 2\gamma + \frac{1}{4} \geq \frac{1}{4}.$$

Adjacent values in the displayed list have squared gap at least $1/4$; nonadjacent values have a larger telescoping gap.

(d) At time w_i, all six gadget vertices are born. At A_i, precisely the edges of G_i have entered. A six-cycle is connected, so it kills five of the six initial components. Two disjoint triangles have two components, so they kill only four. At B_i, every missing intragadget edge enters and the two triangles merge, killing the fifth component.

This is the component count above. In ordinary H_0 persistence, all six classes are born when their vertices enter at w_i. A spanning forest of a connected six-cycle has five edges and yields five deaths at A_i. A spanning forest of two triangles has four edges, yielding four deaths at A_i, followed by one merger at B_i. One gadget component remains for the later cross-gadget stage. □

The bit-independent DTM spectrum is the key device. No restriction such as $\beta \le 1/(6K)$ is imposed. When β extends into other gadgets, the large L-term raises all relevant DTM weights in a bit-independent way.

E.4 Common cross-gadget events and bottleneck separation

Let $w_{\min} = \min_i w_i$ and $w_{\max} = \max_i w_i$. From (28),

$$w_{\max}^2 - w_{\min}^2 = \gamma(a_K^2 - a_1^2) = 2\gamma(K-1) < b_{\max}^2. \tag{32}$$

Let $T_{\max} = B_K$ and define $\eta_K = 1/(8T_{\max})$.

Lemma E.3 (Common cross-gadget barcode and diagram separation)**.** *(a) After the internal events, the remaining K components merge according to edge times determined only by L and the list $w_1, \dots, w_K$. Hence all cross-gadget H_0 bars are identical for every θ.*

(b) If $\theta \neq \theta'$, then

$$d_{\mathrm{B}}(\mathrm{Dgm}_0 \widetilde{\mathcal{X}}_\theta, \mathrm{Dgm}_0 \widetilde{\mathcal{X}}_{\theta'}) \ge \eta_K.$$

Proof. (a) Suppose a cross-gadget edge is present at time t. Its two weighted radii sum to at least L, so at least one radius is at least $L/2$. For that endpoint x,

$$t^2 = w(x)^2 + r_x(t)^2 \ge w_{\min}^2 + \frac{L^2}{4} = w_{\min}^2 + 4b_{\max}^2.$$

On the other hand,

$$B_K^2 = w_{\max}^2 + \frac{b_{\max}^2}{4} < w_{\min}^2 + \frac{5}{4} b_{\max}^2.$$

Therefore every intragadget merger has completed strictly before any cross-gadget edge appears.

Every gadget is connected after B_i. A cross edge between gadgets i, j has the unique entry time determined by

$$\sqrt{t^2 - w_i^2} + \sqrt{t^2 - w_j^2} = L.$$

The distance L and weights w_i, w_j are independent of θ by Lemma E.2(b). Thus the filtered complete graph on the K gadget components is common to the family. Its $K-1$ finite H_0 deaths and final essential component are common as well.

(b) For two distinct internal death levels $s, t \in \{A_i, B_i : i \in [K]\}$, Lemma E.2(c) and $s, t \le T_{\max}$ give

$$|s - t| = \frac{|s^2 - t^2|}{s + t} \ge \frac{1/4}{2T_{\max}} = \eta_K. \tag{33}$$

An internal persistence point cannot cheaply disappear into the diagonal. For an A_i-point,

$$\frac{A_i - w_i}{2} = \frac{A_i^2 - w_i^2}{2(A_i + w_i)} = \frac{a_i^2}{8(A_i + w_i)} \ge \frac{10}{16T_{\max}} = 5\eta_K. \tag{34}$$

For a B_i-point the denominator changes as well, so we check it directly:

$$\frac{B_i - w_i}{2} = \frac{b_i^2}{8(B_i + w_i)} \ge \frac{11}{16T_{\max}} > 5\eta_K.$$

When $K \ge 2$, the first cross-gadget death occurs after every B_i, with a squared gap from B_K larger than $4b_{\max}^2 - \frac{5}{4}b_{\max}^2 = \frac{11}{4}b_{\max}^2$. This also yields a scalar gap larger than η_K. Indeed, every DTM weight is at most the diameter $L = 4b_{\max}$, so the earliest possible cross death $T_\times$ satisfies

$$T_\times - B_K \ge \sqrt{w_{\min}^2 + 4b_{\max}^2} - \sqrt{w_{\min}^2 + \tfrac{5}{4}b_{\max}^2} \ge b_{\max}\left(\sqrt{20} - \sqrt{\tfrac{69}{4}}\right) > 0.31 b_{\max}.$$

The second inequality uses $w_{\min} \le L$ and the fact that the displayed difference decreases with $w_{\min}^2$. Meanwhile

$$\eta_K = \frac{1}{8T_{\max}} \le \frac{1}{4b_1} = \frac{1}{4\sqrt{11}} < 0.076,$$

because $T_{\max} \geq B_1 \geq b_1/2 = \sqrt{11}/2$. As $b_{\max} \geq \sqrt{11}$, the cross-to-internal scalar gap is strictly larger than η_K. For $K = 1$, there are no cross-gadget bars and every cross-bar matching exclusion below is vacuous.

Choose i with $\theta_i \neq \theta_i'$. By Lemma E.2(d), the multiplicity at (w_i, A_i) differs by one, and the multiplicity at (w_i, B_i) differs in the opposite direction. Under a matching of cost strictly below η_K, an unmatched copy cannot:

- move to a different internal death level, by (33);
- move to a cross-gadget bar, by the preceding cross-event gap;
- move to the diagonal, by (34);
- move to an essential bar at finite cost.

The only remaining destinations have exactly the same birth and death coordinates, but their multiplicities are unequal. A perfect bottleneck matching of cost below η_K is impossible. □

E.5 Diameter normalization and the explicit constant

Because DTM weights and power-weighted edge times are homogeneous of degree one in all distances, dividing the metric by L divides every diagram coordinate and bottleneck distance by L. Define the normalized family $\mathcal{X}_\theta = (X, L^{-1}d_\theta, \mu)$. Also,

$$T_{\max} = B_K \leq \sqrt{L^2 + \frac{b_{\max}^2}{4}} = \frac{\sqrt{65}}{2}\, b_{\max},$$

where we used $w_K \leq L$ and $L = 4b_{\max}$. Therefore

$$\frac{\eta_K}{L} = \frac{1}{8T_{\max}L} \geq \frac{1}{16\sqrt{65}\, b_{\max}^2} = \frac{1}{16\sqrt{65}(2K+9)} = \Delta_K.$$

This proves the metric and diagram parts of Theorem E.1.

E.6 Euclidean realization

We next show that the abstract two-distance gadgets and their common cross-distance can be embedded exactly in Euclidean space.

Lemma E.4 (Euclidean six-point gadget)**.** *For $G = C_6$ or $G = C_3 \sqcup C_3$, there are six Euclidean vectors whose pairwise distance is a_i on edges of G and b_i on nonedges. After centering, every vector has squared norm $r_i^2 = (5b_i^2 - 2)/12$, independent of the choice of G.*

Proof. Let $A(G)$ be the 6×6 adjacency matrix and define $H_i = \frac{b_i^2}{2} I_6 + \frac{1}{2}A(G)$. Every eigenvalue of the adjacency matrix of a two-regular graph lies in $[-2, 2]$. Since $b_i^2 \geq 11$, H_i is positive definite and hence is a Gram matrix of six Euclidean vectors $v_1, \ldots, v_6$. For $s \neq t$,

$$\|v_s - v_t\|^2 = H_{ss} + H_{tt} - 2H_{st} = b_i^2 - A(G)_{st}.$$

This equals $b_i^2 - 1 = a_i^2$ on an edge and b_i^2 on a nonedge.

Both graphs are two-regular, so every row sum of H_i equals $b_i^2/2 + 1$. If $\bar{v} = 6^{-1}\sum_s v_s$, then

$$\|v_s - \bar{v}\|^2 = H_{ss} - \frac{1}{6}\left(\frac{b_i^2}{2} + 1\right) = \frac{b_i^2}{2} - \frac{b_i^2/2 + 1}{6} = \frac{5b_i^2 - 2}{12}.$$

Translation preserves pairwise distances. □

Euclidean part of Theorem E.1. Realize each centered gadget in its own subspace E_i of dimension at most five, with the E_i's mutually orthogonal. Introduce a further K-dimensional subspace Z, orthogonal to all E_i, and choose mutually orthogonal center vectors $z_i \in Z$ satisfying $\|z_i\|^2 = L^2/2 - r_i^2$. The right-hand side is positive because

$$r_i^2 < \frac{5b_{\max}^2}{12} < 8b_{\max}^2 = \frac{L^2}{2}.$$

Place gadget i at points $z_i + v$, where $v \in E_i$ ranges over its six centered vectors. For $i \neq j$, orthogonality gives

$$\begin{aligned}\|(z_i + v) - (z_j + v')\|^2 &= \|z_i - z_j\|^2 + \|v\|^2 + \|v'\|^2 \\ &= \left(\frac{L^2}{2} - r_i^2\right) + \left(\frac{L^2}{2} - r_j^2\right) + r_i^2 + r_j^2 = L^2.\end{aligned}$$

All intragadget distances are unchanged. The total dimension is at most $5K + K = 6K$. Scaling by $1/L$ preserves Euclidean realizability. □

E.7 Memory consequences

Memory part of Theorem E.1. The 2^K normalized diagrams are pairwise separated by at least Δ_K. If the decoder error is strictly below $\Delta_K/2$, they form a strict twice-error packing. Apply Theorem D.1 with $L = 2^K$. □

Let $\mathfrak{X}_n$ be the class of finite metric-measure spaces of diameter at most one with at most n atoms, and let Φ_β denote the degree-zero diagram of the order-two DTM-power weighted Rips filtration. Define the deterministic minimax bit budget

$$B_{\mathrm{det}}^{(\beta)}(n,\epsilon) = \inf\left\{ B \in \mathbb{N}_0 : \begin{array}{c}\text{there are a } B\text{-bit encoder } E \text{ and decoder } D \text{ such that} \\ \sup_{X \in \mathfrak{X}_n} d_{\mathrm{B}}\big(D(E(X)), \Phi_\beta(X)\big) \le \epsilon\end{array} \right\}.$$

Define $B_{\mathrm{rand}}^{(\beta,\delta)}(n,\epsilon)$ analogously, requiring the same error with probability at least $1-\delta$ for every input.

Corollary E.5 (Fixed-mass lower-bound rate)**.** *For every fixed $\beta \in (0,1)$, whenever $\min\{n, \epsilon^{-1}\}$ is sufficiently large,*

$$B_{\mathrm{det}}^{(\beta)}(n,\epsilon) = \Omega\Big(\min\{n,\epsilon^{-1}\}\Big), B_{\mathrm{rand}}^{(\beta,\delta)}(n,\epsilon) = \Omega\Big(\min\{n,\epsilon^{-1}\}\Big) \qquad \text{for every fixed } \delta < 1.$$

Equivalently, with input size allowed to grow, a B-bit state has worst-case error $\Omega(B^{-1})$.

Proof. Set

$$K = \left\lfloor \min\left\{ \frac{n}{6}, \frac{1}{256\sqrt{65}\,\epsilon} \right\} \right\rfloor.$$

When $\min\{n,\epsilon^{-1}\}$ is above an absolute constant, $K \ge 9$ and $K = \Omega(\min\{n,\epsilon^{-1}\})$. Moreover $6K \le n$, while $2K + 9 \le 3K$ and therefore

$$\frac{\Delta_K}{2} = \frac{1}{32\sqrt{65}(2K+9)} \ge \frac{1}{96\sqrt{65}K} > \epsilon.$$

Theorem E.1 forces $B \ge K$ deterministically. In the randomized case it forces $B \ge K + \log_2(1-\delta) = K - O_\delta(1)$, which is $\Omega(K)$ for fixed $\delta < 1$ once K is sufficiently large. □

E.8 Why the gadget choices are necessary

- If all K gadgets used the same event scale, their diagram would depend only on the Hamming weight $\sum_i \theta_i$, yielding at most $K+1$ different diagrams. The separated a_i, b_i slots prevent permutation ambiguity.
- Simply requiring all DTM weights to be zero would force $\beta \le 1/(6K)$. Matching the complete radial distance multiset makes the construction valid for every fixed β.
- Birth times are $w_i > 0$, so the distance to the diagonal is $(\text{death} - w_i)/2$, not half the death time. The calculation (34) checks the correct quantity.
- The direct diagram analysis is essential. A packing of metric spaces or measures without separated diagrams would not imply any topology-specific memory lower bound.

F Code constructions, algorithms, and exact memory accounting

The deterministic theorems accept any reconstructed symmetric zero-diagonal dissimilarity satisfying a uniform distortion certificate. This appendix gives concrete instantiations and makes clear which metric each one preserves.

F.1 Random hyperplanes preserve angular distance

Let $x, y \in \mathbb{S}^{d-1}$ and define normalized angular distance $d_{\mathbb{S}}(x, y) = \arccos\langle x, y\rangle/\pi \in [0, 1]$. For $g \sim N(0, I_d)$, rotational symmetry gives the classical identity

$$\Pr[\operatorname{sign}\langle g, x\rangle \neq \operatorname{sign}\langle g, y\rangle] = d_{\mathbb{S}}(x, y) \tag{35}$$

(Charikar, 2002; Plan & Vershynin, 2014).

Draw b independent rows $g_1, \dots, g_b$ and encode $z(x)_\ell = \mathbf{1}\{\langle g_\ell, x\rangle \geq 0\}$. For a fixed pair, the normalized Hamming distance is an average of independent Bernoulli variables with mean $d_{\mathbb{S}}(x, y)$. Hoeffding's inequality gives

$$\Pr\left[\left|b^{-1} d_{\mathrm{H}}(z(x), z(y)) - d_{\mathbb{S}}(x, y)\right| > \eta\right] \leq 2e^{-2b\eta^2}.$$

For M representatives, a union bound over $\binom{M}{2}$ unordered pairs yields

$$\Pr\left[\max_{j<k} \left|b^{-1} d_{\mathrm{H}}(z_j, z_k) - d_{\mathbb{S}}(c_j, c_k)\right| > \eta\right] \leq M(M-1)e^{-2b\eta^2}. \tag{36}$$

Corollary F.1 (Angular code-only persistence)**.** *If $b \geq \log(M(M-1)/\delta)/(2\eta^2)$, then with probability at least $1-\delta$, the max-DTM–Rips filtrations of the normalized angular metric and normalized Hamming metric are η-interleaved at every scale.*

Proof. The bound (36) makes the failure probability at most δ. Apply Corollary 3. □

When $M = 1$, there are no pairs and the code distortion event is vacuous. For $M \geq 2$, the stated bit count is sufficient; ceiling it gives an integer. Sharper binary-embedding results can improve constants or extend from finite sets to structured infinite sets (Yi et al., 2015; Oymak & Recht, 2015). The topological theorem needs only their final uniform event.

F.2 Chordal calibration and norm information

On the unit sphere, chordal Euclidean distance and angular distance satisfy

$$d_E(x, y) = \|x - y\|_2 = 2\sin\left(\frac{\pi d_{\mathbb{S}}(x, y)}{2}\right).$$

Let $h(s) = 2\sin(\pi s/2)$. Since $|h'(s)| = \pi|\cos(\pi s/2)| \leq \pi$, the mean-value theorem gives $|h(s) - h(t)| \leq \pi|s - t|$. Thus the calibrated code distance $\widehat{d}_E(j, k) = h\big(b^{-1} d_{\mathrm{H}}(z_j, z_k)\big)$ obeys uniform error $\pi\eta$ on the event in (36). Theorem 2 then gives a $\pi\eta$-interleaving for chordal max-DTM–Rips persistence.

For unnormalized vectors, sign codes alone are insufficient:

$$\operatorname{sign}(A(cx)) = \operatorname{sign}(Ax) \quad \text{for every } c > 0.$$

Two points on the same positive ray therefore have identical codes regardless of their Euclidean separation. Options include:

- normalize first and state that the target is angular or chordal topology;
- store a quantized norm with every code and calibrate a joint distance;
- use a binary embedding designed for the desired bounded Euclidean domain;
- verify the uniform distortion empirically with retained representatives, then store the resulting certificate with the code-only state.

The last option is a posteriori rather than distribution-free, but Theorem 2 is deterministic and accepts it.

F.3 Centered Euclidean calibration with quantized radii

Let $c_1, \dots, c_M \in \mathbb{R}^d$, choose a translation center o, and write $r_j = \|c_j - o\|_2$ and $R_o = \max_j r_j$. When $r_j > 0$, let $u_j = (c_j - o)/r_j$. For a pair with both radii positive, define its normalized angular distance by $\theta_{jk} = \pi^{-1} \arccos\langle u_j, u_k\rangle$. The normalized Hamming distance is $h_{jk} = b^{-1} d_{\mathrm{H}}(z_j, z_k)$. The following lemma separates angular and radial reconstruction error.

Lemma F.2 (Angular and radial reconstruction errors)**.** *(a) For $r, s \in [0, R]$, define $f_{r,s}(t) = \sqrt{(r-s)^2 + 4rs\sin^2(\pi t/2)}$ for $t \in [0,1]$. Then*

$$|f_{r,s}(t) - f_{r,s}(v)| \le \pi R|t - v| \qquad \text{for all } t, v \in [0,1].$$

(b) For nonnegative $r, s, \widehat{r}, \widehat{s}$ and every $t \in [0,1]$,

$$|f_{r,s}(t) - f_{\widehat{r},\widehat{s}}(t)| \le |r - \widehat{r}| + |s - \widehat{s}|.$$

Proof. (a) Put $q(t) = (\cos(\pi t), \sin(\pi t)) \in \mathbb{R}^2$ and $e_1 = (1, 0)$. The law of cosines gives $f_{r,s}(t) = \|re_1 - sq(t)\|_2$. Hence the reverse triangle inequality and $\|q'(t)\|_2 = \pi$ give

$$|f_{r,s}(t) - f_{r,s}(v)| \le s\|q(t) - q(v)\|_2 \le \pi s|t - v| \le \pi R|t - v|.$$

The same argument covers $r = 0$ or $s = 0$; in that case the left-hand side is in fact zero.

(b) Using the same $q(t)$ and e_1 as above, the reverse triangle inequality yields

$$|f_{r,s}(t) - f_{\widehat{r},\widehat{s}}(t)| \le \|(r - \widehat{r})e_1 - (s - \widehat{s})q(t)\|_2 \le |r - \widehat{r}| + |s - \widehat{s}|.$$

□

Proposition F.3 (Centered Euclidean code distortion)**.** *Let $D_{jk} = \|c_j - c_k\|_2$ and $\epsilon_h, \Delta_{b_r} \ge 0$. Suppose $h = (h_{jk})$ is symmetric, zero-diagonal, and takes values in $[0,1]$. Assume, only for pairs with both exact radii positive, that $|h_{jk} - \theta_{jk}| \le \epsilon_h$. Let the decoded radii be nonnegative and satisfy $\max_j |r_j - \widehat{r}_j| \le \Delta_{b_r}$. Define the decoded centered distance by*

$$\widehat{D}_{jk} = \sqrt{(\widehat{r}_j - \widehat{r}_k)^2 + 4\widehat{r}_j\widehat{r}_k \sin^2\left(\frac{\pi h_{jk}}{2}\right)}.$$

Then

$$\|D - \widehat{D}\|_\infty \le \pi R_o \epsilon_h + 2\Delta_{b_r}. \tag{37}$$

Consequently, on the same labelled weighted vertex set and with the same mass parameter, the exact and decoded max-DTM–Rips filtrations are $(\pi R_o\epsilon_h + 2\Delta_{b_r})$-interleaved and their diagrams have bottleneck distance at most this quantity in every homological degree.

Proof. For $r_j r_k > 0$, the Euclidean law of cosines gives $D_{jk} = f_{r_j,r_k}(\theta_{jk})$, while the decoder gives $\widehat{D}_{jk} = f_{\widehat{r}_j,\widehat{r}_k}(h_{jk})$. Therefore Lemma F.2(a) and Lemma F.2(b) imply

$$\begin{aligned}|D_{jk} - \widehat{D}_{jk}| &\le |f_{r_j,r_k}(\theta_{jk}) - f_{r_j,r_k}(h_{jk})| + |f_{r_j,r_k}(h_{jk}) - f_{\widehat{r}_j,\widehat{r}_k}(h_{jk})| \\ &\le \pi R_o|\theta_{jk} - h_{jk}| + |r_j - \widehat{r}_j| + |r_k - \widehat{r}_k| \le \pi R_o\epsilon_h + 2\Delta_{b_r}.\end{aligned}$$

If either exact radius is zero, the angular term vanishes: the exact distance is independent of the otherwise undefined direction, so only Lemma F.2(b) is needed. Taking the maximum over all pairs proves (37). The interleaving and diagram conclusions then follow from Theorem 2 with $\eta = \pi R_o\epsilon_h + 2\Delta_{b_r}$. □

For $b_r \in \mathbb{N}$ with $b_r \ge 1$, the nearest-grid quantizer with 2^{b_r} levels on $[0, R_o]$ has grid spacing $R_o/(2^{b_r} - 1)$. Thus, when $R_o > 0$, $\Delta_{b_r} \le \frac{R_o}{2(2^{b_r}-1)}$. When $R_o = 0$, every exact and decoded radius is zero and $\Delta_{b_r} = 0$; no division by the grid width is performed.

Corollary F.4 (Translation invariance of the barycenter policy)**.** *Suppose $a_j \ge 0$ and $\sum_j a_j = 1$, and choose $o_w = \sum_{j=1}^M a_j c_j$. Under a common translation $c'_j = c_j + v$, the barycenter-centered vectors, radii, sign codes generated with the same projection, quantized radial payload, and decoded dissimilarity are unchanged. The center itself is not part of the retained state needed to reconstruct within-object distances.*

Proof. The translated barycenter satisfies

$$o'_w = \sum_{j=1}^{M} a_j(c_j + v) = o_w + v,$$

and hence $c'_j - o'_w = c_j - o_w$ for every j. All stated quantities are deterministic functions of these centered vectors, the fixed projection, and the masses. □

The certificate-min policy is a different, frame-local rule. It constructs the origin and barycenter candidates with the same projection, computes each realized uniform distortion in the current coordinate frame, and retains the smaller-distortion candidate, choosing barycenter on an exact tie. It therefore obeys $\eta_{\text{cert}} \le \eta_{\text{origin}}$ in that frame, but it does not inherit the preceding invariance because the origin candidate changes under translation.

F.4 A certificate-first interface

A robust implementation can separate code generation from topology:

- build representatives $c_1, \dots, c_M$ and exact aggregated masses;
- generate candidate codes and a calibration map;
- while representatives are still available, compute or upper-bound $\eta_{\text{cert}} = \max_{j,k} |d(c_j, c_k) - \widehat{d}(j,k)|$;
- record η_{cert}, the metric name, β, the Rips convention, and the mass precision;
- discard representatives only after the certificate is accepted.

This interface records a uniform geometric certificate alongside retrieval-oriented diagnostics. The certificate itself needs one scalar at the chosen precision, in addition to the code state.

F.5 Code-only DTM computation

For each j, the discrete DTM weight can be computed without coordinates. The following conceptual procedure CodeDTM$(\{z_j, a_j\}_{j=1}^{M}, \beta, g)$ uses the finite-sum formula.

- For each j, compute $\widehat{d}(j,k) = g(b^{-1} d_{\text{H}}(z_j, z_k))$ for all k.
- Sort the pairs $(\widehat{d}(j,k), a_k)$ by the first coordinate.
- Sweep cumulative mass A from zero to β. If a distance r receives mass λ, add λr^2 to an accumulator.
- Return $w_j = \sqrt{\text{accumulator}/\beta}$.

Equal distances can be processed as a group. Because normalized Hamming distance has only $b+1$ levels, a histogram indexed by Hamming weight avoids comparison sorting when g is a function of Hamming distance alone:

$$H_{j,\ell} = \sum_{k:\, d_{\text{H}}(z_j, z_k) = \ell} a_k, \qquad \ell = 0, \dots, b.$$

Sweeping ℓ in increasing calibrated distance computes w_j. If g is nonmonotone, levels must instead be ordered by $g(\ell/b)$.

With word size w, one Hamming distance needs $\lceil b/w \rceil$ XOR–popcount word operations. All pairs therefore need $O(M^2 \lceil b/w \rceil)$ word operations. Comparison sorting every row costs $O(M^2 \log M)$; Hamming histograms replace this by the pairwise code cost plus $O(Mb)$ level sweeps. These counts exclude construction of simplices and persistence reduction.

F.6 Filtration construction and complexity

Once the M vertex weights and pair distances are available, every simplex has time

$$\rho(\sigma) = \max\left\{\max_{j\in\sigma} w_j,\ \max_{j,k\in\sigma} \widehat{d}(j,k)\right\}.$$

In a flag implementation, it is enough to add vertices at their weights and edges at $t_{jk} = \max\{w_j, w_k, \widehat{d}(j,k)\}$; higher simplices inherit the maximum of their edges and vertex times.

If only homology through degree k is requested, reduction generally needs the complex through dimension $k+1$. The worst-case number of nonempty simplices considered is $\sum_{r=1}^{k+2}\binom{M}{r}$. This remains polynomial of degree $k+2$ for fixed k, but is exponential when the full complex is formed. Neither $b+1$ Hamming levels nor a small number of distinct entry times changes this simplex count.

F.7 Exact bit accounting

The phrase "Mb-bit sketch" counts only codes. A measure sketch must also store masses and enough metadata to interpret the codes.

For an unweighted raw sample of size n, let n_j be the number of points assigned to cell j. Storing every count independently costs $M\lceil\log_2(n+1)\rceil$ bits. Since $\sum_j n_j = n$, enumerative coding can reduce this to $\left\lceil\log_2\binom{n+M-1}{M-1}\right\rceil$ bits, plus decoding overhead. Either way, exact empirical masses carry a logarithmic dependence on n.

If each mass uses b_a bits and the decoder renormalizes, mass storage is Mb_a bits and independent of n. The sharp theorem then applies to the quantized measure itself; comparing with the exact empirical masses needs an additional mass-perturbation theorem and is not free.

In the Euclidean experiments, $b_a = 64$ and the physical retained state is

$$S_{\mathrm{E}}(M, b, b_r, b_a) = \left\lceil\frac{Mb}{8}\right\rceil + \left\lceil\frac{Mb_r}{8}\right\rceil + \frac{Mb_a}{8} + 8 \quad \text{bytes.} \tag{38}$$

The four terms are the packed sign codes, packed radial indices, float64 masses, and one float64 quantizer maximum. Neither the chosen center nor the rejected candidate is retained. A distortion scalar is not needed to decode the filtration; an optional retained certificate adds another 8 bytes. For a variable-size object, M in (38) is replaced by its realized representative count.

A complete budget also includes:

- calibration parameters for g;
- the DTM mass β;
- representative labels or a canonical code ordering;
- the distortion certificate and its precision;
- hash seed or projection seed if it is not public;
- the maximum requested homology degree and Rips scale convention.

With public randomness and fixed global hyperparameters, the dominant retained angular code state is $Mb + Mb_a$ bits at fixed mass precision. The Euclidean code–radius state instead uses the exact accounting in (38). With exact empirical counts, the angular code state is instead $Mb + M\lceil\log_2(n+1)\rceil$ under the simple representation.

F.8 Construction time and streaming semantics

At fixed mass precision the downstream bit state may be independent of raw n; exact empirical counts retain the logarithmic dependence displayed above. In either case constructing the state is not independent of n. At a minimum, every input must be read, so time is $\Omega(n)$. A streaming implementation maintains:

- a fixed or adaptively managed set of at most M cells;
- a count or weight accumulator per cell;
- a representative update rule;
- a code for the current representative.

If changing a representative changes its code, the uniform distortion certificate must refer to the *final* representatives. If representatives are discarded during the stream, a prospective distributional code guarantee must replace posterior verification. Theorem 4 applies once the final partition radius and final uniform code distortion are certified; it does not prescribe a particular online clustering algorithm.

F.9 A useful design tradeoff

For angular random-hyperplane codes, suppressing constants gives $\eta_{\mathrm{code}} \asymp \sqrt{\log(M/\delta)/b}$. The end-to-end theorem suggests the deterministic-plus-probabilistic budget

$$d_{\mathrm{B}} \lesssim 2r_{\mathrm{part}}(M) + \sqrt{\frac{\log(M/\delta)}{b}}.$$

Increasing M typically reduces the partition radius but increases code storage Mb, distance-table work, and the logarithmic factor in the uniform event. A model for $r_{\mathrm{part}}(M)$ or a measured quantization curve turns this expression into a data-dependent rule for selecting M under the available storage and computation budget.

G Connections to ambient DTM theory and scope of the guarantees

This appendix places the finite code-only construction in the context of ambient DTM sublevel sets, power-distance filtrations, and weighted Čech/Rips complexes. Each construction associates persistence with a specified geometric object; we describe the corresponding assumptions and comparisons with the max-DTM–Rips filtration in Definition 1.

G.1 The classical ambient DTM certificate

Let $(\mathcal{Z}, d)$ be a suitable ambient metric space and μ a Borel probability measure with finite second moment. The order-two DTM is

$$d_{\mu,\beta}(z) = \left(\frac{1}{\beta}\int_0^\beta \delta_{\mu,\ell}(z)^2\, d\ell\right)^{1/2},$$

where $\delta_{\mu,\ell}(z)$ is the smallest radius of a ball about z carrying mass greater than ℓ. For measures μ, ν,

$$\|d_{\mu,\beta} - d_{\nu,\beta}\|_\infty \le \frac{1}{\sqrt{\beta}} W_2(\mu,\nu) \tag{39}$$

(Chazal et al., 2011; Buchet et al., 2016; Chazal et al., 2018). Under the standard tameness/triangulability conditions, function stability gives

$$d_{\mathrm{B}}(\mathrm{Dgm}(d_{\mu,\beta}), \mathrm{Dgm}(d_{\nu,\beta})) \le \frac{1}{\sqrt{\beta}} W_2(\mu,\nu). \tag{40}$$

For a partition of a finite empirical measure $\mu = \sum_{x\in V} u_x \delta_x$ with representatives $c_{\pi(x)}$, define $\nu = \sum_{j=1}^M a_j \delta_{c_j}$. The coupling that transports the mass u_x from x to $c_{\pi(x)}$ has squared cost at most r_{part}^2. Hence

$$W_2(\mu,\nu) \le r_{\mathrm{part}} \quad\Longrightarrow\quad d_{\mathrm{B}}(\mathrm{Dgm}(d_{\mu,\beta}), \mathrm{Dgm}(d_{\nu,\beta})) \le \frac{r_{\mathrm{part}}}{\sqrt{\beta}}. \tag{41}$$

The estimate (41) provides a classical *companion certificate* for distributional topology, obtained by applying established DTM stability to the partition coupling.

The bounds $2r_{\mathrm{part}}$ and $r_{\mathrm{part}}/\sqrt{\beta}$ concern complementary objects and comparison mechanisms:

- $2r_{\text{part}}$ compares two finite max-DTM–Rips filtrations through uniform pairwise metric distortion and a fibre quotient;
- $r_{\text{part}}/\sqrt{\beta}$ compares ambient DTM functions after changing the measure through a transport coupling.

The applicable bound depends on the target filtration. In particular, Theorem 4 retains no $\beta^{-1/2}$ factor because the pullback comparison holds on the same labelled atomic measure before the exact quotient.

G.2 Power distance and weighted Rips bridges

For a set P with DTM weights, the DTM power distance is commonly written

$$d^P_{\mu,\beta}(z) = \min_{p\in P} \sqrt{d_{\mu,\beta}(p)^2 + d(p,z)^2}.$$

Known results compare this function with ambient DTM and realize its sublevel sets by weighted Čech complexes. Weighted Rips complexes provide a flag approximation (Buchet et al., 2016). Under the hypotheses stated in that work, the log-bottleneck comparison between ambient DTM persistence and a weighted Rips construction can be bounded by $\log(2\sqrt{5})$ in a general good-cover setting; Euclidean special cases admit sharper power-distance constants.

These multiplicative/log-scale comparisons can be related to the additive quantity $2r_{\text{part}} + \eta$ through an explicit choice of parameterization:

- state the ambient and finite-code certificates separately;
- put every positive-scale object on a common log parameter and compose multiplicative interleavings;
- restrict scales to $[\tau, T]$ with $\tau > 0$, then convert a multiplicative factor into an explicit additive error on that interval.

Definition 1 uses a maximum of vertex weights and raw edge distances. It was selected because it gives an exact one-Lipschitz simplex-time theorem under metric perturbation. The power-weighted convention in Appendix E instead uses radii $\sqrt{t^2 - w(x)^2}$. Both are robust filtrations, with constants and comparison maps determined by their respective simplex-time definitions.

G.3 DTM-based filtrations in prior work

Anai et al. (2019) study a family in which a vertex with weight $f(x)$ has a scale-dependent radius $r_x(t) = (t^p - f(x)^p)^{1/p}$. Their exponent p parametrizes the weighted filtration; it should not be confused with the order of an L^p-DTM. Their results include stability under perturbations of f, Hausdorff perturbations under a Lipschitz weight, and robust comparisons involving DTM weights.

Our theorem has a different premise and conclusion:

- the atomic weights defining DTM remain fixed;
- the entire pairwise metric is perturbed on the same labelled set;
- the induced DTM vertex weights are recomputed from the perturbed metric;
- the filtration uses the maximum form (3);
- the result is an exact additive all-scale interleaving with constant one.

The same-vertex Rips-only component could also be deduced from Gromov–Hausdorff stability (Chazal et al., 2014). The new work needed here is to carry the metric perturbation through finite weighted radial quantiles, combine it with vertex births, and then prove that aggregated fibres are persistence-equivalent to their representatives.

G.4 The role of binary embedding theory

Binary embedding theory supplies conditions under which $\max_{j,k} |d_C(j,k) - \widehat{d}(j,k)| \le \eta$ holds. Our deterministic theorem then transfers this metric guarantee to an all-scale DTM filtration guarantee. The two results play complementary roles: embedding theory sets the coding budget, while the filtration theorem identifies the persistence accuracy supported by that budget.

For Gaussian signs on the unit sphere, normalized Hamming distance estimates normalized angular distance, with finite-set bit count $O(\eta^{-2}\log(M/\delta))$ (Yi et al., 2015; Oymak & Recht, 2015). A statement about Euclidean distance must either use the chordal calibration on normalized points or store additional norm information. A learned hash can be used if it provides or is checked against the same uniform event; retrieval recall or average pair loss alone is not that event.

G.5 Parameter regime of the lower bound

Theorem 6 is matched to the max-DTM–Rips output and proves $B = \Omega(\min\{n, \beta^{-1}, \epsilon^{-1}\})$ in a concrete parameter regime. Appendix E removes the β^{-1} limitation for a standard power-weighted DTM–Rips convention. Both arguments pack actual H_0 diagrams and allow an arbitrary decoder.

The result is a worst-case uniform lower bound over finite metric-measure spaces of diameter at most one. Distributional assumptions, bounded doubling dimension, margins between critical values, or average-case error can permit smaller sketches. Recent sparsification and hardness results similarly show that geometric assumptions materially affect attainable size (Alonso, 2025; Kolbe & Mayr, 2026; McCabe, 2026).

G.6 Geometric and statistical conditions

A b-bit Hamming cube has $b+1$ distance values, but an M-vertex Rips complex can still contain $2^M - 1$ simplices. Persistent homology algorithms reduce boundary matrices whose size depends on simplices, not merely on the number of distinct filtration timestamps. The correct conditional statement is:

> For fixed representative budget M, code length b, and requested homology degree, downstream work is independent of the discarded raw sample size n.

Here independence from n is conditional on the stated geometric and homological budgets.

The DTM vertex weight depends on enough nearest neighbours to accumulate mass β, while the Rips complex at large scales depends on arbitrary pairs. This motivates the uniform pairwise premise, which controls both radial quantiles and edges. A nearest-neighbour filter instead defines its own sublevel-set persistence and provides a distinct target for approximation.

Suppose every pair individually satisfies $\Pr[|d_{jk} - \widehat{d}_{jk}| \le \eta] \ge 0.99$. With $\Theta(M^2)$ pairs, the probability that all pairs are simultaneously good can be small. Theorem 2 is applied on one uniform event. Concentration, a union bound, a covering argument, or a posterior certificate must supply it.

Random hyperplane hashing satisfies an exact expectation identity, but a particular finite code table fluctuates. Replacing an expectation by the observed Hamming distances requires a concentration statement such as (36).

Theorem 2 keeps equal codes as distinct labels. When $g(0) = 0$, their code dissimilarity is a pseudometric and equal codes lie at zero distance. An implementation that removes duplicates changes the atomic measure and abstract complex. It is valid only after masses are aggregated and a quotient argument, such as Proposition C.3(b), is supplied. When $g(0) > 0$, equal-code labels are not at zero mutual dissimilarity and this exact quotient does not apply.

The constant-one theorem keeps atomic masses fixed. For approximate masses, quantile behavior also depends on the distance between cumulative shell masses and β: near that boundary, a mass perturbation can move the quantile to the next shell. This dependence can be handled in several ways:

- store exact empirical counts;

- assume a cumulative-mass margin;
- compare measures by Wasserstein distance and accept the $\beta^{-1/2}$ factor;
- define the quantized masses as the target distribution rather than claiming zero loss relative to exact masses.

For $c > 0$, $z(cx) = z(x)$ under homogeneous sign hashing. Sign codes therefore describe angular geometry; normalized chordal distance follows by calibration, while unnormalized Euclidean distance additionally uses norm information.

Multiplicative interleavings preserve the common zero complex but $\log 0$ is undefined. Appendix B.3 gives truncation, relative persistence, and extended-parameter solutions.

The statistical limit theory in Chazal et al. (2018) concerns fixed-point or functional DTM estimators under explicit quantile regularity assumptions. Transferring such a limit to persistence diagrams involves an additional regularity and mapping argument, alongside the deterministic stability established here.

An additive error has units of distance, whereas a log-bottleneck error is dimensionless. Expressing both errors in a common parameterization provides a consistent basis for composing the bounds.

Before applying Theorem 2 or 4, the following hypotheses must be verified:

- The metric-perturbation step compares the same labelled vertex set.
- The atomic masses agree exactly and sum to one.
- The common DTM mass parameter is strictly positive.
- The reconstructed dissimilarity is symmetric, nonnegative, and has zero diagonal.
- The uniform error and the target metric use the same physical units.
- The filtration is exactly (3), with edge threshold $d \le r$.
- Duplicate codes remain distinct, unless their masses are explicitly aggregated through the quotient construction.
- The persistence modules are q-tame; this is automatic for the finite filtrations considered here.
- A randomized code is controlled by one simultaneous all-pairs event.
- A bit-complexity statement counts masses, calibration data, and nonpublic random seeds.

These conditions specify the scope of the stated bounds; extensions can be formulated by supplying the corresponding comparison argument for the altered assumption.

H Worked finite examples and a posteriori certification

This appendix instantiates the definitions and bounds on small finite spaces. The calculations illustrate nonuniform DTM weights, the fibre quotient, the diagram packing, and the deterministic certificate $2r_{\mathrm{part}} + \eta_{\mathrm{cert}}$.

H.1 A complete nonuniform DTM calculation

Consider three labelled vertices $V = \{1, 2, 3\}$ with masses $a_1 = 1/2$, $a_2 = 1/3$, and $a_3 = 1/6$, and distances $d(1, 2) = 1$, $d(1, 3) = 2$, and $d(2, 3) = 2$. Take $\beta = 3/4$.

At vertex 1, the ordered distance-mass shells are $(0, 1/2)$, $(1, 1/3)$, and $(2, 1/6)$. The first $1/2$ of the β-integral contributes zero and the next $1/4$ contributes squared radius one.

At vertex 2, the shells are $(0, 1/3)$, $(1, 1/2)$, and $(2, 1/6)$. After the self mass, the remaining $3/4 - 1/3 = 5/12$ units needed to reach β are available at radius one.

At vertex 3, the shells are $(0, 1/6)$ and $(2, 5/6)$. The three weight calculations are therefore

$$w(1)^2 = \frac{(1/4)\cdot 1^2}{3/4} = \frac{1}{3}, \qquad w(1) = \frac{1}{\sqrt{3}},$$
$$w(2)^2 = \frac{(5/12)\cdot 1^2}{3/4} = \frac{5}{9}, \qquad w(2) = \frac{\sqrt{5}}{3},$$
$$w(3)^2 = \frac{(3/4 - 1/6)\cdot 2^2}{3/4} = \frac{(7/12)\cdot 4}{3/4} = \frac{28}{9}, \qquad w(3) = \frac{2\sqrt{7}}{3}.$$

The simplex times are consequently

$$\rho(\{1\}) = \frac{1}{\sqrt{3}}, \quad \rho(\{2\}) = \frac{\sqrt{5}}{3}, \quad \rho(\{3\}) = \frac{2\sqrt{7}}{3},$$
$$\rho(\{1,2\}) = \max\left\{\frac{1}{\sqrt{3}}, \frac{\sqrt{5}}{3}, 1\right\} = 1,$$
$$\rho(\{1,3\}) = \rho(\{2,3\}) = \rho(\{1,2,3\}) = 2,$$

because $2\sqrt{7}/3 < 2$. This example illustrates that DTM births need not coincide with edge-distance levels and can substantially delay a vertex even when its incident edges enter later.

H.2 Checking the sharp additive constant on the example

Perturb only $d(1,2)$ from 1 to $1+\eta$, with $0 < \eta < 1$, and leave the other distances at two. At vertices 1 and 2, every positive quantile used above changes by exactly η, but only over part of $(0, \beta]$. For vertex 1, $\widehat{w}(1) = \sqrt{1/3}(1+\eta)$, so the changes at the two vertices are

$$|\widehat{w}(1) - w(1)| = \frac{\eta}{\sqrt{3}} < \eta, \qquad |\widehat{w}(2) - w(2)| = \frac{\sqrt{5}}{3}\eta < \eta.$$

The edge distance itself changes by η, so the simplex bound of Theorem 2 is attained at the edge-coordinate level even when the DTM norm bound is strict. This confirms why the global theorem has constant one and why a smaller universal constant is impossible for the chosen filtration.

H.3 A two-cell quotient example

Let four unit-mass-normalized vertices be partitioned as $\pi^{-1}(1) = \{x_1, x_2, x_3\}$ and $\pi^{-1}(2) = \{y\}$, with masses u_1, u_2, u_3, u_y. Choose $c_1 = x_1, c_2 = y$, and let $D = d(x_1, y)$. The pullback pseudometric has $\bar{d}(x_r, x_s) = 0$ and $\bar{d}(x_r, y) = D$. Every x_r has radial mass CDF

$$F_{x_r,\bar{d}}(t) = \begin{cases} 0, & t < 0, \\ u_1 + u_2 + u_3, & 0 \le t < D, \\ 1, & t \ge D. \end{cases}$$

The representative vertex 1, with aggregated mass $a_1 = u_1 + u_2 + u_3$, has exactly the same CDF. All three fibre vertices have the same DTM weight.

At any scale, a simplex containing an arbitrary subset of $\{x_1, x_2, x_3\}$ has the same filtration value as the simplex obtained by replacing that nonempty subset by x_1. Geometrically, the fibre spans a simplex whenever it is present and can be contracted to the section vertex. This is the concrete four-vertex instance of the contiguity proof in Proposition C.3(a).

H.4 A complete two-pair lower-bound table

Set $q = 2$, choose $\epsilon \le 1/32$, and $\beta \le 1/4$. The four target diagrams indexed by $\theta \in \{0,1\}^2$ have internal deaths:

Each diagram also contains one common finite bar dying at one and one common essential bar. Any two rows differ in at least one slot by 4ϵ. Cross-slot differences are also at least 4ϵ, the smallest

θ	slot 1 death	slot 2 death
00	8ϵ	16ϵ
01	8ϵ	20ϵ
10	12ϵ	16ϵ
11	12ϵ	20ϵ

persistence is 8ϵ, and every internal death is at most $20\epsilon \le 5/8$. Thus matching to a wrong slot, to the diagonal, or to the common death cannot cost less than 4ϵ. Four distinguishable targets require two bits even if the decoder knows the construction.

H.5 An a posteriori distortion certificate

Proposition H.1 (Certificate computed from a realized code table)**.** *Let (V, d, u), π, $c_1, \dots, c_M$, and the exact aggregated masses $a_1, \dots, a_M$ satisfy the assumptions of Theorem 4. For a realized code table, define*

$$r_{\mathrm{cert}} = \max_{x \in V} d(x, c_{\pi(x)}), \qquad \eta_{\mathrm{cert}} = \max_{j,k \in [M]} |d(c_j, c_k) - \widehat{d}(j,k)|.$$

Then, for every homological degree q,

$$d_{\mathrm{B}}\Big(\mathrm{Dgm}_q\, \mathcal{K}_{d,u,\beta}, \mathrm{Dgm}_q\, \mathcal{K}_{\widehat{d},a,\beta}\Big) \le 2r_{\mathrm{cert}} + \eta_{\mathrm{cert}}.$$

The certificate is deterministic conditional on the realized code table and does not require computing either persistence diagram.

Proof. By definition, r_{cert} is a valid partition-radius bound and η_{cert} is a uniform representative-distance distortion bound. The claim is Theorem 4 with these two computed quantities. □

If masses are rounded before the filtration is constructed, the proposition certifies the filtration of the rounded measure. Comparison with the exact empirical measure additionally requires the mass-perturbation conditions of Appendix K.

I Extensions of the deterministic architecture

The main theorem uses an order-two DTM weight and a maximum edge rule. Its proof exposes a more general principle: if a radial vertex functional is Lipschitz under uniform quantile perturbations and simplex times are Lipschitz in vertex and edge entries, the same all-scale interleaving follows. We record several rigorous extensions here.

I.1 Finite-order discrete DTM

For $1 \le p < \infty$, define

$$w^{(p)}_{d,a,\beta}(x) = \left(\frac{1}{\beta}\int_0^\beta Q_{x,d}(u)^p\, du\right)^{1/p}. \tag{42}$$

For $p = \infty$, define $w^{(\infty)}_{d,a,\beta}(x) = \operatorname{ess\,sup}_{0<u\le\beta} Q_{x,d}(u)$. On a finite atomic space, the latter is the largest radial distance needed before cumulative mass reaches β, modulo the immaterial endpoint convention.

Define the max-L^p-DTM–Rips simplex time

$$\rho^{(p)}_{d,a,\beta}(\sigma) = \max\left\{\max_{x\in\sigma} w^{(p)}_{d,a,\beta}(x), \max_{x,y\in\sigma} d(x,y)\right\}.$$

Theorem I.1 (Stability and aggregation for all L^p-DTM orders)**.** *(a) Fix $p \in [1,\infty]$. Under the same-labelled, same-mass assumptions of Theorem 2, if $\|d - \widehat{d}\|_\infty \le \eta$, then the max-$L^p$-DTM–Rips filtrations are additively η-interleaved. Every homological degree has bottleneck distance at most η. If instead $(1-\epsilon)d \le \widehat{d} \le (1+\epsilon)d$, the multiplicative factor is $(1-\epsilon)^{-1}$.*

(b) For every $p \in [1, \infty]$, the partition construction of Theorem 4 yields

$$d_{\mathrm{B}}\left(\mathrm{Dgm}_q \mathcal{K}^{(p)}_{d,u,\beta}, \mathrm{Dgm}_q \mathcal{K}^{(p)}_{\widehat{d},a,\beta}\right) \le 2r_{\mathrm{part}} + \eta.$$

Proof. (a) Lemma B.1(b) gives $\|Q_{x,d} - Q_{x,\widehat{d}}\|_{L^\infty(0,\beta)} \le \eta$. For $1 \le p < \infty$, the reverse triangle inequality in $L^p(0,\beta)$ yields

$$|w^{(p)}_{d,a,\beta}(x) - w^{(p)}_{\widehat{d},a,\beta}(x)| \le \beta^{-1/p}\|Q_{x,d} - Q_{x,\widehat{d}}\|_{L^p(0,\beta)} \le \beta^{-1/p}(\beta\eta^p)^{1/p} = \eta.$$

For $p = \infty$, the elementary inequality $|\operatorname{ess\,sup} f - \operatorname{ess\,sup} g| \le \|f - g\|_\infty$ gives the same result. The simplex maximum, identity inclusions, q-tameness, and diagram step are identical to Lemma B.1(d) and Theorem 2.

Under a multiplicative metric sandwich, quantiles obey (9). Every L^p norm, including the essential supremum, is monotone and positively homogeneous. Thus the weights and simplex times obey the same multiplicative sandwich, and the proof of Theorem B.2 applies.

(b) Lemma C.2(a) identifies the entire radial CDF, hence the entire quantile function, of a pullback fibre vertex with that of its aggregated representative. Therefore every L^p radial norm agrees exactly. Lemma C.2(b), the contiguity proof, and the error composition are unchanged. Use Theorem I.1(a) for the two metric perturbations. □

Remark (Two different appearances of p). The p in (42) is the order of a radial norm. The p in a power-weighted radius $(t^p - w(x)^p)^{1/p}$ controls how vertex weights and geometric radii combine. They may be chosen independently and should carry different symbols in an implementation.

I.2 General radial vertex functionals

Let Q_β be the set of nonnegative, nondecreasing measurable functions on $(0, \beta]$. Consider a functional $\mathfrak{W} : Q_\beta \to \mathbb{R}_+$.

Assumption I.2 (Uniform radial Lipschitz property)**.** There is $L_{\mathfrak{W}} < \infty$ such that

$$|\mathfrak{W}(f) - \mathfrak{W}(g)| \le L_{\mathfrak{W}}\|f - g\|_{L^\infty(0,\beta)} \quad \text{for all } f, g \in Q_\beta.$$

Define the vertex weight $w^{\mathfrak{W}}_{d,a,\beta}(x) = \mathfrak{W}(Q_{x,d}|_{(0,\beta]})$. Examples include:

- normalized L^p norms, with $L_{\mathfrak{W}} = 1$;
- the essential maximum radial quantile, with $L_{\mathfrak{W}} = 1$;
- a weighted radial average $\int_0^\beta h(u)Q(u)\,du$, when $h \ge 0$ and $\int_0^\beta h = L_{\mathfrak{W}}$;
- a finite maximum of Lipschitz radial summaries, with constant equal to the largest individual constant.

Theorem I.3 (Radial-functional stability and fibre quotient)**.** *(a) Under $\|d - \widehat{d}\|_\infty \le \eta$ and exact common atomic masses, define*

$$\rho^{\mathfrak{W}}_d(\sigma) = \max\left\{\max_{x\in\sigma} w^{\mathfrak{W}}_{d,a,\beta}(x), \max_{x,y\in\sigma} d(x,y)\right\}.$$

Then the d and $\widehat{d}$ filtrations are $\max\{L_{\mathfrak{W}}, 1\}\eta$-interleaved. The same bound holds for bottleneck distance in every degree.

(b) If the vertex functional depends only on the radial quantile function, then the fibre quotient in Proposition C.3(b) is exact. Under Assumption I.2, the end-to-end bound becomes $2\max\{L_{\mathfrak{W}}, 1\}r_{\mathrm{part}} + \max\{L_{\mathfrak{W}}, 1\}\eta$.

Proof. (a) Lemma B.1(b) and Assumption I.2 give $|w^{\mathfrak{W}}_{d,a,\beta}(x) - w^{\mathfrak{W}}_{\widehat{d},a,\beta}(x)| \le L_{\mathfrak{W}}\eta$. Pairwise distance entries change by at most η. Applying Lemma A.2(b) to the combined vector of vertex and pair entries gives

$$|\rho^{\mathfrak{W}}_d(\sigma) - \rho^{\mathfrak{W}}_{\widehat{d}}(\sigma)| \le \max\{L_{\mathfrak{W}}, 1\}\eta.$$

Identity inclusions and Proposition A.3(b) complete the proof.

(b) The exact quotient follows because Lemma C.2(a) identifies quantile functions. The raw-to-pullback metric error is $2r_{\text{part}}$ and the decoded-to-code error is η; apply Theorem I.3(a) twice and compose. □

I.3 Externally supplied vertex weights

Sometimes DTM weights are computed once from decoded geometry and stored separately, while only edge distances are reconstructed from codes. The following elementary result clarifies the resulting error budget.

Proposition I.4 (Independent vertex and edge errors)**.** *Let $v, \widehat{v} : V \to \mathbb{R}_+$ be arbitrary vertex-time functions and $d, \widehat{d}$ arbitrary symmetric zero-diagonal dissimilarities. Define*

$$\rho_{v,d}(\sigma) = \max\left\{\max_{x\in\sigma} v(x),\ \max_{x,y\in\sigma} d(x,y)\right\}.$$

If $\|v - \widehat{v}\|_\infty \le \xi$ and $\|d - \widehat{d}\|_\infty \le \eta$, then the corresponding filtrations are $\max\{\xi,\eta\}$-interleaved.

Proof. Every vertex coordinate changes by at most ξ, every edge coordinate by at most η, and the simplex time is their maximum. Lemma A.2(b) gives a uniform simplex error $\max\{\xi,\eta\}$, after which the standard identity-inclusion proof applies. □

Storing decoded DTM weights may improve computation, but it consumes additional memory and is no longer "codes plus masses only." If the stored weights are quantized, their precision contributes ξ.

I.4 Lipschitz transformations of edge distance

Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be L_ϕ-Lipschitz and define

$$\rho^{\phi}_{d,a,\beta}(\sigma) = \max\left\{\max_{x\in\sigma} w_{d,a,\beta}(x),\ \max_{x,y\in\sigma} \phi(d(x,y))\right\}.$$

Proposition I.5 (Transformed edge times)**.** *Under the premises of Theorem 2, these two transformed-edge filtrations are $\max\{1, L_\phi\}\eta$-interleaved.*

Proof. The DTM vertex weights change by at most η. The mean-value/Lipschitz bound gives $|\phi(d(x,y)) - \phi(\widehat{d}(x,y))| \le L_\phi\eta$. Apply Lemma A.2(b) and the identity-inclusion proof. □

For $\phi(r) = r^2$, no global finite L_ϕ exists on $[0,\infty)$, but on $[0,D]$ one may take $L_\phi = 2D$. This recovers the diameter-dependent warning in Appendix B.

I.5 Fixed sparse flag complexes

Let $G = (V, E)$ be a fixed admissibility graph. Define a sparse flag filtration by admitting only edges in E, and only cliques of G:

$$\rho^{G}_{d,a,\beta}(\sigma) = \max\left\{\max_{x\in\sigma} w_{d,a,\beta}(x),\ \max_{\{x,y\}\subseteq\sigma} d(x,y)\right\}, \quad \sigma \text{ a clique of } G.$$

Proposition I.6 (Fixed-graph sparsification)**.** *If the same graph G is used for d and $\widehat{d}$, then every additive and multiplicative theorem in Appendices B and B.3 remains valid with the same constants.*

Proof. The vertex set and family of admissible simplices are identical on both sides. For each admissible simplex, the same DTM and maximum-entry arguments apply. Identity inclusions never introduce a simplex outside the fixed clique family. □

Remark (Data-dependent sparse graphs). If G is recomputed independently from d and $\widehat{d}$, the admissible simplex families may differ. A small metric perturbation can change a nearest-neighbour graph discontinuously when distances tie. The proposition then requires a graph-stability margin or a separate interleaving map. Merely calling both constructions "sparse Rips" is insufficient.

I.6 Nonmetric reconstructed dissimilarities

Theorem 2 does not use the triangle inequality for $\widehat{d}$; a symmetric, nonnegative, zero-diagonal dissimilarity is enough to define radial CDFs and a flag filtration. The decoded side d is a metric because the partition distortion proof uses its triangle inequality. This asymmetry is useful for learned calibrations that do not produce an exact metric.

If $\widehat{d}$ is asymmetric, two choices are possible:

- symmetrize, for example $\widehat{d}_{\max}(x, y) = \max\{\widehat{d}(x, y), \widehat{d}(y, x)\}$, and certify its uniform error;
- define a directed topology and prove a different theorem.

The undirected Rips theorem cannot be applied to an asymmetric matrix without specifying such a choice.

I.7 A reusable theorem template

The deterministic proof can be reused for a new code-only filtration by checking four items:

- Uniform pairwise distortion implies a uniform bound on the chosen radial summary.
- The simplex entry function is Lipschitz in its vertex and pair entries.
- The same labelled vertices allow identity inclusions, or explicit maps are supplied.
- The resulting persistence modules admit diagrams and the isometry/stability theorem.

The max-DTM–Rips case satisfies these with constants $1, 1, 0$, and finite q-tameness, respectively.

J Information-theoretic variants and lower-bound stress tests

This appendix develops consequences of the diagram packing argument under several sketch models. It also gives an independent minimum-spanning-tree derivation of the matched H_0 family.

J.1 A formal sketch model

A possibly randomized topological sketch consists of:

- an encoder E that receives $X \in \mathfrak{X}$ and produces a retained state S;
- a decoder D that receives S and produces a diagram or collection of diagrams;
- a master random string R, independent of X, containing all encoder and decoder randomness.

For a worst-case B-bit sketch, the state alphabet has cardinality at most 2^B for every realization of R. Public randomness need not be stored, but it is fixed in the counting proof. Private randomness that must persist between encoding and decoding is part of S.

Proposition J.1 (Average-prior success is enough)**.** *Let X be uniform on a strict 2ϵ-packing of size L. If*

$$\Pr_{X,R}[d_{\mathrm{B}}(D(E(X; R); R), \Phi(X)) \le \epsilon] \ge 1 - \delta,$$

then $B \ge \log_2((1 - \delta)L)$.

Proof. Average over R. There exists r for which the deterministic scheme obtained by fixing $R = r$ succeeds on at least $(1 - \delta)L$ packing elements. Two successful elements cannot share a state, by the bottleneck triangle inequality. Hence $2^B \ge (1 - \delta)L$. □

The proposition assumes only average success under one explicit hard prior, which is weaker than the per-input guarantee in Theorem D.1.

J.2 A complete Fano derivation

Let Θ be uniform on $[L]$, let X_Θ be a strict 2ϵ-packing element, and let $\widehat{D}$ be the decoded diagram. Define an estimator $\widehat{\Theta}(\widehat{D}) \in \arg\min_{j\in[L]} d_{\mathrm{B}}(\widehat{D}, \Phi(X_j))$. Whenever $d_{\mathrm{B}}(\widehat{D}, \Phi(X_\Theta)) \le \epsilon$, strict separation makes Θ the unique packing point in the closed ϵ-ball, so $\widehat{\Theta} = \Theta$.

Let $P_e = \Pr[\widehat{\Theta} \neq \Theta] \le \delta$. Fano's inequality gives $H(\Theta \mid \widehat{\Theta}) \le h_2(P_e) + P_e \log_2(L-1)$. Therefore

$$\begin{aligned} I(\Theta;\widehat{\Theta}) &= H(\Theta) - H(\Theta \mid \widehat{\Theta}) \\ &\ge \log_2 L - h_2(P_e) - P_e \log_2(L-1) \ge (1-\delta)\log_2 L - 1, \end{aligned}$$

where the last line uses $h_2(P_e) \le 1$, $\log_2(L-1) \le \log_2 L$, and $P_e \le \delta$. The Markov chain $\Theta \longrightarrow (S,R) \longrightarrow \widehat{D} \longrightarrow \widehat{\Theta}$ and independence of Θ, R imply

$$I(\Theta;\widehat{\Theta}) \le I(\Theta;S,R) = I(\Theta;S \mid R) \le H(S \mid R) \le B.$$

Thus

$$B \ge (1-\delta)\log_2 L - 1. \tag{43}$$

This form extends conveniently to nonuniform priors and mutual-information budgets. For the uniform finite-state model, the fixed-seed counting result is usually stronger.

J.3 Expected bottleneck error

Proposition J.2 (Expected-error lower bound)**.** *Let X be uniform on a strict 4ϵ-packing of size L. If a randomized B-bit sketch satisfies $\mathbb{E}_{X,R} d_{\mathrm{B}}(\widehat{D}, \Phi(X)) \le \epsilon$, then $B \ge \log_2(L/2)$.*

Proof. Markov's inequality gives

$$\Pr[d_{\mathrm{B}}(\widehat{D}, \Phi(X)) \le 2\epsilon] \ge \frac{1}{2}.$$

Because the target packing is strict 4ϵ, its closed 2ϵ-balls are disjoint. Apply Proposition J.1 with accuracy 2ϵ and failure probability $1/2$. □

For a fully explicit application, take the paired family of Appendix D with its construction parameter set to $\epsilon_0 = 2\epsilon$, where $0 < \epsilon \le 1/32$, and put

$$q = \left\lfloor \min\left\{ \frac{n}{2}, \frac{1}{2\beta}, \frac{1}{32\epsilon} \right\} \right\rfloor.$$

Its pairwise diagram separation is at least $4\epsilon_0 = 8\epsilon > 4\epsilon$, so Proposition J.2 with $L = 2^q$ gives $B \ge q - 1$. Thus expected error has the same $\Omega(\min\{n, \beta^{-1}, \epsilon^{-1}\})$ order whenever the displayed terms are nonvacuous up to constants.

J.4 Variable-length states

Suppose a deterministic encoder uses a binary prefix-free string whose length $\ell(X)$ may depend on the input. On a packing of size L, zero-error decoding forces distinct codewords. Kraft's inequality gives $\sum_{j=1}^{L} 2^{-\ell(X_j)} \le 1$. By Jensen's inequality,

$$2^{-\frac{1}{L}\sum_j \ell(X_j)} \le \frac{1}{L}\sum_j 2^{-\ell(X_j)} \le \frac{1}{L}.$$

Therefore the average retained length under the uniform hard prior obeys $\frac{1}{L}\sum_{j=1}^{L} \ell(X_j) \ge \log_2 L$. Thus variable-length prefix coding cannot evade the lower bound on average. Without a prefix-free or self-delimiting requirement, the decoder must be told the length, and that metadata must be counted.

J.5 Covering and packing: an output rate-distortion view

Let $\mathcal{T} = \Phi(\mathfrak{X})$ be the family of target diagrams. Denote by $\mathcal{N}(\epsilon, \mathcal{T}, d_{\mathrm{B}})$ the smallest size of an ϵ-cover and by $\mathcal{M}(2\epsilon, \mathcal{T}, d_{\mathrm{B}})$ the largest strict 2ϵ-packing. Any deterministic B-bit diagram sketch must satisfy

$$B \geq \log_2 \mathcal{M}(2\epsilon, \mathcal{T}, d_{\mathrm{B}}).$$

Conversely, if the encoder is allowed to compute $\Phi(X)$ exactly and an ϵ-cover is available to both parties, it can store the nearest cover index using $\lceil \log_2 \mathcal{N}(\epsilon, \mathcal{T}, d_{\mathrm{B}}) \rceil$ bits. Hence, for unconstrained offline output coding, packing and covering entropy characterize the bit budget up to the usual radius and rounding gaps.

Our upper algorithm is more structured: it stores codes and masses before the diagram is computed and supports reconstruction of the whole filtration. The packing lower bound nevertheless applies because any such structured state is also an output code.

J.6 Minimum spanning tree derivation of the paired barcode

For a finite metric space whose vertices all enter at time zero, degree-zero Rips persistence is governed by Kruskal's algorithm.

Lemma J.3 (H_0 Rips deaths and minimum spanning trees)**.** *Let X be a connected finite metric space with distinct labels and Rips edge threshold $d(x, y) \leq r$. The multiset of finite H_0 death times equals the multiset of edge weights selected by Kruskal's minimum spanning tree algorithm, with any consistent tie breaking.*

Proof. At scale r, the H_0 rank is the number of connected components in the graph containing edges of weight at most r. Kruskal scans edges in nondecreasing order and selects exactly those that merge two current components. Every selected edge decreases the component count by one and therefore kills one H_0 class at its weight. Rejected edges join vertices already connected and do not change H_0. A spanning tree has $|X| - 1$ selected edges, matching the number of finite bars. Under ties, different selected edge identities can occur, but the number of component mergers at each common weight, and hence the multiset of death values, is fixed. □

For the paired family, Kruskal first selects all q within-pair edges because $\ell_i^\theta < 1$. It then selects any $q - 1$ between-pair edges of weight one. Lemma J.3 immediately recovers Lemma D.2(d). The direct component-count proof in the main lower-bound appendix remains useful because it generalizes transparently to weighted vertex births.

J.7 Euclidean status of the two hard families

Lemma D.2(b) realizes the matched max-DTM family in $\mathbb{R}^{2q}$. The stronger mass-uniform power-weighted family embeds in $\mathbb{R}^{6K}$ by the Gram construction in Appendix E. Thus neither lower bound relies on exotic non-Euclidean metrics. Their ambient dimension grows with the number of encoded bits. Characterizing the corresponding rates at fixed Euclidean or doubling dimension is a natural direction for extending these lower bounds.

J.8 Multiple homological degrees and query models

Suppose a sketch promises diagrams in degrees $\{0, \ldots, k\}$, or answers adaptive queries of the form "return degree q at tolerance ϵ." Restricting the decoder to its degree-zero output gives a valid H_0 sketch with the same state. Therefore every lower bound above applies unchanged. Extra output degrees cannot reduce the information already required by H_0.

Likewise, returning Betti numbers at all scales is enough to recover the degree-zero death multiset in the paired family: the downward jumps of $\beta_0(r)$ occur exactly at the encoded values. The memory obstruction therefore applies not only to an explicit diagram data structure but also to any state that uniformly answers all-scale degree-zero rank queries to sufficient resolution.

J.9 Mergeable and distributed sketches

A mergeable summary associates a state $S(A)$ to each data block and a merge operation producing $S(A \cup B)$. If the final merged state has B bits and decodes with the claimed error on every full input, it is a special case of the encoder in Theorem D.1. The construction can be split across blocks arbitrarily, so the same final-state lower bound holds.

The result concerns final-state storage. Total communication additionally depends on intermediate-state sizes and access to raw blocks during decoding, and can be studied within a corresponding communication-complexity model.

J.10 Floor functions and nonvacuity

Theorem 6 sets

$$q = \left\lfloor \min\left\{ \frac{n}{2}, \frac{1}{2\beta}, \frac{1}{16\epsilon} \right\} \right\rfloor .$$

If the minimum lies in $[1, 2)$, the theorem gives a one-bit packing; if it is below one, the displayed asymptotic lower bound is vacuous. A convenient asymptotic reading is: whenever $n \geq 4$, $\beta \leq \frac{1}{4}$, and $\epsilon \leq \frac{1}{32}$, the floor is at least half of the unrounded minimum, up to an absolute constant. Thus $q = \Omega(\min\{n, \beta^{-1}, \epsilon^{-1}\})$. Keeping the exact floor in the formal theorem prevents hidden small-parameter exceptions.

J.11 What auxiliary advice is counted

The state alphabet must include any input-dependent information available to the decoder, including:

- a learned, input-specific calibration table;
- an input-specific ordering or permutation of cells;
- exact masses or counts;
- nonpublic random seeds;
- a list of selected filtration scales;
- side information identifying which hard-family instance occurred.

Global public constants, the common value of β, the error tolerance, and a public random seed need not be charged. Giving the decoder the encoded bit vector θ as "metadata," however, would spend exactly the information the lower bound counts.

K Sharpness, mass-parameter calculus, and convention checks

This appendix examines the sharpness of the constants and extends the bounds to changes in β and in the atomic masses.

K.1 Sharpness of the two geometric constants

Proposition K.1 (Sharp constants for code and partition errors)**.** *(a) For every $\eta > 0$, there are two same-labelled weighted metric spaces with $\|d - \widehat{d}\|_\infty = \eta$ whose degree-zero max-DTM–Rips diagrams have bottleneck distance exactly η.*

(b) For every $r > 0$ and $D > 2r$, there is a partition of a finite subset of the real line with assignment radius r for which $\|d - \bar{d}\|_\infty = 2r$.

Proof. (a) Take two vertices x, y, each of mass $1/2$, and choose $\beta \leq 1/2$. Every DTM vertex weight is zero. Let $d(x, y) = R$ and $\widehat{d}(x, y) = R + \eta$, with $R > 2\eta$. The two degree-zero diagrams each have one essential bar and one finite point, respectively $(0, R)$ and $(0, R + \eta)$. Matching the finite points costs η. Matching either to the diagonal costs at least $R/2 > \eta$, so the bottleneck distance is exactly η.

(b) On the real line, take raw points $x = 0$, $y = D$, and choose cell representatives $c_1 = r$, $c_2 = D - r$, included as additional points of the finite set. Assign x, c_1 to cell one and y, c_2 to cell two. Every point is at most r from its cell representative. Yet

$$d(x, y) = D, \qquad \bar{d}(x, y) = d(c_1, c_2) = D - 2r.$$

The difference is $2r$, matching the upper bound. □

Part (a) shows that the constant in Theorem 2 cannot be improved without additional structure, even though the DTM-weight part of the proof may be strictly smaller than η on many datasets.

Part (b) concerns the metric step. Particular persistence diagrams can have smaller error because topology may be insensitive to the maximally distorted pair.

K.2 The quotient is not generally a simplicial isomorphism

Take two distinct labelled vertices x_1, x_2 in one zero-distance fibre. Whenever their common DTM weight is at most r, the pullback filtration contains both vertices and their edge, while the representative filtration contains a single vertex. The complexes have different numbers of simplices and are not isomorphic. Projection contracts the edge and a section includes one endpoint; contiguity supplies a homotopy equivalence. This example explains why Proposition C.3(b) is stated at the persistence-module level rather than as equality of filtered complexes.

K.3 Dependence on the DTM mass parameter

For a fixed radial quantile Q, write $w_\beta^{(p)} = \left(\frac{1}{\beta}\int_0^\beta Q(u)^p\, du\right)^{1/p}$.

Lemma K.2 (Dependence on the DTM mass parameter)**.** *(a) For every $p \in [1, \infty)$, the function $\beta \mapsto w_\beta^{(p)}$ is nondecreasing and continuous on $(0, 1]$. On intervals avoiding a radial-mass jump, its p-th power obeys*

$$\frac{d}{d\beta}\left(w_\beta^{(p)}\right)^p = \frac{Q(\beta)^p - (w_\beta^{(p)})^p}{\beta} \geq 0.$$

(b) If $0 < \beta \leq \beta' \leq 1$ and $0 \leq Q \leq D$, then for every $p \in [1, \infty)$,

$$0 \leq w_{\beta'}^{(p)} - w_\beta^{(p)} \leq D\left(\frac{\beta' - \beta}{\beta'}\right)^{1/p}.$$

(c) On a finite space of diameter at most D, using $\beta \leq \beta'$ in the same max-L^p-DTM–Rips filtration, for $p \in [1, \infty)$, changes every vertex and simplex time by at most $D((\beta' - \beta)/\beta')^{1/p}$. The persistence diagrams obey the same bottleneck bound.

Proof. (a) Since Q is nondecreasing, $Q(u)^p \leq Q(\beta)^p$ for $u \leq \beta$, so the average of Q^p over $(0, \beta]$ is at most its endpoint value. Away from a jump, the fundamental theorem of calculus and quotient rule give the derivative formula and nonnegativity.

For global monotonicity, if $0 < \beta < \beta'$, the average of a nondecreasing function over the longer initial interval cannot be smaller:

$$\frac{1}{\beta'}\int_0^{\beta'} Q^p \geq \frac{1}{\beta}\int_0^\beta Q^p.$$

The integral $I(\beta) = \int_0^\beta Q(u)^p du$ is continuous even when the step function Q jumps. Dividing by $\beta > 0$ and taking a continuous p-th root proves continuity.

(b) Let $A = (w_\beta^{(p)})^p$. Then

$$(w_{\beta'}^{(p)})^p = \frac{\beta A + \int_\beta^{\beta'} Q(u)^p du}{\beta'} \leq \frac{\beta A + (\beta' - \beta)D^p}{\beta'} \leq A + \frac{\beta' - \beta}{\beta'}D^p.$$

For nonnegative A, C and $p \geq 1$, $(A + C)^{1/p} \leq A^{1/p} + C^{1/p}$, which follows from the triangle inequality in ℓ^p applied to $(A^{1/p}, C^{1/p})$. Combine this with Lemma K.2(a).

(c) Apply Lemma K.2(b) at every vertex. Edge times are unchanged. Lemma A.2(b) and identity inclusions finish the argument. □

This is a worst-case diameter bound; a local radial quantile calculation can be much sharper.

K.4 Approximate masses under a radial margin

Let $a, \widehat{a}$ be strictly positive probability vectors on the same labelled support V and with the same metric d, and write $\mathrm{TV}(a, \widehat{a}) = \frac{1}{2}\sum_{y\in V} |a_y - \widehat{a}_y| \leq \delta$. For every subset $A \subseteq V$, $|a(A) - \widehat{a}(A)| \leq \delta$. In particular,

$$\sup_{x,r} |F_{x,d,a}(r) - F_{x,d,\widehat{a}}(r)| \leq \delta. \tag{44}$$

Define generalized quantiles outside $(0, 1]$ by clipping: $Q_x(u) = Q_x(0^+)$ for $u \leq 0$, and $Q_x(u) = Q_x(1)$ for $u \geq 1$.

Lemma K.3 (Quantile index sandwich under mass error)**.** *Under* (44)*, for* $u \in (0, 1]$,

$$Q_{x,d,a}((u - \delta) \vee 0) \leq Q_{x,d,\widehat{a}}(u) \leq Q_{x,d,a}((u + \delta) \wedge 1).$$

Proof. Let $r = Q_{x,d,a}((u + \delta) \wedge 1)$. When $u + \delta \leq 1$,

$$F_{x,d,a}(r) \geq u + \delta \quad \Longrightarrow \quad F_{x,d,\widehat{a}}(r) \geq u$$

by (44); hence $Q_{x,d,\widehat{a}}(u) \leq r$. If $u + \delta > 1$, strict positivity on the common finite support gives $r = Q_{x,d,a}(1) = \max_{y\in V} d(x, y)$, so $F_{x,d,\widehat{a}}(r) = 1$ and the same conclusion follows. Exchange $a, \widehat{a}$ and rearrange the index for the lower bound; at the clipped lower endpoint both quantiles begin at the common self-distance zero. □

The lemma shows exactly why total variation alone is insufficient: it shifts the *mass index*, and a finite quantile can jump across a large empty annulus.

Assumption K.4 (Uniform radial (δ, κ)-margin)**.** For every vertex x and every $u \in (0, \beta]$,

$$Q_{x,d,a}((u + \delta) \wedge 1) - Q_{x,d,a}((u - \delta) \vee 0) \leq \kappa.$$

Theorem K.5 (Mass perturbation with a radial margin)**.** *If* $\mathrm{TV}(a, \widehat{a}) \leq \delta$ *and Assumption K.4 holds for both* a *and* $\widehat{a}$*, then*

$$\max_x |w_{d,a,\beta}(x) - w_{d,\widehat{a},\beta}(x)| \leq \kappa.$$

The corresponding same-metric max-DTM–Rips filtrations and all persistence diagrams are κ*-close.*

Proof. Lemma K.3, first centered at a and then with the roles exchanged, implies

$$|Q_{x,d,a}(u) - Q_{x,d,\widehat{a}}(u)| \leq \kappa \quad \text{for } u \in (0, \beta].$$

Apply Lemma A.1(c) to obtain the DTM weight bound. Pairwise edge distances are identical, so every simplex maximum changes by at most κ. Identity inclusions and persistence stability complete the proof. □

Assumption K.4 is checkable on a finite representative set. It is strong by design; without it or another transport/geometric condition, the discontinuous atomic example in Appendix A prevents a uniform small-mass-error theorem.

K.5 Combining metric and margin-controlled mass errors

Corollary K.6 (Metric plus mass error)**.** *Let $d, \widehat{d}$ be nonnegative symmetric zero-diagonal dissimilarities on the same labelled set, and let $a, \widehat{a}$ be strictly positive probability vectors on that set. Suppose*

$$\|d - \widehat{d}\|_\infty \le \eta \quad \textit{and} \quad \sup_x |w_{d,a,\beta}(x) - w_{d,\widehat{a},\beta}(x)| \le \kappa.$$

Then for every $q \ge 0$,

$$d_{\mathrm{I}}\Big(H_q\mathcal{K}_{d,a,\beta}, H_q\mathcal{K}_{\widehat{d},\widehat{a},\beta}\Big) \le \eta + \kappa, \qquad d_{\mathrm{B}}\Big(\mathrm{Dgm}_q\,\mathcal{K}_{d,a,\beta}, \mathrm{Dgm}_q\,\mathcal{K}_{\widehat{d},\widehat{a},\beta}\Big) \le \eta + \kappa.$$

Proof. Compose the mass comparison $(d, a) \to (d, \widehat{a})$, whose simplex-time error is at most κ, with Theorem 2 applied to $(d, \widehat{a}) \to (\widehat{d}, \widehat{a})$, whose error is at most η. Equivalently, the endpoint edge errors are at most η, while the endpoint vertex errors are at most $\kappa + \eta$, so Lemma A.2(b) gives the same sum directly. Identity inclusions and algebraic stability give the displayed bounds. □

The two errors add because the quantile change induced by metric motion and the one induced by mass-index motion need not align. A maximum $\max\{\eta, \kappa\}$ would require a joint proof, not separate bounds. If a fixed-precision quantizer creates zero masses, the common-positive-support hypothesis above no longer applies without first deleting and accounting for those cells.

K.6 Closed versus open filtration conventions

We use closed conditions $d(x, y) \le r$ and $\rho(\sigma) \le r$. If every condition is changed consistently to a strict inequality, finite critical values move between left- and right-continuous versions of the same persistence module. Their decorated interval endpoints may differ, while undecorated bottleneck diagrams agree under the standard ephemeral-module convention. The explicit simplex inclusions in Theorem 2 are cleanest with closed sublevels, so all formal statements use that choice.

K.7 A compact constant table

Table 3: Deterministic constants proved in the paper.

perturbation	object	bound
$\|d - \widehat{d}\|_\infty \le \eta$	same-mass max-L^p-DTM–Rips	η
partition radius r_{part}	raw to aggregated representative	$2r_{\mathrm{part}}$
partition + code	raw to code-only	$2r_{\mathrm{part}} + \eta$
$(1-\epsilon)d \le \widehat{d} \le (1+\epsilon)d$	positive log scale	$-\log(1-\epsilon)$
$\beta \le \beta'$, diameter D	same metric, $p \in [1, \infty)$	$D\big(\frac{\beta'-\beta}{\beta'}\big)^{1/p}$
mass TV δ + radial margin κ	same metric	κ

Each row pairs an error bound with its assumptions and target object; these conditions determine which constant applies to a given comparison.

L Experimental methods and supplementary results

The experiments in Section 7 assess the two terms of the error certificate: loss from the weighted quotient and distortion from its binary encoding. This appendix gives the paired protocols and supplementary evidence, distinguishing fidelity to the same max-DTM–Rips filtration from feature recovery across different filtrations.

L.1 Datasets and paired evaluation

The primary synthetic task consists of 84 noisy circle points and 14 uniform square contaminants, with $M \in \{16, 32, 64\}$ representatives and homology in degrees zero and one. The denser parameter studies use 600 noisy circle points and 100 contaminants. They evaluate H_0, since the complete two-skeleton needed for H_1 exceeds the one-million-simplex limit. The optical and range examples are the primary-circle datasets distributed with JavaPlex 4.3.4 (Adams et al., 2014). Each contains 15,000 unit-norm rows, in dimensions 8 and 25, respectively. We use the supplied arrays without centering or renormalization.

Unless a parameter study states otherwise, BMS uses $b = 100$ independent Gaussian sign projections per representative and DTM mass $\beta = 0.100$. Each setting uses five paired repetitions. Data generation or subsampling, partitioning, projection, and the Sparse–Rips starting point use separate random streams; competing methods share the relevant data and partition within a repetition. Reported means and uncertainty are computed across these five repetitions. The notation mean ± half-width denotes a two-sided 95% Student-t confidence interval, with $t_{4,0.975} = 2.776$; figure captions specify the same convention for error bars and bands.

L.2 Comparison targets and resource accounting

The proposed method is BMS with farthest-point sampling (FPS) representatives. Its exact counterpart uses the same representatives and masses but exact distances. Random representatives and snapped k-means representatives, each with exact and BMS distances, isolate the effect of partition choice on the synthetic task. JL-dim stores b float64 projected coordinates. JL-bits stores the nearest feasible whole number of float64 coordinates to the code payload: two coordinates at $b = 100$. Its word-rounded budget is therefore slightly larger than the sign-code payload, rather than exactly equal to it. Exact masses are stored in float64. Code-only storage counts a continuously packed bitstream and these masses; retained representatives and transient distance matrices are accounted for separately.

VR, empirical DTM–Rips, and Sparse–Rips change the filtration. We compare their one-loop recovery, simplex count, runtime, and memory, but do not report bottleneck distance to the max-DTM–Rips reference. Sparse–Rips uses $\epsilon = 0.500$, a randomized greedy permutation, and the edge-time formula of the GUDHI 2.3 implementation. For same-target methods, we report d_B in H_0, H_1, and their maximum.

The one-loop score counts bars whose relative lifetime $(d - b)/d$ is at least $\tau = 0.250$, with one expected loop. If K bars meet this threshold, the score is F1 $= 2\min(K, 1)/(K + 1)$. This post-diagram decision threshold is distinct from the filtration parameter; its sensitivity is examined below. Runtime summaries use only measurements that satisfy the resource-monitoring criteria. Numerical experiments run on CPUs with a single numerical thread per process, and peak memory refers to isolated-process resident set size.

L.3 Separating partition and coding error

Figure 4 tests how representative choice and encoding contribute to error on the contaminated-circle task. At $M = 64$, BMS-FPS has maximum bottleneck error 0.103 ± 0.063, compared with 0.408 ± 0.187 for JL-bits and 0.054 ± 0.031 for JL-dim. The retained states are 2,400, 1,536, and 51,712 bytes, respectively. Exact FPS has error 0.023 ± 0.016, identifying the residual coding gap. Every BMS comparison satisfies its deterministic certificate. At the fixed relative-lifetime threshold, BMS one-loop F1 is 0.933, compared with 1.000 for DTM–Rips and 0.800 for both VR and Sparse–Rips. Sparse–Rips uses 3,925 simplices and 0.021 seconds, compared with 43,744 simplices and 0.065 seconds for BMS.

Across repetitions, FPS has lower mean error than random representatives, with some variation in their ordering: BMS-random has error 0.268 ± 0.173 at $M = 64$. Snapped k-means BMS is comparable in error (0.115 ± 0.069) but takes 1.056 ± 0.020 seconds, compared with a BMS-FPS mean of 0.065 seconds and confidence-interval half-width below 0.001 seconds. Clustering dominates runtime on this small task. These results characterize the representative choices at the evaluated dataset size and budgets.

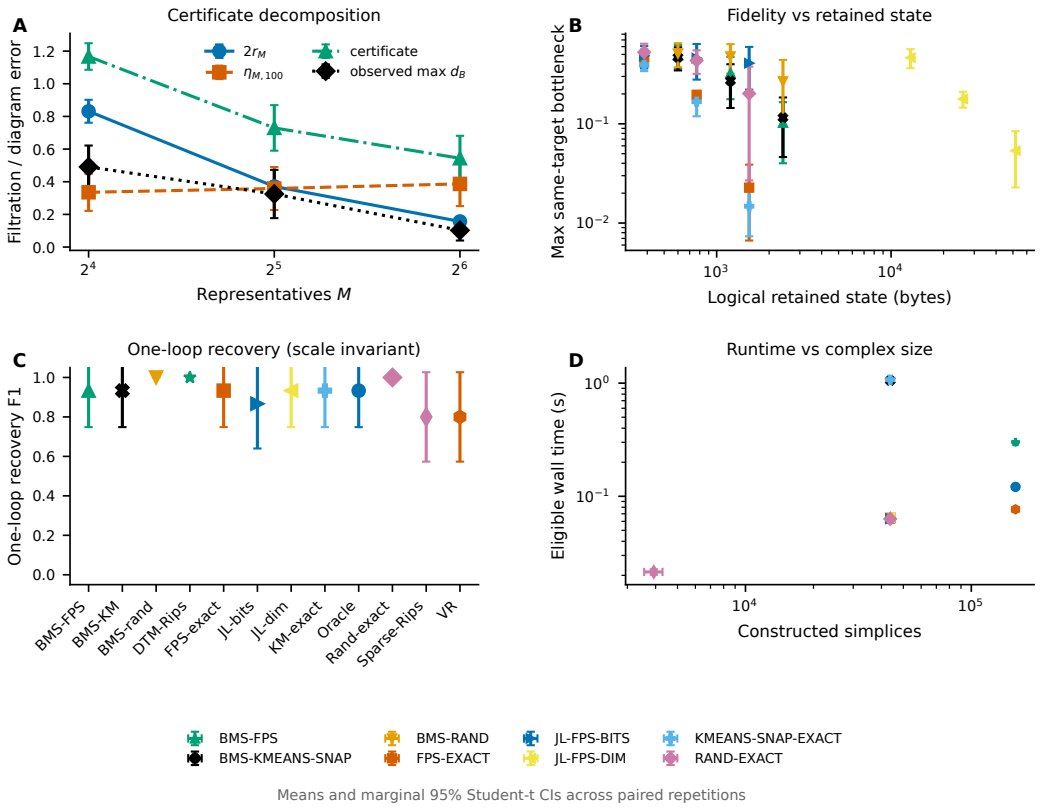


Figure 4: Contaminated-circle comparison over five paired repetitions. Same-target panels show means at each method and representative budget. Cross-filtration methods appear only in topology and complexity panels, which fix $M = 64$ for representative methods and retain the native full-data baselines. Error bars are marginal 95% Student-t confidence-interval half-widths for each plotted quantity.

L.4 Error decomposition and sample-size scaling

To separate coverage, coding precision, and DTM mass, Figure 2 varies M, b, and β over 26, 30, and 11 measured settings, respectively:

$$\begin{aligned} M \in \{&8, 10, 11, 13, 16, 19, 23, 27, 32, 38, 45, 54, 64, 76, 91, 108, 128,\\ &152, 181, 215, 256, 304, 362, 431, 512, 640\},\\ b \in \{&8, 10, 11, 13, 16, 19, 23, 27, 32, 38, 45, 54, 64, 76, 91, 100, 108, 128,\\ &152, 181, 215, 256, 304, 362, 431, 512, 609, 724, 861, 1024\},\\ \beta \in \{&0.010, 0.015, 0.022, 0.032, 0.046, 0.068, 0.100, 0.147, 0.215, 0.316, 0.464\}. \end{aligned}$$

The M-study fixes $b = 100, \beta = 0.100$; the b-study fixes $M = 128, \beta = 0.100$; and the β-study fixes $M = 128, b = 100$. All settings use five paired repetitions, with a shared exact reference whenever the target filtration is unchanged. The 335 comparisons satisfy the partition, coding, and end-to-end bounds $2r_{\text{part}}$, η, and $2r_{\text{part}} + \eta$, respectively. This verifies the deterministic inequalities on the evaluated instances, not a population coverage rate.

At fixed code length, refining the quotient reduces coverage error but leaves a coding floor. From $M = 8$ to 640, the mean cover bound falls from 1.881 to .021, while endpoint code distortion changes from .348 to .538. BMS error falls from .749 to .085 by $M = 128$ and remains .082 at $M = 640$, whereas FPS-EXACT reaches less than 0.001. At fixed $M = 128$, increasing b from 8 to 1,024 reduces mean uniform distortion from 1.303 to .135, coding-only bottleneck error from .258 to .019, and sign-code collision fraction from .884 to .033. A descriptive least-squares fit over the 30 log-

spaced observations gives a log–log slope of $-.481$ for distortion versus b, consistent with, but not a proof of, the $b^{-1/2}$ concentration dependence.

The raw-sample scaling study draws unit-sphere points in dimension 8 at

$$n \in \{128, 181, 256, 362, 512, 724, 1024, 1448, 2048, 2896, 4096, 5793, 8192, 11585, 16384, 23170, 32768\}.$$

It uses blockwise FPS and computes H_0 only on $M = 128$ representatives. At $n = 32768$, a full distance matrix would require 8 GiB; none is allocated. Across this grid, the code-and-mass state is exactly 2,624 bytes, the representative-plus-code state is 10,816 bytes, and the complex contains 8,256 simplices. Thus the post-partition state remains fixed while the raw sample grows. From $n = 128$ to 32,768, mean partition time rises from .009 to .222 seconds, while encoding (.013 to .011 seconds) and persistence (.040 to .041 seconds) remain flat within sampling variation. Constant retained state therefore does not imply constant time to construct the quotient.

These descriptive grids are distinct from configuration searches, which use $\{16, 32, 64, 128, 256, 512\}$ for both M and the base b grid when the filtration fits the resource limit. Complete-H_1 downstream searches restrict M to $\{16, 32, 64, 128\}$. The strict equal-budget comparison additionally admits budget-specific exact-fill configurations for both controlled and operational selection. Downstream radial precision is fixed at $b_r = 8$; the separate radial study below does not reselect a classifier. Lines in the trend figures join measured settings only and do not represent additional hyperparameter optimization.

L.5 Real-data fidelity and the cost of quotient construction

Figure 5 evaluates both raw-to-code and quotient-to-code fidelity on the real sphere data. In subset mode, all seven methods use the same 128 sampled rows, whose exact max-DTM–Rips filtration is the reference. In full mode, blockwise FPS scans all 15,000 rows, computes exact bucket masses, and constructs persistence on 128 representatives. The reference is the exact filtration on those representatives, not an oracle on all 15,000 points.

BMS-code has lower diagram error than the word-rounded JL-bits baseline; JL-dim reduces error further at substantially greater storage cost. The 2,624-byte BMS state contains 1,600 packed code bytes and 1,024 mass bytes. With original-space representatives retained, storage becomes 10,816 bytes for optical and 28,224 bytes for range, respectively 17% and 6% above their exact representative states. BMS one-loop F1 is 1.000 on optical and 0.800 on range; the best cross-filtration baseline reaches 1.000 on each, but differs by dataset. One-loop recovery depends on the dataset and filtration, so the evidence supports a fidelity–memory trade-off rather than uniform topological dominance. Threshold sensitivity is analyzed in Figure 7.

Code collisions occur only in the range subsets, between representatives whose exact distances range from 0.008 to 0.114. No collision occurs among the full-mode FPS representatives. Colliding codes remain labelled vertices with their own masses, as required by the code-only theorem; they are not merged.

A full 15000×15000 float64 distance matrix would require 1.800 GB, and the full two-skeleton would contain 562,500,012,500 simplices. Neither is constructed. Full mode instead uses 1,920,000 blockwise point–representative distance evaluations, a 128×128 distance matrix, and 349,632 closed simplices. These counts describe computational cost separately from the retained representation payload.

L.6 Radial precision and the angular floor

To isolate the radial term in Proposition F.3, we use the same 600-circle-point and 100-contaminant model, fix $M = 128, b = 512, \beta = 0.100$, and sweep $b_r \in \{1, \ldots, 32\}$. Each of five paired repetitions fixes its source cloud, FPS quotient, exact quotient H_0 reference, and angular projection across all 32 radial precisions. Both the fixed barycenter and certificate-minimizing center policies are evaluated. This isolates quantization error without changing the angular sketch or performing downstream model selection.

Figure 6 shows the predicted transition from radial-limited to angular-limited reconstruction. For the barycenter policy, mean total metric distortion falls from 1.948 ± 0.103 at one radial bit to $0.175 \pm$

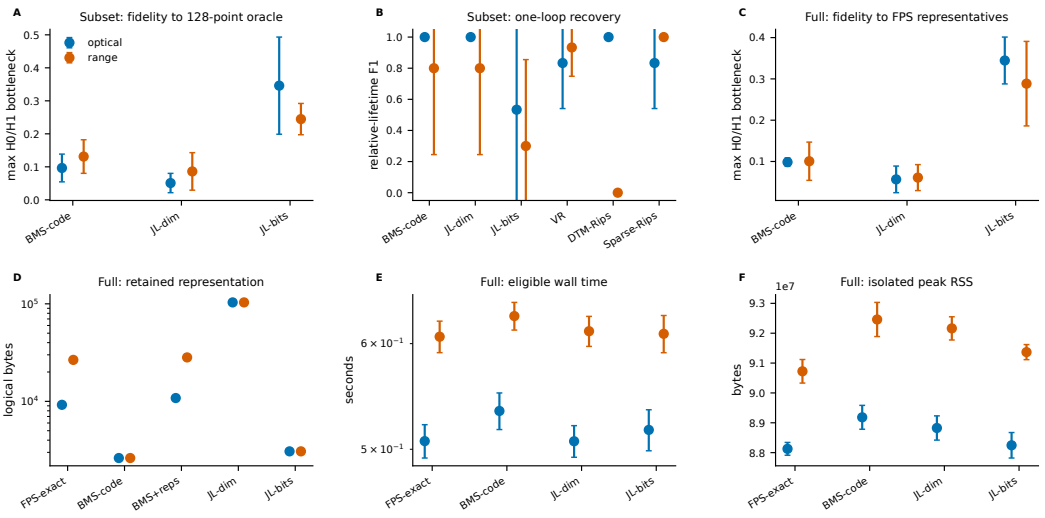


Figure 5: JavaPlex fidelity, topology, and cost over five paired repetitions. Points are means and error bars are 95% Student-t confidence-interval half-widths. A–B use the 128-row subset; C–F use all rows for assignment but compute persistence only on 128 representatives. D separates code-only storage from storage that also retains representatives. No panel reports bottleneck distance to a full-data oracle.

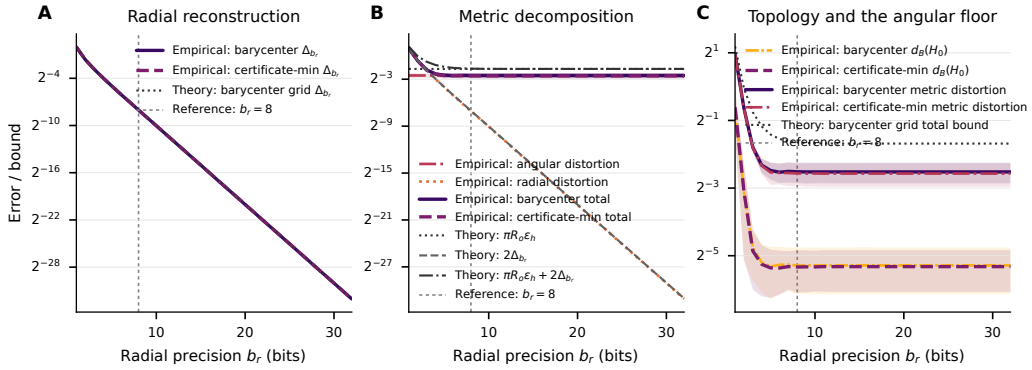


Figure 6: Radial precision reveals an angular approximation floor over five paired repetitions. A compares realized radial error Δ_{b_r} with its nearest-grid bound. B separates angular, radial, and total metric distortion and shows the sufficient terms $\pi R_o \epsilon_h$, $2\Delta_{b_r}$, and their sum. C compares H_0 bottleneck error with the metric envelopes. Curves are means with 95% Student-t confidence bands for empirical quantities; bound curves are shown without bands. The horizontal axis is linear in b_r, vertical axes are $\log_2$, and the dotted reference marks the downstream choice $b_r = 8$.

0.036 at eight bits, and remains 0.174 ± 0.036 at 32 bits. Mean H_0 bottleneck error falls from 0.655 ± 0.293 to 0.026 ± 0.011 by eight bits. Further radial refinement leaves the angular approximation floor essentially unchanged. Small nonmonotonic changes in realized diagram error remain compatible with stability, which controls an upper bound rather than requiring monotone error. Both center policies satisfy all radial, angular, total-distortion, and H_0 inequalities throughout the study. Code-only state grows from 9,240 to 9,736 bytes between $b_r = 1$ and 32, matching $\lceil Mb/8 \rceil + \lceil Mb_r/8 \rceil + 8M + 8$.

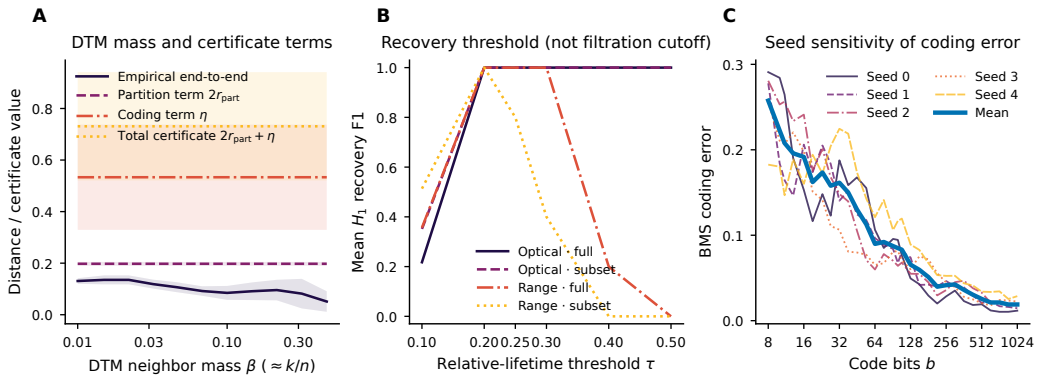


Figure 7: DTM-mass, loop-threshold, and random-seed sensitivity. A varies β, showing mean end-to-end error and partition, coding, and total certificate terms with 95% Student-t confidence bands. B shows mean one-loop F1 as the post-diagram relative-lifetime threshold varies for the optical and range examples in full and subset modes. C shows the five individual random-seed curves and their mean across 30 code lengths, making repetition variability explicit in this sensitivity analysis. Lines join measured settings; A uses a logarithmic β axis and C a $\log_2$ code-length axis.

L.7 Neighbor mass, recovery scale, and random-seed sensitivity

Figure 7 distinguishes three sources of sensitivity. In a uniform n-point cloud, DTM mass β integrates the first $k_\beta = \lceil \beta n \rceil$ radial order statistics, including the zero self-distance, with a fractional final atom when βn is noninteger. Thus k_β counts included order statistics, not neighbors after excluding the query point. For the 700-point study, $\beta = 0.010$–0.464 spans $k_\beta = 7$–325. A weighted quotient instead uses cumulative bucket mass, so no single integer neighbor count need apply. The relative-lifetime threshold $\tau \in \{0.100, 0.200, 0.250, 0.300, 0.400, 0.500\}$ is applied after persistence computation; it is not a filtration cutoff.

Changing β changes the DTM filtration and realized diagram error, even when the uniform metric terms $2r_{\text{part}}$ and η are fixed. Changing τ can add or remove counted loops without changing the diagram itself, explaining why feature-recovery conclusions depend on the decision scale. The final panel displays the joint effect of data, partition, and projection randomness across the five experimental seeds. These analyses characterize sensitivity rather than select parameters using the reported outcomes.

M Downstream prediction certificates

For a fixed feature map, scaler, and classifier, diagram stability yields a prediction certificate. This appendix derives that conditional guarantee for the sampled landscapes used in the experiments, then specifies the grouped evaluation of predictive utility. The certificate concerns evaluation of a fixed learned classifier and its fixed outer-training scaler; it implies neither stability under retraining nor population generalization.

M.1 Sampled persistence landscapes

Let $\mathcal{H}$ be a finite set of homological degrees. For each $r \in \mathcal{H}$, let D_r and $\widehat{D}_r$ denote the finite persistence diagrams presented to the vectorizer. The experiments use the following finite-bar convention: bars with infinite death are removed before vectorization, while every finite bar participates in the bottleneck matching and may be matched to the diagonal. Write $\lambda_{D_r,k}(t)$ for landscape level k of D_r (Bubenik, 2015). Fix K landscape levels and a common grid $\mathcal{T} = \{\tau_1, \dots, \tau_G\}$. In the coordinate order $i = (r, k, g)$, define

$$\Phi(D) = \big(\lambda_{D_r,k}(\tau_g) : r \in \mathcal{H},\ 1 \le k \le K,\ 1 \le g \le G\big).$$

Theorem M.1 (Sampled landscape and linear-score certificate)**.** *Suppose that $d_{\mathrm{B}}(D_r, \widehat{D}_r) \le \epsilon_r$ for every $r \in \mathcal{H}$. Then, for any fixed affine score $s(x) = w^\top x + a$,*

$$\left\|\Phi(D) - \Phi(\widehat{D})\right\|_2 \le \sqrt{KG \sum_{r\in\mathcal{H}} \epsilon_r^2}, \qquad \left|s(\Phi(D)) - s(\Phi(\widehat{D}))\right| \le \|w\|_2 \sqrt{KG \sum_{r\in\mathcal{H}} \epsilon_r^2}.$$

Proof. Persistence-landscape stability gives, degree by degree,

$$\sup_{k\ge 1} \sup_{t\in\mathbb{R}} \left|\lambda_{D_r,k}(t) - \lambda_{\widehat{D}_r,k}(t)\right| \le d_{\mathrm{B}}(D_r, \widehat{D}_r) \le \epsilon_r.$$

If the bottleneck infimum is not taken as attained, apply the stability inequality to a matching of cost at most $\epsilon_r + \gamma$ and let $\gamma \downarrow 0$. Evaluating on the fixed grid and retaining only the first K levels are coordinate projections, so neither operation enlarges the bound. Hence, for every retained coordinate $i = (r, k, g)$,

$$\left|\Phi_i(D) - \Phi_i(\widehat{D})\right| = \left|\lambda_{D_r,k}(\tau_g) - \lambda_{\widehat{D}_r,k}(\tau_g)\right| \le \epsilon_r. \tag{45}$$

Summing the KG squared coordinate bounds within each degree yields

$$\left\|\Phi(D) - \Phi(\widehat{D})\right\|_2^2 = \sum_{r\in\mathcal{H}} \sum_{k=1}^{K} \sum_{g=1}^{G} \left|\lambda_{D_r,k}(\tau_g) - \lambda_{\widehat{D}_r,k}(\tau_g)\right|^2 \le KG \sum_{r\in\mathcal{H}} \epsilon_r^2.$$

Taking square roots proves the feature bound. Cauchy–Schwarz then gives

$$\left|w^\top\left(\Phi(D) - \Phi(\widehat{D})\right)\right| \le \|w\|_2 \left\|\Phi(D) - \Phi(\widehat{D})\right\|_2,$$

which proves the score statement because the intercept cancels.

If a finite diagram is empty, all of its landscape levels are identically zero, and the same coordinate argument applies through the diagonal part of the bottleneck matching. Under the stated finite-bar convention, essential bars retain their infinite deaths and contribute no sampled coordinate. Equivalently, when starting from extended diagrams at finite bottleneck distance, essential points must match essential points; restricting that matching to finite points preserves its cost bound. □

The theorem concerns evaluation with a fixed trained pipeline. Comparing parameters after replacing training diagrams and refitting the scaler or classifier additionally calls for stability assumptions on the learning procedure.

M.2 A standardized multiclass margin certificate

Let the outer-training split determine a coordinate mean μ_i and a scale $\sigma_i > 0$. As in the implementation, a zero-variance training coordinate is assigned $\sigma_i = 1$. Standardize a raw landscape vector v by $x_i = (v_i - \mu_i)/\sigma_i$. For coordinate $i = (r, k, g)$, choose either the realized difference $\Delta_i^{\mathrm{real}} = |\Phi_i(D) - \Phi_i(\widehat{D})|$ or its degreewise bottleneck upper bound $\Delta_i^{\mathrm{cert}} = \epsilon_r$, and define

$$B(\Delta) = \left(\sum_i \frac{\Delta_i^2}{\sigma_i^2}\right)^{1/2}. \tag{46}$$

Equation (45) ensures that $B(\Delta^{\mathrm{real}}) \le B(\Delta^{\mathrm{cert}})$.

Corollary M.2 (Fixed-classifier sampled margin certificate)**.** *Let a fixed multiclass affine classifier have scores $s_c(x) = w_c^\top x + a_c$, and suppose its unique prediction on D is y. For either valid choice of Δ above, the prediction on $\widehat{D}$ is also y whenever*

$$s_y(x) - s_c(x) > \left\|w_y - w_c\right\|_2 B(\Delta) \qquad \text{for every } c \ne y. \tag{47}$$

Proof. Let $\widehat{x}$ be the standardized feature vector of $\widehat{D}$. Because the same outer-training mean is subtracted from both vectors,

$$\|x - \widehat{x}\|_2^2 = \sum_i \frac{|\Phi_i(D) - \Phi_i(\widehat{D})|^2}{\sigma_i^2} \le B(\Delta)^2.$$

For every competitor $c \neq y$, Cauchy–Schwarz gives

$$\begin{aligned} s_y(\widehat{x}) - s_c(\widehat{x}) &= s_y(x) - s_c(x) + (w_y - w_c)^\top(\widehat{x} - x) \\ &\geq s_y(x) - s_c(x) - \left\|w_y - w_c\right\|_2 \|\widehat{x} - x\|_2 \\ &\geq s_y(x) - s_c(x) - \left\|w_y - w_c\right\|_2 B(\Delta). \end{aligned}$$

Condition (47) makes the last line strictly positive for every competitor, so y remains the unique maximizer. □

For the raw-to-code BMS comparison, Theorem 4 supplies the degreewise premise without an additional learning assumption. In the notation above one may set $\epsilon_r = 2r_{\text{part}} + \eta_r$ with $\eta_r := \eta$ for $r \in \mathcal{H}$, where η is the single uniform decoded-distance distortion. A sharper degree-specific verified diagram bound may replace $2r_{\text{part}} + \eta$ in a realized diagnostic, but it must not be reported as the all-scale geometric certificate of Theorem 4. Corollary M.2 is therefore a sufficient sample-level robustness statement for a fixed trained pipeline, not a claim of population accuracy or retraining stability.

M.3 Development selection and grouped confirmation

The graph study in Section 7.1 asks whether the compressed representation retains predictive utility under a fixed encoding protocol. The development candidates were IMDB-BINARY, IMDB-MULTI, and COLLAB. Development-only screening required an informative BMS representation, including balanced accuracy above chance; IMDB-BINARY and COLLAB were eligible and proceeded to independent confirmation. Dataset selection and the encoding protocol were frozen before confirmation evaluation. This selection limits the scope of the confirmation evidence to the two eligible datasets, rather than arbitrary graph tasks.

The confirmation populations contain 792 graphs with class counts 400/392 and 4,032 graphs with class counts 2,072/621/1,339, respectively. Their 426 and 3,254 exact-isomorphism groups are kept intact across development/confirmation, outer folds, and inner folds. The original TU collections contain 1,000 and 5,000 graphs (Morris et al., 2020). Ten repeats of stratified grouped 10-fold cross-validation give 100 outer folds per dataset. Five-fold grouped inner validation selects logistic-regression $C \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ and unweighted or balanced class weights by mean balanced accuracy, breaking ties by smaller C and then unweighted classes. Feature scaling is fitted on training data only.

The graph metric is the median-normalized resistance embedding described in Section M.4. The per-object caps are $\{32, 64, 128, 256, 512\}$ KiB for IMDB-BINARY and $\{64, 128, 256, 512\}$ KiB for COLLAB. The nominal representative count is $M_B = 16$ at the smallest cap and 32 otherwise, with effective count $M_i = \min(M_B, n_i)$. BMS uses $b = 512$ sign bits and 8-bit quantized norms, choosing between the origin and mass barycenter to minimize the certificate term. Gaussian JL uses dimension 12 on IMDB-BINARY and 32 on COLLAB. Thus the five and four nominal caps, respectively, yield only two distinct nominal encoder configurations per method: $M_B = 16$ at the first cap and $M_B = 32$ thereafter, with code length and JL dimension fixed. The resulting curves compare these encoders over nominal budget caps; they are not optimally tuned frontiers in achieved payload bytes. All three methods compute max-DTM persistence with $\beta = 0.100$ in H_0 and H_1. The sampled landscapes of Theorem M.1 retain five levels and 128 grid points per degree, excluding infinite-death bars. Each degree's grid runs from the smallest finite birth to the largest finite death in the current training fold; both grids and coordinate scaling are fitted anew within inner validation and on the complete outer-training fold before testing.

Paired inference uses a separate estimand for each comparison. For BMS versus Gaussian JL, each outer fold contributes the difference between normalized areas of right-continuous balanced-accuracy curves on the fixed $\log_2 B$ support, without interpolation. For BMS versus FPS, it contributes the equal-weight mean of balanced-accuracy differences across retained caps. Let these paired differences be $d_1, \ldots, d_N$, with $N = 100$, mean $\bar{d}$, and sample variance s_d^2. To account for dependence in repeated cross-validation, the Nadeau–Bengio correction uses (Nadeau & Bengio, 2003)

$$\rho = \frac{1}{N}\sum_{i=1}^{N} \frac{n_{\text{test},i}}{n_{\text{train},i}}, \qquad \widehat{\text{SE}}_{\text{corr}} = \sqrt{\left(\frac{1}{N} + \rho\right) s_d^2}, \qquad T = \frac{\bar{d} - d_0}{\widehat{\text{SE}}_{\text{corr}}}. \tag{48}$$

Table 4: Task-level graph references provide a separate predictive comparison. Balanced accuracy is mean ± sample SD across all 100 outer folds for each dataset and method. The SD is descriptive fold-to-fold spread, not a standard error or confidence interval. Chance is the analytical uniform-class expectation. These methods do not enter the byte-budget curves or the four confirmatory tests.

Reference	IMDB-BINARY	COLLAB
HKS-PL	0.697 ± 0.080	0.713 ± 0.025
WL	0.714 ± 0.077	0.727 ± 0.043
Degree	0.717 ± 0.074	0.744 ± 0.034
NoG	0.639 ± 0.104	0.700 ± 0.030
Chance	0.500 ± 0.000	0.333 ± 0.000

Here $n_{\text{test},i}$ and $n_{\text{train},i}$ count unique exact-isomorphism groups. The four predeclared one-sided tests use 99 degrees of freedom: the superiority boundary is $d_0 = 0$, and the non-inferiority boundary is $d_0 = -0.020$. Holm's procedure is applied once across all four tests at familywise level 0.050 (Holm, 1979). The shaded bands in panels C/D of Figure 3 are method-specific 95% Nadeau–Bengio corrected intervals. In panels E/F of Figure 3, points mark paired means, horizontal lines give two-sided 95% corrected intervals, and arrows start at one-sided 95% corrected lower bounds. Panel F marks the non-inferiority boundary at −0.020. All corrected 95% intervals in the figure are unadjusted for multiplicity; they are not Holm-adjusted intervals. The separate compression condition compares maximum BMS and FPS payloads over all objects and retained caps, requiring $\max S_{\text{BMS}}/\max S_{\text{FPS}} \leq 1/4$. It is not a maximum of objectwise ratios. Payload includes packed codes, quantized norms, masses, and the stored norm scale for BMS, or coordinates and masses for FPS/JL. Shared landscapes, scaler/classifier state, working memory, and training time are outside this comparison.

BMS improves the normalized log-budget accuracy area over Gaussian JL by 0.065 on IMDB-BINARY and 0.063 on COLLAB, with globally Holm-adjusted $p = 0.035$ and $p < 0.001$, respectively. The BMS–FPS equal-cap mean differences are −0.003 and −0.001; their one-sided 95% corrected lower bounds, −0.032 and −0.021, fall below the −0.020 non-inferiority margin. The corresponding globally Holm-adjusted values are $p = 0.169$ and $p = 0.124$. Thus the mean accuracies are close in these comparisons, while the corrected uncertainty leaves non-inferiority unresolved. The supported advantage is the normalized log-budget accuracy area relative to the fixed Gaussian encoder on the two evaluated datasets; the exact-geometry comparison characterizes the remaining accuracy–memory balance.

To examine predictive utility beyond the shared metric, Table 4 compares task-specific graph representations on the same outer folds, separating predictive utility from fidelity to the shared geometric target. It reports heat-kernel signature persistence landscapes (HKS-PL), structure-only Weisfeiler–Lehman (WL) features, degree histograms, basic graph-count features (NoG), and analytical chance performance. NoG comprises node count, edge count, component count, density, and cycle rank. Each mean gives equal weight to all 100 outer-fold scores; the accompanying sample standard deviation (SD) uses divisor 99 and describes fold-to-fold spread, not uncertainty of the mean. These exploratory references are neither byte-axis methods nor members of the four-test family.

Degree histograms have the largest descriptive mean among these references on both datasets. These results situate the compressed topological representation among task-specific alternatives; their descriptive comparison complements the same-target fidelity and byte-budget evaluations.

M.4 The fixed graph metric

The structure-only graph tasks use the canonical square-root pseudoinverse embedding

$$\Psi_G = (L^\dagger)^{1/2} = U \operatorname{diag}\big(\lambda_k^{-1/2}\mathbf{1}\{\lambda_k > 0\}\big)U^\top,$$

where L is the combinatorial Laplacian and each row represents a vertex. For a connected graph it stores n_i coordinates but has rank $n_i - 1$. For a disconnected graph, all vertices are retained and κ_i scaled one-hot component coordinates separate its components, so $d_i = n_i + \kappa_i$; byte accounting uses this stored input dimension, not spectral rank. Distances are divided by their median nonzero value.

The BMS guarantee conditions on this fixed metric map. Stability with respect to graph edits would additionally involve the response of the map itself. The shared representation comparison uses graph structure without node attributes.

M.5 Predictive transfer across datasets

Persistence fidelity and predictive accuracy assess complementary aspects of the representation. Earlier auxiliary studies used 212 Tumor-C cell-pattern regions of interest (ROIs) (Vipond et al., 2021), 188 MUTAG graphs, and 1,113 PROTEINS graphs (Morris et al., 2020). They used distinct protocols and are not pooled with the grouped confirmation study: Tumor-C used 20 stratified 80/20 splits for uncapped evaluation and 100 for strict equal-budget evaluation; the graph tasks used 10×10 stratified cross-validation. Selection used training data only. Tumor-C concerns ROI-level pattern classification, not patient diagnosis or cross-patient generalization.

At each dataset's largest common cap, BMS had lower controlled mean H_1 error than Gaussian JL. Operational accuracy under separately selected representations varied across tasks: FPS had higher means on Tumor-C and MUTAG, whereas BMS had a higher descriptive mean on PROTEINS, with the same ordering in the uncapped graph comparisons. Spatial statistics and QuPID achieved higher means on the reported Tumor-C tasks (Van Huffel et al., 2025), while HKS, WL, and structural graph features provided additional predictive references. These comparisons concern task-specific representation quality, complementing the same-target bottleneck analysis.

After global Holm adjustment, the six two-sided equal-budget BMS–FPS/JL comparisons and the eight uncapped graph comparisons against FPS, JL-bits, HKS-PL, and linear structure-only WL remained statistically inconclusive. Both families used the repeated-cross-validation correction in Equation (48), with object-count train/test ratios. These tests leave equivalence and non-inferiority open. Together, the auxiliary studies motivate evaluating persistence fidelity, retained-state compression, and predictive utility as distinct outcomes. Retained payload measures the representation itself; total model memory and training cost involve additional components.