SHARP CERTIFICATES FOR WORST-CASE CONTROLLED COST SENSITIVITY

Anonymous authors
Paper under double-blind review

ABSTRACT

We study exact identification and finite-sample certification of the worst-case controlled contrast $M_i = \sup_z |\Delta_i(z)|$. The access contract is explicit: fresh bounded joint interventions fix every nuisance coordinate, and a valid global Lipschitz constant is supplied externally. For any feasible collection of simultaneous grid-value intervals, McShane–Whitney envelopes give the exact identified interval for M_i; at zero trace its unavoidable geometric width is $\min\{L_i\rho_i, 1\}$. Thresholding its confidence refinement yields family-wise valid SENSITIVE, INSENSITIVE, or UNRESOLVED reports. For adaptive placement we prove a KL-weighted cap- packing lower bound and a doubling-tree upper bound. On $[-1, 1]^d$, a converse valid for every L-Lipschitz insensitive base, without the lower comparison's strict-slope or interiority assumptions, bounds the packing functional by $2(3/2)^{d+2} \max\{2L, 1\}^d \int (\tau - |\Delta|)^{-(d+2)}$. The reverse bound, and hence the two-sided comparison, still require an $(L/16)$-Lipschitz base and interior contrasts; corrected constants improve by factors 8, 16, and 32 for $d = 1, 2, 3$ but remain loose. On the declared synthetic bank, Lipschitz placement remains the strongest equal-cap comparator (1,155 versus 1,037 stagewise decisions at 1,048,576 pairs), and understating L causes false-low reports. A retained $d = 2$ execution is reported only as an aggregate theorem audit, not a competitive evaluation. The result certifies controlled sensitivity under this oracle contract; it does not learn L, identify causal parents, improve a controller, or establish policy safety.

1 INTRODUCTION

A pointwise intervention can show that changing candidate X_i affects cost at one reference setting. It cannot show that the effect is uniformly below a tolerance over all feasible references. Failure to reject is especially unsafe: it supplies no upper bound. We study the worst-case controlled sensitivity

$$M_i = \sup_{z \in \mathcal{Z}_i} |\Delta_i(z)|,$$

where $\Delta_i(z)$ is a bounded joint-intervention contrast. The report is SENSITIVE, INSENSITIVE, or UNRESOLVED, so missing evidence is not converted into evidence of low sensitivity.

For a fixed finite reference net and a known valid Lipschitz constant, we combine simultaneous bounded-mean intervals with the classical McShane–Whitney Lipschitz extension (McShane, 1934; Whitney, 1934). This gives lower and upper confidence endpoints for every M_i and controls the family event of any wrong decisive report. Certification requires both known valid L and paired joint interventions; missing either forces withholding. We provide no estimator or validator for L.

For arbitrary simultaneous grid-value intervals, we characterize exactly all possible values of M_i among bounded Lipschitz extensions. The resulting trace-specific upper endpoint never exceeds the coarse $L_i\rho_i$ certificate. At the noiseless zero trace, the zero function and $\min\{Ld(z, G), 1\}$ recover the worst-case gap $\min\{L\rho, 1\}$. Thus additional samples at unchanged references cannot remove geometric uncertainty. A separate bounded Bernoulli mechanism shows that finite grids also cannot identify direct parents over unrestricted mechanisms.

This is a partial-identification statement under a smoothness restriction: the data and shape constraint identify a set of possible values rather than a single effect. That perspective is classical in causal

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.

bounds (Balke & Pearl, 1997; Tamer, 2010); our specific object is the supremum of an interval-valued Lipschitz trace, not an average treatment effect.

Evaluation details and assumption controls are in Appendix E.

1.1 THEOREM-LEVEL RELATION TO PRIOR GUARANTEES

Theorems 1–2A: fixed-trace identification, not a new extension theorem. McShane–Whitney already supplies the extremal real-valued Lipschitz extensions (McShane, 1934; Whitney, 1934); DIRECT and LIPO likewise use Lipschitz envelopes for sequential optimization (Jones et al., 1993; Malherbe & Vayatis, 2017). We do not claim either construction as new. Theorem 1 instead puts simultaneous bounded-pair confidence intervals around the continuous supremum M_i and controls the family event of any wrong decisive report. Theorem 2A applies the classical extensions to a *fixed* feasible vector of interval-valued grid observations and identifies exactly the attainable set of $\sup_z |f(z)|$, including both endpoints and every intermediate value. Accordingly, "exact" and "sharp" here mean set identification conditional on that fixed trace, the known valid L, and the range cap $[-1, 1]$; they do not mean sequential minimax optimality, a sharper extension theorem, or learning L. At the fixed zero trace the identified interval is $[0, \min\{L\rho, 1\}]$.

Noisy GP level-set estimation already supports above/below/deferred labels on a finite domain (Gotovos et al., 2013). Allocating its failure probability and thresholding valid envelopes can recover the same three labels. Theorem 1's delta is therefore not the labels: it is the family-wise interval for a continuous worst-case controlled contrast under the bounded paired-intervention oracle.

A certified lower/upper pair already gives the three-way decision. Any valid $LB \leq M_i \leq UB$ already implies SENSITIVE when $LB > \tau$, INSENSITIVE when $UB < \tau$, and UNRESOLVED otherwise; certified multifidelity zeroth-order optimization (de Montbrun & Gerchinovitz, 2024) and instance-dependent zeroth-order Lipschitz optimization with error certificates (Bachoc et al., 2021b) already supply such a lower/upper pair, so naming the same three outputs SENSITIVE/INSENSITIVE/UNRESOLVED is not by itself a contribution. What must be compared item by item instead is: (i) whether the interval is the tightest set identified by one already-collected, fixed trace, as in Theorem 2A, or only some valid but possibly loose interval; (ii) whether the access oracle is an exact or purchasable-fidelity zeroth-order query, as assumed in those two references, or the bounded stochastic joint-intervention mean used here; and (iii) whether adaptive query complexity is governed by the gap to a global maximizer, as in standard best-point identification, or by structure local to one fixed threshold τ, as in Theorem 4D below. A tighter fixed-trace interval does not by itself imply fewer adaptive queries under any policy; the query policy and its computational cost are compared on their own terms in Theorems 4C–4D, not inferred from interval tightness alone.

Theorems 4C–4D: the lower/upper instance pair. Finite-arm thresholding has inverse-gap complexity $\sum_k \Delta_k^{-2}$ (Locatelli et al., 2016), and exact-KL allocation characterizes fixed-confidence best-arm identification (Garivier & Kaufmann, 2016). Theorem 4C uses the same change-of-measure logic but its Q ranges over disjoint Lipschitz-consistent cap alternatives for a sign-symmetric continuum supremum decision. It claims neither a new KL inequality nor an inherited optimal constant. Lipschitz-bandit KL lower bounds address asymptotic regret (Magureanu et al., 2014), not this fixed-confidence three-way decision.

For the upper side, certified continuous Lipschitz optimization gives high-probability near-maximizer bounds (de Montbrun & Gerchinovitz, 2024), while Lipschitz level-set approximation with exact evaluations assumes a nonempty target level set (Bachoc et al., 2021a). At Theorem 4C's insensitive base the τ-level set of $|\Delta|$ is empty. Theorem 4D's specific delta is a noisy, family-wise threshold-margin spend bound $L \int (\tau - |\Delta(z)|)^{-3} dz$ and its conditional comparison with Q. The worst-case exponent agrees after specialization, and no rate separation from classical level-set or Lipschitz-optimization methods is asserted.

Contributions. The paper centers the simultaneous supremum interval and its exact fixed-trace identified set (Theorems 1–2A), then the cap-packing lower bound and adaptive margin-integral upper bound (Theorems 4C–4D). Theorem 6 gives the dimension-explicit comparison. Its converse upper bound holds without the lower comparison's strict-slope and interiority assumptions; equiv-

alence still requires those assumptions. The adaptive claim is limited to the distinctions above, not uniqueness of three-way reporting or a different worst-case exponent.

The result is not observational conditioning, structural-parent identification, measured-data transfer, controller improvement, deployment, or policy safety.

2 TARGET AND ACCESS MODEL

Let $C \in [0,1]$ be a one-step cost and $X = (X_i, X_{-i})$ observed candidate coordinates. The historical causal-interface target was $P_C = \mathrm{Pa}(C) \cap X$, but this paper does not estimate or identify P_C. Its active estimand requires a simulator or experiment that can execute the joint operations

$$\Delta_i(z) = \mathbb{E}[C \mid \mathrm{do}(X_i = 1, X_{-i} = z)] - \mathbb{E}[C \mid \mathrm{do}(X_i = -1, X_{-i} = z)], \tag{1}$$

for z in a compact feasible reference domain $\mathcal{Z}_i$. Thus $X_{-i} = z$ is intervened on, not conditioned on after intervening on X_i; descendants in X_{-i} are overwritten. If this operation is unavailable, the procedure withholds.

Our target is worst-case controlled sensitivity

$$M_i = \sup_{z \in \mathcal{Z}_i} |\Delta_i(z)| \in [0,1]. \tag{2}$$

Given a declared tolerance τ, the report is SENSITIVE when a simultaneous lower bound exceeds τ, INSENSITIVE when a simultaneous upper bound is at most τ, and UNRESOLVED otherwise. "Insensitive" means only $M_i \le \tau$ under the access and smoothness contract; it does not mean nonparent, irrelevant under another intervention, or safe for a policy.

Assumption 1 (confirmation contract). For candidate i, a finite grid $G_i = \{z_{ik}\}_{k=1}^{K_i}$ and tolerance τ are fixed before confirmation. At each grid point, n_{ik} independent pairs produce $D_{ikr} = C^+_{ikr} - C^-_{ikr} \in [-1,1]$ with $\mathbb{E}D_{ikr} = \Delta_i(z_{ik})$. The function Δ_i is L_i-Lipschitz under the declared metric, and G_i has covering radius $\rho_i = \sup_{z \in \mathcal{Z}_i} \min_{g \in G_i} d(z,g)$.

Independence is required across replicate pairs, not across grid points or candidates. The family of candidates and grids is predeclared. Pilot data may select a reference only when every final confirmation pair is drawn fresh and independently of the pilot. Unknown feasibility, unknown valid L_i, adaptive reuse, dependent trajectories, and partial coordinate control are outside this contract.

Adaptive procedures use the following stronger clause. Relative to the pre-query filtration $\mathcal{F}_{t-1}$, the next location and batch size are predictable. Given $\mathcal{F}_{t-1}$, the new observations have the product kernel associated with that now-fixed query: they are conditionally mutually independent, bounded, and each has conditional mean equal to the contrast at the queried location. No marginal independence between the adaptive query and the earlier transcript is asserted. The requirement is that no observation from an earlier batch is reused in this new product draw. Thus dependence across fixed grid points is allowed only for the nonadaptive certificate; the adaptive theorem requires fresh conditional product batches.

The local cost-mechanism factorization $C = f_C(P_C, A, U^C)$ can justify nuisance invariance when the true parents and noise law are unchanged, but it does not make P_C recoverable. Structural interpretation additionally needs causal sufficiency, path blocking, faithfulness or interventional-identifiability assumptions, and appropriate support Maathuis et al. (2009); Peters et al. (2016); Hauser & Bühlmann (2015). Hidden or unfixed paths may produce nonzero Δ_i, while direct effects may disappear on a finite grid. These identification limits are preserved throughout.

3 FIXED-TRACE AND ADAPTIVE WORST-CASE SENSITIVITY CERTIFICATES

Let $N = \sum_{i=1}^{p} K_i$, let $\overline{D}_{ik} = n_{ik}^{-1} \sum_r D_{ikr}$, and define simultaneous radii

$$r_{ik} = \sqrt{\frac{2\log(2N/\alpha)}{n_{ik}}}. \tag{3}$$

The earlier pointwise procedure rejects $\Delta_i(z_{ik}) = 0$ when $|\overline{D}_{ik}| > r_{ik}$, or applies Holm to the valid two-sided values $\min\{1, 2\exp(-n_{ik}\overline{D}^2_{ik}/2)\}$. Bonferroni and Holm control FWER over the

N point nulls. Point-null rejection, however, supplies no upper bound on M_i and cannot certify low sensitivity.

Define

$$\underline{M}_i = \max_k \big(|\overline{D}_{ik}| - r_{ik}\big)_+, \tag{4}$$

$$\overline{M}_i = \min\left\{1,\ \max_k \big(|\overline{D}_{ik}| + r_{ik}\big) + L_i\rho_i\right\}. \tag{5}$$

Theorem 1 (simultaneous worst-case sensitivity certificate). Under Assumption 1,

$$\Pr\big(\forall i:\ \underline{M}_i \le M_i \le \overline{M}_i\big) \ge 1 - \alpha. \tag{6}$$

Consequently, reporting SENSITIVE if $\underline{M}_i > \tau$, INSENSITIVE if $\overline{M}_i \le \tau$, and UNRESOLVED otherwise has family-wise probability at most α of making any wrong decisive report.

Proof. For $D_{ikr} \in [-1, 1]$, two-sided Hoeffding gives $\Pr(|\overline{D}_{ik} - \Delta_i(z_{ik})| > r_{ik}) \le \alpha/N$. A union bound yields an event of probability at least $1 - \alpha$ on which all grid errors are bounded. On this event, each $|\Delta_i(z_{ik})| \ge (|\overline{D}_{ik}| - r_{ik})_+$, proving the lower bound. For arbitrary z, choose z_{ik} within ρ_i. Lipschitz continuity gives $|\Delta_i(z)| \le |\Delta_i(z_{ik})| + L_i\rho_i \le |\overline{D}_{ik}| + r_{ik} + L_i\rho_i$; maximizing proves the upper bound. A decisive report cannot be wrong while M_i lies in its interval. □

Theorem 2 (identified width at the fixed zero trace). For the exact grid trace $\Delta|_G = 0$ and contrasts in $[-1, 1]$, the identified set of $M = \sup_z |\Delta(z)|$ over the L-Lipschitz class is exactly $[0, \min\{L\rho, 1\}]$. Thus any interval rule valid for every extension of this fixed trace must contain both endpoints. This is a fixed-trace identification claim, not sequential minimax optimality and not pointwise optimality of the coarse $L\rho$ allowance for a nonzero trace.

Proof. The zero function attains the lower endpoint. The function $\Delta_1(z) = \min\{L\, d(z, G), 1\}$ is bounded, L-Lipschitz, agrees with the zero trace, and attains $\min\{L\rho, 1\}$ because compactness makes $d(\cdot, G)$ attain ρ. Scaling Δ_1 by any factor in $[0, 1]$ attains every intermediate value. □

Theorem 2A (exact fixed-trace Lipschitz identification).

Let $(\mathcal{Z}, d)$ be compact, $G = \{g_k\}_{k=1}^K$ with $K \ge 1$, and $I_k = [a_k, b_k] \subseteq [-1, 1]$. Let $\mathcal{F}(I)$ contain all L-Lipschitz $f : \mathcal{Z} \to [-1, 1]$ satisfying $f(g_k) \in I_k$. This class is nonempty exactly when $a_j \le b_k + Ld(g_j, g_k)$ for every j, k. When it is nonempty, define

$$m_I = \max_k \operatorname{dist}(0, I_k),$$

$$U_I = \max_{z\in\mathcal{Z}} \max\left\{\min\left(1, \min_k[b_k + Ld(z, g_k)]\right), \min\left(1, \min_k[-a_k + Ld(z, g_k)]\right)\right\}. \tag{7}$$

Then the identified set is exactly

$$\left\{\sup_{z\in\mathcal{Z}} |f(z)| : f \in \mathcal{F}(I)\right\} = [m_I, U_I]. \tag{8}$$

Both endpoints and every value between them are attainable. For Theorem 1's clipped intervals $I_{ik} = [\max\{-1, \overline{D}_{ik} - r_{ik}\}, \min\{1, \overline{D}_{ik} + r_{ik}\}]$, $m_{I_i} = \underline{M}_i$. Replacing $\overline{M}_i$ by U_{I_i} when the intervals are feasible and returning UNRESOLVED otherwise preserves simultaneous coverage and never enlarges the original interval. For exact grid values this is the sharp trace-specific envelope; at the zero trace it equals $\min\{L\rho, 1\}$ and recovers Theorem 2. The complete proof is in Appendix Q.

The proof uses the classical real-valued McShane–Whitney extremal extensions (McShane, 1934; Whitney, 1934); it does not claim a new extension theorem. The delta is the exact identified interval of the sup-norm functional under a fixed feasible interval trace and the range cap, plus its simultaneous-confidence refinement of Theorem 1.

Auxiliary consequences and budget bracket. Appendix D gives the variance-adaptive and resolution corollaries, the allocation-free lower bound (Theorem 3), the family-wise slot scan (Theorem 4), the adaptive fresh-batch construction (Theorem 4A), and the unknown-geometry comparison (Theorem 4B). The details retain the adverse boundaries: empirical-Bernstein radii are not

uniformly smaller; the $\Theta(L/\gamma^3)$ reading needs interiority; and the expected- spend comparison in Theorem 4B assumes growth conditions that are not proved for arbitrary ambiguity geometry. The next two theorems are the paper's principal adaptive lower/upper pair.

Theorem 4C (unconditional cap-packing lower bound). Fix an insensitive instance Δ_0 with $M(\Delta_0) \le \tau - \gamma$. For $z \in [-1,1]$ let $s_z = \operatorname{sign}\Delta_0(z)$ (with $s_z = 1$ if $\Delta_0(z) = 0$), $T_z(z') = \tau + \gamma - L|z' - z|$, and define the sign-symmetric *cap*

$$\Delta^{(z)}(z') = s_z \max\{s_z\Delta_0(z'), T_z(z')\}.$$

Let its support be $S_z = \{z' : \Delta^{(z)}(z') \neq \Delta_0(z')\}$, and its divergence $\kappa_z = \sup_{z'\in S_z} \mathrm{kl}\big(\frac{1+\Delta_0(z')}{2} \,\big\|\, \frac{1+\Delta^{(z)}(z')}{2}\big)$. Every cap is L-Lipschitz, $[-1,1]$-valued, equals Δ_0 off S_z, and has $M(\Delta^{(z)}) \ge \tau + \gamma$; no location or flippability assumption is needed. Call $Z \subset [-1,1]$ a cap packing if its supports are pairwise disjoint and set $Q(\Delta_0) = \sup_Z \sum_{z\in Z} \kappa_z^{-1}$. Every adaptive three-way procedure meeting Theorem 3's error criterion satisfies

$$\mathbb{E}_0 T \;\ge\; \log\big(1/(2\epsilon)\big)\, Q(\Delta_0).$$

With $h_z = \tau + \gamma - |\Delta_0(z)|$, $|\Delta^{(z)} - \Delta_0| \le h_z$ on S_z; if all contrasts lie in $[-(1-c), 1-c]$ then $\kappa_z \le h_z^2/(c(2-c))$, so $\mathbb{E}_0 T \ge c(2-c)\log(1/(2\epsilon))\sum_{z\in Z} h_z^{-2}$ for every cap packing Z. When Theorem 4B's alternatives are caps with disjoint supports, Theorem 4C also implies its common-divergence lower bound.

Theorem 4D (doubling-batch branch-and-bound). Assume $L > \gamma/4$ and run Theorem 4A's tree with $D = \lceil \log_2(4L/\gamma)\rceil \ge 1$; every opened node receives fresh batches with predeclared cumulative sizes $n_0, 2n_0, \dots, 2^{K-1}n_0$, $r(N) = \sqrt{2\log(2pHK/\alpha)/N}$, and $2^{K-1}n_0 \ge 32\log(2pHK/\alpha)/\gamma^2$. After each opening or doubling, prune and stop as in Theorem 4A; otherwise select the active unfrozen cell with the largest upper endpoint and *split* it if it is not a leaf and $r(N) \le L\,\mathrm{rad}(I)$; otherwise *double* it, unless it has already received stage K, in which case *freeze* it (abstain when no unfrozen active cell remains). The procedure halts after at most HK selections and never spends more than $T_{\max} = H2^{K-1}n_0$. All reports over the p candidates are simultaneously correct with probability at least $1-\alpha$; on that event every instance with $|M - \tau| \ge \gamma$ is resolved, and an insensitive instance with local margin $m(z) = \tau - |\Delta_0(z)| \ge \gamma$ is resolved with

$$T \;\le\; \log(2pHK/\alpha)\left(CL\int_{-1}^{1} m(z)^{-3}\,dz + \frac{36}{m(0)^2}\right) + n_0 H, \qquad C \le 3.4\cdot 10^3,$$

hence $\mathbb{E}_0 T$ is at most this bound plus $\alpha T_{\max}$. The fixed-batch tree of Theorem 4A spends $n = \Theta(\log/\gamma^2)$ on every opened node, so it adapts only through the number of opened nodes, never through per-node spend. If $L \le \gamma/4$ the tree is a single leaf and $T \le 64\log(2pK/\alpha)/m(0)^2 + n_0 \le 256\log(2pK/\alpha)Q(\Delta_0)/(c(2-c)) + n_0$ directly.

Conditional comparison for Theorems 4C–4D. Let $\log = \log(2pHK/\alpha)$ and $C \le 3.4\cdot 10^3$. The comparison requires (C1) $\alpha < 1/6$; (C2) $L > \gamma/4$; (C3) an insensitive base with $m \ge \gamma$ that is $(L/16)$-Lipschitz; (C4) base and cap contrasts in $[-(1-c), 1-c]$; and (C5) $2^{K-1}n_0 \ge 32\log/\gamma^2$, $n_0 \le C\gamma\log/16$, and $\alpha 2^{K-1}n_0 \le C\gamma\log/16$. Under these conditions,

$$Q(\Delta_0) \ge \frac{c(2-c)L}{288}\int m^{-3}, \qquad \mathbb{E}_0 T \le \frac{(576C+144)\log}{c(2-c)} Q(\Delta_0).$$

The last two schedule inequalities can coexist with the first only if $\alpha \le C\gamma^3/512$. If they fail, $n_0 H + \alpha T_{\max}$ remains unabsorbed and no expected-spend match is claimed; if (C3)–(C4) fail, both functionals remain valid but their equivalence is not claimed. The complete proof, compatible point, loose numerical ratio, and adverse audit are in Appendix G.

Theorem 5 (metric-entropy form). The one-dimensional bracket has a metric-entropy form on a compact metric space whose necessary and sufficient sides are stated at different levels (per candidate versus family-wise) and share an order only on $[-1,1]^d$; the statement and its explicit ratios are in Appendix I.

Theorem 6 (dimension-explicit instance law). Let $\mathcal{Z} = [-1,1]^d$ carry the max metric, $d \ge 1$, and read each $|z' - z|$ in Theorem 4C as $\|z' - z\|_\infty$. Write $m(z) = \tau - |\Delta_0(z)|$, $h_z = m(z) + \gamma$, and $J = \int_{\mathcal{Z}} m(z)^{-(d+2)}dz$. Parts (a) and (d) use (C3)–(C4); part (c) does not:

(a) $Q(\Delta_0) \geq c(2-c)L^d J/(9 \cdot 2^{5d})$;

(b) the dyadic tree with 2^d children per node, $D = \lceil \log_2(4L/\gamma) \rceil$ and $H = (2^{d(D+1)} - 1)/(2^d - 1)$ retains Theorem 4D's validity, halting and resolution, and on its confidence event an insensitive instance obeys $T \leq \log(2pHK/\alpha)\big(C_d L^d J + 36/m(0)^2\big) + n_0 H$ with $C_d = 2^{3d+10}/(2^{d+2} - 1) + 2^{d+6} + 2^{3d+8}$;

(c) without (C3)–(C4), every cap satisfies $\kappa_z \geq h_z^2/2$ and $\{z' : \|z' - z\|_\infty < h_z/(2L)\} \subseteq S_z$; consequently $Q(\Delta_0) \leq 2(3/2)^{d+2} \max\{2L, 1\}^d J$;

(d) under (C1)–(C4) and the dimensional schedule condition (C5$_d$), $\mathbb{E}_0 T \leq (576 \cdot 32^{d-1} C_d + 144) \log Q(\Delta_0)/(c(2-c))$.

At $d = 1$ the lower denominator is 288 and the comparison coefficient is $576C + 144$. Parts (a) and (c) give $Q(\Delta_0) = \Theta_d(L^d J)$ when (C3)–(C4) hold and $L \geq 1/2$. The upper converse alone needs neither C3 nor C4 and retains the factor $\max\{2L, 1\}^d$; it does not establish a two-sided comparison. The integral is an explicit oracle functional, not a data-computable certificate for an unknown surface. The proof and its numerical audit are in Appendix R.

Proposition 1 (bounded finite-grid blindness). For any finite references, let $h(z) = \prod_k \|z - z^k\|^2$, let $\sigma(t) = (1 + e^{-t})^{-1}$ be the logistic link, which maps $\mathbb{R}$ into $(0, 1)$ and is strictly increasing, and generate bounded cost as $C \sim \text{Bernoulli}(\sigma(X_i h(z)))$. Because σ is strictly increasing, $\sigma(h(z)) \neq \sigma(-h(z))$ whenever $h(z) > 0$, so the mechanism depends on X_i there; but at every grid point $h(z^k) = 0$, both values collapse to $\sigma(0) = 1/2$, and the tested contrast is exactly zero. Thus finite-grid blindness persists inside the bounded-cost access model. Theorem 1 certifies M_i only under the Lipschitz contract; it does not identify parents.

4 Empirical scope and conclusion

The synthetic evaluation compares rules under the same known smoothness constants, tolerance, pair budgets, and per-seed family error level. At a cap of 1,048,576 pairs, Lipschitz placement resolves 1,155 of 1,200 candidates, versus 1,037 for the stagewise rule and 926 for the fixed net. The stagewise rule uses 309,842 pairs per candidate on average; at matched average spend, Lipschitz placement still resolves more candidates (1,093). These aggregate comparisons do not establish a universal allocation advantage. The full curves, seed-level accounting, and scope of the declared evaluation split are in Appendix A.

The access assumptions are substantive: on 100 between-grid tents, declaring one quarter of the valid Lipschitz constant produces eight false-low reports, whereas the valid constant yields abstention on all 100. No empirical check here learns or certifies that constant. The contribution is exact fixed-trace identification and a conditional adaptive query-cost bracket under the stated oracle, not causal-parent identification, policy safety, or external deployment validation. The synthetic bank and the aggregate multidimensional theorem audit do not establish transport to measured intervention data.

REFERENCES

François Bachoc, Tommaso Cesari, and Sébastien Gerchinovitz. The sample complexity of level set approximation. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *PMLR*, 2021a. URL https://proceedings.mlr.press/v130/bachoc21a.html.

François Bachoc, Tommaso R. Cesari, and Sébastien Gerchinovitz. Instance-dependent bounds for zeroth-order lipschitz optimization with error certificates. In *Advances in Neural Information Processing Systems*, volume 34, 2021b. URL https://arxiv.org/abs/2102.01977.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997. doi: 10.1080/01621459.1997.10474074.

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12:1655–1695, 2011.

Étienne de Montbrun and Sébastien Gerchinovitz. Certified multifidelity zeroth-order optimization. *SIAM/ASA Journal on Uncertainty Quantification*, 2024. doi: 10.1137/23M1591086.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7:1079–1105, 2006.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory*, volume 49 of *PMLR*, pp. 998–1027, 2016. URL https://proceedings.mlr.press/v49/garivier16a.html.

Alkis Gotovos, Nathalie Casati, Gregory Hitz, and Andreas Krause. Active learning for level set estimation. In *Proceedings of IJCAI*, pp. 1344–1350, 2013.

Alain Hauser and Peter Bühlmann. Jointly interventional and observational data: Estimation of interventional markov equivalence classes of directed acyclic graphs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):291–318, 2015. doi: 10.1111/rssb.12071.

Donald R. Jones, Cary D. Perttunen, and Bruce E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, 1993.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020. doi: 10.1017/9781108571401.

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *PMLR*, pp. 1690–1698, 2016. URL https://proceedings.mlr.press/v48/locatelli16.html.

Marloes H. Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A), 2009. doi: 10.1214/09-AOS685.

Stefan Magureanu, Richard Combes, and Alexandre Proutière. Lipschitz bandits: Regret lower bounds and optimal algorithms. In *Proceedings of the 27th Conference on Learning Theory*, volume 35 of *PMLR*, pp. 975–999, 2014. URL https://proceedings.mlr.press/v35/magureanu14.html.

Cédric Malherbe and Nicolas Vayatis. Global optimization of lipschitz functions. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 2314–2323, 2017.

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. In *Proceedings of the Conference on Learning Theory*, 2009.

E. J. McShane. Extension of range of functions. *Bulletin of the American Mathematical Society*, 40 (12):837–842, 1934. doi: 10.1090/S0002-9904-1934-05978-0.

Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):947–1012, 2016. doi: 10.1111/rssb.12167.

Elie Tamer. Partial identification in econometrics. *Annual Review of Economics*, 2:167–195, 2010. doi: 10.1146/annurev.economics.050708.143401.

Hassler Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934. doi: 10.1090/S0002-9947-1934-1501735-3.

Table 1: Equal-data decision curve on the predeclared 1,200-function synthetic bank. Each entry is (insensitive-certified)/(sensitive-certified) out of 1,200 candidates; $0/0$ means the rule certified nothing in either direction at that budget. Wrong-decision counts, both false-safe and false-sensitive, are zero throughout and are therefore not shown; unresolved cases remain unlisted, and neither the counts nor the generator-specific zero-event bound estimate accuracy on broader mechanisms.

Pairs/candidate	Hoeffding	Empirical Bernstein	Point band
2,048	0/64	0/0	0/0
4,096	1/289	0/229	0/229
8,192	277/299	101/296	0/296
16,384	300/300	300/300	0/300

APPENDIX

A FOCUSED EVALUATION

Fixed function bank and declared evaluation split. The local protocol declares seeds 0–99 for method development and the original curve, and disjoint seeds 200–299 for the matched five-method rerun. The manuscript verifies the disjoint identifiers but cannot independently certify that the second range was untouched during development. No statement here should be read as an externally certified holdout: "declared evaluation split" means this local protocol declaration and the disjointness of identifiers, nothing stronger. Each seed has 12 candidates. For each seed, three families (random Fourier, radial-basis, and piecewise-linear) are crossed with supremum targets $M \in \{.12, .34, .46, .75\}$ around tolerance $\tau = .40$. Fourier functions have three harmonics, coefficients sampled uniformly then divided by j^3, and phases uniform on $[-\pi, \pi]$. RBF functions sum three signed Gaussian features with centers uniform on $[-.9, .9]$ and widths on $[.4, .75]$. Piecewise-linear functions use seven equally spaced anchors with values uniform on $[-1, 1]$. The deterministic key is $810000 + 100s + 7f + t$, for seed s, family index f, and target index t.

Functions are normalized on a 2,001-point mesh. Analytical derivative or segment bounds give L_i; below-threshold truth is certified by mesh maximum plus $0.0005L_i$, while above-threshold truth has a mesh witness. At reference z, the paired Bernoulli means are $(1 + \Delta(z))/2$ and $(1 - \Delta(z))/2$, yielding $D \in [-1, 1]$ with expectation $\Delta(z)$. The fixed-grid and allocation baselines use the same 17 uniform references. Every method uses the same candidate family, tolerance, pair budget, and per-seed 12-candidate family error level .05; Appendix E gives the formal error accounting and rendered decision/error ledgers.

Decision curve and fair comparator. Table 1 reports the complete equal-data curve. Hoeffding decisions increase from 64/1,200 at 2,048 pairs to 600/1,200 at 16,384. Empirical Bernstein makes no decision at 2,048, remains more conservative at 4,096 and 8,192, and ties Hoeffding only at 16,384. Its finite-n linear term therefore outweighs variance adaptation in the intermediate regime. This is a negative comparator result. The no-smoothness point band certifies some high cases but, correctly, no low case: pointwise uncertainty alone cannot upper-bound the off-grid supremum.

No rule makes an observed wrong decision on this fixed bank. Rather than report a zero-width bootstrap wrong-rate interval, we use the exact one-sided binomial bound for zero wrong seed events: $1 - .05^{1/100} = .0295$. The sampling model is the generator itself: each of the 100 seeds draws that seed's Fourier coefficients and phases, RBF weights, centres and widths, or piecewise-linear anchor values from the declared distributions through an independent stream, so the 100 seed-level wrong-decision indicators are independent draws from this generator. The bound therefore concerns this specified generator only. It neither estimates error on a broader mechanism population nor replaces the simultaneous analytic guarantee.

Adaptive and Lipschitz-optimization comparators. Against the uniform 17-point net we ran successive-elimination allocation with a union bound over points, rounds, and attainable cumulative counts (Even-Dar et al., 2006), Lipschitz-envelope placement in the LIPO/DIRECT family (Malherbe & Vayatis, 2017; Jones et al., 1993), level-set estimation (Gotovos et al., 2013), slot scan, and Theorem 4A's stagewise branch-and-bound, all on the declared evaluation seeds 200–299 with the

Pair budget	Fixed	Successive	Lipschitz	Level set	Slot scan	Stagewise
16,384	600	578	600	599	**603**	591
65,536	646	606	**731**	669	663	630
262,144	870	874	**1,078**	1,069	838	788
1,048,576	926	933	**1,155**	1,148	1,056	1,037

Table 2: Decisive candidates on declared evaluation seeds 200–299 under equal pair budgets and one per-seed, 12-candidate family-wise $\alpha = .05$ scheme. Every upper endpoint bounds the continuous supremum, including the $L/400$ off-mesh allowance for the two envelope rules, and successive elimination additionally unions over every attainable cumulative count. Stagewise uses a complete-tree union bound and fresh predictable node batches. Every method makes zero observed wrong decisions. The slot scan's headline gain is three decisions; it is dominated at every larger audited budget.

same labels, threshold, budget, and per-seed family error level; both envelope rules carry the $L/400$ allowance that bounds the continuous supremum rather than a mesh maximum. At the headline budget the six counts are 600, 578, 600, 599, 603, and 591 decisions out of 1,200 with zero observed wrong decisions at every budget. With $N = 204$ grid contrasts, $\alpha = .05$ and $n \approx 963$ pairs per reference, Eq. (3) gives $r = .137$, more than twice the $.06$ distance of the two near-boundary families from the tolerance, so the fixed net cannot resolve them at this budget and its 600 decisions are exactly the two far-boundary families.

The fixed-net radius constrains that method only, not every allocation rule. Theorem 3's lower bound holds for every adaptive procedure but never exceeds 4,708 pairs over the 600 near-threshold development mechanisms (medians 1,046 at $M = .34$ and 1,465 at $M = .46$) against a headline of 16,384, so the clustering of fixed net, Lipschitz placement, and level set at 599–600 is a property of the allocation rules, not an information-theoretic barrier; the slot scan resolves 603, all three extra decisions in the $M = .46$ family (Appendix E). The radius crosses the $.06$ margin between 65,536 and 262,144 pairs, where Table 2 shows separation and then domination; Lipschitz placement leads at every larger budget, reaching 1,155 against 926 for the fixed net at 1,048,576 pairs. Under the anytime-valid successive-elimination radii of Appendix E it resolves 606 and trails the fixed net. These budgets show that simultaneous validity under known L, rather than allocation efficiency, is the fixed certificate's contribution.

The sufficiency side, corrected and executed. At budget N, slot scan picks k to minimise $2\sqrt{2\log(2kp/\alpha)/\lfloor N/k \rfloor} + L/k$ for $p = 12$, queries the k centres, and applies the same three-way rule. Under the same per-seed FWER as every baseline, the declared-split counts are 603, 663, 838, and 1,056, with zero observed wrong decisions. At 16,384 pairs, slot scan gains three decisions over the fixed net, all of them in the near-threshold family $M = .46$. At every larger audited budget it is dominated. Thus the valid result is not a material allocation advantage; it is an executable family-wise upper-bound companion to Theorem 3.

Adaptive fresh-batch certificate. Theorem 4A predeclares a complete dyadic tree, opens only cells whose valid Lipschitz envelope can still matter, and draws a fresh fixed batch whenever a node opens. Pruned subtrees release their reserved budget to later fresh batches; no cumulative mean is retested at an observation-dependent count. The executable stores the resulting heterogeneous n_v, r_v pairs and uses Theorem 4A's minimum envelope $U(I)$ verbatim; Appendix E gives the formula and exact pruning predicates. The recycled-cap specialization now also has a finite-budget theorem: every opened node receives at least the complete-tree baseline batch, so fixed-batch resolution and the deterministic ambiguity-node bound carry over while total spend remains below the cap. At caps 16,384, 65,536, 262,144, and 1,048,576 it makes respectively 591, 630, 788, and 1,037 decisions with zero observed wrong decisions, using 11,804, 38,175, 111,542, and 309,842 pairs per candidate on average. At those matched average spends the fixed net makes 595, 603, 815, and 875 decisions, while Lipschitz placement makes 600, 607, 949, and 1,093. These matched-mean-spend comparator counts are aggregate-only and have no rendered per-seed ledger. Thus stagewise allocation beats both matched-spend comparators at the 65,536 cap and beats the fixed net by 162 decisions at the largest cap while using 29.5% of that cap on average; it does not dominate Lipschitz placement and

loses at the other two matched spends. These are measured efficiency differences, not a minimax constant claim.

Deterministic ambiguity-geometry audit. The rendered manifest enumerates 54 functions crossing $M \in \{.34, .46\}$ with peak counts $1, 2, 4$, widths $.08, .16, .32$, and shifts $-.15, 0, .15$ (`f17_ambiguity_manifest.json`, SHA-256 `e4dc0438...`), each run with the common valid $L = 4.25$, the fixed-fresh-batch specialization of Theorem 4A at a 1,048,576 cap, and 50 deterministic replicates. Across 2,700 trials there are zero wrong decisions and mean spend is 42.9% of the cap. On the insensitive half the trace-independent node bound correlates .993 with measured opened nodes and holds in all trials (widths $.08, .16, .32$ give $69.94, 76.49, 90.02$ opened nodes); on the sensitive half wider regions are easier to witness ($254, 336, 415$ decisions of 450). This validates the instance-dependent geometry on a synthetic factor grid, not on an external domain.

Assumption controls. The empirical-Bernstein radius is capped at 2, the maximum possible $|\overline{D} - \mathbb{E}D|$: exact enumeration for $n \in \{1, 2, 3, 5, 10, 20, 50\}$ over 1,001 Bernoulli parameters gives worst failure .499 for the invalid cap 1 versus 3.46×10^{-14} for cap 2 at per-contrast $\alpha = 2.45 \times 10^{-4}$ (the $n = 1$ row uses the cap itself, since Corollary 1 needs $n > 1$); this is a regression audit, not a distribution-free proof. No robustness range for a misdeclared L is claimed: on 100 bounded between-grid tents with $M = .46$, declaring $.25L$ makes 8 false-low decisions while valid L abstains on all 100, so nothing here validates L from data. Appendix E renders the complete per-seed ledgers; local provenance commands remain in the released source rather than the PDF, and no external or private data are used. Superseded implementation details are consolidated in Appendix S.

B External Evidence Status

This submission reports no measured-data or external-artifact result: Assumption 1 simultaneously requires a compact candidate-specific domain, paired joint interventions fixing every other coordinate, and an independently valid Lipschitz constant, and no external surface used here supplies all three, so the synthetic banks of Appendix A are the entire empirical basis and no transfer, deployment, controller, or policy-safety claim is made (Appendix L).

C Limitations and Conclusion

Under bounded independent joint-intervention pairs, predeclared nets, and valid known Lipschitz constants, all M_i intervals cover simultaneously with probability at least $1 - \alpha$, decisive reports have family-wise error control, and the fixed-grid ambiguity is $\min\{L_i \rho_i, 1\}$. The upper certificate fails under understated L_i, an incorrect domain or cover, unsupported dependence, or incomplete coordinate control; none is learned or repaired here. Covering growth limits high dimensions. The retained two-dimensional execution is an aggregate theorem audit, not a competitive evaluation; neither it nor the one-dimensional synthetic bank establishes transport. Hidden paths may create sensitivity and direct parents may remain invisible. A lower endpoint can certify material sensitivity; only a valid smoothness-aware upper endpoint can certify low worst-case sensitivity, and the adaptive doubling tree attains that certificate at a cost matching the cap-packing lower bound up to a logarithm and an explicit, loose constant on $(L/16)$-Lipschitz instances with interior contrasts, $\alpha < 1/6$, and an absorbable batch schedule; outside those listed conditions we claim only that both bounds are valid. This is a finite-sample controlled-sensitivity certificate with abstention, not a generic confidence band, causal-discovery algorithm, or safety certificate.

Data Statement

No human subjects or private data are used. The entire empirical basis is the synthetic function bank specified in Appendix A and Appendix E; no external or public benchmark environment is evaluated.

D AUXILIARY CERTIFICATE AND BUDGET RESULTS

These results support the main Theorems 1–2A and 4C–4D. They are placed here to keep the main narrative on the identified interval and the cap-packing/margin- integral pair. Their hypotheses, adverse cases, and proofs are unchanged.

Corollary 1 (variance-adaptive simultaneous certificate). Let S_{ik}^2 be the unbiased sample variance of the bounded differences and $n_{ik} > 1$. Replacing r_{ik} by the clipped empirical-Bernstein radius

$$r_{ik}^{\mathrm{EB}} = \min\left\{2, \sqrt{\frac{2S_{ik}^2 \log(4N/\alpha)}{n_{ik}}} + \frac{14\log(4N/\alpha)}{3(n_{ik}-1)}\right\} \tag{9}$$

in Eqs. (4)–(5) preserves the simultaneous $1-\alpha$ coverage and three-way decision guarantee.

Justification. Apply the two-sided empirical-Bernstein bound of Maurer & Pontil (2009) to $(D+1)/2 \in [0,1]$, rescale by two, and union-bound the N grid contrasts. The cap at 2, not 1, is coverage-preserving because both the sample mean and expectation lie in $[-1,1]$, so their maximum possible deviation is 2. □

This comparator is variance-adaptive but not uniformly tighter: its finite-n linear term can exceed Hoeffding's radius. Capping its deviation radius at 1 would be invalid, even though the target magnitude M_i itself is at most 1.

Corollary 2 (resolution and sample separation). On the simultaneous coverage event of Theorem 1, if $M_i \geq \gamma_i > L_i\rho_i$, some grid point has contrast magnitude at least $\gamma_i - L_i\rho_i$. It is certified sensitive whenever $\gamma_i - L_i\rho_i > \tau + 2\max_k r_{ik}$. Conversely, if $M_i + 2\max_k r_{ik} + L_i\rho_i \leq \tau$, it is certified insensitive. Thus sample size, multiplicity, and reference resolution enter both sides of the decision rather than only detection power.

Theorem 3 (allocation-free budget lower bound). Fix a tolerance τ and a margin $\gamma \in (0,\tau)$ with $\tau+\gamma<1$, and let $L \geq 2\gamma$. Consider any procedure that may query any $z \in [-1,1]$, adaptively and off the grid, observes $D \in \{-1,+1\}$ with $\mathbb{E}[D \mid z] = \Delta(z)$, spends n pairs in total, and reports SENSITIVE, INSENSITIVE, or UNRESOLVED. Write α_{err} for a bound on wrong decisive reports over every L-Lipschitz $\Delta : [-1,1] \to [-1,1]$. Require the failure-to-resolve probability to be at most α_{abs} only on the margin-separated subclass $|M(\Delta)-\tau| \geq \gamma$, and put $\epsilon = 2\alpha_{\mathrm{err}} + \alpha_{\mathrm{abs}}$. If $\epsilon < 1/2$ then

$$n \;\geq\; m\,\frac{\log\big(1/(2\epsilon)\big)}{\mathrm{kl}\Big(\frac{1+\tau-\gamma}{2}\,\Big\|\,\frac{1+\tau+\gamma}{2}\Big)}, \qquad m = \Big\lfloor \frac{L}{2\gamma} \Big\rfloor, \tag{10}$$

where kl is the Bernoulli Kullback–Leibler divergence. The bound holds throughout $\tau+\gamma<1$, but it is informative only away from the range boundary: as $\tau+\gamma \uparrow 1$ the divergence diverges and the right-hand side falls to zero. At equality the displayed KL denominator is infinite, so we make no finite lower-bound claim there. Stating the condition explicitly, if $\tau+\gamma \leq 1-c$ for a fixed $c>0$ then $\mathrm{kl} = \Theta\big(\gamma^2/(1-(\tau+\gamma)^2)\big)$ and the necessary budget is $\Theta(L/\gamma^3)$ with constants controlled by c. At $\tau = .40$, $\gamma = .06$, $\mathrm{kl}/\gamma^2 = 2.43$, against 57.06 at $\tau+\gamma = .99$. Every rate statement below assumes this interiority; without it only the displayed inequality, not the rate, is claimed.

Proof. Let $\Delta_0 \equiv \tau - \gamma$. Choose m intervals in $[-1,1]$ with disjoint interiors, each of width $4\gamma/L$ and let Δ_j add to Δ_0 the tent of height 2γ and slope L supported on the j-th interval. Each Δ_j is L-Lipschitz with values in $[-1,1]$, and $M(\Delta_0) = \tau-\gamma < \tau$ while $M(\Delta_j) = \tau+\gamma > \tau$. Let A be the event that the report is INSENSITIVE. Under Δ_0 that report is correct, so $P_0(A^c) \leq \alpha_{\mathrm{err}} + \alpha_{\mathrm{abs}}$: the report is either the wrong decisive one or no decision at all. Under Δ_j the same report is the wrong decisive one, so $P_j(A) \leq \alpha_{\mathrm{err}}$. Adding,

$$P_0(A^c) + P_j(A) \;\leq\; 2\alpha_{\mathrm{err}} + \alpha_{\mathrm{abs}} \;=\; \epsilon.$$

Bretagnolle–Huber (Lattimore & Szepesvári, 2020, Thm. 14.2) bounds the same sum below by $\frac{1}{2}\exp(-\mathrm{KL}(P_0\|P_j))$, so $\mathrm{KL}(P_0\|P_j) \;\geq\; \log(1/(2\epsilon))$. The two laws differ only on the j-th interval, so the divergence decomposition for the realized adaptive query sequence (Lattimore & Szepesvári, 2020, Lem. 15.1) gives $\mathrm{KL}(P_0\|P_j) \;=\; \mathbb{E}_0\big[\sum_{t\leq n} \mathrm{kl}\Big(\frac{1+\Delta_0(z_t)}{2}\,\Big\|\,\frac{1+\Delta_j(z_t)}{2}\Big)\big] \;\leq$

$\mathbb{E}_0[N_j]\,\mathrm{kl}\big(\frac{1+\tau-\gamma}{2}\|\frac{1+\tau+\gamma}{2}\big)$, where N_j counts queries in the region where the j-th tent is strictly positive, the summand vanishes outside that region, and the divergence is maximized at the tent peak: with $p=(1+\tau-\gamma)/2$ fixed and $q\in[p,(1+\tau+\gamma)/2]$, $\partial_q\mathrm{kl}(p\|q)=(q-p)/(q(1-q))\ge 0$. The intervals are disjoint in their positive tent regions, so $\sum_{j\le m}\mathbb{E}_0[N_j]\le n$, and averaging the m inequalities gives the display. □

Theorem 3 is a necessary-budget statement over the declared class. The next result supplies the sufficient side for this margin-separated resolution requirement, with family-wise multiplicity in its upper bound, so the two together bracket the budget rather than leaving a one-sided claim.

Theorem 4 (family-wise slot-scan sufficiency). Fix $L>0$, $\alpha\in(0,1)$, a target margin $\gamma>0$, and p candidates. For each candidate take $k=\max\{1,\lceil 4L/\gamma\rceil\}$ slot centres $z_j=-1+(2j-1)/k$, $j=1,\dots,k$, query each centre $n=\lceil 32\log(2kp/\alpha)/\gamma^2\rceil$ times, and set $r=\sqrt{2\log(2kp/\alpha)/n}$, the same range-2 radius as Eq. (3), $\widehat{M}_{\mathrm{lo}}=\max_j(|\overline{D}_j|-r)_+$ and $\widehat{M}_{\mathrm{hi}}=\min\{1,\max_j|\overline{D}_j|+r+L/k\}$. Report SENSITIVE if $\widehat{M}_{\mathrm{lo}}>\tau$, INSENSITIVE if $\widehat{M}_{\mathrm{hi}}\le\tau$, and UNRESOLVED otherwise. Then with family-wise probability at least $1-\alpha$ every interval covers its M_i, so no decisive report is wrong, and every instance with $|M_i-\tau|\ge\gamma$ is resolved, boundary cases included. The per-candidate budget is $kn=O\big(K_\gamma[1+\gamma^{-2}\log(2pK_\gamma/\alpha)]\big)$, where $K_\gamma=\max\{1,L/\gamma\}$. This retains the single-slot cost and positive logarithm when L/γ is small.

Proof. Hoeffding and a union bound over the kp centres put every $|\overline{D}_j-\Delta(z_j)|$ below r with probability at least $1-\alpha$; on that event $\widehat{M}_{\mathrm{lo}}\le\max_j|\Delta(z_j)|\le M$ and, because the centres have covering radius $1/k$ and Δ is L-Lipschitz, $M\le\max_j|\Delta(z_j)|+L/k\le\widehat{M}_{\mathrm{hi}}$. If $4L/\gamma\ge 1$, the ceiling in k gives $L/k\le\gamma/4$; otherwise $k=1$ and $L<\gamma/4$, so the same inequality holds. The ceiling in n gives $r\le\gamma/4$, so both endpoints lie within $L/k+2r\le 3\gamma/4$ of M. That slack is strict: if $M\ge\tau+\gamma$ then $\widehat{M}_{\mathrm{lo}}\ge M-3\gamma/4\ge\tau+\gamma/4>\tau$, and if $M\le\tau-\gamma$ then $\widehat{M}_{\mathrm{hi}}\le M+3\gamma/4\le\tau-\gamma/4\le\tau$. Both rounding operations only increase k and n, so the guarantee survives them. □

Theorem 4A (adaptive fresh-batch branch-and-bound). Under Assumption 1's adaptive filtration clause, fix a dyadic tree of depth D on $[-1,1]$ and let $H=2^{D+1}-1$ be its complete node count. Every possible node is predeclared. When an active cell is opened, its centre receives a fresh batch of n_v independent pairs, where n_v may depend on the preceding transcript but not on that batch. Set

$$r_v=\sqrt{2\log(2pH/\alpha)/n_v}.$$

For a current cell I, combine all opened centres by

$$U(I)=\min\left\{1,\min_v\left[|\overline{D}_v|+r_v+L\max_{z\in I}|z-z_v|\right]\right\}.$$

The global lower endpoint is $\underline{M}=\max_v(|\overline{D}_v|-r_v)_+$. The upper endpoint is the maximum of $\underline{M}$, the $U(I)$ values over active cells, and τ if any cell was previously pruned under $U(I)\le\tau$. Cells with $U(I)<\underline{M}$ may be pruned, as may cells with $U(I)\le\tau$ while unresolved; otherwise split a cell with largest upper endpoint. The usual three-way rule may stop after any opening. With probability at least $1-\alpha$, all reports over all p candidates are correct simultaneously, including every data-dependent early stop and budget recycling step.

For the fixed-fresh-batch specialization, set $D=\max\{0,\lceil\log_2(4L/\gamma)\rceil\}$ and give every node exactly $n=\lceil 32\log(2pH/\alpha)/\gamma^2\rceil$ pairs. Then every instance with $|M-\tau|\ge\gamma$ is resolved by the complete tree. Thus its worst-case budget is at most Hn and hence $O((L/\gamma^3)\log(Lp/(\gamma\alpha)))$ when $L>\gamma/4$, the same rate as Theorem 4; when $L\le\gamma/4$, the one-node bound is $O(\gamma^{-2}\log(p/\alpha))$. Pruning can reduce realized spend.

There is also a trace-independent instance bound. Let $\mathcal{P}_d$ be the depth-d dyadic cells, with centre c_I and radius $\mathrm{rad}(I)=2^{-d}$. For the fixed-batch radius r, define

$$A_d(\Delta)=\#\{I\in\mathcal{P}_d:|\Delta(c_I)|+2r+L\,\mathrm{rad}(I)>\tau\},\qquad H_\Delta=\min\{H,\,1+2\sum_{d=0}^{D-1}A_d(\Delta)\}.$$

On the simultaneous event, the algorithm opens at most H_Δ nodes and uses at most nH_Δ pairs. Unlike H, this deterministic quantity contracts when few parent cells can reach the threshold.

The executed recycled-cap specialization has a corresponding finite-budget guarantee. Let $B \geq H$, start the root with $n_0 = \lfloor B/H \rfloor$ pairs, and, before each split, let R be the complete number of unopened descendants that could still be reached from the active cells. Give each of the two new children $\lfloor B_{\mathrm{rem}}/R \rfloor$ fresh pairs. Then every opened node receives at least n_0 pairs, total spend is at most B, and the complete-tree union bound remains valid. If

$$n_0 \geq \left\lceil 32 \log(2pH/\alpha)/\gamma^2 \right\rceil,$$

every instance with $|M - \tau| \geq \gamma$ resolves and at most H_Δ nodes open, where H_Δ uses $r_0 = \sqrt{2\log(2pH/\alpha)/n_0}$. This does not claim the pair bound $n_0 H_\Delta$: recycling intentionally spends more than n_0 on some opened nodes, while the deterministic cap remains B.

Proof outline. Condition on each predictable fresh query and batch size, apply Hoeffding, and union-bound the pH potential node events. The retained lower endpoint and threshold cover pruned cells. Only parents counted by A_d can split; the recycling invariant $B_{\mathrm{rem}} \geq n_0 R$ gives the per-node minimum and deterministic total cap. The complete filtration, resolution and recycling proof is in Appendix O.

The role of $(A_d)_d$ is analogous to complexity measures in Lipschitz optimization and level-set estimation. Their packing exponents agree, while Theorem 4A replaces H by the threshold-specific H_Δ within the three-way criterion when the geometry is favourable.

The next two lower bounds retain the binary observation experiment of Theorem 3: at query z, $(Y+1)/2$ is Bernoulli with success probability $(1+\Delta(z))/2$. Their transcript-KL identities are not asserted for every bounded paired oracle; the upper bounds retain their stated bounded-observation model.

Theorem 4B (unknown-geometry local instance bound). Fix a base Δ_0 with $M(\Delta_0) \leq \tau - \gamma$ and margin γ. Call disjoint intervals $I_1, \ldots, I_q$ locally flippable if, for each j, the class contains an L-Lipschitz Δ_j equal to Δ_0 off I_j with $M(\Delta_j) \geq \tau + \gamma$. The locations are not revealed to the procedure. Let

$$\bar{\kappa} = \max_j \sup_{z \in I_j} \mathrm{kl}\left(\frac{1+\Delta_0(z)}{2} \,\middle\|\, \frac{1+\Delta_j(z)}{2} \right), \qquad \epsilon = 2\alpha_{\mathrm{err}} + \alpha_{\mathrm{abs}}.$$

Every adaptive three-way procedure satisfying Theorem 3's error criterion on $\Delta_0, \ldots, \Delta_q$ obeys

$$\mathbb{E}_0 T \geq q \, \frac{\log(1/(2\epsilon))}{\bar{\kappa}}.$$

For the fixed-batch procedure of Theorem 4A, write $\bar{H} = \max_{0 \leq j \leq q} H_{\Delta_j} \leq H$. Its bound $T \leq n H_{\Delta_j}$ holds on the simultaneous confidence event, not on every transcript; the deterministic cap is nH. Hence

$$\max_{0 \leq j \leq q} \mathbb{E}_j T \leq n\big((1-\alpha)\bar{H} + \alpha H\big) \leq n(\bar{H} + \alpha H).$$

For an expected-spend comparison take $\alpha_{\mathrm{err}} = \alpha_{\mathrm{abs}} = \alpha < 1/6$, so $\epsilon = 3\alpha$. Assume, as hypotheses of this comparison rather than as properties established here, the growth conditions (G1) $\bar{H} \leq Cq \log H$ and (G2) $\alpha H \leq C_{\mathrm{fail}} q \log H$ with fixed constants C, C_{fail}, together with the regularity conditions (G3) all contrasts lie in $[-(1-c), 1-c]$ for fixed $c > 0$ and (G4) $\max_j \|\Delta_j - \Delta_0\|_\infty \leq C_\Delta \gamma$ with fixed C_Δ; then (G3)–(G4) give $\bar{\kappa} \leq C_\Delta^2 \gamma^2 / [c(2-c)]$. With $n = \lceil 32 \log(2pH/\alpha)/\gamma^2 \rceil$, these assumptions give matching $q\gamma^{-2}$ dependence in *expected* spend, up to tree-depth and confidence logarithms. Neither (G1) nor (G2) is established here for the ambiguity geometry of an arbitrary locally flippable family; both are assumed. If the failure-event condition (G2) fails, the term $\alpha n H$ remains unabsorbed and no expected-spend match is claimed. The perturbation locations are not revealed to the algorithm. The proof is in Appendix F.

E SELF-CONTAINED FAMILY-WISE EXPERIMENT SPECIFICATION

Generator and declared local split. The local protocol assigns seeds 0–99 to development and seeds 200–299 to the comparison. Their identifiers are disjoint, but the manuscript cannot externally certify that the latter range was untouched during development. Each seed creates $p = 12$ candidates

by crossing three families with targets $M \in \{.12, .34, .46, .75\}$ and $\tau = .40$. The deterministic generator key is $810000 + 100s + 7f + t$. For Fourier candidates,

$$g(z) = \sum_{j=1}^{3} a_j j^{-3} \sin(j\pi z + \phi_j), \qquad a_j \sim \mathrm{Unif}[-1, 1], \quad \phi_j \sim \mathrm{Unif}[-\pi, \pi],$$

and the raw Lipschitz bound is $\sum_j |a_j| j^{-2}\pi$. RBF candidates use $g(z) = \sum_{j=1}^{3} w_j \exp(-((z - c_j)/h_j)^2)$ with $w_j \sim \mathrm{Unif}[-1, 1]$, $c_j \sim \mathrm{Unif}[-.9, .9]$, $h_j \sim \mathrm{Unif}[.4, .75]$, and bound $\sum_j |w_j|\sqrt{2/e}/h_j$. Piecewise-linear candidates use seven anchors $-1, -2/3, \ldots, 1$, independent values in $[-1, 1]$, and the largest segment slope. Each raw function and its bound are multiplied by the same scale that sets the maximum on the 2,001-point mesh to the target. A below-threshold label is accepted only if mesh maximum plus $.0005L$ is below τ; an above-threshold label has a mesh witness above τ.

Normalization fixes the maximum on the mesh, so boundedness off the mesh needs its own argument. The same Lipschitz allowance supplies it: the continuous supremum is at most the mesh maximum plus $.0005L$, and over all 1,200 mechanisms that certified bound attains its maximum 0.758 at seed 25, family `random_rbf`, target .75, $L = 16.03$, leaving slack 0.242. Hence $|\Delta(z)| \le 1$ on the whole interval, not only at mesh points.

At a query z, one pair draws independent Bernoulli variables with means $(1 + \Delta(z))/2$ and $(1 - \Delta(z))/2$ and records their difference $D \in \{-1, 0, 1\}$, whose expectation is $\Delta(z)$. The bound just certified is what makes both means lie in $[0, 1]$ at every adaptively chosen query point. Method m uses random key $10^7 s + 10^5 q + c_m$ under the local runners. For f16, a PCG64 stream seeded by 20260902 supplies one 63-bit draw per candidate, XORed with $10^7 s + 10^5 q$; left children open before right children and smaller node IDs break equal-upper-endpoint ties.

One FWER for every method. Every table cell is a separate run with pair budget B per candidate. The family event is all tests made for all $p = 12$ candidates of one seed, and its error level is $\alpha = .05$. A mean based on n pairs and one of H possible tested events uses

$$r(n, H) = \sqrt{2\log(2H/\alpha)/n}.$$

Thus fixed, Lipschitz, and level-set rules use $H = 17p$, and a k-slot scan uses $H = kp$: each takes one fresh batch of predetermined size at a predictably chosen location, so conditional Hoeffding at each query followed by this finite union bound covers the whole transcript. Successive elimination is not of that form. It re-tests *cumulative* means whose count depends on which points survived earlier rounds and on how many points were still active when the round budget was divided, and a fixed-sample radius evaluated only at that random count is not valid. For candidate c, point j, round s, and deterministic prefix length $t \in \{1, \ldots, B\}$, let $E_{c,j,s,t}$ be the fixed-time Hoeffding event for the first t i.i.d. observations at that point. A union bound over all $17 \cdot 4pB$ such events gives probability at least $1 - \alpha$ that every prefix mean is covered simultaneously. Every cumulative mean selected by every possible survival and reallocation path is one of those prefixes, so evaluating the radius at the realised count is valid on this single path-independent event; no optional-stopping assumption is used. Thus successive elimination uses $H = 17 \cdot 4pB$. That costs one extra logarithm rather than a change of algorithm. Every method reports SENSITIVE when its lower endpoint exceeds τ, INSENSITIVE when its upper endpoint is at most τ, and UNRESOLVED otherwise.

The stagewise tree uses a different finite union: for complete depth D it sets $H_{\mathrm{nodes}} = 2^{D+1} - 1$ per candidate and $H_{\mathrm{total}} = pH_{\mathrm{nodes}}$ overall. Theorem 4A's H denotes H_{nodes}, so its $\log(2pH/\alpha)$ equals $\log(2H_{\mathrm{total}}/\alpha)$. An opened node receives a fresh batch whose size is fixed from the preceding transcript. Conditional Hoeffding is therefore valid at that predictable location and size; unioning every possible node, including those never opened, covers all pruning, early stopping, and recycled-budget paths. Every synthetic runner generates a newly requested batch from fresh Bernoulli variates after the query and size are chosen. Conditional on the transcript, those observations are mutually independent and have the queried contrast as conditional mean. The adaptive rules therefore satisfy Assumption 1's stronger filtration clause rather than relying on the fixed-grid allowance for cross-location dependence.

Executable rules.

1. **Fixed 17-point net.** Query $G = \{-1, -7/8, \dots, 1\}$, splitting B as evenly as integer rounding permits. Use $H = 17p$, lower endpoint $\max_{z\in G}(|\bar{D}_z| - r_z)_+$, and upper endpoint $\min\{1, \max_{z\in G}(|\bar{D}_z| + r_z) + L/16\}$.
2. **Successive elimination.** Start with the same 17 points and four rounds. In each round divide $\lfloor B/4 \rfloor$ over active points, update their cumulative means, and retain point j iff its upper confidence value is at least the largest lower confidence value. Use $H = 17 \cdot 4pB$ throughout, the anytime-valid choice justified above. After round four, apply the fixed-net endpoints to the surviving measured points with geometric allowance $L/16$. This preserves the upper endpoint: on the simultaneous prefix event, if point j is discarded in a round then its true value is at most its upper endpoint, which is below the largest lower endpoint of a point retained in that round and hence below that retained point's true value. If the witness is discarded later, the same argument replaces it by a retained point with still larger true value. Induction over the four rounds therefore maps every discarded point to a final survivor whose true value is no smaller. The maximum over final survivors equals the maximum over the original grid on the coverage event, so adding the original $L/16$ geometric allowance still bounds the continuous supremum.
3. **Lipschitz-envelope placement.** Allocate $\lfloor B/17 \rfloor$ pairs to each query. Start at $-1, 0, 1$; until 17 points are measured, select from the 401-point mesh $\mathcal{M} = \{-1 + j/200\}_{j=0}^{400}$ the unqueried z maximizing $u(z) = \min_j\{|\bar{D}_j| + r + L|z - z_j|\}$, breaking ties toward the smallest mesh index. Use $H = 17p$. The lower endpoint is $\max_j(|\bar{D}_j| - r)_+$. Because u is L-Lipschitz and $\mathcal{M}$ has covering radius $1/400$ on $[-1, 1]$, a mesh maximum can understate the continuous maximum by at most $L/400$; the upper endpoint is therefore $\min\{1, \max_{z\in\mathcal{M}} u(z) + L/400\}$, which bounds $\sup_{z\in[-1,1]} |\Delta(z)|$ on the coverage event.
4. **Level-set placement.** Use the same initial points, $\lfloor B/17 \rfloor$ pairs per query, and $H = 17p$. Write $u(z) = \min_j\{|\bar{D}_j| + r + L|z - z_j|\}$ and $\ell(z) = \max_j(|\bar{D}_j| - r - L|z - z_j|)_+$. After each query form $\underline{M} = \max_j(|\bar{D}_j| - r)_+$ and $\overline{M} = \min\{1, \max_{z\in\mathcal{M}} u(z) + L/400\}$, the same off-mesh allowance as above, and stop with the three-way rule as soon as $\underline{M} > \tau$ or $\overline{M} \le \tau$. Otherwise query the unmeasured mesh point maximizing the ambiguity score $a(z) = \min\{u(z) - \tau,\ \tau - \ell(z)\}$, breaking ties toward the smallest mesh index, and stop after 17 measured points whatever the outcome.
5. **Slot scan.** For each candidate choose

$$k = \arg \min_{1\le k\le \min(B,20001)} \left\{ 2\sqrt{\frac{2\log(2kp/\alpha)}{\lfloor B/k \rfloor}} + \frac{L}{k} \right\},$$

   breaking ties toward smaller k. Query the k slot centres $-1 + (2j - 1)/k$ equally, use $H = kp$, and add L/k only to the upper endpoint. This is the executed finite-budget version of Theorem 4.
6. **Stagewise branch-and-bound.** Choose the complete-tree depth that minimises $2r + L/2^D$ under the common cap when every potential node receives the baseline batch $\lfloor B/(2^{D+1} - 1) \rfloor$. Start at the root. For every active cell I, the executable retains each opened centre's actual pair count n_v and its own radius $r_v = \sqrt{2\log(2pH_{\text{nodes}}/\alpha)/n_v}$, and evaluates exactly

$$U(I) = \min\left\{1, \min_v \left[|\overline{D}_v| + r_v + L \max_{z\in I} |z - z_v| \right] \right\}.$$

   It uses $\underline{M} = \max_v(|\overline{D}_v| - r_v)_+$, prunes exactly when $U(I) < \underline{M}$ or $U(I) \le \tau$, with no positive numerical tolerance, and retains τ in the global upper endpoint after any threshold pruning. It splits the remaining cell with largest upper endpoint, breaking ties by node ID. Before opening its left then right child, assign each $\lfloor B_{\text{remaining}}/H_{\text{remaining}} \rfloor$ fresh pairs, where $H_{\text{remaining}}$ is the complete number of descendants that could still open; this predictable allocation recycles pruned-subtree budget into fresh batches. Stop under the same three-way rule or when depth D is exhausted.

Aggregate result on the declared evaluation split.

Pairs	Fixed	Successive	Lipschitz	Level set	Slot scan	Stagewise
16,384	600	578	600	599	603	591
65,536	646	606	731	669	663	630
262,144	870	874	1,078	1,069	838	788
1,048,576	926	933	1,155	1,148	1,056	1,037

All 24 method–budget cells have zero observed wrong decisions. This is an empirical count, not a zero-error claim; validity comes from the simultaneous construction above. Three earlier series are withdrawn. The slot-scan counts $611, 717, 886, 1116$ used $H = k$ rather than $H = kp$; the seed 100–199 series was measured against envelope comparators whose upper endpoint was a mesh maximum without the $L/400$ off-mesh allowance, so those decisions were not certified for the continuous supremum; and the successive-elimination counts $600, 798, 911, 954$ used $H = 17 \cdot 4p$, a fixed-sample radius evaluated at a data-dependent cumulative count. Correcting the last of these is what removes that rule's reported lead at 65,536 pairs, from 798 down to 606.

Decisive counts by target at the headline budget. Each target family holds 300 of the 1,200 candidates.

Method	$M = .12$	$M = .34$	$M = .46$	$M = .75$
Fixed net	300	0	0	300
Successive elimination	281	0	0	297
Lipschitz envelope	300	0	0	300
Level set	300	0	0	299
Slot scan	300	0	3	300
Stagewise	296	0	0	295

Every rule except the slot scan decides only the two far-boundary families and nothing near the tolerance, so the slot scan's three-decision margin over the fixed net is exactly its three $M = .46$ resolutions. Level set's 599 is a loss of one at $M = .75$ rather than a near-threshold gain, and successive elimination's 578 is a loss at both far families under the anytime-valid radii.

Per-seed decision and error counts. The following tables contain every declared evaluation seed needed to reconstruct the aggregate table. "Wrong" sums wrong decisions over the five legacy methods for that seed and budget.

Table 3: Per-seed decisions on the declared evaluation split at 16,384 pairs per candidate. The final column is the total wrong decisions across all five methods.

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
200	6	6	6	6	6	0
201	6	6	6	6	6	0
202	6	6	6	6	6	0
203	6	6	6	6	6	0
204	6	6	6	6	6	0
205	6	5	6	6	6	0
206	6	6	6	6	6	0
207	6	5	6	6	6	0
208	6	6	6	6	6	0
209	6	6	6	6	6	0
210	6	6	6	6	6	0
211	6	6	6	6	6	0
212	6	6	6	6	6	0
213	6	6	6	6	6	0
214	6	6	6	6	6	0
215	6	5	6	6	6	0
216	6	6	6	6	6	0
217	6	6	6	6	6	0
218	6	6	6	6	7	0
219	6	6	6	6	6	0
220	6	6	6	6	6	0
221	6	6	6	6	6	0
222	6	6	6	6	6	0
223	6	6	6	6	6	0
224	6	6	6	6	6	0
225	6	6	6	6	6	0
226	6	5	6	6	6	0

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
227	6	6	6	6	6	0
228	6	6	6	6	6	0
229	6	4	6	6	6	0
230	6	6	6	6	7	0
231	6	6	6	6	6	0
232	6	6	6	6	6	0
233	6	6	6	6	6	0
234	6	5	6	6	6	0
235	6	6	6	6	6	0
236	6	5	6	6	6	0
237	6	6	6	6	6	0
238	6	6	6	6	6	0
239	6	6	6	6	6	0
240	6	5	6	6	6	0
241	6	5	6	6	6	0
242	6	6	6	6	6	0
243	6	6	6	6	6	0
244	6	5	6	6	6	0
245	6	6	6	6	6	0
246	6	6	6	6	7	0
247	6	3	6	6	6	0
248	6	6	6	6	6	0
249	6	6	6	6	6	0
250	6	6	6	6	6	0
251	6	6	6	6	6	0
252	6	6	6	6	6	0
253	6	6	6	6	6	0
254	6	5	6	6	6	0
255	6	6	6	6	6	0
256	6	5	6	6	6	0
257	6	6	6	5	6	0
258	6	5	6	6	6	0
259	6	5	6	6	6	0
260	6	6	6	6	6	0
261	6	6	6	6	6	0
262	6	6	6	6	6	0
263	6	6	6	6	6	0
264	6	6	6	6	6	0
265	6	6	6	6	6	0
266	6	6	6	6	6	0
267	6	6	6	6	6	0
268	6	6	6	6	6	0
269	6	6	6	6	6	0
270	6	6	6	6	6	0
271	6	6	6	6	6	0
272	6	6	6	6	6	0
273	6	5	6	6	6	0
274	6	6	6	6	6	0
275	6	5	6	6	6	0
276	6	6	6	6	6	0
277	6	6	6	6	6	0
278	6	5	6	6	6	0
279	6	6	6	6	6	0
280	6	6	6	6	6	0
281	6	6	6	6	6	0
282	6	6	6	6	6	0
283	6	6	6	6	6	0
284	6	6	6	6	6	0
285	6	6	6	6	6	0
286	6	6	6	6	6	0
287	6	6	6	6	6	0
288	6	6	6	6	6	0
289	6	6	6	6	6	0
290	6	6	6	6	6	0
291	6	6	6	6	6	0
292	6	6	6	6	6	0
293	6	6	6	6	6	0
294	6	5	6	6	6	0
295	6	6	6	6	6	0
296	6	6	6	6	6	0
297	6	6	6	6	6	0
298	6	6	6	6	6	0
299	6	6	6	6	6	0

Table 4: Per-seed decisions on the declared evaluation split at 65,536 pairs per candidate. The final column is the total wrong decisions across all five methods.

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
200	8	6	7	6	6	0
201	6	6	8	7	7	0
202	6	6	8	6	6	0
203	6	6	8	7	7	0
204	6	6	9	7	7	0
205	7	6	9	7	6	0
206	7	6	6	7	6	0
207	7	6	8	7	8	0
208	7	6	7	7	6	0
209	6	6	6	6	7	0
210	6	6	7	6	7	0
211	7	6	7	7	6	0
212	8	6	7	6	6	0
213	7	6	7	7	6	0
214	6	6	7	7	7	0
215	7	6	8	6	6	0
216	6	6	7	6	6	0
217	6	6	8	7	7	0
218	6	6	6	7	7	0
219	6	6	9	7	6	0
220	6	6	8	8	7	0
221	7	6	9	8	7	0
222	6	6	7	6	6	0
223	7	6	6	7	6	0
224	6	6	6	6	6	0
225	6	6	7	7	8	0
226	6	6	8	7	6	0
227	7	6	8	6	8	0
228	6	6	7	6	7	0
229	6	6	6	6	7	0
230	6	6	7	7	6	0
231	7	6	7	6	6	0
232	6	6	7	7	7	0
233	6	6	8	6	6	0
234	8	6	9	7	6	0
235	6	6	7	6	6	0
236	6	6	8	6	8	0
237	6	6	6	7	6	0
238	6	6	8	7	7	0
239	7	6	7	8	7	0
240	6	6	7	6	6	0
241	7	6	8	7	6	0
242	6	6	6	6	6	0
243	7	6	8	6	6	0
244	6	7	7	7	7	0
245	6	6	7	6	8	0
246	6	6	7	6	6	0
247	6	6	7	6	7	0
248	7	6	8	7	6	0
249	6	6	7	7	6	0
250	8	6	6	7	6	0
251	6	6	7	6	7	0
252	6	6	6	6	7	0
253	7	6	8	6	6	0
254	6	6	8	6	6	0
255	7	6	8	6	6	0
256	6	6	7	6	8	0
257	7	7	9	7	8	0
258	8	6	6	6	6	0
259	6	6	9	6	7	0
260	6	6	6	7	6	0
261	7	6	8	7	6	0
262	6	6	7	8	8	0
263	6	6	7	7	6	0
264	6	6	7	6	6	0
265	7	6	8	8	7	0
266	6	6	7	7	7	0
267	7	7	6	8	7	0
268	6	6	7	6	7	0
269	7	6	7	6	7	0
270	7	6	6	6	7	0
271	6	6	7	6	6	0
272	6	6	6	7	7	0
273	6	6	7	6	7	0
274	8	6	8	8	7	0
275	6	6	7	7	7	0

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
276	6	6	9	7	8	0
277	6	6	9	8	8	0
278	7	6	8	7	6	0
279	7	6	8	6	7	0
280	6	6	8	7	7	0
281	6	6	7	7	6	0
282	6	6	6	7	8	0
283	6	6	9	7	6	0
284	7	6	7	6	6	0
285	6	7	7	6	6	0
286	6	6	7	7	7	0
287	7	6	7	8	6	0
288	7	7	7	8	7	0
289	7	6	9	8	6	0
290	7	6	7	8	6	0
291	6	6	7	6	7	0
292	6	6	8	6	7	0
293	9	7	8	6	7	0
294	6	6	7	8	7	0
295	6	6	8	7	6	0
296	6	6	7	7	7	0
297	6	6	6	6	7	0
298	6	6	7	7	6	0
299	7	6	6	6	7	0

Table 5: Per-seed decisions on the declared evaluation split at 262,144 pairs per candidate. The final column is the total wrong decisions across all five methods.

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
200	8	9	11	11	9	0
201	8	8	11	10	8	0
202	9	8	10	10	9	0
203	9	9	10	10	9	0
204	9	9	11	11	9	0
205	10	9	10	10	9	0
206	9	9	10	10	8	0
207	9	9	10	10	8	0
208	8	9	11	12	8	0
209	9	9	12	12	9	0
210	9	9	11	11	8	0
211	9	9	10	10	8	0
212	8	9	11	11	7	0
213	9	9	11	11	8	0
214	8	8	11	11	7	0
215	9	9	10	10	9	0
216	8	8	10	10	9	0
217	9	9	11	10	8	0
218	8	8	11	11	7	0
219	9	9	11	11	9	0
220	9	9	11	11	7	0
221	9	9	10	10	9	0
222	9	9	10	11	9	0
223	9	9	12	11	8	0
224	8	8	11	11	8	0
225	9	9	11	11	8	0
226	8	9	10	10	7	0
227	9	9	10	10	9	0
228	8	8	12	12	9	0
229	8	8	10	11	8	0
230	9	9	10	10	8	0
231	9	9	11	10	8	0
232	9	9	11	11	8	0
233	8	8	10	10	8	0
234	9	9	10	11	9	0
235	8	8	10	9	8	0
236	9	9	10	10	9	0
237	9	9	10	10	9	0
238	8	8	11	10	8	0
239	9	9	12	12	8	0
240	8	8	11	11	8	0
241	9	9	11	11	10	0
242	8	8	11	11	8	0
243	9	9	12	11	8	0
244	9	9	10	10	8	0
245	9	9	11	11	9	0
246	9	9	11	11	8	0

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
247	8	8	12	11	8	0
248	9	9	10	11	8	0
249	10	10	11	11	9	0
250	9	9	11	11	8	0
251	8	8	11	10	9	0
252	8	9	11	11	8	0
253	9	9	11	11	8	0
254	8	8	11	11	8	0
255	9	9	10	10	8	0
256	8	8	12	12	8	0
257	10	9	11	11	9	0
258	9	9	9	9	8	0
259	8	8	11	10	8	0
260	9	9	11	11	8	0
261	9	9	12	12	8	0
262	8	8	10	10	8	0
263	8	8	12	12	8	0
264	9	9	11	10	9	0
265	9	9	11	11	10	0
266	8	8	10	10	8	0
267	9	9	12	12	9	0
268	9	9	11	11	9	0
269	9	9	10	10	9	0
270	8	8	11	11	8	0
271	8	8	11	11	9	0
272	8	9	10	10	8	0
273	9	9	11	11	8	0
274	9	10	11	10	8	0
275	8	8	10	10	9	0
276	9	9	11	11	10	0
277	9	9	11	11	10	0
278	8	8	12	12	8	0
279	9	9	11	10	8	0
280	9	9	10	10	8	0
281	9	9	12	12	9	0
282	9	9	10	10	9	0
283	9	9	12	11	8	0
284	9	9	10	11	9	0
285	9	9	10	10	8	0
286	9	9	11	10	9	0
287	9	9	11	11	9	0
288	9	9	11	11	7	0
289	9	9	11	11	8	0
290	9	9	11	11	10	0
291	8	8	11	11	9	0
292	9	9	11	12	9	0
293	9	9	10	10	8	0
294	9	9	12	12	9	0
295	9	9	11	11	8	0
296	9	9	12	11	8	0
297	8	8	10	10	8	0
298	8	8	10	10	8	0
299	9	9	10	10	8	0

Table 6: Per-seed decisions on the declared evaluation split at 1,048,576 pairs per candidate. The final column is the total wrong decisions across all five methods.

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
200	9	9	12	12	12	0
201	9	9	11	11	10	0
202	10	10	11	11	9	0
203	9	9	11	11	9	0
204	10	10	12	11	11	0
205	10	10	12	11	11	0
206	9	9	12	12	10	0
207	9	9	11	10	10	0
208	9	9	12	12	12	0
209	9	10	12	12	11	0
210	9	9	11	11	10	0
211	9	9	10	11	10	0
212	10	10	11	12	10	0
213	10	10	12	12	10	0
214	10	10	12	12	10	0
215	9	9	11	11	11	0
216	9	9	10	10	9	0
217	10	10	11	11	10	0

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
218	8	8	12	12	11	0
219	10	10	12	12	11	0
220	9	9	11	11	10	0
221	9	9	12	12	11	0
222	9	9	12	12	12	0
223	9	9	12	12	10	0
224	10	10	11	11	11	0
225	10	11	12	12	12	0
226	10	10	12	12	12	0
227	9	9	11	11	10	0
228	10	10	12	12	12	0
229	9	9	11	11	10	0
230	10	10	10	10	9	0
231	9	9	12	10	11	0
232	10	10	12	12	11	0
233	9	9	12	11	11	0
234	9	9	12	12	11	0
235	8	8	11	11	10	0
236	9	10	12	11	10	0
237	9	9	12	12	10	0
238	9	9	11	11	10	0
239	10	10	12	12	11	0
240	9	9	12	12	12	0
241	9	9	12	11	11	0
242	9	10	12	12	11	0
243	9	9	12	12	10	0
244	9	9	12	12	10	0
245	9	9	12	12	11	0
246	9	9	11	11	10	0
247	8	8	12	12	11	0
248	9	9	12	12	9	0
249	10	10	11	11	11	0
250	10	10	11	11	11	0
251	8	8	12	12	10	0
252	9	9	12	12	9	0
253	9	9	12	11	10	0
254	9	9	12	12	11	0
255	9	9	11	11	9	0
256	9	9	12	12	11	0
257	11	11	12	12	12	0
258	9	9	10	10	8	0
259	9	8	12	12	11	0
260	11	11	12	12	12	0
261	10	10	12	12	11	0
262	9	9	11	11	9	0
263	8	8	12	12	11	0
264	9	9	11	11	11	0
265	10	10	12	12	11	0
266	8	8	11	11	10	0
267	9	10	12	12	11	0
268	9	9	12	12	11	0
269	9	10	12	12	11	0
270	9	9	12	11	11	0
271	8	8	11	12	11	0
272	9	9	12	12	11	0
273	10	10	12	12	11	0
274	10	10	11	11	10	0
275	8	8	11	10	10	0
276	10	10	12	12	12	0
277	9	10	11	11	11	0
278	8	8	12	12	12	0
279	9	9	11	11	9	0
280	9	9	10	11	11	0
281	10	10	12	12	11	0
282	9	9	11	11	11	0
283	10	10	12	12	10	0
284	9	9	11	11	10	0
285	9	9	12	11	11	0
286	9	10	12	12	10	0
287	11	11	11	11	12	0
288	9	9	12	12	9	0
289	9	9	12	12	10	0
290	10	10	12	12	11	0
291	9	9	11	11	10	0
292	10	9	12	12	12	0
293	10	10	10	10	10	0
294	9	9	12	12	11	0
295	9	9	12	12	9	0
296	10	10	12	12	11	0

Seed	Fixed	Succ.	Lipschitz	Level-set	Slot	Wrong
297	9	10	11	11	11	0
298	10	10	11	11	10	0
299	9	9	11	12	11	0

Table 7: Complete stagewise per-seed ledger: decisive (D) and wrong (W) reports at each cap.

Seed	D_{16k}	W_{16k}	D_{65k}	W_{65k}	D_{262k}	W_{262k}	D_{1m}	W_{1m}
200	5	0	6	0	7	0	11	0
201	6	0	6	0	8	0	10	0
202	6	0	7	0	8	0	10	0
203	6	0	6	0	7	0	10	0
204	6	0	7	0	8	0	9	0
205	6	0	6	0	9	0	10	0
206	5	0	6	0	9	0	11	0
207	6	0	6	0	7	0	9	0
208	6	0	7	0	9	0	11	0
209	6	0	6	0	8	0	10	0
210	6	0	6	0	8	0	11	0
211	6	0	6	0	7	0	10	0
212	6	0	6	0	7	0	12	0
213	6	0	6	0	8	0	10	0
214	6	0	7	0	8	0	10	0
215	5	0	6	0	8	0	11	0
216	6	0	6	0	8	0	10	0
217	6	0	6	0	8	0	11	0
218	6	0	6	0	7	0	10	0
219	6	0	6	0	8	0	10	0
220	6	0	7	0	8	0	9	0
221	6	0	6	0	8	0	10	0
222	6	0	6	0	7	0	12	0
223	6	0	6	0	7	0	10	0
224	6	0	6	0	8	0	11	0
225	5	0	7	0	7	0	11	0
226	5	0	6	0	8	0	10	0
227	6	0	6	0	8	0	9	0
228	6	0	6	0	9	0	12	0
229	6	0	6	0	8	0	11	0
230	6	0	7	0	8	0	9	0
231	6	0	6	0	8	0	11	0
232	6	0	7	0	7	0	10	0
233	6	0	6	0	7	0	10	0
234	6	0	6	0	8	0	12	0
235	6	0	6	0	7	0	10	0
236	6	0	6	0	9	0	10	0
237	6	0	6	0	8	0	11	0
238	6	0	6	0	7	0	10	0
239	6	0	7	0	8	0	10	0
240	6	0	6	0	8	0	10	0
241	6	0	7	0	7	0	11	0
242	6	0	7	0	8	0	11	0
243	6	0	6	0	8	0	10	0
244	6	0	6	0	8	0	9	0
245	6	0	6	0	9	0	11	0
246	6	0	6	0	8	0	10	0
247	5	0	6	0	7	0	11	0

Seed	D_{16k}	W_{16k}	D_{65k}	W_{65k}	D_{262k}	W_{262k}	D_{1m}	W_{1m}
248	6	0	6	0	7	0	9	0
249	6	0	6	0	8	0	10	0
250	6	0	6	0	8	0	11	0
251	6	0	6	0	7	0	11	0
252	6	0	6	0	7	0	11	0
253	6	0	6	0	7	0	10	0
254	6	0	6	0	7	0	10	0
255	6	0	6	0	7	0	9	0
256	6	0	7	0	8	0	11	0
257	6	0	6	0	9	0	11	0
258	6	0	6	0	8	0	8	0
259	6	0	6	0	8	0	11	0
260	6	0	7	0	9	0	11	0
261	6	0	6	0	8	0	10	0
262	6	0	6	0	7	0	9	0
263	6	0	7	0	8	0	12	0
264	6	0	6	0	8	0	10	0
265	6	0	6	0	9	0	11	0
266	6	0	6	0	8	0	8	0
267	6	0	7	0	9	0	11	0
268	6	0	7	0	9	0	11	0
269	6	0	7	0	8	0	10	0
270	6	0	6	0	7	0	11	0
271	6	0	6	0	8	0	11	0
272	6	0	6	0	8	0	9	0
273	6	0	6	0	8	0	10	0
274	6	0	7	0	7	0	10	0
275	6	0	6	0	9	0	10	0
276	5	0	6	0	10	0	12	0
277	6	0	7	0	10	0	11	0
278	6	0	6	0	8	0	12	0
279	6	0	6	0	8	0	9	0
280	6	0	7	0	8	0	10	0
281	6	0	7	0	7	0	11	0
282	6	0	6	0	8	0	11	0
283	6	0	6	0	6	0	11	0
284	6	0	6	0	6	0	11	0
285	6	0	7	0	8	0	9	0
286	5	0	6	0	9	0	11	0
287	6	0	6	0	8	0	12	0
288	6	0	6	0	7	0	9	0
289	6	0	6	0	8	0	9	0
290	6	0	8	0	10	0	12	0
291	6	0	7	0	9	0	11	0
292	6	0	7	0	10	0	11	0
293	6	0	6	0	7	0	10	0
294	5	0	6	0	8	0	11	0
295	6	0	7	0	7	0	10	0
296	6	0	7	0	8	0	10	0
297	6	0	6	0	7	0	11	0
298	6	0	7	0	8	0	10	0
299	6	0	7	0	7	0	10	0

Ambiguity-geometry factor grid. The rendered f17 manifest crosses two targets, three peak counts, three widths, and three location shifts, for 54 designs. Each contrast is a maximum of triangular peaks over baseline .12 and has valid common bound $L = 4.25$. Fifty deterministic replicates per design give 2,700 trials. The audit records trace-independent ambiguity-

node bounds, opened nodes, pair spend, decisions, and errors. The full manifest SHA-256 is `e4dc04381f14c9477e8ffa792604f3252cd0223091886b9a6f988f4beb46b0c8`. Table 8 renders each design's 50-trial satisfaction count and error/spend summaries; all satisfaction counts are 50 and all wrong counts are zero.

Table 8: Complete f17 design ledger. Sat is the number of 50 trials satisfying the trace-independent node bound.

M	q	w	shift	H_Δ	Sat	Wrong	Dec.	mean nodes	mean pairs
0.34	1	0.08	-0.15	67	50	0	0	64.96	536310
0.34	1	0.08	+0.00	67	50	0	0	67.00	553152
0.34	1	0.08	+0.15	67	50	0	0	65.04	536970
0.34	1	0.16	-0.15	71	50	0	0	69.00	569664
0.34	1	0.16	+0.00	71	50	0	0	68.64	566692
0.34	1	0.16	+0.15	71	50	0	0	69.00	569664
0.34	1	0.32	-0.15	79	50	0	0	74.64	616228
0.34	1	0.32	+0.00	79	50	0	0	74.96	618870
0.34	1	0.32	+0.15	79	50	0	0	74.28	613256
0.34	2	0.08	-0.15	71	50	0	0	70.48	581883
0.34	2	0.08	+0.00	71	50	0	0	67.04	553482
0.34	2	0.08	+0.15	71	50	0	0	70.20	579571
0.34	2	0.16	-0.15	79	50	0	0	74.12	611935
0.34	2	0.16	+0.00	79	50	0	0	75.00	619200
0.34	2	0.16	+0.15	79	50	0	0	73.80	609293
0.34	2	0.32	-0.15	95	50	0	0	86.24	711997
0.34	2	0.32	+0.00	95	50	0	0	85.68	707374
0.34	2	0.32	+0.15	95	50	0	0	86.80	716621
0.34	4	0.08	-0.15	79	50	0	0	75.48	623163
0.34	4	0.08	+0.00	79	50	0	0	73.32	605330
0.34	4	0.08	+0.15	79	50	0	0	75.96	627126
0.34	4	0.16	-0.15	95	50	0	0	86.16	711337
0.34	4	0.16	+0.00	95	50	0	0	86.72	715960
0.34	4	0.16	+0.15	95	50	0	0	86.04	710346
0.34	4	0.32	-0.15	117	50	0	0	109.40	903206
0.34	4	0.32	+0.00	119	50	0	0	108.24	893629
0.34	4	0.32	+0.15	117	50	0	0	109.92	907500
0.46	1	0.08	-0.15	69	50	0	1	65.60	541594
0.46	1	0.08	+0.00	67	50	0	40	14.20	117235
0.46	1	0.08	+0.15	69	50	0	2	64.72	534328
0.46	1	0.16	-0.15	73	50	0	21	52.48	433275
0.46	1	0.16	+0.00	71	50	0	37	19.20	158515
0.46	1	0.16	+0.15	73	50	0	13	58.52	483141
0.46	1	0.32	-0.15	81	50	0	39	37.64	310756
0.46	1	0.32	+0.00	79	50	0	38	21.20	175027
0.46	1	0.32	+0.15	81	50	0	41	36.12	298207
0.46	2	0.08	-0.15	71	50	0	42	16.16	133417
0.46	2	0.08	+0.00	75	50	0	6	65.64	541924
0.46	2	0.08	+0.15	71	50	0	43	17.64	145636
0.46	2	0.16	-0.15	79	50	0	46	15.68	129454
0.46	2	0.16	+0.00	83	50	0	29	50.40	416102
0.46	2	0.16	+0.15	79	50	0	45	17.60	145306
0.46	2	0.32	-0.15	97	50	0	48	9.72	80248
0.46	2	0.32	+0.00	99	50	0	50	30.00	247680
0.46	2	0.32	+0.15	97	50	0	49	11.72	96760
0.46	4	0.08	-0.15	81	50	0	46	14.60	120538
0.46	4	0.08	+0.00	83	50	0	25	61.68	509230
0.46	4	0.08	+0.15	81	50	0	49	9.88	81569
0.46	4	0.16	-0.15	97	50	0	50	16.40	135398
0.46	4	0.16	+0.00	99	50	0	46	42.04	347082
0.46	4	0.16	+0.15	97	50	0	49	12.48	103035
0.46	4	0.32	-0.15	119	50	0	50	12.08	99732
0.46	4	0.32	+0.00	119	50	0	50	29.84	246359
0.46	4	0.32	+0.15	119	50	0	50	11.20	92467

F Proof of Theorem 4B

Let E be the event that the procedure reports INSENSITIVE. Under Δ_0, $P_0(E) \geq 1 - \alpha_{\mathrm{err}} - \alpha_{\mathrm{abs}}$; under every Δ_j, $P_j(E) \leq \alpha_{\mathrm{err}}$. The same binary change-of-measure inequality used in Theorem 3 gives $\mathrm{KL}(P_0, P_j) \geq \log(1/(2\epsilon))$. The adaptive KL chain rule and equality of Δ_0, Δ_j off I_j give

$$\mathrm{KL}(P_0, P_j) \leq \bar{\kappa}\, \mathbb{E}_0 N(I_j),$$

where $N(I_j)$ counts queries in I_j. Summing over disjoint intervals yields the lower bound because $\sum_j N(I_j) \leq T$. For the upper statement, let G_j be the simultaneous confidence event under Δ_j,

with $P_j(G_j^c) \le \alpha$. The fixed-batch tree always satisfies $T \le nH$ and on G_j satisfies $T \le nH_{\Delta_j}$. Since $H_{\Delta_j} \le H$, splitting the expectation gives

$$\mathbb{E}_j T \le nH_{\Delta_j} P_j(G_j) + nH P_j(G_j^c) \le n\big((1-\alpha)H_{\Delta_j} + \alpha H\big).$$

Maximise over j to obtain the stated expected-spend upper bound. Theorem 4A's correctness and margin-resolution guarantee meets the lower bound's criterion with $\alpha_{\mathrm{err}} = \alpha_{\mathrm{abs}} = \alpha$. Under Theorem 4B's assumed growth conditions (G1) $\bar{H} \le Cq \log H$ and (G2) $\alpha H \le C_{\mathrm{fail}} q \log H$, which are hypotheses of that comparison rather than properties established here, this expectation is at most $n(C + C_{\mathrm{fail}})q \log H$. Interiority and the perturbation-amplitude condition give $\bar{\kappa} \le C_\Delta^2 \gamma^2/[c(2-c)]$. Substituting $n = O(\log(pH/\alpha)/\gamma^2)$ now compares expected expenditure on both sides. Without failure-event absorption, the high-probability ambiguity bound alone does not establish this comparison; $\alpha n H$ must be retained separately.

G Proofs of Theorems 4C and 4D and the corollary

Caps are admissible flips. Both $s_z\Delta_0$ and $T_z = \tau+\gamma-L|\cdot - z|$ are L-Lipschitz, and the pointwise maximum of two real L-Lipschitz functions is L-Lipschitz:

$$|\max\{f(x), g(x)\} - \max\{f(y), g(y)\}| \le \max\{|f(x)-f(y)|, |g(x)-g(y)|\}.$$

Moreover $s_z\Delta_0 \in [-1,1]$, $T_z \le \tau+\gamma \le 1$, and their maximum is at least $s_z\Delta_0 \ge -1$; multiplying by s_z preserves the range $[-1,1]$. This argument fixes s_z at the cap centre and does not assume that Δ_0 keeps that sign elsewhere. At the centre, $s_z\Delta^{(z)}(z) \ge T_z(z) = \tau+\gamma$, hence $M(\Delta^{(z)}) \ge \tau+\gamma$; off S_z the two functions agree by definition. On S_z the sign-transformed difference satisfies

$$0 < s_z[\Delta^{(z)}(z') - \Delta_0(z')] \le (\tau+\gamma-L|z'-z|) - [s_z\Delta_0(z) - L|z'-z|] = h_z$$

by the Lipschitz lower bound on $s_z\Delta_0$. If Δ_0 is L'-Lipschitz with $L' < L$, then $T_z(z') - s_z\Delta_0(z') \le h_z - (L-L')|z'-z| \le 0$ once $|z'-z| \ge h_z/(L-L')$, so $S_z \subseteq (z - h_z/(L-L'), z + h_z/(L-L'))$.

Proof of Theorem 4C. Let E be the event that the procedure reports INSENSITIVE. Because $M(\Delta_0) \le \tau - \gamma$, validity gives $P_0(E) \ge 1 - \alpha_{\mathrm{err}} - \alpha_{\mathrm{abs}}$, and because $M(\Delta^{(z)}) \ge \tau+\gamma$, $P_z(E) \le \alpha_{\mathrm{err}}$. The change-of-measure inequality used in the proof of Theorem 3 therefore gives $\mathrm{KL}(P_0, P_z) \ge \log(1/(2\epsilon))$ for the transcript laws. Under the adaptive filtration clause the transcript divergence is $\mathbb{E}_0 \sum_{t \le T} \mathrm{kl}\big(\frac{1+\Delta_0(z_t)}{2} \| \frac{1+\Delta^{(z)}(z_t)}{2}\big) \le \kappa_z \, \mathbb{E}_0 N(S_z)$, because the two instances coincide off S_z and each fresh observation contributes at most κ_z. For a cap packing Z the supports are disjoint, so $\sum_{z\in Z} N(S_z) \le T$ and $\log(1/(2\epsilon)) \sum_{z\in Z} \kappa_z^{-1} \le \mathbb{E}_0 T$; take the supremum over Z. For the divergence bound, both $\Delta_0(z')$ and $\Delta^{(z)}(z')$ lie in $[-(1-c), 1-c]$, so the Bernoulli parameters lie in $[c/2, 1-c/2]$ and $\mathrm{kl}(p\|q) \le (p-q)^2/(q(1-q)) \le 4(p-q)^2/(c(2-c))$ with $|p-q| = |\Delta_0 - \Delta^{(z)}|/2 \le h_z/2$.

Proof of Theorem 4D. *Validity.* The pHK (candidate, node, stage) triples and the cumulative sizes $2^k n_0$ are predeclared. Under the adaptive clause, fresh batches are independent of the past given their predictable location, so the observations at node v may be realised as the successive entries of a pre-generated i.i.d. stream $\xi_{v,1}, \xi_{v,2}, \ldots$ with mean $\Delta(z_v)$, of which the algorithm's choices only decide how long a prefix is revealed; stage k uses the prefix of length $2^k n_0$. Hoeffding applied to each fixed prefix, whether or not it is ever reached, fails with probability at most $\alpha/(pHK)$, and a union bound over all pHK prefixes gives an event G with $P(G) \ge 1-\alpha$ on which $|\overline{D}_v - \Delta(z_v)| \le r(N_v)$ at every reached stage, with no appeal to the stopping rule. Since $\mathrm{rad}(I) = 2^{-d}$ and $2^D \ge 4L/\gamma$, every leaf has $L\,\mathrm{rad} \le \gamma/4$. On G the endpoints are valid exactly as in Theorem 4A, whatever the data-dependent order of doublings, splits, freezes, and stops, so no decisive report is wrong. *Halting.* Every selection doubles a node (at most K times per node), splits a node (at most once per node, depth at most D), freezes a node (at most once per node), or stops; hence at most HK selections and at most $T_{\max} = H2^{K-1}n_0$ pairs. *Resolution.* Leaves have $L\,\mathrm{rad} \le \gamma/4$ and stage-K nodes have $r \le \gamma/4$. Work on G. If $M(\Delta_0) \ge \tau+\gamma$ at $z^\star$, the cell containing $z^\star$ has upper endpoint at least $\tau+\gamma$ and remains active; it is selected as soon as no unfrozen cell has a larger endpoint, and each other cell is frozen or split after finitely many selections, so the cell containing $z^\star$ is refined to a stage-K leaf whose centre satisfies $|\Delta_0(c)| \ge \tau + 3\gamma/4$

and whose lower endpoint $|\overline{D}| - r \geq |\Delta_0(c)| - 2r \geq \tau + \gamma/4 > \tau$ forces SENSITIVE before any abstention. If $M(\Delta_0) \leq \tau - \gamma$, a leaf with centre c has $m(c) \geq 3\gamma/4$ and is pruned once $2r \leq m(c) - \gamma/4$, which holds at stage K since $2r \leq \gamma/2$; for an internal cell $d < D$, $2^d \leq 2^{D-1} < 4L/\gamma$, hence $L2^{-d} > \gamma/4 \geq r$ at stage K, so when selected at stage K it satisfies the split test and is split (or has been pruned) rather than frozen, because the rule tests splitting before freezing. Thus no cell is ever frozen, and the procedure ends with every cell pruned, i.e. INSENSITIVE. *Spend.* Work on G with Δ_0 insensitive, so $m \geq \gamma$ and m is L-Lipschitz; write $\log$ for $\log(2pHK/\alpha)$. Each node spends $\max\{n_0, N\} \leq n_0 + N$ with N the size below, which yields the $n_0 H$ term. (i) A cell I at depth $d < D$ with centre c and radius $\rho = 2^{-d}$ is split only when $r(N_I) \leq L\rho$; then $U(I) \leq |\Delta_0(c)| + 2r + L\rho \leq |\Delta_0(c)| + 3L\rho$, and the split requires $U(I) > \tau$, so $m(c) < 3L\rho$. (ii) Doubling of an internal cell stops at the first cumulative size with $r \leq L\rho$ (or earlier, by pruning), and the previous size had $r > L\rho$, so $N_I < 4\log/(L\rho)^2$. (iii) A cell at depth $d \geq 1$ is opened only if its parent was split, so $m(c_{\text{parent}}) < 6L\rho$ and every z in the cell has $m(z) < 8L\rho$; opened depth-d cells therefore lie in $A_d = \{m < 8L2^{-d}\}$ and number at most $2^{d-1}|A_d|$. (iv) Summing (ii) over (iii) for $1 \leq d < D$ gives at most $(2\log/L^2)\sum_d 8^d|A_d| \leq (2\log/L^2)\int \sum_{d:8L2^{-d}>m(z)} 8^d\, dz \leq (8192/7)\, L\log\int m^{-3}$. (v) *Root.* On G the observed upper endpoint is at most $|\Delta_0(0)| + 2r + L$, so $2r \leq m(0) - L$ suffices for pruning; it is not an if-and-only-if characterization of the observed decision. For $m(0) < 3L$, the root is pruned or split no later than its first stage with $r \leq L$. Earlier pruning only reduces cost. Apart from the n_0 term already counted, doubling gives $N < 4\log/L^2$. Since $\int m^{-3} \geq \int_{-1}^{1}(3L + L|z|)^{-3}dz = 7/(144L^3)$, this cost is at most $(576/7)L\log\int m^{-3}$. For $m(0) \geq 3L$, a split at $r \leq L$ is impossible on G because the upper endpoint is at most τ. Pruning occurs by the first stage with $2r \leq m(0) - L$, at cost at most $16\log/(m(0) - L)^2 \leq 36\log/m(0)^2$, apart from n_0. Retain this term separately: the earlier inequality $\int m^{-3} \geq 7/(16Lm(0)^2)$ is false, for example at constant $m = .5$, $L = .001$, where its two sides are 16 and 1750. (vi) *Leaves.* A leaf with centre c is pruned once $2r \leq m(c) - \gamma/4$, hence once $N \geq 8\log/(m(c) - \gamma/4)^2$, and doubling overshoots by at most a factor of two. For z in the leaf, $m(c) \geq m(z) - \gamma/4$ and $m(z) \geq \gamma$ give $m(c) - \gamma/4 \geq m(z) - \gamma/2 \geq m(z)/2$, so the leaf spends at most $64\log/m(z)^2 \leq 64\log/\gamma^2$. Leaves lie in $A_D \subseteq \{m < 2\gamma\}$ and number at most $2^{D-1}|A_D| \leq (4L/\gamma)|A_D|$, so their spend is at most $256\, L\log\gamma^{-3}|A_D| \leq 2048\, L\log\int m^{-3}$ because $m^{-3} > (2\gamma)^{-3}$ on A_D. Adding (iv)–(vi) gives $C \leq 8192/7 + 576/7 + 2048 < 3302$. *Expectation.* Off G the spend is at most $T_{\max}$, so $\mathbb{E}_0 T \leq C\log L\int m^{-3} + 36\log/m(0)^2 + n_0 H + \alpha T_{\max}$. *Single leaf.* If $L \leq \gamma/4$ the root is a leaf with allowance $L \leq \gamma/4$; it is pruned once $2r \leq m(0) - \gamma/4$, so $T \leq 16\log/(m(0) - \gamma/4)^2 + n_0 \leq 64\log/m(0)^2 + n_0$, and the one-cap packing $\{0\}$ gives $Q \geq c(2-c)/h_0^2 \geq c(2-c)/(4m(0)^2)$ since $h_0 = m(0) + \gamma \leq 2m(0)$.

Proof of the corollary. By the lemma with $L' = L/16$, $S_z \subseteq B_z := (z - w_z, z + w_z)$, $w_z = 16h_z/(15L)$. The radii are bounded, so the Vitali covering lemma yields a disjoint subfamily $\{B_{z_j}\}$ with $[-1, 1] \subseteq \bigcup_j 5B_{z_j}$; disjoint B's give disjoint supports, hence a cap packing. On $5B_{z_j}$, whose length is $32h_{z_j}/(3L)$, the $(L/16)$-Lipschitz function h satisfies $h \geq h_{z_j} - \frac{1}{3}h_{z_j} = 2h_{z_j}/3$, because every point is at radius at most $5w_{z_j}$ (not the full side length) from the centre. Thus $\int_{5B_{z_j}} h^{-3} \leq (32h_{z_j}/(3L))\cdot(3/(2h_{z_j}))^3 = 36/(Lh_{z_j}^2)$ and $\sum_j h_{z_j}^{-2} \geq (L/36)\int_{-1}^{1} h^{-3}$. Since $h = m + \gamma \leq 2m$, $\int h^{-3} \geq \frac{1}{8}\int m^{-3}$, and Theorem 4C's divergence bound gives $Q(\Delta_0) \geq c(2-c)\sum_j h_{z_j}^{-2} \geq c(2-c)\, L\int m^{-3}/288$. The doubling procedure is wrong or abstains with probability at most α each on Δ_0 and on every cap (a sensitive instance with margin γ): Theorem 4D resolves on G, so both the wrong-decision event and the abstention event are subsets of G^c, whose probability is at most α. Hence $\epsilon = 2\alpha_{\text{err}} + \alpha_{\text{abs}} \leq 3\alpha$, and (C1) $\alpha < 1/6$ makes $\epsilon < 1/2$ so that Theorem 4C applies with $\log(1/(2\epsilon)) \geq \log(1/(6\alpha)) > 0$. Combining with Theorem 4D's expectation bound gives the ratio. For the additive term, $H < 2^{D+1} \leq 16L/\gamma$, so $n_0 H \leq 16n_0 L/\gamma$ and $\alpha T_{\max} \leq 16\alpha 2^{K-1} n_0 L/\gamma$; under (C5) each is at most $CL\log$, and $2CL\log \leq C\log L\int m^{-3}$ because $m \leq 1$ forces $\int_{-1}^{1} m^{-3} \geq 2$; the margin-integral term is at most doubled, giving $576C$. The one-cap packing at zero satisfies $Q \geq c(2-c)/(m(0)+\gamma)^2 \geq c(2-c)/(4m(0)^2)$, so the separate root cost is at most $144\log Q/(c(2-c))$. This gives $576C + 144$. Feasibility of (C5): $\alpha 2^{K-1} n_0 \geq 32\alpha\log/\gamma^2$, so the two absorption inequalities can hold only if $32\alpha\log/\gamma^2 \leq C\gamma\log/16$, i.e. $\alpha \leq C\gamma^3/512$.

A compatible parameter point and the audit values. A compatible point is $p = 12$, $L = 2$, $\gamma = .25$, $\alpha = .05$, $c = .5$, $K = 10$, $n_0 = 16$ ($D = 5$, $H = 63$, $\log \approx 12.6$, $T_{\max} = 516{,}096$). Take $\tau = .25$, $\Delta_0 \equiv 0$, and $C = 3400$: the base margin is .25 and every cap lies in $[-.5, .5]$. Writing $\ell = \log(2pHK/\alpha)$ and $N_K = 2^{K-1}n_0 = 8192$, the three schedule checks are $32\ell/\gamma^2 \approx 6462 < N_K$, $n_0 = 16 < C\gamma\ell/16 \approx 670$, and $\alpha N_K = 409.6 < 670$; it is αN_K, not N_K, in the last check. The displayed ratio is larger than $2 \cdot 10^7$. The comparison is therefore a rate statement with explicit but loose constants, not a sharp characterization. The audit of Appendix K runs at $\gamma = .06$, $\alpha = .05$, where (C5) fails; the corrected implementation reports grid-proxy ratios, not a certified comparison to Theorem 4C's continuous functional. Those numerical values cannot establish the corollary's constants or necessity.

H EXECUTED COMPATIBLE-POINT SCALING AUDIT

We froze p=12, L=2, γ=.25, α=.05, K=10, n_0=16 (D=5, H=63, $\ell = 12.6195$, $T_{\max} = 516{,}096$) and the bank before execution. Its fifteen comparisons use common $c = .25$: the largest cap is $\tau+\gamma = .75$, so C4 requires $c \le .25$; the former common $c = .5$ was valid only at $\tau = .25$. We executed the procedure with the literal schedule of the theorem statement—cumulative batch sizes $n_0, 2n_0, \ldots, 2^{K-1}n_0$, so per-node spend never exceeds $2^{K-1}n_0$ and run spend never exceeds $T_{\max}$. The F18 audit implementation used the coarser cumulative convention $n_0(2^{k+1} - 1)$ and is not used here. The three (C5) schedule checks hold at the frozen point: $32\ell/\gamma^2 = 6{,}461.2 < 8{,}192 = N_K$, $n_0 = 16 \le C\gamma\ell/16 = 670.4$, and $\alpha N_K = 409.6 \le 670.4$; the joint feasibility condition $\alpha \le C\gamma^3/512$ holds as $.05 < .104$.

Fifteen insensitive settings vary the margin (constant zero, $\tau \in \{.25, \ldots, .50\}$) and geometry at $\tau = .40$ (a shifted constant, two sawtooths, a descent, two plateaus, and three fixed sines); C1–C5 apply to these comparisons. Three sensitive tents test only Theorem 4D resolution: their clipped negative flanks are not C4 packing alternatives and receive no corollary lower bound. Every setting ran 128 predeclared replicates (seed `setting`$\times 100000 +$ `replicate`) under the binary oracle: 1,920 insensitive and 384 sensitive runs per kernel, or 4,608 runs across both kernels. The reporting repair retains the same settings and seeds, reproduces every original mean and extremum, and records each replicate's decision and spend. Both kernels have zero wrong resolutions and zero abstentions. The exact one-sided zero-event bound 0.0231 applies separately to each setting's 128 replicate pairs of runs, not simultaneously to the eighteen settings. Implementation: `simulation/code/f20_compatible_scaling.py` with regressions in `simulation/code/test_f20_compatible_scaling.py`; evidence is retained in the accompanying artifact record. Its hash-bound C4 accounting note derives from all 2,304 paired records without replay and changes only the c-dependent column. The 09-11 result is retained as the original summary. It omitted paired abstentions and used a t critical value for 19 rather than 127 degrees of freedom; its interval and paired-zero-abstention claims are superseded. The final repair also removes a duplicated Lipschitz factor from the margin terms: the enclosure routine already returns $J = L\int m^{-3}$. The original implementation multiplied J by L again. The corrected corollary lower term is $\log(1/(2\epsilon))c(2-c)J/288$ and the upper margin term is ℓCJ. The finite-packing bound and observed trajectories are unchanged.

Same comparison on the paired-difference kernel. The same frozen settings were rerun under Appendix N's complementary paired kernel: each observation is $D = U - V$ with independent $U \sim \mathrm{Bern}((1+\Delta)/2)$ and $V \sim \mathrm{Bern}((1-\Delta)/2)$, at the identical predeclared schedule and confidence level. Replicate seeds agree, but primitive samples do not: binomial and uniform generators consume randomness differently. The paired kernel uses 0.849–0.990 times the binary observation count on insensitive settings. Across all eighteen settings the range is 0.849–1.100: the double-cap setting costs 3,748.625 pairs versus 3,407.75 binary observations. At constant-0, $\tau = .40$, the counts are 13,552 versus 14,704. Each pair costs two primitive draws, so the observed primitive-cost ratio is 1.697–2.200, not exactly two and not a tie. Binary is cheaper under this executed procedure and accounting rule; this does not compare optimal procedures. The paired packing lower bound halves in pair units and coincides in primitive units, while the common bounded-observation Hoeffding upper bound retains its native-observation form (Appendix N).

Setting	τ	$m(0)$	Spend [t95]	Pack. lower	Cor. lower	Upper (complete)
constant 0	.25	.2500	59,992 [±1,020]	14.45	0.468	11,018,100
constant 0	.30	.3000	30,168 [±372]	11.10	0.252	6,388,353
constant 0	.35	.3500	21,200 [±132]	8.56	0.146	4,033,444
constant 0	.40	.4000	14,704 [±140]	6.58	0.088	2,711,298
constant 0	.45	.4500	12,454 [±76]	3.76	0.055	1,912,462
constant 0	.50	.5000	7,618 [±183]	2.81	0.034	1,401,633
constant .10	.40	.3000	24,558 [±335]	9.19	0.209	6,388,353
sawtooth $1/8$	.40	.2578	42,890 [±810]	11.12	0.343	10,553,743
sawtooth $1/4$	.40	.2656	38,987 [±721]	11.02	0.328	10,095,672
descent	.40	.3125	23,468 [±343]	8.77	0.190	5,836,355
plateau $1/20$	.40	.2500	25,736 [±415]	8.77	0.212	6,533,393
plateau $3/10$	.40	.2500	32,458 [±589]	9.88	0.270	8,299,138
sine 0/1/2	.40	.2630–.2675	34,331–35,395	10.48–10.60	0.297–0.302	9.10–9.28M
cap at 0	.40	—	505	—	—	—
cap at .5	.40	—	256	—	—	—
cap double	.40	—	3,408	—	—	—

Table 9: Executed compatible-point scaling audit (128 predeclared replicates per setting; zero wrong, zero abstain in both kernels). Spend and displayed pointwise $t95$ half-widths are for the binary kernel, with 127 degrees of freedom and no multiplicity adjustment. "Pack. lower" is the certified finite-packing lower bound $\log(1/(2\epsilon))\sum_z \bar{\kappa}_z^{-1}$ of Appendix N recomputed at this point by the exact rational machinery; "Cor. lower" is the closed form $\log(1/(2\epsilon))\, c(2-c)L\int m^{-3}/288$ at the common $c = .25$, with the certified integral enclosure's lower end; "Upper (complete)" is Theorem 4D's full $\ell(CL\int m^{-3} + 36/m(0)^2) + n_0 H + \alpha T_{\max}$ with the enclosure's upper end. The unabsorbed additive terms are $n_0 H = 1{,}008$ and $\alpha T_{\max} = 25{,}804.8$ on every row.

What the audit establishes and what it does not. First, larger margins reduce observed spend on the constant sweep. The endpoint ratio $59{,}992/7{,}618 = 7.87$ is near the cubic ratio 8, but this does not identify an exponent: intermediate ratios do not all match a cubic law, and an upper bound is not an expected-spend equality. Second, observed mean spend sits strictly inside the certified bracket on every insensitive setting: it exceeds the instance-specific packing lower bound by a factor of 2.2–$4.2 \cdot 10^3$ and is exceeded by the complete upper bound by a factor of approximately 154–265, so the two certified quantities order correctly on executed runs, with the additive terms reported unabsorbed. Third, the sensitive-side resolution the theorem promises on G is exercised: all 384 cap runs report SENSITIVE. These last three controls use the clipped contrast $\max\{-1, \min\{1, \Delta_{\text{tent}}(z)\}\}$ implied by the implemented Bernoulli probabilities. Negative flanks can attain larger absolute sensitivity than the positive tent peak. They are not the localized base-to-cap alternatives used to construct the insensitive packing bounds. The bracket remains two to three orders of magnitude wide on both sides, as the compatible-point discussion already states; the audit does not narrow the constants, the packing is a conservative finite container rather than the continuous optimum, and the corollary's closed form undershoots the instance-specific packing bound because the Vitali step loses a factor the explicit containers do not. Nothing here measures spend on real submissions or certifies a deployment property.

I METRIC-ENTROPY FORM OF THE BRACKET

Theorem 5 (metric-entropy form). Let $(\mathcal{Z}, \rho)$ be a compact metric space, write $M(\cdot)$ for its packing number and $N(\cdot)$ for its covering number at the stated radii, and let Δ be L-Lipschitz with respect to ρ. The two sides are stated at different levels and must not be combined into a single per-candidate claim. *Necessary, per candidate.* Planting one tent of height 2γ and slope L at each point of a $4\gamma/L$-packing keeps the supports disjoint, so Theorem 3 applies verbatim and a procedure meeting $(\alpha_{\text{err}}, \alpha_{\text{abs}})$ *on that one candidate* needs $n \geq M(4\gamma/L)\log(1/(2\epsilon))/\text{kl}$ pairs on it. *Sufficient, family-wise over p candidates.* Querying the centres of a $\gamma/(4L)$-cover keeps the geometric allowance at $\gamma/4$, and widening the union bound to Np tests keeps the deviation radius at $\gamma/4$ at *family-wise* level α, so

$$M\big(\tfrac{4\gamma}{L}\big)\,\frac{\log(1/(2\epsilon))}{\text{kl}} \;\leq\; n_i, \qquad \sum_{i\leq p} n_i \;\leq\; p\,N\big(\tfrac{\gamma}{4L}\big)\left\lceil \frac{32\log(2Np/\alpha)}{\gamma^2}\right\rceil. \tag{11}$$

The two radii differ by a factor of 16, so on a general compact metric space Eq. (11) does *not* by itself force the two sides to share an order. On $[-1,1]^d$ with the max metric, a $\lfloor L/(2\gamma)\rfloor^d$

tent packing and a $(\lceil 4L/\gamma \rceil + 1)^d$ grid cover give both sides order $(L/\gamma)^d \gamma^{-2}$ with ratio at most $\left((\lceil 4L/\gamma \rceil + 1)/\lfloor L/(2\gamma) \rfloor\right)^d$ times one logarithm, a per-dimension multiplier that converges to 8 as $L/\gamma \to \infty$ (`simulation/code/cover_packing_ratio.py`); the $d = 1$ necessary row uses the explicit construction of Eq. (10). Under interiority and fixed error levels, Theorems 3 and 4 share the order $\Theta(L/\gamma^3)$; their explicit ratio $256\,\mathrm{kl}/\gamma^2 \cdot \log(2kp/\alpha)/\log(1/(2\epsilon))$ grows only logarithmically in L and p (`simulation/code/bracket_ratio.py`). Neither statement is a minimax-optimality, transfer, or safety claim.

J DIRECT-SPECIALIZATION BOUNDARY, THEOREM BY THEOREM

Theorem 1. GP level-set estimation gives noisy confidence-based above/below labels and deferred decisions on a finite domain (Gotovos et al., 2013). Allocating its failure probability across candidates gives family-wise point labels. Its GP-sample, Gaussian-noise statement does not specialize to a uniform continuous Lipschitz class or return both endpoints of the bounded-pair supremum; Theorem 1 does. Thresholding those simultaneous supremum intervals gives the three candidate reports as a direct corollary. *Theorem 2.* DIRECT and LIPO give Lipschitz envelopes, and LIPO gives a covering-controlled optimization guarantee (Jones et al., 1993; Malherbe & Vayatis, 2017). Applying the envelope to Δ and $-\Delta$ on a ρ-net directly yields the $L\rho$ upper allowance. Their sequential, noiseless optimization-loss statements do not yield the exact bounded fixed-trace lower gap $\min\{L\rho, 1\}$. *Theorems 3 and 4.* Noisy $\mathcal{X}$-armed regret and its simple-regret conversion use near-optimality packings (Bubeck et al., 2011); the one-dimensional Lipschitz exponent agrees with L/γ^3. Theorem 4 follows from the displayed finite cover and family-wise concentration. The regret statement does not, by parameter substitution, impose separate wrong-decision and failure-to-resolve probabilities for Theorem 3's threshold alternatives. *Theorem 4C.* The closest fixed-confidence continuum-Lipschitz result, the stochastic guarantee of de Montbrun & Gerchinovitz (2024), upper-bounds the samples for a certified ε-best arm and leaves an instance-dependent stochastic lower bound open. The exact-evaluation Lipschitz level-set bounds of Bachoc et al. (2021a) assume a nonempty target level set, whereas 4C starts from an insensitive base with an empty τ-level set and uses bandit change of measure (Lattimore & Szepesvári, 2020). Thus neither cited statement specializes to its KL-weighted $Q(\Delta_0)$ lower bound. *Theorem 4D.* Running the fixed-confidence optimizer on Δ and $-\Delta$, splitting confidence across candidates, and thresholding its envelopes gives the same three-way family-wise validity. Its published sample bound is organized by near-maximizer layers, however, and parameter substitution does not yield 4D's threshold-margin integral or its conditional comparison to Q. Bachoc et al.'s exact-value level-set result and GP-LSE's finite-domain, GP-information-gain result (Gotovos et al., 2013) do not yield that bound either. The supported separation is limited to 4C's noisy instance lower bound and 4D's threshold-margin upper comparison under (C1)–(C5); the reporting transform and worst-case exponent are shared statistical content.

K CONDITIONAL CAP-PACKING COMPARISON AUDIT

`simulation/code/f18_instance_optimal_branch_bound.py` was re-executed on twelve fixed insensitive functions with three seeds per function. The previous random generator normalized by the maximum of each query batch, so scalar and vector calls were not evaluations of one function. It now uses a fixed analytic sine envelope; the cap implementation also uses the centre's sign as required by Theorem 4C. The old random-instance comparisons and the descent ratio 184 are withdrawn.

Write $\widehat{Q}_{\mathrm{grid}}$ for the greedy finite-grid cap sum and $\widehat{J}_{\mathrm{grid}}$ for trapezoidal quadrature. Support endpoints and KL maxima are sampled on a grid, without continuum error enclosures. Thus $\widehat{Q}_{\mathrm{grid}}$ is neither the exact supremum Q nor a certified lower bound on it, and $\log(1/(2\epsilon))\widehat{Q}_{\mathrm{grid}}$ is only a scaled proxy. The corrected proxy ratios $\widehat{J}_{\mathrm{grid}}/\widehat{Q}_{\mathrm{grid}}$ range from 4.50 to 9.64 across all instances and from 6.01 to 9.36 in the slow-slope subset. Both slope groups overlap; this audit does not establish necessity of the corollary's slope condition. Mean spend divided by the scaled proxy ranges from 658 to 1,534, not a ratio to the proven continuous lower bound. All 36 runs resolve as insensitive without a wrong or frozen-cell report. These finite calculations do not validate an asymptotic rate or sharp instance-optimality constant.

From the paper root, the source command is `python simulation/code/f18_instance_optimal_branch_bound.py --seeds 3 --L 3 --output f18_rerun.json`.

Table 10: Corrected fixed-function audit at $L = 3$, $\tau = .40$, $\gamma = .06$, $\alpha = .05$, $p = 12$, $n_0 = 32$, with three seeds. Q_g is a greedy sampled-support/KL proxy, not the exact cap-packing supremum or a certified lower bound; J_g is trapezoidal quadrature for $L \int_{-1}^{1} \max\{\tau - |\Delta_0(z)|, \gamma/4\}^{-3}\, dz$. $S_g = \log(1/(2\epsilon))Q_g$ with $\epsilon = .15$. Ratios are numerical diagnostics only. The original batch-dependent random curves and unsigned negative caps were corrected.

Instance	param.	caps	Q_g	J_g/Q_g	spend	spend/S_g	spend/J_g
constant	–	26	2969	9.36	5.404e+6	1512	195
sawtooth	0.05	15	1713	5.21	1.760e+6	853	197
sawtooth	0.2	6	600	4.50	4.755e+5	658	176
descent	–	20	2284	9.64	4.218e+6	1534	191
plateau	0.05	16	924	6.01	1.147e+6	1031	207
plateau	0.3	19	1563	7.90	2.427e+6	1290	197
random 1	–	17	1177	6.55	1.602e+6	1130	208
random 2	–	20	1727	7.03	2.767e+6	1331	228
random 3	–	20	1698	6.99	2.754e+6	1347	232
random 4	–	21	1932	7.56	3.326e+6	1430	228
random 5	–	20	1624	6.71	2.256e+6	1154	207
random 6	–	21	1966	7.89	3.454e+6	1460	223

L External evidence scope details

Certification requires a compact candidate-specific reference domain, paired joint interventions of the form in Eq. (1) that fix every other coordinate while toggling X_i, and a Lipschitz constant L_i that is valid rather than fitted. Observational logs do not establish the paired intervention contrast. Single-factor experiments supply the first two only when every remaining coordinate is held fixed by design, and no procedure here certifies L_i from data, as the misspecified-L witness shows it must not. A surface missing any requirement therefore leaves the certificate UNRESOLVED; selective withholding is the supported status until an external surface supplies all three together.

M Why the earlier multidimensional audit is withdrawn

The 3D/4D runner used an axis-cross set but treated .25 as though it were the covering radius of the whole cube. This is false: points with multiple nonzero coordinates can be much farther than .25 from every axis-cross point. Moreover, a maximum finite difference on restricted coordinate lines is an empirical derivative estimate, not a certified upper bound. Consequently the reported intervals were not established for $\sup_{z\in[-1,1]^d} |\Delta(z)|$. We withdraw the decision counts rather than reinterpret them as a full-cube result. The metric-entropy theorem applies only after a genuine cover and valid Lipschitz bound are supplied.

N Finite packing certificates and oracle-specific lower bounds

Certified scope. The local certified-packing artifact records twelve corrected F18 functions, interpreted using their declared rational constants and recorded binary-float sine parameters. It certifies conservative finite packings, not the exact continuous packing supremum Q. F18 uses a binary Bernoulli oracle, unlike the complementary paired-difference oracle in the main evaluation. The implementation `simulation/code/f19_certified_cap_packing.py` uses rational arithmetic and 40-digit interval enclosures. The regressions in `simulation/code/tests/test_f19_certified_cap_packing.py` independently check analytic integrals, small finite packing optima, overlapping support containers, and the oracle KL identity. This artifact is local preparation, not publicly released data.

Conditional proposition. Let $\Delta_0 : [-1, 1] \to [-1, 1]$ be L-Lipschitz, where $L > 0$, with $M = \sup_x |\Delta_0(x)| \le \tau - \gamma$, $0 < \gamma < \tau$, and $\tau + \gamma < 1$. For each centre z in a finite set Z, fix

$s_z = \operatorname{sign}(\Delta_0(z))$, taking $s_z = 1$ at zero, and define

$$\Delta^{(z)}(x) = s_z \max\{s_z \Delta_0(x), \tau + \gamma - L|x - z|\}.$$

Assume each cap is admissible in the declared class, its change support $S_z = \{x : \Delta^{(z)}(x) \neq \Delta_0(x)\}$ lies in a certified container I_z, and the containers are disjoint. Let $0 < \bar{\kappa}_z < \infty$ bound the base-to-cap single-observation KL divergence throughout I_z. Assume fresh conditional product observations at predictable queries. A procedure has wrong-decision probability at most α_{err} throughout the class and abstention probability at most α_{abs} on margin-separated instances. Put $\epsilon = 2\alpha_{\mathrm{err}} + \alpha_{\mathrm{abs}} \in (0, 1/2)$ and let T count oracle observations. Then

$$\mathbb{E}_0 T \geq \log(1/(2\epsilon)) \sum_{z \in Z} \bar{\kappa}_z^{-1}.$$

Proof. For the event E of reporting insensitive, $P_0(E^c) + P_z(E) \leq \epsilon$. Binary change of measure therefore gives $\mathrm{KL}(P_0 \| P_z) \geq \log(1/(2\epsilon))$, where P_0, P_z are transcript laws. The adaptive KL chain rule bounds this divergence by $\bar{\kappa}_z \mathbb{E}_0 N(I_z)$, with $N(I_z)$ the number of queries in I_z. Outside the support the kernels agree. Summing and using $\sum_z N(I_z) \leq T$ proves the claim; infinite expected spend is trivial. Neither optimal packing nor slope absorption is required.

Oracle conversion. For contrasts a, b, set $p = (1 + a)/2$ and $q = (1 + b)/2$. The binary oracle has divergence $\mathrm{kl}(p\|q)$, where kl denotes Bernoulli KL. For independent costs $U \sim \mathrm{Bernoulli}(p)$ and $V \sim \mathrm{Bernoulli}(1 - p)$, the joint divergence is $2\mathrm{kl}(p\|q)$. Moreover, $D = U - V$ satisfies $D + 1 \sim \mathrm{Binomial}(2, p)$; conditional on $D = 0$, the two merged outcomes have equal, parameter-independent probabilities. Thus observing the difference preserves that joint divergence. Corresponding conservative packing weights, and hence lower bounds at fixed ϵ, halve when counting these pairs. This conversion applies only to the specified complementary independent kernel, not arbitrary paired costs, and is not applied to F18.

Same-oracle fixed-confidence bracket. Let Q_{bin} be Theorem 4C's packing functional, and let Q_{pair} use the complementary paired kernel just specified. For every admissible cap its KL is doubled, so $Q_{\mathrm{pair}} = Q_{\mathrm{bin}}/2$. Assume all of (C1)–(C5), including the fresh conditional-product clause, and set $\epsilon = 3\alpha$. The paired observation $D \in [-1, 1]$ has the same mean Δ and range as the binary observation; hence every Hoeffding event and geometric step in Theorem 4D remains valid when T_{pair} counts pairs. Writing $\ell = \log(2pHK/\alpha)$, its complete expected-spend upper bound is

$$\mathbb{E}_0 T_{\mathrm{pair}} \leq \ell \left(CL \int_{-1}^{1} m(z)^{-3}\, dz + 36/m(0)^2 \right) + n_0 H + \alpha T_{\max}.$$

The same paired experiment has lower bound $\log(1/(6\alpha)) Q_{\mathrm{pair}}$. Under the absorption conditions, replacing Q_{bin} by $2Q_{\mathrm{pair}}$ in the binary corollary gives

$$\log(1/(6\alpha)) Q_{\mathrm{pair}} \leq \mathbb{E}_0 T_{\mathrm{pair}} \leq \frac{2(576C + 144)\ell}{c(2 - c)} Q_{\mathrm{pair}}.$$

Thus both sides refer to one fixed oracle and one error/abstention criterion; the upper bound does not exploit the paired variance reduction. In primitive-draw units both paired endpoints are multiplied by two: the lower endpoint equals the binary lower endpoint, while the stated upper bound doubles. This is a conditional bracket, not equality of optimal costs across oracles, and it does not extend the KL identity to arbitrary paired cost distributions.

Numerical ledger. The ledger specializes to $\tau = .4$, $\gamma = .06$, $L = 3$, and $\epsilon = .15$, using a certified lower enclosure of $\log(10/3)$. The table rounds the exact binary lower bounds downward to two decimal places; repeated family names follow artifact order.

Instance	Caps	Binary lower bound
Constant	24	1582.02
Sawtooth 1	4	253.94
Sawtooth 2	4	235.52
Descent	4	263.67
Plateau 1	15	476.02
Plateau 2	18	836.98
Random 0	17	638.86
Random 1	20	982.61
Random 2	20	970.26
Random 3	21	1105.28
Random 4	19	882.41
Random 5	22	1181.18

For the positive true margin $m(x) = \tau - |\Delta_0(x)|$, the ledger encloses $J = L\int_{-1}^{1} m(x)^{-3}\,dx$, without clipping. Relative widths $(J_+ - J_-)/J_-$ range from zero to 0.00986417, with J_-, J_+ the enclosure endpoints.

Unchanged limitations. The absorption condition at $\gamma = .06$, $\alpha = .05$ remains false. Historical three-seed means remain descriptive, not twenty-seed results or expectation certificates. Negative Bernstein results, mixed allocation efficiency, radius-cap and understated-L failures remain. Withdrawn multiplicity, mesh-only, adaptive-count, original F18 ratio, and axis-cross claims stay withdrawn. These certificates establish neither sharp optimality nor unknown smoothness, external access, transfer, or policy safety; model review is not independent human validation.

Reproduction record. The entry point is `simulation/code/f19_certified_cap_packing.py`. JSON result paths in this appendix are relative to the project evidence bundle (frozen 2026-09-08) at the repository root. Source and configuration bindings accompany the local records; this does not establish independent external replication.

O PROOF OF THE ADAPTIVE FRESH-BATCH CERTIFICATE

This is the full proof of Theorem 4A, including its recycled-cap specialization.

Conditional on the preceding filtration, each opened node uses a predictable location and size, and its new observations follow the corresponding conditional product kernel: they are mutually independent, bounded, and have the queried contrast as conditional mean. No marginal independence between the chosen query and the earlier transcript is required. Formally, let E_v be the event that node v is opened and its single selected fresh-batch mean violates radius r_v. Conditional on the sigma-field immediately before that batch, both n_v and the query are fixed, so

$$\Pr(E_v \mid \mathcal{F}_{v^-}) \le \mathbf{1}\{v \text{ opened}\}\alpha/(pH).$$

Taking expectations gives $\Pr(E_v) \le \alpha/(pH)$, despite the continuum of possible histories and integer batch sizes. Thus no union over those histories or sizes is needed: only one post-selection fresh batch is used at each node. A union bound over the pH potential node events, with unopened-node events empty, covers every adaptive transcript. Repeated cumulative prefixes are instead covered by Theorem 4D's explicit stage index. On this event $\underline{M} \le M$, and Lipschitz continuity makes every $U(I)$ a valid upper bound on its cell. A cell pruned below $\underline{M}$ remains covered by the retained $\underline{M}$, and a threshold-pruned cell remains covered by the retained τ; hence the stated upper endpoint continues to cover M after every pruning step. If $U(I) < \underline{M}$, then $\sup_{z\in I}|\Delta(z)| \le U(I) < \underline{M} \le M$, so I cannot contain a global maximizer. If $U(I) \le \tau$, that cell cannot witness sensitivity. Hence either pruning rule preserves the three-way report. At depth D, the cell allowance is at most $\gamma/4$ and each radius is at most $\gamma/4$, so a sensitive instance has a leaf lower endpoint above τ unless already resolved; for an insensitive instance, every remaining leaf has upper endpoint below τ, and threshold-pruned cells contribute only τ. Thus every margin-separated instance resolves. Finally $H < 16L/\gamma$ when $L > \gamma/4$ gives the displayed rate; the $D = 0$ case uses the one-node bound stated above. For the instance bound, the opened centre of any depth-d parent cell gives $U(I) \le |\Delta(c_I)| + 2r + L\,\mathrm{rad}(I)$ on the simultaneous event. Therefore a parent outside $A_d(\Delta)$

is pruned before splitting, independently of priority order. If a sensitive lower endpoint is attained, the algorithm stops before any further pruning or splitting. Thus, on every transcript that continues to split, only cells counted by A_d can add their two children, proving the node and fixed-batch pair bounds. For the recycled-cap rule, initially $B - n_0 \geq n_0(H-1)$. More generally, if $B_{\mathrm{rem}} \geq n_0 R$, where $R \geq 2$ whenever a split occurs, then $q = \lfloor B_{\mathrm{rem}}/R \rfloor \geq n_0$, $2q \leq B_{\mathrm{rem}}$, and

$$B_{\mathrm{rem}} - 2q \geq B_{\mathrm{rem}}(R-2)/R \geq n_0(R-2).$$

Pruning only decreases the future value of R, so induction proves both the per-node lower bound and the total cap. Hence every recycled radius is at most r_0; the fixed-batch resolution argument and the parent inequality defining H_Δ apply verbatim with r_0. □

P CERTIFIED FINITE-FAMILY QUERY BRACKETS

To measure the remaining finite-instance separation, we freeze a restriction of the continuous observation model, not the earlier graph-search oracle. Set $\tau = 2/5$, $\gamma = 1/4$, and use the constant null $\Delta_0 = 3/20$ with $m = \lfloor L/(2\gamma) \rfloor$ alternatives obtained by adding height-$1/2$, slope-L tents with pairwise-disjoint positive regions. The range is $[3/20, 13/20] \subset [-3/4, 3/4]$, so common $c = 1/4$ satisfies C4. At any predictable location, including an off-grid location, a fresh binary contrast has success probability $(1 + \Delta(z))/2$. One observation costs one pair.

Let $V_m^\star$ be the infimum of expected null spend $\mathbb{E}_0 T$ over legal procedures having wrong-decision probability at most $1/20$ under every member of this finite family and zero abstention. Theorem 3 states a fixed-total budget bound. Its stopped-transcript extension gives the expected-spend bound used here: truncate the transcript at $T \wedge k$, apply the conditional KL chain rule, and pass to the stopped transcript and expected query counts. Pairwise-disjoint positive tent regions ensure $\sum_j N_j \leq T$; zero-height shared endpoints contribute no information. Thus the same change-of-measure argument yields

$$\ell_m = \frac{m \log 5}{\mathrm{kl}(23/40 \| 33/40)} \leq V_m^\star.$$

This is a lower bound over all legal adaptive schedulers, not only scans.

A legal upper construction queries each tent centre n times and declares sensitivity if any positive-observation count reaches h. Writing $F_p(h-1) = \Pr\{\mathrm{Bin}(n,p) < h\}$, its exact null and alternative errors are

$$1 - F_{23/40}(h-1)^m, \qquad F_{33/40}(h-1)F_{23/40}(h-1)^{m-1}.$$

Both must be at most $1/20$. Exhaustive rational enumeration in increasing n finds the minimum within this symmetric fixed-centre scan class; its deterministic spend $u_m = mn$ gives $V_m^\star \leq u_m$. The known $L = 2$ instance is the primary case; $L = 13/4, 17/4$ were fixed before evaluating the two larger confirmation families. Development used only $m = 1, 2, 3$. These deterministic families are not population replicates.

m	L	(n, h)	ℓ_m (rounded down)	u_m	u_m/ℓ_m (approx.)
4	2	$(50, 37)$	37.976	200	5.27
6	$13/4$	$(55, 41)$	56.964	330	5.79
8	$17/4$	$(59, 44)$	75.952	472	6.21

The corresponding exact error pairs, shown approximately, are $(.046416, .042796)$, $(.039163, .045849)$, and $(.038738, .042058)$. Independent rational enumeration reproduces every scheduler and error. Logarithms and propagated endpoints use outward enclosures, checked against higher precision before execution. The certified bracket widths are approximately 162.024, 273.036, and 396.048 pairs. These widths are not the scheduler's true optimality gaps: neither endpoint is proved globally tight. The prespecified exact-match condition fails in all three cases. The result quantifies a remaining constant-factor uncertainty rather than establishing finite-instance optimality. The executable protocol and complete rational fractions are retained in the query-gap freeze record and the query-gap result record; the earlier 2,304 paired records are unchanged and are not reused as observations in this calculation.

Q PROOF OF THE TRACE-EXACT LIPSCHITZ ENVELOPE

Feasibility. Necessity of the pairwise condition follows from any $f \in \mathcal{F}(I)$:

$$a_j \le f(g_j) \le f(g_k) + Ld(g_j, g_k) \le b_k + Ld(g_j, g_k).$$

For sufficiency, define the upper and lower extremal extensions

$$p(z) = \min\left\{1, \min_k [b_k + Ld(z, g_k)]\right\},$$

$$q(z) = \max\left\{-1, \max_k [a_k - Ld(z, g_k)]\right\}.$$

These are the classical McShane–Whitney extremal extensions of the grid constraints (McShane, 1934; Whitney, 1934), truncated to the range $[-1, 1]$; the verification below specializes the standard argument to interval data. The reverse triangle inequality makes every cone in these displays L-Lipschitz. A finite pointwise minimum or maximum of real-valued L-Lipschitz functions is again L-Lipschitz, and truncation by a constant preserves that property. Also $p, q \in [-1, 1]$. At every grid point, $p(g_j) \le b_j$, while the pairwise condition gives $p(g_j) \ge a_j$; hence $p \in \mathcal{F}(I)$. Similarly, $q(g_j) \ge a_j$ and the same condition with the indices exchanged gives $q(g_j) \le b_j$, so $q \in \mathcal{F}(I)$. This proves sufficiency and also constructs both extremal extensions.

Sharp upper endpoint. For any $f \in \mathcal{F}(I)$ and every z, k, Lipschitz continuity and the grid intervals give

$$f(z) \le b_k + Ld(z, g_k), \qquad f(z) \ge a_k - Ld(z, g_k).$$

Together with the range constraint, these inequalities say $q(z) \le f(z) \le p(z)$. Therefore

$$|f(z)| = \max\{f(z), -f(z)\} \le \max\{p(z), -q(z)\},$$

and taking suprema proves $\sup_z |f(z)| \le U_I$. The functions p and q are continuous, so compactness of $\mathcal{Z}$ makes the maximum defining U_I attainable. If its first branch is active at a maximizer z_+, then $\sup_z |p(z)| \ge p(z_+) = U_I$; if its second branch is active at z_-, then $\sup_z |q(z)| \ge -q(z_-) = U_I$. The preceding universal upper bound supplies the reverse inequality in either case. Thus some member of $\mathcal{F}(I)$ attains U_I.

Sharp lower endpoint and the full identified set. Every $f \in \mathcal{F}(I)$ satisfies

$$\sup_z |f(z)| \ge \max_k |f(g_k)| \ge \max_k \operatorname{dist}(0, I_k) = m_I.$$

Fix any $h \in \mathcal{F}(I)$ and let $\Pi_m(x) = \max\{-m_I, \min\{x, m_I\}\}$. The map Π_m is nonexpansive, so $f_{\min} = \Pi_m \circ h$ is L-Lipschitz and has supremum magnitude at most m_I. Each I_k intersects $[-m_I, m_I]$ by the definition of m_I; because I_k is an interval, clipping any of its points to $[-m_I, m_I]$ remains in I_k. Hence $f_{\min} \in \mathcal{F}(I)$, and the lower bound above shows that its supremum magnitude is exactly m_I.

Let $f_{\max}$ be the extremal extension attaining U_I. For $t \in [0, 1]$, the convex combination $f_t = (1 - t)f_{\min} + tf_{\max}$ remains in $\mathcal{F}(I)$. Moreover $f \mapsto \sup_z |f(z)|$ is continuous in the uniform norm. Its values along $(f_t)_{t\in[0,1]}$ therefore include every number between m_I and U_I. The two sharp bounds exclude every other number, proving Eq. (8).

Confidence refinement and zero-trace specialization. On Theorem 1's simultaneous concentration event, the true Δ_i belongs to $\mathcal{F}(I_i)$ for every candidate, so every such class is feasible and the exact identified interval covers M_i. Since $\overline{D}_{ik} \in [-1, 1]$,

$$\operatorname{dist}(0, I_{ik}) = (|\overline{D}_{ik}| - r_{ik})_+,$$

which proves $m_{I_i} = \underline{M}_i$. For any z, choose a grid point g_k within ρ_i. Its two cone bounds obey

$$p(z) \le \min\{1, |\overline{D}_{ik}| + r_{ik} + L_i\rho_i\},$$

$$-q(z) \le \min\{1, |\overline{D}_{ik}| + r_{ik} + L_i\rho_i\}.$$

Consequently $U_{I_i} \le \overline{M}_i$. Returning UNRESOLVED when the observed intervals are infeasible cannot create a wrong decisive report; on the simultaneous event this fallback is never invoked. Theorem 1's family-wise guarantee is therefore preserved.

Finally, for the exact zero trace $I_k = \{0\}$, $p(z) = -q(z) = \min\{1, Ld(z, G)\}$. Compactness gives $m_I = 0$ and $U_I = \min\{1, L\rho\}$, which is precisely the worst-case witness of Theorem 2. □

R DIMENSION-EXPLICIT INSTANCE LAW

Throughout, $\mathcal{Z} = [-1,1]^d$ carries the max metric, $d \geq 1$, and every $|z' - z|$ in Theorem 4C is read as $\|z' - z\|_\infty$. Let $\Delta_0 : \mathcal{Z} \to [-1,1]$ be L-Lipschitz with $M(\Delta_0) \leq \tau - \gamma$ and $\tau + \gamma < 1$, and write $m(z) = \tau - |\Delta_0(z)| \geq \gamma$, $h_z = \tau + \gamma - |\Delta_0(z)| = m(z) + \gamma$, and $J = \int_{\mathcal{Z}} m(z)^{-(d+2)}\, dz$. Caps $\Delta^{(z)}$, supports S_z, divergences κ_z, cap packings and $Q(\Delta_0)$ are as in Theorem 4C. At $d = 1$ every constant below reduces to one already published in Appendix G; those reductions are stated inline and are the arithmetic check on the general-d derivation.

Caps and their containers. The admissibility argument of Appendix G uses only that $s_z\Delta_0$ and $T_z = \tau + \gamma - L\|\cdot - z\|_\infty$ are L-Lipschitz and that the pointwise maximum of two L-Lipschitz functions is L-Lipschitz; both hold in any metric space. Hence every cap is L-Lipschitz, $[-1,1]$-valued, agrees with Δ_0 off S_z, satisfies $M(\Delta^{(z)}) \geq \tau + \gamma$, and obeys $0 < s_z[\Delta^{(z)} - \Delta_0] \leq h_z$ on S_z. Under (C3) the base is $(L/16)$-Lipschitz, so $T_z(z') - s_z\Delta_0(z') \leq h_z - \frac{15L}{16}\|z' - z\|_\infty$ and, from below, $T_z(z') - s_z\Delta_0(z') \geq h_z - \frac{17L}{16}\|z' - z\|_\infty$. The two estimates give an outer container and an inner ball,

$$S_z \subseteq B_z := \{z' : \|z' - z\|_\infty < w_z\}, \qquad w_z = \frac{16h_z}{15L},$$

$$\{z' : \|z' - z\|_\infty < \rho_z\} \subseteq S_z, \qquad \rho_z = \frac{16h_z}{17L}.$$

The outer container restates the one-dimensional lemma; the inner ball is used only for the optional C3-specific refinement in part (c).

Part (a): the lower bound. The change-of-measure and adaptive chain-rule steps of Theorem 4C's proof never use the domain, so $\mathbb{E}_0 T \geq \log(1/(2\epsilon))\, Q(\Delta_0)$ holds verbatim on $\mathcal{Z}$. For the closed form, the $5r$ covering lemma is valid in any metric space: the radii w_z are bounded, so a countable disjoint subfamily $\{B_{z_j}\}$ satisfies $\mathcal{Z} \subseteq \bigcup_j 5B_{z_j}$. Disjoint containers give disjoint supports, so $\{z_j\}$ is a cap packing. Each $5B_{z_j}$ is a cube of side $32h_{z_j}/(3L)$, and on it the $(L/16)$-Lipschitz function h satisfies $h \geq h_{z_j} - (L/16)5w_{z_j} = 2h_{z_j}/3$. The displacement is the radius $5w_{z_j}$, not the cube's full side length. Hence

$$\int_{5B_{z_j}} h^{-(d+2)} \leq \Big(\frac{32h_{z_j}}{3L}\Big)^d \Big(\frac{3}{2h_{z_j}}\Big)^{d+2} = \frac{9 \cdot 2^{4d-2}}{L^d}\, h_{z_j}^{-2}.$$

Summing over the covering and using $h \leq 2m$ gives

$$\sum_j h_{z_j}^{-2} \;\geq\; \frac{L^d J}{9 \cdot 2^{5d}}.$$

With Theorem 4C's divergence ceiling under (C4), $Q(\Delta_0) \geq c(2-c)L^d J/(9 \cdot 2^{5d})$. The denominator is 288 at $d = 1$, improving the former value by 2^{d+2} (factors 8, 16, and 32 for $d = 1, 2, 3$).

Part (b): the upper bound. Run Theorem 4A's rule on the dyadic tree of $\mathcal{Z}$ with 2^d children per node, $\mathrm{rad}(I) = 2^{-\mathrm{depth}}$ in the max metric, $D = \lceil \log_2(4L/\gamma) \rceil \geq 1$, and complete node count $H = \sum_{j=0}^{D} 2^{dj} = (2^{d(D+1)} - 1)/(2^d - 1)$. Validity, halting and resolution are Theorem 4D's arguments verbatim: the Hoeffding union bound runs over pHK predeclared prefixes and does not see the domain; halting counts at most HK selections and $T_{\max} = H2^{K-1}n_0$ pairs; and resolution uses only $L\,\mathrm{rad} \leq \gamma/4$ at leaves together with $L2^{-j} > \gamma/4 \geq r$ at stage K for $j < D$, both statements about radii.

For the spend, work on the confidence event with Δ_0 insensitive and write $\log$ for $\log(2pHK/\alpha)$. Steps (i) and (ii) are unchanged: a cell of radius ρ splits only when $r \leq L\rho$ and $U(I) > \tau$, which forces $m(c) < 3L\rho$, and its doubling stops with $N_I < 4\log/(L\rho)^2$. (iii) A depth-j cell is opened only if its parent, of radius 2ρ with $\rho = 2^{-j}$, was split, so $m(c_{\mathrm{parent}}) < 6L\rho$ and every z in the cell has $m(z) < 8L\rho$; opened depth-j cells are disjoint cubes of volume $2^{d(1-j)}$ inside

$A_j = \{m < 8L2^{-j}\}$ and hence number at most $2^{d(j-1)}|A_j|$. (iv) Summing (ii) over (iii) for $1 \le j < D$,

$$\frac{4\log}{2^d L^2}\sum_j 2^{(d+2)j}|A_j| \le \frac{4\log}{2^d L^2}\int_{\mathcal{Z}} \sum_{j:\,2^j<8L/m(z)} 2^{(d+2)j}\,dz \le \frac{2^{3d+10}}{2^{d+2}-1}\,L^d \log\, J,$$

using $\sum_{j\le J_0} 2^{(d+2)j} \le 2^{(d+2)J_0}2^{d+2}/(2^{d+2}-1)$. At $d=1$ this constant is $2^{13}/7 = 8192/7$, the published one.

(v) *Root.* Since m is L-Lipschitz, $m(z) \le m(0) + L\|z\|_\infty$. When $m(0) < 3L$ the root is pruned or split no later than its first stage with $r \le L$, costing $N < 4\log/L^2$ beyond the n_0 already counted. The layer-cake identity on $\mathcal{Z}$, where $\{\|z\|_\infty \le t\}$ has volume $(2t)^d$, gives

$$J \ge \int_{\mathcal{Z}} (3L + L\|z\|_\infty)^{-(d+2)}dz = \frac{2^d d}{L^{d+2}}\int_0^1 \frac{t^{d-1}}{(3+t)^{d+2}}\,dt \ge \frac{1}{2^{d+4}L^{d+2}},$$

bounding $(3+t)^{-(d+2)} \ge 4^{-(d+2)}$ and using $\int_0^1 t^{d-1}dt = 1/d$. Hence $L^{-2} = L^d L^{-(d+2)} \le 2^{d+4}L^d J$ and the root doubling costs at most $2^{d+6}L^d \log J$. When $m(0) \ge 3L$ a split is impossible on the confidence event, and pruning occurs by the first stage with $2r \le m(0) - L$ at cost at most $36\log/m(0)^2$; that term is retained separately, exactly as in one dimension. At $d=1$ the general bound gives $2^7 = 128$ against the published $576/7 \approx 82.3$: the general-d root constant is deliberately the looser of the two, and C_1 below is quoted accordingly.

(vi) *Leaves.* A depth-D leaf is pruned once $N \ge 8\log/(m(c) - \gamma/4)^2$, and doubling overshoots by at most two. Its radius satisfies $L2^{-D} \le \gamma/4$, so $m(c) \ge m(z) - \gamma/4$ and $m(z) \ge \gamma$ give $m(c) - \gamma/4 \ge m(z)/2$ and a spend of at most $64\log/m(z)^2 \le 64\log/\gamma^2$. Leaves lie in $A_D \subseteq \{m < 2\gamma\}$ and number at most $2^{d(D-1)}|A_D| \le (4L/\gamma)^d|A_D|$ because $2^D < 8L/\gamma$; and $m^{-(d+2)} > (2\gamma)^{-(d+2)}$ on A_D gives $\gamma^{-(d+2)}|A_D| \le 2^{d+2}J$. The leaf spend is therefore at most $2^{3d+8}L^d\log J$, which is $2^{11} = 2048$ at $d=1$.

Adding (iv)–(vi) and the $n_0 H$ term from $\max\{n_0, N\} \le n_0 + N$,

$$T \le \log\Big(C_d L^d J + \frac{36}{m(0)^2}\Big) + n_0 H, \qquad C_d = \frac{2^{3d+10}}{2^{d+2}-1} + 2^{d+6} + 2^{3d+8},$$

and $\mathbb{E}_0 T$ adds $\alpha T_{\max}$. At $d=1$, $C_1 = 8192/7 + 128 + 2048 < 3347 \le 3.4\cdot 10^3$, so Theorem 4D's declared constant is recovered.

Part (c): a C3/C4-free upper converse. Fix any cap packing Z; only the base assumptions at the start of this appendix are used. At a cap centre the contrast gap is h_z, so Pinsker gives $\kappa_z \ge h_z^2/2$ and $\kappa_z^{-1} \le 2h_z^{-2}$, without C4.

For $r = \|x - z\|_\infty$, Lipschitzness of the base gives $T_z(x) - s_z\Delta_0(x) \ge h_z - 2Lr$. Therefore the *open* clipped cube

$$B_z = \{x \in \mathcal{Z} : \|x - z\|_\infty < r_z\}, \qquad r_z = \min\{h_z/(2L), 1\},$$

lies in S_z. Closed-ball containment is not asserted: equality can occur at the boundary. On B_z, $h(x) < 3h_z/2$, and domain clipping leaves volume at least r_z^d. Consequently

$$h_z^{-2} \le (3/2)^{d+2}\max\{2L, 1\}^d \int_{B_z} h(x)^{-(d+2)}\,dx,$$

where $h_z/r_z = \max\{2L, h_z\} \le \max\{2L, 1\}$. The supports, hence these open cubes, are pairwise disjoint. Summing, using $h \ge m$, and taking the packing supremum proves

$$Q(\Delta_0) \le 2(3/2)^{d+2}\max\{2L, 1\}^d J.$$

This is only an upper bound on the cap-packing functional. It removes neither C3 from the reverse Vitali bound nor C4 from its KL ceiling.

Under C3 one may retain the sharper inner radius $\min\{16h_z/(17L), 1\}$ and the corresponding bound $Q \le \frac{648}{289}(\frac{18}{17})^d \max\{17L/16, 1\}^d J$; this refinement is not needed for the broader converse.

Part (d): the conditional match, and what (a) with (c) says. Under (C5$_d$), $H \le 2(8L/\gamma)^d$ and $n_0 \le C_d\gamma^d \log/2^{2d+2}$ give $n_0 H \le 2^{d-1}C_d L^d \log$, and the same computation bounds $\alpha T_{\max}$; since $m \le \tau < 1$ forces $J > 2^d$, the two additive terms together are at most $C_d L^d \log J$, so the margin term is at most doubled. Part (a) converts $L^d J \le 9 \cdot 2^{5d} Q(\Delta_0)/(c(2-c))$, and the one-cap packing at the origin gives $Q \ge c(2-c)/h_0^2 \ge c(2-c)/(4m(0)^2)$, so the retained root term is at most $144 \log Q/(c(2-c))$. Hence

$$\mathbb{E}_0 T \le \frac{(576 \cdot 32^{d-1} C_d + 144)\log}{c(2-c)}\, Q(\Delta_0),$$

which at $d = 1$ is the tightened $576C + 144$ form. Feasibility of (C5$_d$) is as in one dimension: $\alpha 2^{K-1} n_0 \ge 32\alpha \log/\gamma^2$ forces $\alpha \le C_d\gamma^{d+2}/2^{2d+7}$, which at $d = 1$ reads $\alpha \le C\gamma^3/512$.

Parts (a) and (c) give a two-sided order comparison only when the lower bound's C3–C4 assumptions also hold and $L \ge 1/2$ (so $\max\{2L, 1\} = 2L$). Without those assumptions, part (c) remains a valid upper converse but proves no equivalence. Thus no cap packing can exceed the order of the explicit oracle functional on the restricted class. The integral depends on the unknown full surface and is not itself a data-computable certificate. The constants remain large, and no minimax or finite-instance optimality is claimed.

Retained theorem audit. F21 used true dyadic cubes and analytic cosine-mixture Lipschitz bounds. Its executed branch-and-bound component contains 100 insensitive and 30 sensitive $d = 2$ runs plus 30 one-dimensional controls: all 290 upper-bound checks passed, with minimum bound/spend ratio 160.97. Separate geometry checks covered $d \in \{1, 2, 3\}$ under the former, looser constants. These aggregate checks are a theorem/implementation audit, not evidence of competitive multidimensional performance: no equal-cost baseline, runtime comparison, or per-run record was retained. They are not offered as an empirical proof of the tightened constant or the broader converse above.

S Correction and withdrawal log

This appendix records superseded implementation claims; it is not a scientific result. Current values and ledgers appear in Appendices E and A.

Consolidated supersession record. Current tables use the 12-candidate family-wise radius, include the off-mesh Lipschitz allowance in envelope maxima, and use anytime-valid radii at data-dependent cumulative counts. Earlier values lacking any of these controls are superseded. Earlier 3D/4D counts are also excluded: their axis-cross set is not a full-cube cover and restricted-line finite differences do not certify a global Lipschitz constant. The retained F21 $d = 2$ result instead uses full dyadic cubes and analytic Lipschitz bounds, and is described only as a theorem audit in Appendix R.