THE PRICE OF A BLIND JUMP: HOW ROUNDS AND DISTRIBUTIONS CREATE LEARNING PATHS

Anonymous authors
Paper under double-blind review

ABSTRACT

How do adaptive rounds and input distributions change the cost of learning hierarchical structure? We answer this question through the blind-jump width, the number of unknown coordinates a learner must locate together. For fixed-link product juntas with square-integrable likelihood ratios and full detectability, we identify the exact ambient exponent of R-batch full statistical-query (SQ) cost q/τ^2 as the R-wave carrier-closure width BLeap_R. Equal nonadaptive and sequential endpoints can nevertheless yield reversed difficulty rankings at intermediate rounds. A hidden-rung block staircase exhibits three regimes for the same support-recovery task. Under Rademacher inputs, depth D and block width b give exponent $b\lceil D/R\rceil$, with explicit binomial bounds for growing structural parameters. Bounded amplitude perturbations at dispersion $u = \Theta(1/(bD))$ create unary response-energy witnesses: n nonadaptive full SQs recover the support, with an IID implementation using $O_D(b \log n)$ rows and $O(nN)$ arithmetic at fixed signal-to-noise ratio. Every fixed symmetric standardized nonconstant-modulus law instead requires exponentially many passive samples in b, at fixed depth and signal-to-noise ratio. These results quantify how rounds build useful carriers and how distribution design supplies them immediately. Lean 4 formalizations accompany the carrier calculus, lower-bound success estimates, and selected analytic identities.

1 INTRODUCTION

Learning a hierarchical target hinges on finding useful intermediate interactions. Staircase analyses describe how early discoveries expose later features (Abbe et al., 2021; 2022; 2023a), and hierarchical polynomial learning shows how intermediate representations reduce dimension-dependent costs (Nichani et al., 2023; Wang et al., 2024). We characterize the number of unknown coordinates that must be located simultaneously, and how rounds and input distributions change it.

For parity characters, one identity captures the opportunity. If an interaction is supported on $S \cup B$ and the carrier S is already known, then

$$\chi_{S\cup B}\chi_S = \chi_B. \tag{1}$$

The known factor cancels, leaving the fresh block B. Without the carrier, the learner must search over placements of the larger interaction. We call the number of coordinates guessed together the *blind-jump width*. A width-k search has order n^k candidates; if the normalized witness margin is λ, estimation adds a factor of order λ^{-2}. The schematic price

$$\mathrm{Price}(k, \lambda) \asymp n^k \lambda^{-2} \tag{2}$$

separates search from measurement. The SQ results quantify these two factors, and the passive-information results characterize the cost of obtaining the observations themselves.

Adaptive rounds acquire carriers from earlier answers. Input distribution design can expose a unary witness before any target coordinate is known. Favorable product distributions and curricula can aid parity learning (Arpe & Mossel, 2008; Cornacchia & Mossel, 2023; Abbe et al., 2023b); positive distribution shift captures the benefit of selecting training covariates (Medvedev et al., 2026). We

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.

quantify these mechanisms on one hidden target and one recovery task. Rounds shorten blind jumps, while critical dispersion exposes the active union through coordinatewise response energy.

Let $B_1, \dots, B_D$ be disjoint unknown blocks of width b, and write $A_t = B_1 \cup \cdots \cup B_t$. Independently on each observation, draw an unobserved $T \sim \mathrm{Unif}[D]$ and set

$$Y = a\chi_{A_T}(X) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2). \tag{3}$$

The task is recovery of the unordered union A_D. Selecting one rung per row removes the cross-rung products of a deterministic sum, preserving the hierarchy under arbitrary nonlinear label queries.

1.1 Results and contributions

Our first result extends the detectable-set framework of Joshi et al. (2024) to a fixed number of frozen batches. For a known link and fixed latent dimension, let Γ_k add every detectable set requiring at most k fresh coordinates relative to the carrier available at the start of a batch. The minimum k for which R waves reveal the whole support is BLeap_R. Theorem 2.2 identifies this invariant as the exact ambient exponent of full-SQ query–precision cost q/τ^2. Discoveries inside a frozen batch become carriers only in the next batch. Equal Cover and Leap endpoints can conceal a strict reversal of difficulty between two and three rounds (Proposition D.7).

Theorem 1.1 (Two ways to obtain a learning path, informal)**.** *For the hidden-rung model with $n \geq 2bD$ and fixed nonzero signal-to-noise ratio:*

(i) *Under Rademacher inputs, R frozen full-SQ batches have fixed-parameter ambient exponent $b\lceil D/R \rceil$. Explicit binomial bounds also apply when block width grows.*

(ii) *Under one two-radius product design with dispersion $u = \Theta(1/(bD))$, n nonadaptive full SQs recover the same union. An IID implementation uses $O_D(b \log(n/\delta))$ rows and $O(nN)$ arithmetic.*

(iii) *Under every fixed symmetric standardized law with nonconstant modulus, distinguishing adjacent unions requires exponentially many passive samples in b, at fixed depth and signal-to-noise ratio.*

An antisymmetric likelihood slice controls full-SQ search; exact unary moments give critical recovery for bounded amplitudes (Theorem K.2); Gaussian affinity yields the fixed-dispersion obstruction. Appendix A evaluates the resulting learning paths across round budgets, dimensions, depths, noise levels, and amplitude laws.

We provide a Lean 4 formalization using mathlib (de Moura & Ullrich, 2021; The mathlib Community, 2020). It covers the carrier calculus, the general and rigid lower-bound success estimates, and selected moment, optimization, and Hellinger arguments. Appendix B maps these results to their formal statements and assumptions.

2 Carriers, Queries, and Hidden Rungs

2.1 The fixed-link detectable-set interface

We first recall the interface on which the carrier invariant is general. Work on standard Borel sample spaces. Fix a latent dimension P independent of n. Let $(Y, Z_1, \dots, Z_P)$ have a known joint law ρ, with $Z \sim \mu_x^P$, and hide the latent coordinates through an injection $\iota : [P] \hookrightarrow [n]$. The target law ν^ι draws $X \sim \mu_x^n$ and applies the link $\rho_{Y|Z}$ to $(X_{\iota(1)}, \dots, X_{\iota(P)})$. Write ρ_y for the label marginal and

$$\nu_0 = \mu_x^n \otimes \rho_y.$$

A full SQ is any $\phi \in L_2(\nu_0)$; a relative-L_2 oracle of tolerance τ may perturb its expectation by at most $\tau \|\phi\|_{L_2(\nu_0)}$.

Write $L_{2,0}(\mu_x)$ for the centered subspace. A latent set $U \subseteq [P]$ is *detectable* when there exist $T \in L_2(\rho_y)$ and $T_i \in L_{2,0}(\mu_x)$ such that

$$\mathbb{E}\left[T(Y) \prod_{i \in U} T_i(Z_i)\right] \neq 0. \tag{4}$$

Let $\mathcal{C}_{\mathrm{SQ}}(\rho)$ be the resulting set system. This is the full-SQ specialization of the detectable-set interface of Joshi et al. (2024): it separates the analytic question "which latent interactions have witnesses?" from the combinatorial question "how can those witnesses be scheduled?"

2.2 One-wave carrier closure

Let $\mathcal{C} \subseteq 2^{[P]}$ be a detectable set system whose union is $[P]$. For a known latent set A and an integer $k \geq 0$, define one parallel closure wave by

$$\Gamma_k(A) = A \cup \bigcup \{U \in \mathcal{C} : |U \setminus A| \leq k\}. \tag{5}$$

Each frozen batch applies one closure wave; its discoveries become carriers at the start of the next batch.

Definition 2.1 (Batch-leap width)**.** The R-batch blind-jump width is

$$\mathrm{BLeap}_R(\mathcal{C}) = \min\{k : \Gamma_k^{\circ R}(\varnothing) = [P]\}.$$

For the nested chain $\{[1], \ldots, [D]\}$, one closure wave advances by exactly k coordinates, so $\mathrm{BLeap}_R = \lceil D/R \rceil$. If every rung adds a block of b coordinates, the width is multiplied by b.

The invariant characterizes arbitrary fixed detectable set systems, as the next theorem establishes (Appendix E). For $B \subseteq [P]$, let ν_B^ι retain the joint marginal of (Y, Z_B) and decouple all other latent coordinates. A legal oracle may depend on the target and public query history but must answer every coordinate within its true tolerance interval.

Theorem 2.2 (General one-batch–one-closure law)**.** *Fix P, R, and a known fixed link ρ. Assume $\bigcup_{U \in \mathcal{C}_{\mathrm{SQ}}(\rho)} U = [P]$ and*

$$\rho \ll \rho_y \otimes \mu_x^P, \qquad \frac{d\rho}{d(\rho_y \otimes \mu_x^P)} \in L_2(\rho_y \otimes \mu_x^P).$$

Set $k_R = \mathrm{BLeap}_R(\mathcal{C}_{\mathrm{SQ}}(\rho))$. There are positive constants C, c_0, c_1 and n_0, depending on (P, R, ρ) but not on n, such that for every $n \geq n_0$:

- *a deterministic R-batch full-SQ learner recovers $\iota([P])$ using tolerance c_0 and at most Cn^{k_R} queries;*
- *every randomized R-batch full-SQ learner that recovers $\iota([P])$ with probability at least $2/3$ for every injection and every legal oracle satisfies*

$$\frac{q}{\tau^2} \geq c_1 n^{k_R}.$$

Thus BLeap_R is the exact fixed-parameter ambient exponent of query–precision cost, and of query count at constant tolerance.

The lower-bound success estimate is formalized in Lean (Appendix B). For smallest normalized witness correlation β, the proof permits $c_0 = \beta/4$ and $C \leq R2^P P^P$. The finite-round extension of the Cover and Leap endpoints of Joshi et al. (2024) uses conditional truncation after arbitrary earlier frozen transcripts.

The endpoints do not order finite-round difficulty. Consider two prefix chains with block widths $(1, 2, 1, 2)$ and $(1, 1, 2, 2)$. Both have Cover 6 and Leap 2, yet their exact profiles are

Block widths	$R = 1$	$R = 2$	$R = 3$	$R \geq 4$
$(1, 2, 1, 2)$	6	3	3	2
$(1, 1, 2, 2)$	6	4	2	2

The first is easier at two rounds, the second at three. Proposition D.7 proves the table and realizes both families with the same Rademacher inputs, Gaussian label channel, and latent dimension. Their full-SQ costs therefore reverse from $\Theta(n^3)$ versus $\Theta(n^4)$ to $\Theta(n^3)$ versus $\Theta(n^2)$.

2.3 HIDDEN-RUNG SUPPORT RECOVERY

Let $B_1, \dots, B_D \subseteq [n]$ be pairwise disjoint with $|B_t| = b$ and set $A_t = \bigcup_{s \le t} B_s$. A fresh hidden selector $T \sim \mathrm{Unif}[D]$ is drawn on every observation. Given a symmetric standardized coordinate law μ, draw $X \sim \mu^n$ and

$$Y = a \prod_{j \in A_T} X_j + \epsilon, \qquad \epsilon \sim N(0, \sigma^2). \tag{6}$$

The coefficient $a \neq 0$, noise level $\sigma > 0$, and structural parameters (b, D) are known. The output is the unordered active union A_D.

Proposition 2.3 (Cross-rung screening in a deterministic sum)**.** *Under Rademacher inputs, replace (6) by*

$$Y = a \sum_{t=1}^{D} \chi_{A_t}(X) + \epsilon.$$

Then every block is visible nonadaptively through a squared-label query:

$$\mathbb{E}[Y^2 \chi_{B_t}(X)] = 2a^2, \qquad t \in \{2, \dots, D\}.$$

Thus width-b enumeration recovers the blocks in one full-SQ batch. The hidden-rung model removes these cross-rung products and supports the round hierarchy under the full query class.

Proof. Consecutive prefixes contribute $2a^2 \chi_{A_{t-1}} \chi_{A_t} = 2a^2 \chi_{B_t}$ to Y^2; Fourier orthogonality removes all other terms. The first block is visible from $\mathbb{E}[Y \chi_{B_1}(X)] = a$. □

2.4 RIGID FULL-SQ ACCESS

At the rigid endpoint $\mu = \mathrm{Rad}$, every parity is a fair sign. Specialize the preceding full-SQ model to $Q_+ = N(a, \sigma^2)$, $Q_- = N(-a, \sigma^2)$, $\mu_y = (Q_+ + Q_-)/2$, and

$$\nu_0 = \mathrm{Rad}^n \otimes \mu_y.$$

The preceding full-SQ and frozen-batch conventions apply. The oracle may be target-specific and stateful, and sees private learner randomness through the public query history.

2.5 TWO-RADIUS AMPLITUDE PATH

For $0 < u < 1$, independently draw signs $\xi_j \sim \mathrm{Rad}$ and amplitude bits $R_j \sim \mathrm{Rad}$, and set

$$X_j = \xi_j \sqrt{1 + \sqrt{u}\, R_j}. \tag{7}$$

Then $\mathbb{E}X_j^2 = 1$ and $\mathbb{E}X_j^4 = 1 + u$. The value u is the variance of X_j^2. The critical result uses a single fixed design with $u = \Theta(1/(bD))$ for all observations.

For a general standardized coordinate law μ, define

$$M_\alpha(\mu) = \mathbb{E}_\mu |X|^\alpha, \qquad D_\alpha(\mu) = \mathbb{E}|X - X'|^\alpha,$$

where X' is independent. Every fixed nonconstant-modulus law satisfies $M_\alpha(\mu) < 1$ for $0 < \alpha < 2$.

3 MAIN RESULTS

3.1 RIGID INPUTS: DEPTH BECOMES WIDTH

Set

$$L = \left\lceil \frac{D}{R} \right\rceil, \qquad \ell = L - 1, \qquad J = \left\lceil \frac{D}{L} \right\rceil.$$

For $r = 0, \dots, J-1$, let $p_r = rL$ and $s_r = \min\{L, D - p_r\}$. Define

$$K_{n,b,D,R} = \sum_{r=0}^{J-1} \sum_{s=1}^{s_r} \binom{n - bp_r}{bs}, \tag{8}$$

and

$$B_{n,b,D,R} = \binom{n - b(R-1)(L-1)}{bL}. \tag{9}$$

Define the ideal-transcript guessing probability

$$g_{n,b,D,R} = \binom{n - bR(L-1)}{b(D - R(L-1))}^{-1}. \tag{10}$$

Let

$$\beta_{\mathrm{lr}}^2 = \mathbb{E}_{\mu_y} \tanh^2\left(\frac{aY}{\sigma^2}\right), \qquad \gamma_{\mathrm{sgn}} = |2\Phi(a/\sigma) - 1|.$$

Theorem 3.1 (Rigid full-SQ round law)**.** *Assume $n \geq 2bD$ and $1 \leq R \leq D$. At every tolerance*

$$0 < \tau < \frac{\gamma_{\mathrm{sgn}}}{4D},$$

a deterministic learner recovers A_D in at most R batches using exactly $K_{n,b,D,R}$ pointwise-bounded SQs. Conversely, fix $s \in (g_{n,b,D,R}, 1)$. Suppose a randomized R-batch full-SQ learner outputs A_D with probability at least s over its private coins for every fixed target and every legal target-specific stateful oracle. If it uses at most q queries, then

$$\frac{q}{\tau^2} \geq \frac{D(s - g_{n,b,D,R})}{\beta_{\mathrm{lr}}^2} B_{n,b,D,R}. \tag{11}$$

In particular, success $s = 2/3$ requires at least $DB_{n,b,D,R}/(6\beta_{\mathrm{lr}}^2)$ because $g_{n,b,D,R} \leq 1/2$. For fixed (b, D, R) and $n \to \infty$, the query–precision currency q/τ^2 has exact ambient exponent $b\lceil D/R \rceil$; when τ is bounded below independently of n, the query count q has the same exponent. For growing parameters, the explicit binomial bounds retain the dependence on active dimension and density.

The finite-dimensional success estimate yielding the $s = 2/3$ lower bound has a Lean formalization (Appendix B).

Remark 3.2 (Finite-parameter comparison). Let $k = bL$ and $\Delta = b(R-1)(L-1)$. Binomial monotonicity under $n \geq 2bD$ gives

$$\frac{K_{n,b,D,R}}{B_{n,b,D,R}} \leq RL \frac{\binom{n}{k}}{\binom{n-\Delta}{k}} \leq RL \left(\frac{n}{n - \Delta - k + 1}\right)^k.$$

The ratio is therefore at most $RL \exp(O(k(k+\Delta)/n))$ in the sparse regime. At constant active density, the displayed binomial ratio tracks the structural dependence explicitly.

This is the block version of the batch-leap law:

$$\mathrm{BLeap}_R = b \left\lceil \frac{D}{R} \right\rceil.$$

3.2 Critical dispersion: the distribution buys a unary path

Theorem 3.3 (Critical response-energy recovery)**.** *Under the two-radius law (7), for every coordinate j,*

$$\mathbb{E}[Y^2 R_j] = \frac{a^2\sqrt{u}}{D} |\{t : j \in A_t\}|. \tag{12}$$

Split $N = mK$ IID samples into K blocks. In each block, coordinate j votes positive when the block sum of $Y^2 R_j$ is positive. Exact recovery with probability at least $1 - \delta$ holds when

$$m \geq \frac{3D^2}{u}\left((1+u)^{bD} + 6\frac{\sigma^2}{a^2} + 3\frac{\sigma^4}{a^4}\right),$$
$$K \geq 32 \log \frac{2n}{\delta}.$$

At $u = c/(bD)$ with $0 < c < bD$ and fixed $(D, c, |a|/\sigma)$,

$$N = O\left(\frac{bD^3}{c}\left(e^c + 6\frac{\sigma^2}{a^2} + 3\frac{\sigma^4}{a^4}\right)\log\frac{2n}{\delta}\right), \qquad \textit{work} = O(nN).$$

The learner uses one fixed passive design and one frozen batch of unary statistics.

The same positive result can be stated in the rigid theorem's oracle currency. Let $\mu_{y,u}$ be the label marginal under the two-radius model; symmetry makes it common to every embedding, and put $\nu_{0,u} = \mu_u^n \otimes \mu_{y,u}$ and $V_u = \mathbb{E}_{\mu_{y,u}} Y^4$.

Corollary 3.4 (Critical recovery as one-batch relative-L_2 SQ)**.** *At the two-radius design, the n fixed queries*

$$\phi_j(x, y) = y^2 \frac{x_j^2 - 1}{\sqrt{u}}$$

recover A_D in one batch whenever the relative-$L_2(\nu_{0,u})$ tolerance satisfies

$$\tau < \frac{a^2\sqrt{u}}{4D\sqrt{V_u}}.$$

Thus the rigid and critical upper bounds both belong to the same full-SQ model; Theorem 3.3 is its IID implementation.

Proof. The product-reference norm is $\sqrt{V_u}$. Equation (12) separates inactive and active expectations by at least $\Delta_u = a^2\sqrt{u}/D$, while the oracle error is less than $\Delta_u/4$. Thresholding at $\Delta_u/2$ recovers the union. □

An amplitude class, with the same critical rate. Theorem K.2 extends the unary path to $X = \xi\sqrt{1 + \sqrt{u}\, Z}$ for any centered unit-variance Z with $|Z| \le M$ and $M\sqrt{u} < 1$, including continuous and asymmetric amplitude laws. The observable $G = (X^2 - 1)/\sqrt{u}$ has the same exact signal and reference norm, with $\mathrm{Var}(Y^2 G_j) < 4V_u$. Median-of-means recovery retains $O_D(b \log(n/\delta))$ rows at $u = c/(bD)$ whenever $bD > cM^2$.

Proposition 3.5 (Critical-design information lower bound)**.** *Assume $n \ge 2bD$ and any symmetric standardized product design, including the critical two-radius law. Even if the learner is given $B_2, \dots, B_D$, every IID procedure recovering B_1 with worst-case success at least $3/4$ requires*

$$N \ge c\frac{\sigma^2}{a^2}\, b \log\left(\frac{n - b(D-1)}{b}\right) \tag{13}$$

for a universal constant $c > 0$. At fixed depth, SNR, and failure probability, the upper bound is $O(b \log n)$ and the lower bound is $\Omega(b \log(n/b))$. They match up to constants in every regime $b \le n^{1-\varepsilon}$ with fixed $\varepsilon > 0$, where $\log(n/b) \asymp_\varepsilon \log n$.

Proof sketch. Fix the later blocks and vary B_1 over the b-subsets of the remaining coordinates. For any two choices, condition on $T = t$. Their Gaussian regression means are distinct standardized monomials and have squared L_2 distance two, so component KL is a^2/σ^2. Joint convexity of KL under the common hidden-selector mixture preserves this upper bound. Fano over $\binom{n-b(D-1)}{b}$ hypotheses gives (13). □

3.3 Fixed dispersion: passive information collapses

Compare two targets obtained by swapping one coordinate of B_1 with one inactive coordinate. Let P and P' be their one-sample joint laws.

Theorem 3.6 (Passive adjacent-target collapse)**.** *For every standardized symmetric coordinate law μ and $0 < \alpha \le 2$,*

$$H^2(P, P') \le \left(\frac{|a|}{\sqrt{8}\sigma}\right)^\alpha D_\alpha(\mu) \frac{1}{D} \sum_{t=1}^{D} M_\alpha(\mu)^{bt-1}. \tag{14}$$

Consequently, for every fixed standardized nonconstant-modulus law, fixed depth, and fixed nonzero signal-to-noise ratio, there are constants $C_{\mu,a/\sigma,D} < \infty$ and $c_\mu > 1$ such that passive testing with worst-case success at least $3/4$ requires

$$N \ge C_{\mu,a/\sigma,D}^{-1} c_\mu^b.$$

The lower bound holds even if the learner is told the two candidate targets.

Theorems 3.3 and 3.6 identify the role of distribution scaling. At $u = \Theta(1/(bD))$, the unary estimator achieves the stated sample guarantee as the coordinate law changes with total degree. A fixed nonconstant-modulus law has $M_\alpha(\mu) < 1$, which yields the exponential passive obstruction.

Proposition 3.7 (Quantitative two-radius Goldilocks curve)**.** *For the two-radius family and $0 < u < 1$, the unary learner's sample bound is, up to universal constants,*

$$N_{\text{unary}}(u) \le \frac{D^2}{u}\left((1+u)^{bD} + 6\frac{\sigma^2}{a^2} + 3\frac{\sigma^4}{a^4}\right)\log\frac{2n}{\delta}. \tag{15}$$

For adjacent targets, passive one-sample information obeys

$$H^2(P_u, P_u') \le C(a/\sigma, D)\exp\left(-\frac{u(b-1)}{8}\right). \tag{16}$$

Consequently, $ub \to \infty$ is passively information-poor, while $u = \Theta(1/(bD))$ keeps the Hellinger exponent constant and makes (15) linear in b at fixed $(D, a/\sigma)$. The signal-dominated factor $(1+u)^{bD}/u$ is minimized exactly at $u = 1/(bD-1)$ when $bD > 2$; for $bD \le 2$, its infimum over $0 < u < 1$ occurs as $u \uparrow 1$.

Appendix I derives this curve from the exact response fourth moment and the bound $M_1(\mu_u) \le e^{-u/8}$. It also optimizes this sufficient sample bound including noise and the boundary cases $bD \le 2$; the optimizer is specific to the displayed unary sample certificate.

3.4 One sample, many evaluations

Theorem 3.8 (One-reused-sample rigid path)**.** *Fix the rigid hidden-rung model and an R-batch block-enumeration schedule. There is a single-sample implementation that computes every requested correlation from the same IID dataset. With probability at least $1-\delta$, it returns the exact active union using the population query count $K_{n,b,D,R}$ and*

$$N = O\left(\frac{D^2}{\gamma_{\text{sgn}}^2}\left(bD\log\frac{en}{bD} + \log\frac{1}{\delta}\right)\right) \tag{17}$$

labeled examples. At full adaptivity and $b = 1$, the algorithm evaluates $O(nD)$ correlations. Sample reuse amortizes observations behind the candidate family; it does not remove the candidate evaluations. If each candidate parity is formed from its known carrier and at most bL fresh coordinates, direct evaluation uses

$$O(NbLK_{n,b,D,R}) \tag{18}$$

arithmetic operations and $O(Nn)$ storage when the reusable dataset is retained explicitly. Thus at full adaptivity and $b = 1$ the end-to-end work is $O(NnD)$, up to constant-cost updates of the known carrier.

4 Constructive Recovery Algorithms

The upper bounds give streaming unary screening and adaptive carrier enumeration, with observations and arithmetic accounted for separately.

Algorithm 4.1 (CriticalUnary)**.** Choose m, K from Theorem 3.3. Decode $R_{i,j} = (X_{i,j}^2 - 1)/\sqrt{u}$ and accumulate $V_{k,j} = \sum_{i \in k} Y_i^2 R_{i,j}$ in each block. Return coordinates whose fraction of positive block sums exceeds $5/8$.

This uses $O(n)$ streaming accumulators and $O(nN)$ arithmetic. For rigid inputs, Theorem 3.8 controls the full candidate family before adaptive selection, allowing all batches to reuse one dataset; Appendix J gives the implementation.

The finite-sample evaluation of these learning paths appears in Appendix A.

5 Related Work and Scope

Statistical queries and the direct predecessor. The SQ model separates expectation access from individual samples (Kearns, 1998). Fourier characterizations (Blum et al., 1994) and statistical

dimension (Feldman et al., 2017; Feldman, 2017) quantify query hardness and precision tradeoffs. Building on the fixed-link support task of Joshi et al. (2024), we characterize every finite frozen-batch budget between their Cover and Leap endpoints. Equal-endpoint ranking reversal demonstrates the profile's additional information; conditional projection after earlier grants supplies the finite-round lower bound.

Full SQ and hierarchical features. Full SQ and label-linear CSQ can differ (Kiani et al., 2024; Joshi et al., 2024); our likelihood slice controls arbitrary coupled label queries, while the positive result uses $Y^2 G_j$ in the same query class. Staircase and leap analyses describe successive feature acquisition (Abbe et al., 2021; 2022; 2023a). Information exponents identify an initial search phase (Ben Arous et al., 2021), with a two-phase account for XOR (Glasgow, 2024). Low-dimensional polynomial learning (Damian et al., 2022), one-step feature learning (Ba et al., 2022; Dandi et al., 2024a), and hierarchical three-layer constructions (Nichani et al., 2023; Wang et al., 2024) show benefits of intermediate structure. These mechanisms extend the fixed-kernel NTK description (Jacot et al., 2018); feature learning at infinite width also depends on parametrization (Yang & Hu, 2021). Our frozen rounds quantify the information dependencies between certified carriers.

Distribution design and nonlinear screening. Magnitude statistics aid sparse-polynomial learning (Andoni et al., 2014), and squared-response screening reveals variables absent from main effects (Fan et al., 2016). Biased product inputs (Arpe & Mossel, 2008), parity curricula (Cornacchia & Mossel, 2023; Abbe et al., 2023b), and positive distribution shift (Medvedev et al., 2026) provide related distributional benefits. Weight initialization also affects high-degree parity learning (Abbe et al., 2025). We connect unary screening to an exact round hierarchy for the same hidden target, identify a critical amplitude class, and establish the passive-information obstruction at fixed dispersion.

Samples, computation, and information. Sample reuse (Dandi et al., 2024b) interacts with adaptive selection (Dwork et al., 2015; Russo & Zou, 2016); simultaneous concentration over our fixed candidate family enables reuse while retaining explicit evaluation costs. This separates sample and computational complexity, as in nearly linear-sample procedures with potentially exponential computation (Mousavi-Hosseini et al., 2025). Block amplification uses robust mean estimation and bounded-vote concentration (Lugosi & Mendelson, 2019; Hoeffding, 1963). Fano and two-point reductions (Yu, 1997; Tsybakov, 2009) convert the model-specific divergence bounds into sample lower bounds. Product-basis conventions follow O'Donnell (2014).

6 Conclusion

Under full detectability and square-integrable likelihood ratios, blind-jump width gives the exact ambient exponent of full-SQ query–precision cost for fixed latent dimension, round budget, and link. Its round profile contains information absent from the nonadaptive and sequential endpoints: two tasks can share both endpoints yet reverse their difficulty ordering at intermediate rounds.

The hidden-rung model makes two learning mechanisms explicit. Under Rademacher inputs, adaptive carriers reduce the width to $b\lceil D/R\rceil$. Bounded amplitude perturbations at critical dispersion $u = \Theta(1/(bD))$ create coordinatewise response-energy witnesses, enabling one-batch full-SQ recovery. An IID implementation uses $O_D(b\log(n/\delta))$ rows at fixed signal-to-noise ratio. Every fixed symmetric standardized nonconstant-modulus law incurs an exponential passive obstruction as width grows at fixed depth and signal-to-noise ratio. Lean proofs certify the carrier calculus, lower-bound success estimates, and selected analytic identities. These results distinguish the roles of adaptive information, distribution design, and estimation cost.

The hidden-rung comparison holds the support and recovery objective fixed while varying how evidence becomes accessible. In the rigid model, reusing one dataset controls observation cost across the adaptive path, while the candidate family still determines the arithmetic cost of search. Critical dispersion changes the available witnesses themselves: coordinatewise response energy replaces the need to enumerate large unknown interactions. This separates the precision needed to resolve each witness from the number of candidate interactions that must be evaluated.

REFERENCES

Emmanuel Abbe, Enric Boix-Adserà, Matthew S. Brennan, Guy Bresler, and Dheeraj Nagaraj. The staircase property: How hierarchical structure can guide deep learning. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/e2db7186375992e729165726762cb4c1-Abstract.html.

Emmanuel Abbe, Enric Boix Adserà, and Theodor Misiakiewicz. The merged-staircase property: A necessary and nearly sufficient condition for SGD learning of sparse functions on two-layer neural networks. In *Proceedings of the 35th Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 4782–4887, 2022. URL https://proceedings.mlr.press/v178/abbe22a.html.

Emmanuel Abbe, Enric Boix Adserà, and Theodor Misiakiewicz. SGD learning on neural networks: Leap complexity and saddle-to-saddle dynamics. In *Proceedings of the 36th Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 2552–2623, 2023a. URL https://proceedings.mlr.press/v195/abbe23a.html.

Emmanuel Abbe, Elisabetta Cornacchia, and Aryo Lotfi. Provable advantage of curriculum learning on parity targets with mixed inputs. In *Advances in Neural Information Processing Systems*, volume 36, 2023b. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/4c8ce3c63f6b66d6811c6d67c68e487b-Abstract-Conference.html.

Emmanuel Abbe, Elisabetta Cornacchia, Jan Hązła, and Donald Kougang Yombi. Learning high-degree parities: The crucial role of the initialization. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/31455446488c433ef29495ab44d4f53c-Abstract-Conference.html.

Alexandr Andoni, Rina Panigrahy, Gregory Valiant, and Li Zhang. Learning sparse polynomial functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2014. URL https://www.mit.edu/~andoni/papers/learnpoly.pdf.

Jan Arpe and Elchanan Mossel. Multiple random oracles are better than one. *arXiv preprint arXiv:0804.3817*, 2008. URL https://arxiv.org/abs/0804.3817.

Jimmy Ba, Murat A. Erdogdu, Taiji Suzuki, Zhichao Wang, Denny Wu, and Greg Yang. High-dimensional asymptotics of feature learning: How one gradient step improves the representation. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/f7e7fabd73b3df96c54a320862afcb78-Abstract-Conference.html.

Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Online stochastic gradient descent on non-convex losses from high-dimensional inference. *Journal of Machine Learning Research*, 22 (106):1–51, 2021. URL https://jmlr.org/papers/v22/20-1288.html.

Avrim Blum, Merrick Furst, Jeffrey Jackson, Michael Kearns, Yishay Mansour, and Steven Rudich. Weakly learning DNF and characterizing statistical query learning using Fourier analysis. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*, pp. 253–262, 1994. doi: 10.1145/195058.195147.

Elisabetta Cornacchia and Elchanan Mossel. A mathematical model for curriculum learning for parities. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6402–6423, 2023. URL https://proceedings.mlr.press/v202/cornacchia23a.html.

Alexandru Damian, Jason D. Lee, and Mahdi Soltanolkotabi. Neural networks can learn representations with gradient descent. In *Proceedings of the 35th Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pp. 5413–5452, 2022. URL https://proceedings.mlr.press/v178/damian22a.html.

Yatin Dandi, Florent Krzakala, Bruno Loureiro, Luca Pesce, and Ludovic Stephan. How two-layer neural networks learn, one (giant) step at a time. *Journal of Machine Learning Research*, 25(349): 1–65, 2024a. URL https://jmlr.org/papers/v25/23-1543.html.

Yatin Dandi, Emanuele Troiani, Luca Arnaboldi, Luca Pesce, Lenka Zdeborova, and Florent Krzakala. The benefits of reusing batches for gradient descent in two-layer networks: Breaking the curse of information and leap exponents. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 9991–10016, 2024b. URL https://proceedings.mlr.press/v235/dandi24a.html.

Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction—CADE 28*, pp. 625–635. Springer, 2021. doi: 10.1007/978-3-030-798 76-5_37. URL https://doi.org/10.1007/978-3-030-79876-5_37.

Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. Preserving statistical validity in adaptive data analysis. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pp. 117–126, 2015. doi: 10.1145/2746539.2746580.

Yingying Fan, Yinfei Kong, Daoji Li, and Jinchi Lv. Interaction pursuit with feature screening and selection. *arXiv preprint arXiv:1605.08933*, 2016. URL https://arxiv.org/abs/1605 .08933.

Vitaly Feldman. A general characterization of the statistical query complexity. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pp. 785–830, 2017. URL https://proceedings.mlr.press/v65/feldman17c.html.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S. Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM*, 64(2):8:1–8:37, 2017. doi: 10.1145/3046674.

Margalit Glasgow. SGD finds then tunes features in two-layer neural networks with near-optimal sample complexity: A case study in the XOR problem. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/pape r/2024/hash/e6d37cc5723e810b793c834bcb6647cf-Abstract-Conference. html.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/5a4be1fa3 4e62bb8a6ec6b91d2462f5a-Abstract.html.

Nirmit Joshi, Theodor Misiakiewicz, and Nathan Srebro. On the complexity of learning sparse functions with statistical and gradient queries. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files /paper/2024/hash/baf0fab890edc9dce805d7c518058712-Abstract-Confe rence.html.

Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM*, 45(6): 983–1006, 1998. doi: 10.1145/293347.293351.

Bobak T. Kiani, Thien Le, Hannah Lawrence, Stefanie Jegelka, and Melanie Weber. On the hardness of learning under symmetries. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/6e51 12eaa45f8c30b242c5f576213a92-Abstract-Conference.html.

Gábor Lugosi and Shahar Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19:1145–1190, 2019. doi: 10.1007/s10208-019-09427-x.

Marko Medvedev, Idan Attias, Elisabetta Cornacchia, Theodor Misiakiewicz, Gal Vardi, and Nathan Srebro. Positive distribution shift as a framework for understanding tractable learning. *arXiv preprint arXiv:2602.08907*, 2026. URL https://arxiv.org/abs/2602.08907.

Alireza Mousavi-Hosseini, Denny Wu, and Murat A. Erdogdu. Learning multi-index models with neural networks via mean-field Langevin dynamics. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/61d8577984e4ef0cba20966eb3ef2ed8-Abstract-Conference.html.

Eshaan Nichani, Alex Damian, and Jason D. Lee. Provable guarantees for nonlinear feature learning in three-layer neural networks. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/236b6a814a1d2c0ff504ca7bf380f7ff-Abstract-Conference.html.

Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi: 10.1017/CBO9781139814782.

Daniel Russo and James Zou. Controlling bias in adaptive data analysis using information theory. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pp. 1232–1240, 2016. URL https://proceedings.mlr.press/v51/russo16.html.

The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pp. 367–381. ACM, 2020. doi: 10.1145/3372885.3373824. URL https://doi.org/10.1145/3372885.3373824.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009. doi: 10.1007/b13794.

Zihao Wang, Eshaan Nichani, and Jason D. Lee. Learning hierarchical polynomials with three-layer neural networks. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/af536fa4312b5b2e5c541156cd31e93b-Abstract-Conference.html.

Greg Yang and Edward J. Hu. Tensor programs IV: Feature learning in infinite-width neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11727–11737, 2021. URL https://proceedings.mlr.press/v139/yang21c.html.

Bin Yu. Assouad, Fano, and Le Cam. In David Pollard, Erik Torgersen, and Grace L. Yang (eds.), *Festschrift for Lucien Le Cam*, pp. 423–435. Springer, 1997. doi: 10.1007/978-1-4612-1880-7_29.

AI USE STATEMENT

Generative AI tools assisted with literature discovery, manuscript drafting and editing, mathematical cross-checks, and the preparation of figures and LaTeX source. The authors take full responsibility for verifying all AI-assisted content and for the accuracy of the final claims, references, implementation, and results.

Appendix

A Experiments and Recovery Mechanisms 12

B Formal Verification in Lean 4 20

C Proof Strategy and Notation 26

D Carrier Closure and Batch Scheduling 26

E General SQ Round Complexity 30

F Rigid Staircase Round Complexity 34

G Unary Recovery at Critical Dispersion 39

H Passive Information Lower Bounds 41

I Dispersion Tuning and Learning Regimes 45

J Sample Reuse and Computational Cost 47

K Extensions of the Hidden-Rung Model 49

L Auxiliary Probability and Product-Space Results 53

M Model Assumptions and Resource Regimes 55

A Experiments and Recovery Mechanisms

Finite-round recovery separates chains with identical Cover and Leap endpoints. On the same observation model, critical dispersion makes coordinatewise response energy sufficient for union and hierarchy recovery. We evaluate these two mechanisms across dimensions, widths, depths, noise levels, and amplitude laws, using the hidden-rung observation law (6) and a separately labeled noiseless control.

A.1 Experimental Design

The evaluation comprises 8,800 generated datasets and 224,200 recorded learner–sample outcomes. Rademacher inputs test frozen-batch carrier enumeration; bounded-amplitude inputs test unary median-of-means recovery. The parameter grids and replication counts are given below.

Experimental coverage		Recorded scale	
Input design	**Parameter grid**	**Datasets**	**Outcomes**
Adaptive carrier construction; shared data across round budgets			
Rademacher	4 dimensions; 7 learners; 100 replicates	400	19,600
Rademacher	3 dimensions $\times$ 3 noise levels; 7 learners; 200 replicates	1,800	113,400
Unary recovery; independent trials within each configuration			
Bounded amplitude	20 dispersion/width/law configurations; 200 replicates	4,000	60,000
Critical amplitude	13 dimension/depth/noise configurations; 200 replicates	2,600	31,200
Total	Generated datasets and recorded outcomes	**8,800**	**224,200**

A round-study dataset is shared by seven family/round combinations, and sample budgets are nested within each trial. Replicates within each configuration are independent; shared algorithms, checkpoints, and coordinates are not additional independent replicates. Configurations appearing in both study grids use separate observation seeds. Pilot and final streams are disjoint.

For a compact comparison across the full parameter grid, define N_{90}^{union} and $N_{90}^{\text{hierarchy}}$ as the first *tested* allocations with at least 90% observed union or ordered-block recovery. These summarize the recorded grid, with no interpolation or fitted transition. NR means the level was not reached at any tested allocation; the corresponding largest-budget outcome is reported. Fixed-budget recovery intervals are pointwise 95% Wilson intervals. Carrier quartiles and coordinate standard deviations describe empirical variation rather than confidence intervals.

A.2 Identical Endpoints, Reversed Recovery Order

Chains $\mathrm{A} = (1, 2, 1, 2)$ and $\mathrm{B} = (1, 1, 2, 2)$ share six active coordinates, depth four, Cover six, and Leap two (Proposition D.7). Figure 1 maps their recovery-rate differences across the complete dimension/noise grid and compares the corresponding query costs. The ordering contrast becomes pronounced at larger dimensions: two rounds favor A, while three rounds favor B in the recovery transition. Table 2 reports the associated empirical recovery crossings.

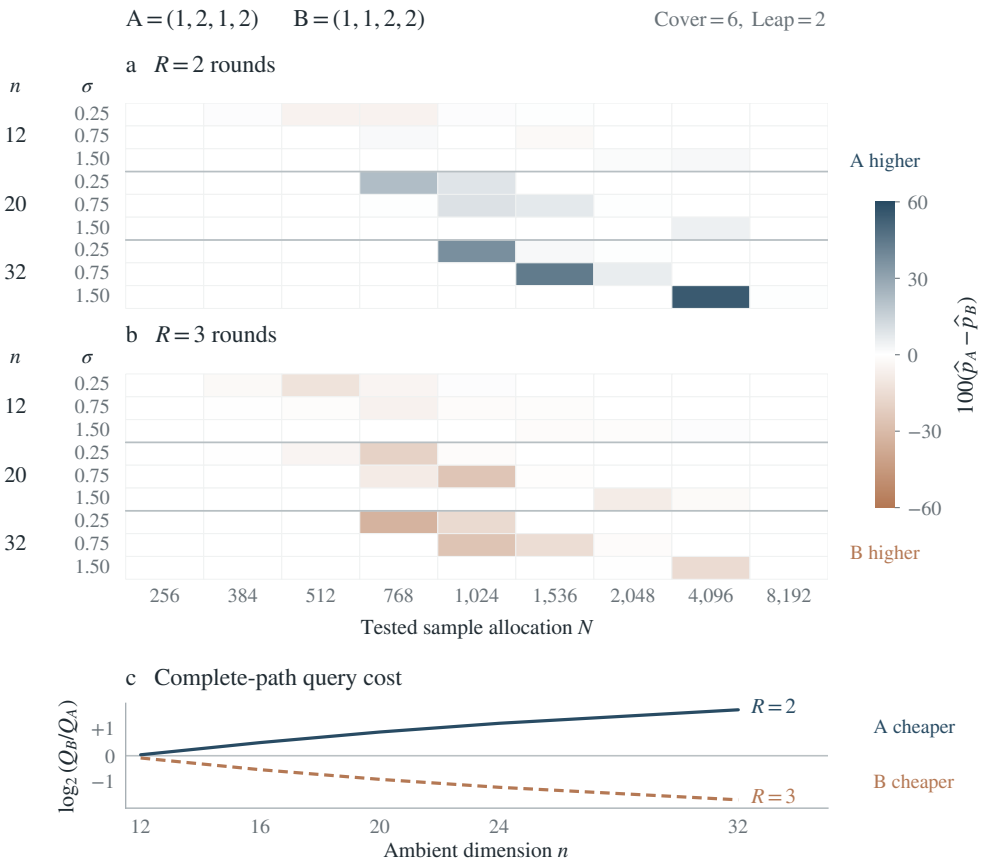

Figure 1: **Finite-round ordering across the full recovery grid. (a,b)** Observed exact-support recovery difference $100(\widehat{p}_A - \widehat{p}_B)$, in percentage points, for $R = 2, 3$. Each panel includes all nine dimension/noise conditions and nine tested budgets, with 200 independent paired datasets per condition (1,800 total). The heatmaps share a fixed symmetric color scale and discrete, equally spaced budget columns. White denotes equal observed recovery, including joint failure or joint success. **(c)** The complete-path query ratio $\log_2(Q_B/Q_A)$ across five dimensions. Positive values favor A; negative values favor B. These schedule costs are deterministic and match the recorded complete runs. Absolute recovery rates and their intervals remain in the numerical companion.

At $n = 32$ and $N = 1{,}024$, two-round recovery is $107/200$ for A and $32/200$ for B; three-round recovery is $155/200$ for A and $188/200$ for B. The complete two-round query budgets are $8{,}675$ and $28{,}368$, while the three-round budgets are $4{,}180$ and $1{,}341$. Thus the same endpoint descriptors support opposite recovery and search-cost orderings, as predicted by the finite-round width profile.

Table 1 gives the exact counts for every tested dimension, including the equal-width chain. Complete-path counts are kept separate from the smaller number of queries a failed run may submit before termination. Table 2 summarizes all 63 configurations in the $n \in \{12, 20, 32\}$, $\sigma \in \{1/4, 3/4, 3/2\}$ grid. Each uses $a = 1$, 200 paired datasets, and $N \in \{256, 384, 512, 768, 1024, 1536, 2048, 4096, 8192\}$. At the largest allocation, every configuration recovers in $199/200$ or $200/200$ trials. These finite-dimensional measurements complement the ambient exponents of Theorem 2.2.

Table 1: **Complete frozen-batch query budgets.** All five tested dimensions are shown. Bold values identify the smaller A/B budget at each round count. Counts include every candidate on a complete scheduled path; early stopping can reduce the number actually submitted.

Ambient	Two rounds		Three rounds		Equal-width chain		
n	A	B	A	B	$R = 1$	$R = 2$	$R = 4$
Scheduled scalar correlations; all tested dimensions							
12	**325**	333	160	**151**	793	133	42
16	**875**	1 228	420	**293**	2 516	241	58
20	**1 857**	3 423	888	**483**	6 195	381	74
24	**3 399**	7 846	1 628	**721**	12 950	553	90
32	**8 675**	28 368	4 180	**1 341**	41 448	993	122

Table 2: **Recovery across the complete round–noise grid.** Each entry is the first tested N with at least $180/200$ exact support recoveries. These empirical grid crossings summarize all 63 dimension–noise–learner settings; no interpolation is used. At $N = 8{,}192$, 62 settings recover $200/200$ times; B at $(n, \sigma, R) = (32, 3/2, 2)$ recovers $199/200$ times.

Condition		Two rounds		Three rounds		Equal-width chain		
n	σ	A	B	A	B	$R = 1$	$R = 2$	$R = 4$
Ambient dimension $n = 12$; 200 paired datasets per noise level								
12	1/4	1 024	1 024	768	768	1 024	768	768
12	3/4	1 536	1 536	1 536	1 536	1 536	1 536	1 024
12	3/2	4 096	4 096	4 096	4 096	4 096	4 096	4 096
Ambient dimension $n = 20$; 200 paired datasets per noise level								
20	1/4	1 536	1 536	1 024	1 024	1 536	1 024	768
20	3/4	1 536	2 048	1 536	1 536	2 048	1 536	1 024
20	3/2	8 192	8 192	4 096	4 096	8 192	4 096	4 096
Ambient dimension $n = 32$; 200 paired datasets per noise level								
32	1/4	1 536	1 536	1 536	1 024	1 536	1 024	768
32	3/4	2 048	4 096	2 048	1 536	4 096	1 536	1 024
32	3/2	8 192	8 192	8 192	4 096	8 192	4 096	4 096

A.3 Adaptive Rounds and Carrier Formation

For the equal-width chain $b = 1$, $D = 4$, additional rounds construct shorter fresh-coordinate searches. Table 3 records this mechanism at the pilot-fixed diagnostic $N = 512$, $\sigma = 1/4$, using the same data for $R = 1, 2, 4$. Define $H_r = t$ when the current carrier is the true prefix A_t, with $A_0 = \varnothing$, and $H_r = 0$ when it is not a true prefix. A stopped run retains its last carrier. The table reports the mean normalized depth after every allowed batch, terminal replicate quartiles, and exact recovery.

At $n = 32$, the four-round mean carrier depth progresses through 0.209, 0.390, 0.550, and 0.681 of the hierarchy. Four rounds recover the full union in $105/200$ trials at this diagnostic budget, while one and two rounds yield $0/200$. At $N = 8{,}192$, all three recover in $200/200$ trials, using complete-path budgets of $41{,}448$, 993, and 122 queries. The comparison separates the observations supplied from the search saved by constructing carriers.

Table 3: **Carrier formation at a common data budget.** Equal-width chains with $b = 1$, $D = 4$, $\sigma = 1/4$, and $N = 512$. Entries under H_r/D are means over 200 paired trials; the terminal interval contains the 25th–75th percentiles. $H_0 = 0$. A terminated run retains its last carrier. Dashes indicate rounds outside the learner's budget.

Condition		**Mean carrier depth**				**Terminal outcome**	
n	R	H_1/D	H_2/D	H_3/D	H_4/D	**Terminal IQR**	**Recovered**
Ambient dimension $n = 12$; shared observations across round budgets							
12	1	0.005	—	—	—	$[0.00, 0.00]$	$1/200$
12	2	0.333	0.588	—	—	$[0.00, 1.00]$	$102/200$
12	4	0.233	0.450	0.662	0.866	$[1.00, 1.00]$	$163/200$
Ambient dimension $n = 20$; shared observations across round budgets							
20	1	0.000	—	—	—	$[0.00, 0.00]$	$0/200$
20	2	0.165	0.228	—	—	$[0.00, 0.50]$	$25/200$
20	4	0.226	0.430	0.623	0.802	$[0.75, 1.00]$	$144/200$
Ambient dimension $n = 32$; shared observations across round budgets							
32	1	0.000	—	—	—	$[0.00, 0.00]$	$0/200$
32	2	0.045	0.045	—	—	$[0.00, 0.00]$	$0/200$
32	4	0.209	0.390	0.550	0.681	$[0.25, 1.00]$	$105/200$

A.4 The Finite-Sample Dispersion Region

Generate $X_j = \xi_j \sqrt{1 + \sqrt{u}\, Z_j}$ with independent fair signs, $\mathbb{E}Z_j = 0$, and $\mathbb{E}Z_j^2 = 1$. The observable $G_j = (X_j^2 - 1)/\sqrt{u} = Z_j$ has unary response-energy levels

$$\mathbb{E}[Y^2 G_j] = \Delta_u h_j, \qquad \Delta_u = \frac{a^2\sqrt{u}}{D}, \qquad h_j = |\{t : j \in A_t\}|.$$

For $n = 128$, $b = 8$, $D = 2$, $a = 1$, $\sigma = 1/4$, and $Z_j \sim \mathrm{Rad}$, Table 4 reports every one of the 13 tested dispersion settings. Each uses 200 trials and 15 sample budgets up to $N = 58{,}646$. All are admissible because $u = c/16 < 1$, where $c = ubD$.

The observed crossings identify an intermediate-dispersion recovery region. Critical tuning $c = 1$ and nearby designs separate the unary levels with substantially fewer rows than the small- and large-dispersion edges. At $c = 1/16$, 8, and 12, the largest-budget support counts are $101/200$, $155/200$, and $0/200$, respectively; the table retains these boundary regimes alongside the successful crossings. The fixed-dispersion passive obstruction is established separately by Theorem 3.6.

Table 4: **The full dispersion sweep.** Two-point amplitudes at $(n, b, D, \sigma) = (128, 8, 2, 1/4)$. N_{90}^{union} and $N_{90}^{\text{hierarchy}}$ are first tested allocations attaining at least $180/200$ recoveries. The last two columns retain outcomes at the common largest budget $N = 58{,}646$; NR denotes a 90% level not reached on the tested grid. All 13×15 checkpoints and their intervals are included in the numerical companion.

Design	First 90% grid crossing		Largest-budget recovery	
$c = ubD$	N_{90}^{union}	$N_{90}^{\text{hierarchy}}$	Final union	Final hierarchy
Two-point amplitudes; $n = 128$, $b = 8$, $D = 2$, $\sigma = 1/4$				
$1/16$	NR	NR	$101/200$	$94/200$
$1/8$	44 014	58 646	$191/200$	$191/200$
$1/4$	29 323	29 323	$200/200$	$200/200$
$1/2$	22 007	22 007	$200/200$	$200/200$
$3/4$	14 691	14 691	$200/200$	$200/200$
1	11 033	11 033	$200/200$	$200/200$
$3/2$	9 204	9 204	$200/200$	$200/200$
2	9 204	9 204	$200/200$	$200/200$
3	11 033	11 033	$200/200$	$200/200$
4	14 691	14 691	$200/200$	$200/200$
6	29 323	44 014	$200/200$	$200/200$
8	NR	NR	$155/200$	$136/200$
12	NR	NR	$0/200$	$0/200$

The scale $ubD = 1$ is a theoretical reference, not an empirical minimizer.

A.5 Width, Dimension, Depth, and Noise Scaling

At $u = 1/(bD)$, the fixed-depth grid crosses $n \in \{128, 256\}$ with $b \in \{4, 8, 16\}$ at $D = 2$, $a = 1$, $\sigma = 1/4$, and $\delta = 0.1$. Table 5 reports all six configurations. Their first observed 90% support crossings satisfy $N_{90}^{\text{union}}/[b \log(n/\delta)] \in [160.3, 193.8]$, organizing the transitions across a fourfold change in block width.

Table 5: **Width and dimension scaling at critical dispersion.** All six width–dimension configurations, with $D = 2$, $\sigma = 1/4$, and $\delta = 0.1$. Recovery counts are evaluated at the first tested allocation with at least $180/200$ exact recoveries.

Condition		Sample requirement		Recorded outcome and budget	
n	b	N_{90}^{union}	$N_{90}/[b \log(n/\delta)]$	Recovered	Largest N
Ambient dimension $n = 128$*; fixed depth* $D = 2$					
128	4	5 546	193.8	190/200	29 323
128	8	11 033	192.8	192/200	58 646
128	16	18 349	160.3	180/200	117 233
Ambient dimension $n = 256$*; fixed depth* $D = 2$					
256	4	6 048	192.7	185/200	32 193
256	8	12 096	192.7	186/200	64 323
256	16	24 129	192.2	188/200	128 583

The dimension/depth/noise grid fixes $b = 4$, $c = 1$, $a = 1$, and $\delta = 0.1$. It varies $n \in \{64, 128, 256, 512\}$ at $D = 2$, $\sigma = 1/4$; varies $D \in \{2, 3, 4, 6\}$ at $n = 128$, $\sigma = 1/4$; and crosses $D \in \{2, 4\}$ with $\sigma \in \{0, 1/4, 3/4, 3/2\}$ at $n = 128$. Shared configurations are evaluated once within this grid, giving 13 configurations with 200 trials each. The $\sigma = 0$ settings are noiseless decoder controls.

For these two-point amplitudes, define the analytic sample scale

$$V = \frac{a^4}{D}\sum_{t=1}^{D}(1+u)^{bt} + 6a^2\sigma^2 + 3\sigma^4, \qquad S = \frac{bD^3V}{a^4}\log(n/\delta).$$

Since $\Delta_u^{-2} = bD^3/a^4$ at $c = 1$, S combines the exact response fourth moment with the unary signal gap. The frozen grid uses $N = K\lceil tS/K\rceil$ for

$$t \in \{1/2, 1, 3/2, 2, 3, 4, 6, 8, 12, 16, 24, 32\}.$$

Table 6 gives the raw and normalized crossings for every configuration, with no fitted normalization. At $n = 128$, $\sigma = 1/4$, depths $D = 2, 3, 4, 6$ give $N_{90}^{\text{union}} = 4{,}543$, $21{,}830$, $50{,}327$, and $165{,}259$, but $N_{90}^{\text{union}}/S = 8.03$, 12.03, 12.01, and 12.00. This scale organizes a greater than 36-fold change in raw sample requirements. Every configuration attains $200/200$ support and ordered-hierarchy recovery at its largest tested budget.

Table 6: **Dimension, depth, and noise under one analytic scale.** All 13 configurations use $b = 4$, $ubD = 1$, $a = 1$, and $\delta = 0.1$. The analytic scale is $S = bD^3V\log(n/\delta)/a^4$. Recovery counts are evaluated at the first observed 90% union crossing. At the largest allocation, every configuration attains $200/200$ union and hierarchy recoveries. The $\sigma = 0$ rows are noiseless controls.

Condition			Sample requirement			Outcome and budget	
n	D	σ	N_{90}^{union}	$N_{90}^{\text{hierarchy}}$	N_{90}^{union}/S	Recovered	Largest N
Critical dispersion $ubD = 1$; block width $b = 4$							
64	2	1/4	4 134	4 134	8.09	184/200	16 377
128	2	0	5 782	5 782	12.12	198/200	15 281
128	2	1/4	4 543	6 844	8.03	181/200	18 113
128	2	3/4	17 641	17 641	12.02	200/200	46 964
128	2	3/2	84 547	84 547	12.00	195/200	225 498
128	3	1/4	21 830	21 830	12.03	198/200	58 115
128	4	0	41 831	41 831	12.01	199/200	111 510
128	4	1/4	50 327	50 327	12.01	196/200	134 166
128	4	3/4	136 880	136 880	12.00	198/200	364 915
128	4	3/2	672 364	672 364	12.00	198/200	1 792 892
128	6	1/4	165 259	165 259	12.00	199/200	440 612
256	2	1/4	7 497	7 497	12.08	197/200	19 908
512	2	1/4	8 142	8 142	12.06	197/200	21 666

A.6 Amplitude Laws and Unary-Level Calibration

Fix $n = 128$, $b = 8$, $D = 2$, $u = 1/16$, $a = 1$, and $\sigma = 1/4$. Compare the three designs from Appendix K.5: two-point $Z \sim \text{Rad}$, continuous uniform $Z \sim \text{Unif}[-\sqrt{3}, \sqrt{3}]$, and asymmetric Z taking $-1/2$ and 2 with probabilities $4/5$ and $1/5$. They share $\mathbb{E}Z = 0$, $\mathbb{E}Z^2 = 1$, the coordinate fourth moment, the population unary levels, and the response fourth moment. All use the same decoder and parameter choices.

Table 7 compares union and ordered-block recovery at common sample budgets, with fixed-budget Wilson intervals. At $N = 14{,}691$, the union is recovered in $197/200$, $200/200$, and $195/200$ trials; the ordered hierarchy is recovered in $197/200$, $200/200$, and $187/200$. All three designs reach $200/200$ for both outcomes at $N = 58{,}646$.

Table 7: **Amplitude-law robustness at common allocations.** $(n, b, D, \sigma) = (128, 8, 2, 1/4)$ and $u = 1/16$. Four shared allocations from the frozen sample grid use multipliers $64, 128, 256, 1024$. Cells give recovered trials out of 200 and pointwise 95% Wilson intervals in percent. Every law is shown at every selected allocation; all 15 checkpoints remain in the numerical companion.

Design and allocation		**Exact recovery: count [95% CI]**	
Amplitude law	**Samples N**	**Union**	**Hierarchy**
Two-point amplitude law; 200 independent trials			
Two-point	3 717	$5/200$ $[1.1, 5.7]$	$3/200$ $[0.5, 4.3]$
Two-point	7 375	$137/200$ $[61.8, 74.5]$	$132/200$ $[59.2, 72.2]$
Two-point	14 691	$197/200$ $[95.7, 99.5]$	$197/200$ $[95.7, 99.5]$
Two-point	58 646	$200/200$ $[98.1, 100.0]$	$200/200$ $[98.1, 100.0]$
Uniform amplitude law; 200 independent trials			
Uniform	3 717	$11/200$ $[3.1, 9.6]$	$8/200$ $[2.0, 7.7]$
Uniform	7 375	$134/200$ $[60.2, 73.1]$	$129/200$ $[57.7, 70.8]$
Uniform	14 691	$200/200$ $[98.1, 100.0]$	$200/200$ $[98.1, 100.0]$
Uniform	58 646	$200/200$ $[98.1, 100.0]$	$200/200$ $[98.1, 100.0]$
Asymmetric amplitude law; 200 independent trials			
Asymmetric	3 717	$22/200$ $[7.4, 16.1]$	$5/200$ $[1.1, 5.7]$
Asymmetric	7 375	$124/200$ $[55.1, 68.4]$	$82/200$ $[34.4, 47.9]$
Asymmetric	14 691	$195/200$ $[94.3, 98.9]$	$187/200$ $[89.2, 96.2]$
Asymmetric	58 646	$200/200$ $[98.1, 100.0]$	$200/200$ $[98.1, 100.0]$

Table 8 summarizes all 76,800 normalized coordinate estimates $\widehat{\mu}_j/\Delta_u$ from these 600 trials at $N = 14{,}691$, grouped by true level. The means, standard deviations, and root mean squared errors quantify separation and calibration directly. Ground-truth levels are used for scoring and grouping; the learner uses the fixed union threshold $\Delta_u/2$ and nearest-level rounding. Coordinates within a trial share observations, so their pooled standard deviations describe estimate variability rather than uncertainty from independent trials.

Table 8: **Calibration of the recovered unary levels.** At the common allocation $N = 14{,}691$, pool $\widehat{\mu}_j/\Delta_u$ by the true population level h_j. All 76,800 coordinate estimates from 600 trials are included. SD describes the pooled coordinate distribution, not a standard error: coordinates within a trial share data. Ground-truth levels are used for this diagnostic only.

Coordinate group			Normalized estimate		
Amplitude law	**Level h_j**	**Coordinates**	**Mean ± SD**	**Bias**	**RMSE**
Two-point amplitude law; common allocation $N = 14{,}691$					
Two-point	0	22 400	0.000 ± 0.130	+0.000	0.130
Two-point	1	1 600	0.993 ± 0.130	-0.007	0.130
Two-point	2	1 600	1.989 ± 0.132	-0.011	0.133
Uniform amplitude law; common allocation $N = 14{,}691$					
Uniform	0	22 400	0.001 ± 0.129	+0.001	0.129
Uniform	1	1 600	0.979 ± 0.132	-0.021	0.133
Uniform	2	1 600	1.981 ± 0.131	-0.019	0.132
Asymmetric amplitude law; common allocation $N = 14{,}691$					
Asymmetric	0	22 400	-0.038 ± 0.128	-0.038	0.134
Asymmetric	1	1 600	0.931 ± 0.159	-0.069	0.173
Asymmetric	2	1 600	1.929 ± 0.167	-0.071	0.181

A.7 Estimators, Frozen Batches, and Reproduction

Unary estimator. Each trial draws a uniformly random ordered list of bD distinct coordinates and partitions it into D blocks. The target remains fixed, and every row draws a fresh hidden selector, independent amplitudes and signs, and Gaussian noise. With

$$K = \min\{k \in \mathbb{N} : k \text{ odd},\ k \geq 8\log(n/\delta)\}, \qquad \widehat{\mu}_j = \operatorname{median}_{1 \leq k \leq K} \left\{ \frac{1}{m} \sum_{i \in I_k} Y_i^2 G_{i,j} \right\},$$

use $N = mK$ observations in disjoint equal blocks. Decode the union by $\widehat{\mu}_j > \Delta_u/2$ and the ordered hierarchy by rounding $\widehat{\mu}_j/\Delta_u$ to the nearest integer in $\{0, \ldots, D\}$. This is the median-of-means decoder of Theorem K.2, with its logarithmic block count fixed throughout. The estimator receives observable magnitudes and labels; it receives neither selectors nor target coordinates. No top-k selection or law-specific threshold fitting is used.

The dispersion/width/law grid uses $m = \lceil tb\log(n/\delta)/K \rceil$, with multipliers

$$t \in \{16, 24, 32, 48, 64, 80, 96, 128, 160, 192, 256, 384, 512, 768, 1024\}.$$

The dimension/depth/noise grid uses S. Each block grows with fresh IID rows between checkpoints. Target placements are matched across amplitude laws or noise levels at fixed (n, b, D) and replicate, with separate observation streams. The population unary identity and fourth-moment calculation are checked against independent simulations.

Frozen-batch learner. The known widths determine a partition into R nonempty consecutive segments. Minimize the largest segment weight, then the complete candidate count, with lexicographic cuts resolving ties. This rule uses the known widths and (n, R). At each batch, freeze every extension at each rung's cumulative fresh size relative to the pre-batch carrier. Evaluate the entire vector of empirical sign–parity correlations and apply threshold $\gamma_{\text{sgn}}/(2D)$ from Algorithm J.2. Each requested size must have exactly one hit. Uniqueness failure terminates the run; otherwise the largest hit supplies the next carrier. No answer changes the queries inside its own batch.

Within a paired dataset, chains A and B share their six active coordinates, covariates, hidden selectors, and noise; the equal-width chain uses the first four coordinates. All round budgets receive the same data for their chain. Every submitted correlation counts, including those in a failed batch.

Complete scheduled budgets and actual submitted counts are recorded separately. Packed XOR and population counts compute exact integer sign–parity sums, validated against direct multiplication. The sample-reuse guarantee appears in Appendix J.

Exact sampling and audit. Unary sampling uses the exact joint law of (G, Y). Conditional on amplitudes and the selected rung, the product of its independent coordinate signs is a fresh fair sign η, giving

$$Y = a\eta \exp\left(\frac{1}{2}\sum_{j\in A_T} \log(1+\sqrt{u}\,G_j)\right) + \epsilon.$$

A coupled full-covariate implementation checks this identity. Unary estimates use float64 arithmetic. The unary scaling study uses $653{,}203{,}400$ IID rows in total, with nested checkpoints reusing rows within each trial. The companion provides experiment configurations, code, random seeds, coordinate-calibration estimates, trial-level outcomes, and complete numerical summaries. For the round experiments, code and seeds regenerate observations, query masks, integer response sums, batch boundaries, and stopping reasons. Independent checks recompute all $417{,}698{,}941$ submitted query sums and verify 206,170 frozen batches across the dimension–noise grid. Unary observations regenerate from pinned code and seeds. The figure and tables are derived from the recorded summaries and coordinate estimates included with the numerical companion. The numerical companion provides outcomes at every evaluated checkpoint, covering the full parameter grids summarized in the figure and tables.

B Formal Verification in Lean 4

We verify the mathematical mechanisms behind the round and distribution tradeoffs using Lean 4. The artifact connects carrier schedules, probability-space constructions, and adaptive oracle arguments to explicit recovery-success bounds. It also verifies the unary moment identities, exact dispersion optimization, and Hellinger tools used in the distributional analysis. All 911 named theorems, including supporting lemmas across 113 modules, pass the kernel and axiom audit.

The verification checks schedule attainability, agreement of probability laws and normalizations, and validity of the adaptive lower-bound construction. In the final lower bounds, an actual legal oracle acts against a randomized learner, and the success estimate holds for one fixed target. The following results connect these guarantees to explicit Lean declarations and reproducible audit outputs.

B.1 Verification Setup and Result Map

The development uses Lean 4.24.0 and mathlib v4.24.0 (de Moura & Ullrich, 2021; The mathlib Community, 2020). Finite carrier families are represented by finite sets; probability laws, conditional expectations, and query norms use mathlib's measure and L_2 constructions. In particular, the product reference defining an SQ norm remains distinct from the target law defining its expectation. Each theorem records its parameter restrictions and measurability assumptions explicitly.

The following map lists the principal entry points, with names relative to the `BlindJump` namespace. The artifact's `Audit.lean` lists every full theorem name; the source modules contain their statements and proofs. The 911 declarations count the reusable proof infrastructure together with the final results.

Paper result	Verified statement	Lean entry point
Appendix D	Closure, certificates, stabilization, and exact equal/weighted block widths	`feasible_iff_certificate`; `stationary_at_card`; `paper_equal_block_width`; `paper_weighted_chain_minimax`
Lemma E.1, Eq. (4)	Actual product projections and equivalence with centered L_2 product witnesses	`LabeledDetectability.detectable_iff_nonzero`
Theorem 2.2, lower bound	Conditional injection energy, legal oracle, and fixed-target success estimate	`BLeapSuccess.randomized`
Appendix F, hybrid projection	Actual joint Gaussian density and its conditional expectation	`RigidProductProjection.actual_density_condExp`
Theorem 3.1, $2/3$ lower bound	Arbitrary joint L_2 queries, legal oracle, and finite-dimensional success estimate	`RigidJointProtocol.randomized`
Lemma G.1	Unary expectation, exact fourth moment, variance, and query normalization	`UnaryMoments.unary_expectation`; `UnaryBounds.unary_variance`
Appendix I, exact optimization	Unique optimizer, noise threshold, and boundary infimum	`Goldilocks.minimizer_iff`; `Goldilocks.boundary_infimum`
Appendix H, information tools	Gaussian affinity, mixture contraction, tensorization, and testing	`GaussianAffinity.gaussian_affinity`; `Hellinger.iid_sample_lower_bound`

B.2 Structural Certificates and Exact Examples

The structural development starts from the paper's frozen-batch update: every witness in a batch uses the carrier available before that batch. It proves that budget feasibility is equivalent to a successful schedule certificate, and that the minimum certificate load is attained. Closure monotonicity and finite stabilization then identify the round and width endpoints.

For disjoint blocks of common positive width b and $R > 0$, the formalization proves

$$\mathrm{BLeap}_R = b \left\lceil \frac{D}{R} \right\rceil .$$

The Lean statement uses exactly the prefix family $\{A_1, \dots, A_D\}$. A separate lemma proves that inserting the empty prefix changes neither closure nor width. For unequal block widths, the verified minimax theorem supplies both an attaining partition and a lower bound valid for every partition. These results establish the staircase and weighted scheduling laws for all admissible parameters.

Exact examples make the scheduling distinction concrete. For the branching family

$$\mathcal{C}_{\mathrm{br}} = \big\{\{1\}, \{1,2,3\}, \{1,4,5\}, \{2,4,6,7\}\big\},$$

Lean verifies widths 4, 3, and 2 at one, two, and three rounds. It also proves width 2 for every $R \geq 3$ and infeasibility of budget 1 for every round count. For consecutive blocks of widths $(1, 9, 1)$, two rounds require width 10, whereas the rounded average load is 6. The declarations in `Examples` verify these claims using kernel-checked finite computations and quantified proofs. Together, they show why the certificate must retain the carrier structure and indivisible blocks.

For the equal-endpoint pair in Proposition D.7, Lean verifies both full width profiles, the strict ranking reversal between two and three rounds, and equality of the endpoint widths. In the `Examples` namespace, `reversal_strict_crossing` and `reversal_same_endpoints` combine finite kernel computations with quantified stabilization for every $R \geq 4$. The two families have the same six-coordinate support. The Gaussian-link realization is established by the likelihood and detectability argument in Appendix D.5.

B.3 Analytic Identities and Parameter Boundaries

The moment development constructs the hidden-rung expectation by averaging the selector, independent coordinate signs, and Gaussian noise. For $0 < u < 1$, the observable amplitude bit is $R_j = (X_j^2 - 1)/\sqrt{u}$. With $h_j = |\{t : j \in A_t\}|$, the verified identities are

$$\mathbb{E}[Y^2 R_j] = \frac{a^2\sqrt{u}}{D} h_j, \qquad V_u := \mathbb{E}Y^4 = \frac{a^4}{D}\sum_{t=1}^{D}(1+u)^{bt} + 6a^2\sigma^2 + 3\sigma^4.$$

The variance and reference-norm results use $R_j^2 = 1$ to connect these identities to the normalized unary signal. Thus the formal statements preserve both the signal introduced by dispersion and its estimation cost.

The optimization proofs cover the complete parameter domain of the sufficient-bound objective

$$g_{q_0,C}(u) = \frac{(1+u)^{q_0} + C}{u}, \qquad q_0 = bD, \quad C \geq 0, \quad 0 < u < 1.$$

For $q_0 > 2$, a unique interior minimizer exists precisely when $C < 2^{q_0-1}(q_0 - 2)$. At $C = 0$, it equals $1/(q_0 - 1)$. At and above the threshold, the infimum is $2^{q_0} + C$ as $u \uparrow 1$ and is not attained inside the domain. Separate theorems cover $q_0 = 1, 2$. The verified boundary statements distinguish an admissible optimizer from a limiting parameter value.

For the information analysis, the formalization integrates the actual Gaussian densities to obtain

$$\mathrm{Aff}\big(\mathcal{N}(m,\sigma^2), \mathcal{N}(m',\sigma^2)\big) = \exp\left(-\frac{(m-m')^2}{8\sigma^2}\right), \qquad \sigma > 0.$$

It then proves common-weight mixture contraction and the exact product identity

$$H^2(P^{\otimes N}, Q^{\otimes N}) = 1 - \big(1 - H^2(P,Q)\big)^N.$$

The testing reduction turns a small single-observation Hellinger distance into a necessary sample size for distinguishing the two targets. Each step is proved for the represented densities and measures, connecting the Gaussian calculation to the statistical testing statement.

B.4 Verified Statistical Recovery and Ordered Levels

The recovery development separates the stochastic accuracy guarantee from the deterministic decoding step. Both are verified, so the same decoding theorem applies to an SQ answer, a block-vote procedure, or a median-of-means estimate once its stated accuracy condition has been established.

Block estimation and voting. For independent square-integrable observations with common mean μ and variance at most V, the actual sample mean $\overline{W}_m$ satisfies

$$\mathbb{E}\overline{W}_m = \mu, \qquad \mathrm{Var}(\overline{W}_m) \leq V/m, \qquad \mathbb{P}\{|\overline{W}_m - \mu| \geq \eta\} \leq \frac{V}{m\eta^2}, \quad \eta > 0.$$

The proof derives the variance of the finite sum from independence. A separate Cantelli argument, using an optimized shift before applying the second-moment bound, yields

$$\mathbb{P}\{W - \mathbb{E}W \leq -t\} \leq \frac{\mathrm{Var}(W)}{\mathrm{Var}(W) + t^2}, \qquad t > 0.$$

Consequently, a positive block mean with signal at least Δ has positive sign with probability at least $3/4$ when $m \geq 3V/\Delta^2$.

The bounded-vote proof uses independent $[0,1]$-valued votes within each coordinate. At threshold $5/8$, inactive vote means at most $1/2$ and active vote means at least $3/4$ give error at most $e^{-K/32}$ in either case. A union bound gives at most $ne^{-K/32}$ probability of any coordinate error. The allocation $K \geq 32\log(2n/\delta)$ makes this at most $\delta/2$. Coordinates may share the same rows: the argument requires independence between the blocks used for one coordinate, not independence between coordinates.

Median accuracy and exact decoding. An odd-block median outside the interval $[\mu - \eta, \mu + \eta]$ requires a majority of inaccurate block estimates. With $m \geq 4V/\eta^2$, the verified block-error bound is at most $1/4$. Bounded-indicator concentration gives the median error rate $e^{-K/8}$ and simultaneous rate $ne^{-K/8}$. The median is constructed by sorting the block means, and its two majority-side inequalities are proved, including tied values.

The decoder is the explicit integer-valued map

$$\widehat{h}_j = \mathrm{round}(\widehat{\mu}_j/\Delta), \qquad \Delta > 0.$$

Lean proves that $|\widehat{\mu}_j - \Delta h_j| < \Delta/2$ implies $\widehat{h}_j = h_j$. Simultaneous accuracy therefore recovers every ordered level set and their active union. The paper's choice $\eta = \Delta/3$ lies strictly inside this decoding cell. The generic concentration results are instantiated on the actual nested IID product measure. Row and block independence, the block-mean moments, and the bad-block probabilities are derived from this sampling law. For bounded amplitudes, the observable statistic is $Y^2(X_j^2 - 1)/\sqrt{u}$; its almost-sure equality with $Y^2 G_j$ connects the estimator to the model-specific moments. The final declaration in `BoundedAmplitudeRecovery`, `bounded_amplitude_allocated_success`, proves simultaneous level recovery with probability at least $1 - \delta$, using the explicit integer allocations

$$m = \max\left\{1, \left\lceil \frac{36 r_u^2 V_u}{\Delta_u^2} \right\rceil\right\}, \qquad K = 2\left\lceil 4 \log \frac{n}{\delta} \right\rceil + 1.$$

The chain-count identities identify each recovered level set with its ordered block. The theorem covers arbitrary admissible bounded amplitude laws, nonzero signal, and $0 < \delta < 1$, including zero Gaussian noise variance. The sorted median, integer decoder, and final success and failure events are also proved measurable. No concentration conclusion is supplied as an assumption to this final theorem.

B.5 Verified Amplitude and Passive-Information Extensions

An arbitrary bounded amplitude law. The amplitude development uses a probability measure on $\mathbb{R}$, including discrete, continuous, and asymmetric laws. For $\mathbb{E}Z = 0$, $\mathbb{E}Z^2 = 1$, $|Z| \leq M$ almost surely, and $s = \sqrt{u} > 0$ with $sM < 1$, it derives $M \geq 1$, positivity of the radius, and the observable identity

$$X = \xi\sqrt{1 + sZ}, \qquad (X^2 - 1)/s = Z.$$

All polynomial and mixed-moment integrability follows from boundedness. The exact scalar moments are $\mathbb{E}X^2 = 1$, $\mathbb{E}X^4 = 1 + s^2$, and $\mathrm{Var}(X^2) = s^2$. The local mixed-moment estimate

$$\mathbb{E}[(1 + sZ)^k Z^2] \leq (1 + sM)^k$$

provides the variance control needed for actual IID observations.

Product factorization, independent fair signs, the hidden selector, and the Gaussian response then give the unary mean $a^2 s h_j/D$ and the exact fourth moment V_u. Writing $r_u = 1 + sM < 2$, the verified mixed-moment chain establishes

$$\mathbb{E}[Y^4 G_j^2] \leq r_u^2 V_u, \qquad \mathrm{Var}(Y^2 G_j) \leq r_u^2 V_u < 4V_u.$$

For nonzero signal, $V_u > 0$, giving the strict final inequality. For the product of the input law and the represented target's label marginal, the reference calculation separately gives $\|y^2 G_j\|^2_{L_2(\nu_{0,u})} = V_u$. This separates the target-law variance used for IID recovery from the reference norm used for an SQ tolerance. The parameter lemmas translate $0 < u < M^{-2}$ and $u = c/q$ with $q > cM^2$ into the radius-positivity conditions used by these proofs.

The three experimental amplitude designs are instantiated as actual measures: the fair sign, the uniform law on $[-\sqrt{3}, \sqrt{3}]$, and the asymmetric two-point law with masses $4/5$ at $-1/2$ and $1/5$ at 2. Lean verifies their normalization, first two moments, and bounds. Their entry points are `AmplitudeExamples.fair_law`, `continuous_law`, and `asymmetric_law`.

Selector and noise extensions. For an arbitrary known probability selector π on a finite rung set, the actual weighted target law gives

$$\mathbb{E}[Y^2 G_j] = a^2\sqrt{u}\sum_{t:j\in A_t}\pi_t, \qquad \mathbb{E}Y^4 = a^4\sum_t \pi_t(1+u)^{|A_t|} + 6a^2\sigma^2 + 3\sigma^4,$$

with the corresponding mixed-moment variance bound. These identities include zero selector weights; positivity of the weights needed for a particular recovery task is a separate condition. For independent centered noise with finite fourth moment m_4, the verified response identity replaces $3\sigma^4$ by m_4. Each rung is nonempty, so averaging its independent fair signs cancels the term involving the noise's third moment. No symmetry assumption on the noise is imposed. The entries `weighted_unary_variance_bound` and `noise_response_fourth` reside in the `BoundedAmplitude` namespace.

Fractional clipping for actual regression laws. For every real displacement z and every real exponent $0 < \alpha \le 2$, Lean verifies

$$1 - e^{-z^2} \le |z|^\alpha.$$

Applying this inequality to the Gaussian affinity integral yields, under any common covariate probability law,

$$H^2(P_m, P_{m'}) \le \frac{\mathbb{E}|m(X) - m'(X)|^\alpha}{(8\sigma^2)^{\alpha/2}}, \qquad \sigma > 0.$$

This generic bound assumes the displayed fractional difference moment is integrable. The verified common-weight contraction and IID testing tools then apply to the represented Gaussian regression mixtures. Thus clipping acts on actual densities and integrals, including covariate laws with atoms.

The `PassiveCollapse` module also constructs adjacent hidden-rung targets on one common ambient product space. With U_t the common coordinates in rung t, it proves the exact factorization

$$\mathbb{E}|m_t(X) - m'_t(X)|^\alpha = |a|^\alpha D_\alpha(\mu) M_\alpha(\mu)^{|U_t|}.$$

Unused coordinates integrate to one. Applying the mixture theorem gives the full fractional bound in Eq. (14). From $\mathbb{E}X^2 = 1$ and nonconstant modulus, the development derives $0 < r := \mathbb{E}|X| < 1$. Its final `hidden_rung_testing_lower_bound` proves that an IID test with success at least $3/4$ under each adjacent mixture needs

$$N \ge \frac{\sqrt{8}\,\sigma}{16|a|}\exp\left(q\log\frac{1}{r}\right), \qquad q \le \min_t |U_t|, \quad a \ne 0, \quad \sigma > 0.$$

For a first-block swap, $q = b - 1$. The strict moment contraction, product integrals, common-selector mixture, and binary-testing reduction are all derived within this chain.

Critical scale and rare-event energy. For the two-radius design, the exact first absolute moment satisfies

$$M_1(u)^2 = \frac{1+\sqrt{1-u}}{2} \le 1 - \frac{u}{4}, \qquad M_1(u) \le e^{-u/8}, \qquad 0 \le u \le 1.$$

The formal product calculation proves $\mathbb{E}|A| = M_1(u)^q$ for a q-coordinate monomial A. For $\varepsilon > 0$ this implies the paired statements

$$\mathbb{P}\{|A| > \varepsilon\} \le \varepsilon^{-1}e^{-uq/8}, \qquad \mathbb{E}[A^2\mathbf{1}\{|A| > \varepsilon\}] \ge 1 - \varepsilon^2.$$

These bounds verify the rare-energy mechanism at finite degree. A separate critical-scale theorem proves $(1 + c/q)^q \le e^c$ for $q > 0$ and $c \ge 0$, including every lower-degree rung and its uniform average. For the sufficient-bound objective, the verified finite-degree comparison is

$$g_{q,C}(u) \ge q, \qquad q \le g_{q,C}(1/q) \le q(e + C), \quad C \ge 0,$$

where $u > 0$ for the lower bound and $1/q \in (0, 1)$ when $q \ge 2$. This establishes the degree-linear scale of the objective without identifying it with a minimax sample complexity.

B.6 End-to-End Adaptive Lower Bounds

For the general model, the formal development identifies detectability with nonzero ANOVA components in the actual product L_2 space. It proves conditional uniformity of injection completions, derives the query-energy bound, and constructs the online oracle from true and truncated expectations. For $k = \mathrm{BLeap}_R > 0$ and $n \geq P + 2$, one fixed target satisfies

$$\mathbb{P}(\text{success}) \leq \frac{C_\rho q}{\tau^2 (n-P)^k} + \frac{1}{3}, \qquad C_\rho = 2^P P! \, \|L\|_2^2 > 0. \tag{19}$$

The probability averages the learner's private seed. The target is selected after this averaging, and the oracle is legal on every submitted history.

For the rigid model, the formalization constructs the actual joint Gaussian mixture and a norm-preserving correspondence between joint L_2 queries and their finite input sections. The hybrid density equals the conditional expectation retaining the label and disclosed block prefix. Distinct Fourier levels give the precise conditional energy bound. With $\ell = \lceil D/R \rceil - 1$ and $n \geq 2bD$, one fixed target satisfies

$$\mathbb{P}(\text{success}) \leq \frac{\beta_{\mathrm{lr}}^2 q}{D\tau^2 \binom{n-b(R-1)\ell}{b(\ell+1)}} + \frac{1}{2}. \tag{20}$$

This yields the $2/3$-success specialization of Theorem 3.1. Both final statements derive conditional energy and residual guessing bounds from the model. Their interfaces fix the batch size in each round and allow normalized queries to depend on the private seed, preceding transcript, and public grants.

B.7 Formal Assumptions and Verified Scope

The general lower bound assumes a square-integrable likelihood, full detectable support, positive rounds and tolerance, and a measurable success event under an independent private seed. The rigid bound uses disjoint positive-width blocks and nonzero Gaussian variance. Lean represents variance by $v = \sigma^2$ and structural ceilings by natural-number division with positive denominator.

The bounded-amplitude IID recovery and adjacent-mixture testing results connect their actual models to final probability bounds. The voting interface instead records its vote means and independence explicitly. Weighted selectors extend moments and variance; general centered noise extends the unary mean and response fourth moment. Constructive enumeration and sample-reuse programs, complete detectable-prefix classification, label-signal asymptotics, KL/Fano packing, and large-degree expansions retain their mathematical proofs in the following appendices. The source modules state the definitions and hypotheses of each verified declaration explicitly, giving its precise mathematical scope.

B.8 Reproducing the Build and Axiom Audit

The audit establishes a reproducible record for the statements summarized above:

Verification check	**Result**
Proof modules imported and built	113
Named theorems with audited axiom dependencies	911
Unproved placeholders or custom axioms	0
Uses of `native_decide`	0
Mathematical appendices with verified statements	8

Run `bash check.sh` with the pinned toolchain to build every module and audit every named theorem using `lean -trust=0`. Accepted foundational dependencies are `propext`, `Classical.choice`, and `Quot.sound`. The archive contains the Lean sources, dependency lockfile, and verification script. The script checks that every theorem is audited and rejects unproved placeholders, proof bypasses, and dependencies outside these foundational axioms.

C Proof Strategy and Notation

The appendix develops the argument in three stages: the combinatorics of a frozen batch, the analytic cost of locating a detectable interaction, and the statistical cost of estimating or distinguishing the resulting signals. All results concern the model stated in the main text. In particular, the hidden selector is redrawn independently for every sample. It is never revealed to either upper-bound algorithm.

C.1 Proof Dependencies

Theorem 2.2 combines a deterministic closure schedule with Lemma E.4. Its upper bound needs finitely many nonzero witnesses; its lower bound needs a square-integrable likelihood ratio and a legal oracle construction. Theorem 3.1 uses the sharper hidden-rung identities in Lemmas F.3 and F.4. Theorem 3.3 follows from Lemma G.1, a one-sided variance bound, and independent block votes. Proposition 3.5 and Theorem 3.6 use different divergences: an upper bound on KL for a support packing, and an upper bound on Hellinger distance for adjacent targets. Theorem 3.8 is a uniform-concentration argument over a fixed family of bounded parities. Its guarantee follows directly from simultaneous control of the candidate correlations. Proposition D.7 realizes equal-endpoint ranking reversal through weighted chains; Theorem K.2 combines exact unary moments with a uniform mixed-moment bound and median-of-means decoding.

C.2 Dimensions and Query Resources

Throughout, n is the ambient dimension. The general theorem fixes the latent dimension P as n grows. The hidden-rung specialization writes $P = bD$, where b is block width and D is the number of rungs. Its finite bounds allow bD to grow provided $n \geq 2bD$. The number of frozen batches is R, the number of scalar queries is q, and their relative tolerance is τ. The number of IID rows is N. An empirical implementation may evaluate many queries on the same N rows; q counts those scalar evaluations.

C.3 Signal and Distribution Parameters

The coefficient a is nonzero and the noise standard deviation σ is positive. Their ratio $|a|/\sigma$ is the SNR parameter used in the asymptotic statements. We write $\chi_C(x) = \prod_{j \in C} x_j$, with $\chi_\varnothing = 1$; this is a Boolean character only at the rigid endpoint. The two-radius dispersion is $u = \mathrm{Var}(X_j^2) \in (0, 1)$. The reference measure for an SQ is always the product of the input marginal and the target-independent label marginal. It is generally different from the true joint law of (X, Y).

C.4 Testing and Measure Conventions

Our squared Hellinger distance is

$$H^2(P, Q) = \frac{1}{2} \int (\sqrt{p} - \sqrt{q})^2 = 1 - \int \sqrt{pq}.$$

Thus its range is $[0, 1]$. Total variation is $\mathrm{TV}(P, Q) = \frac{1}{2} \int |p - q|$. All logarithms are natural. Product-space conditional expectations and likelihood ratios are understood up to null sets. For the general theorem we use standard Borel coordinate and label spaces, so the L_2 spaces admit countable bases and the finite conditional constructions are measurable.

D Carrier Closure and Batch Scheduling

The carrier calculus expresses the combinatorial structure through the detectable family $\mathcal{C}$, independently of the particular input and label laws. We take its ground set to be $\bigcup_{U \in \mathcal{C}} U = [P]$. Empty witnesses leave every closure unchanged. The seven numbered structural results and worked examples in this appendix have Lean counterparts listed in Appendix B.

D.1 Closure Monotonicity and Schedule Certificates

For a set $A \subseteq [P]$ known before a batch, recall

$$\Gamma_k(A) = A \cup \bigcup_{U \in \mathcal{C}:\, |U \setminus A| \le k} U.$$

An individual query template can guess at most k fresh coordinates. There may be many templates in a batch. Consequently k is a maximum fresh load per template, rather than a bound on the total number of coordinates discovered in parallel.

Lemma D.1 (Monotone one-wave calculus)**.** *If $A \subseteq A'$ and $k \le k'$, then $\Gamma_k(A) \subseteq \Gamma_{k'}(A')$. In particular, iterated closures are increasing in the number of waves and in the width budget, and $\mathrm{BLeap}_{R+1}(\mathcal{C}) \le \mathrm{BLeap}_R(\mathcal{C})$.*

Proof. For every U contributing to $\Gamma_k(A)$,

$$|U \setminus A'| \le |U \setminus A| \le k \le k'.$$

It therefore also contributes to $\Gamma_{k'}(A')$. The old set A is contained in A', proving the inclusion. Iteration preserves inclusion, and $A \subseteq \Gamma_k(A)$ gives monotonicity in the wave count. A width feasible in R waves remains feasible in $R+1$ waves. □

A useful alternative to closure is a *schedule certificate*. At stage r, choose a subfamily $\mathcal{U}_r \subseteq \mathcal{C}$, and set

$$A_0 = \varnothing, \qquad A_r = A_{r-1} \cup \bigcup_{U \in \mathcal{U}_r} U.$$

The load of the certificate is $\max_{r, U \in \mathcal{U}_r} |U \setminus A_{r-1}|$. Empty stages have zero load. The certificate succeeds if $A_R = [P]$.

Proposition D.2 (Certificate characterization)**.** $\mathrm{BLeap}_R(\mathcal{C})$ *equals the smallest load of a successful certificate with at most R stages.*

Proof. If a certificate has load at most k, induction using Lemma D.1 gives

$$A_r \subseteq \Gamma_k^{\circ r}(\varnothing).$$

Indeed, every $U \in \mathcal{U}_r$ needs at most k new coordinates relative to A_{r-1}, and hence at most k relative to the larger closure set. A successful certificate therefore implies $\Gamma_k^{\circ R}(\varnothing) = [P]$. Conversely, choose $\mathcal{U}_r$ to contain every set eligible at the start of wave r. The resulting certificate is precisely the iterated closure, and its load is at most k. Minimizing k proves the result. □

Every template in stage r uses the same pre-batch carrier A_{r-1}. This common reference makes the certificate faithful to frozen parallel access and assigns each new dependency to the next round.

D.2 Endpoint Widths and Structural Operations

Proposition D.3 (Endpoint formulas)**.** *Assume $\bigcup_{U \in \mathcal{C}} U = [P]$. Then*

$$\mathrm{BLeap}_1(\mathcal{C}) = \max_{i \in [P]} \min_{U \in \mathcal{C}:\, i \in U} |U|.$$

For any fixed k, closure either reaches $[P]$ within P waves or stabilizes at a strict subset. Hence the unrestricted-wave width is achieved by $R = P$.

Proof. One wave from the empty set includes coordinate i exactly when some detectable set containing i has size at most k. This proves the first formula. For the second claim, every nonstationary wave adds at least one coordinate. There can be at most P such waves. If a wave adds nothing, applying the same deterministic map again also adds nothing; no later wave can restart growth. □

These formulas explain the relationship with the nonadaptive Cover and sequential Leap endpoints of Joshi et al. (2024). A one-wave guarantee must provide every coordinate with a small detectable set immediately. Multiple waves can use intersections with earlier sets to reduce the fresh load. The definition quantifies how much of that sequential advantage survives a finite round budget.

Proposition D.4 (Adding witnesses and parallel components)**.** *If $\mathcal{C} \subseteq \mathcal{C}'$, then* $\mathrm{BLeap}_R(\mathcal{C}') \leq \mathrm{BLeap}_R(\mathcal{C})$. *Relabeling latent coordinates does not change the width. If $[P]$ is partitioned into disjoint components $V_1, \ldots, V_m$ and every detectable set lies in one component, then*

$$\mathrm{BLeap}_R(\mathcal{C}) = \max_j \mathrm{BLeap}_R(\mathcal{C}_j), \qquad \mathcal{C}_j = \{U \in \mathcal{C} : U \subseteq V_j\}.$$

Proof. Adding witnesses can only enlarge each closure wave. A permutation preserves all differences $|U \setminus A|$ and commutes with closure, proving relabeling invariance. For disjoint components, the closure restricted to V_j is exactly the closure generated by $\mathcal{C}_j$ on $A \cap V_j$. The components evolve independently in the same rounds. They all finish by round R exactly when the common budget is at least the largest component width. □

The maximum in the last formula, rather than a sum, reflects frozen parallel access. Query counts do add across components, but with fixed latent dimension this affects constants rather than the largest ambient exponent. This is one reason to keep the closure invariant distinct from an arithmetic operation count.

D.3 Equal-Width and Weighted Block Chains

Let $A_t = \bigcup_{s \leq t} B_s$ and $\mathcal{C} = \{A_1, \ldots, A_D\}$. Since the family is nested, the closure of a prefix is another prefix.

Lemma D.5 (Exact progress on a block chain)**.** *If all blocks have size b, then*

$$\Gamma_k(A_p) = A_{\min\{D, p + \lfloor k/b \rfloor\}}, \qquad A_0 = \varnothing.$$

Consequently $\mathrm{BLeap}_R = b\lceil D/R \rceil$.

Proof. For $t > p$, the set A_t requires exactly $b(t-p)$ fresh coordinates. It is eligible precisely when $t - p \leq \lfloor k/b \rfloor$. The union of eligible nested sets is the largest eligible prefix. Iterating R times reaches depth $\min\{D, R\lfloor k/b \rfloor\}$. Full recovery is therefore equivalent to $R\lfloor k/b \rfloor \geq D$. The smallest integer k satisfying this is $b\lceil D/R \rceil$. □

For unequal block sizes $b_1, \ldots, b_D$, the invariant is the minimum maximum segment weight over R ordered segments. Write $w(p,t) = \sum_{s=p+1}^{t} b_s$.

Proposition D.6 (Weighted chain scheduling)**.** *For the unequal-block chain,*

$$\mathrm{BLeap}_R = \min_{0 = t_0 \leq t_1 \leq \cdots \leq t_R = D} \max_{1 \leq r \leq R} w(t_{r-1}, t_r).$$

For a proposed budget k, the greedy rule that advances to the largest prefix with $w(p,t) \leq k$ reaches at least as far as every other feasible schedule after each wave.

Proof. Any sequence of cuts gives a certificate: in round r, query the detectable prefix A_{t_r} using $A_{t_{r-1}}$ as its carrier. Its fresh load is the displayed segment weight. Conversely, closure itself produces a sequence of prefix endpoints whose segment weights are at most k, so Proposition D.2 gives the reverse inequality. For greedy optimality, suppose the greedy endpoint is at least the endpoint of another schedule before a wave. Positivity of block sizes implies that any later prefix reachable from the smaller endpoint is also reachable from the larger endpoint whenever it lies beyond that endpoint. Thus the greedy endpoint remains at least as large after the wave, completing the induction. □

For example, widths $(1, 9, 1)$ with two rounds give width 10: either cut leaves a segment of weight 10, whereas $\lceil (1 + 9 + 1)/2 \rceil = 6$. Equal-size replication is therefore a special case of weighted scheduling. The explicit theorem in the main text uses equal widths so that both the schedule and the conditional combinatorial layers have a simple closed form.

D.4 Branching Carrier Families

Consider the detectable sets

$$\mathcal{C} = \{\{1\}, \{1,2,3\}, \{1,4,5\}, \{2,4,6,7\}\}.$$

With budget $k = 2$, the successive carriers are

$$\varnothing \longmapsto \{1\} \longmapsto \{1,2,3,4,5\} \longmapsto \{1,2,3,4,5,6,7\}.$$

The final set needs only $\{6,7\}$ after both branches have been exposed. With budget $k = 3$, the first wave already covers $\{1,2,3,4,5\}$, and the second finishes. In one wave, coordinates $6,7$ require the size-four set, so the profile is

$$\mathrm{BLeap}_1 = 4, \qquad \mathrm{BLeap}_2 = 3, \qquad \mathrm{BLeap}_3 = 2.$$

Budget one cannot finish at any number of waves: after discovering $\{1\}$, both size-three branches still require two fresh coordinates. This example distinguishes a finite-round profile from a simple division of total latent dimension by the number of rounds. It also illustrates why the closure retains a union of certified coordinates instead of a single selected interaction.

D.5 Identical Endpoints, Reversed Finite-Round Difficulty

Endpoint widths do not determine even the ordering of intermediate-round costs. Reordering the same four block sizes already gives a strict reversal.

Proposition D.7 (Equal-endpoint reversal)**.** *Let $\mathcal{C}_{\mathrm{A}}$ and $\mathcal{C}_{\mathrm{B}}$ be prefix families with respective block widths $(1,2,1,2)$ and $(1,1,2,2)$. Both have support size six, depth four, Cover 6, and Leap 2, but*

R	1	2	3	$R \geq 4$
$\mathrm{BLeap}_R(\mathcal{C}_{\mathrm{A}})$	6	3	3	2
$\mathrm{BLeap}_R(\mathcal{C}_{\mathrm{B}})$	6	4	2	2

Thus $\mathcal{C}_{\mathrm{A}}$ is easier at two rounds and $\mathcal{C}_{\mathrm{B}}$ is easier at three rounds, in the ambient exponent of Theorem 2.2.

Proof. By Proposition D.6, the width is the minimum largest segment sum over at most R contiguous segments. One segment has weight six for both chains. For $\mathcal{C}_{\mathrm{A}}$, the cut $(1,2) \mid (1,2)$ attains width three in two rounds, which is minimal since two segments must cover total weight six. A budget of two requires all four blocks to be separate: every adjacent pair has weight three. Hence three rounds still require width three. For $\mathcal{C}_{\mathrm{B}}$, the possible nontrivial two-segment cuts have maximum loads 5, 4, and 4; width four is attained by $(1,1) \mid (2,2)$. Three segments $(1,1) \mid (2) \mid (2)$ attain width two. In both families a size-two block forces width at least two for every round count, and four rounds suffice by taking individual blocks. These arguments give the whole table, including the unrestricted-round endpoints. □

Realization with the same input law and label channel. The table describes actual full-SQ learning problems. In both cases use Rademacher inputs, a fresh uniform selector on four prefixes, and the same known $a \neq 0$ and $\sigma > 0$ in (6), now with the indicated unequal block widths. Every prefix parity is a fair sign, so both links have the identical label marginal $\mu_y = \frac{1}{2}N(a,\sigma^2) + \frac{1}{2}N(-a,\sigma^2)$. Their respective likelihood ratios are

$$L(x,y) = 1 + \frac{w(y)}{4}\sum_{t=1}^{4}\chi_{A_t}(x), \qquad w(y) = \tanh(ay/\sigma^2).$$

This follows by conditioning on the selected prefix, exactly as in (31). Since $|w| \leq 1$, the likelihood lies in $[0,2]$ and is square-integrable. Every centered function of a Rademacher coordinate is a scalar multiple of that coordinate. Fourier orthogonality therefore makes every nonprefix, nonempty interaction undetectable, while

$$\mathbb{E}[w(Y)\chi_{A_t}(X)] = \frac{1}{4}\mathbb{E}_{\mu_y} w(Y)^2 > 0$$

detects each prefix. Thus the two detectable families are exactly $\mathcal{C}_{\mathrm{A}}$ and $\mathcal{C}_{\mathrm{B}}$. Theorem 2.2 yields optimal query–precision costs $\Theta(n^3)$ versus $\Theta(n^4)$ at two rounds and $\Theta(n^3)$ versus $\Theta(n^2)$ at three rounds, for fixed signal and noise. The reversal holds with the same latent dimension, depth, block-size multiset, input law, and label marginal. It identifies information that neither endpoint records.

E General SQ Round Complexity

We prove Theorem 2.2. The key point beyond the nonadaptive and fully sequential endpoints is that the energy bound must remain valid after conditioning on all earlier frozen batches.

E.1 Product Projections and Detectable Supports

Let $\zeta = \rho_y \otimes \mu_x^P$ and $L = d\rho/d\zeta$. Both specified marginals are fixed, so

$$\mathbb{E}_{Z\sim\mu_x^P} L(y, Z) = 1 \quad \rho_y\text{-a.e.}, \qquad \mathbb{E}_{Y\sim\rho_y} L(Y, z) = 1 \quad \mu_x^P\text{-a.e.}$$

In particular $\mathbb{E}_\zeta L = 1$. We use the orthogonal product decomposition in the covariates, leaving the label variable unexpanded. This is the product-space analogue of the Boolean Fourier decomposition (O'Donnell, 2014).

For $V \subseteq [P]$, define $\mathsf{E}_V f = \mathbb{E}[f(Y, Z) \mid Y, Z_V]$, with the missing coordinates integrated under the product reference. For $U \subseteq [P]$, let

$$\Pi_U f = \sum_{V\subseteq U} (-1)^{|U|-|V|} \mathsf{E}_V f. \tag{21}$$

This function depends only on (Y, Z_U) and has zero integral in each coordinate indexed by U when $U \neq \varnothing$. Write $L_U = \Pi_U L$.

Lemma E.1 (Orthogonal product decomposition)**.** *For every $f \in L_2(\zeta)$,*

$$f = \sum_{U\subseteq[P]} \Pi_U f, \qquad \|f\|_2^2 = \sum_{U\subseteq[P]} \|\Pi_U f\|_2^2 .$$

Moreover $L_\varnothing = 1$, and a nonempty U is detectable if and only if $L_U \neq 0$ in $L_2(\zeta)$.

Proof. Let E_i integrate the ith coordinate, keeping Y and all other coordinates fixed. The operators E_i are commuting orthogonal projections under the product measure. Expanding

$$I = \prod_{i=1}^{P} \big(E_i + (I - E_i)\big)$$

gives 2^P projections with mutually orthogonal ranges. The term using $(I - E_i)$ exactly for $i \in U$ is $(\prod_{i\in U}(I - E_i))(\prod_{i\notin U} E_i)$, which expands to (21). The norm identity follows by orthogonality. The marginal identity above gives $L_\varnothing = 1$.

If a centered product $T(Y)\prod_{i\in U} T_i(Z_i)$ has nonzero correlation with L, it lies in the range of Π_U, so its inner product with L_U is nonzero. Conversely, centered tensor products span that range densely: take an orthonormal basis consisting of the constant and centered coordinate functions, together with a basis in the label variable. If $L_U \neq 0$, at least one such tensor has nonzero inner product with L_U. It supplies the witnesses in (4). Every inner product exists by Cauchy–Schwarz under ζ; the product of the witness L_2 norms is exactly its reference norm. □

Lemma E.2 (Truncated links are probability laws)**.** *For $B \subseteq [P]$, the density of the link retaining only (Y, Z_B) is*

$$L^{(B)} = \mathsf{E}_B L = \sum_{U\subseteq B} L_U .$$

It is nonnegative, integrates to one, and has the same label marginal as L. After embedding, it defines the reference-decoupled law ν_B^ι used in the main theorem.

Proof. Conditional expectation of a nonnegative density is nonnegative and preserves its integral. Integrating all covariates yields the unchanged label marginal. In the orthogonal expansion, integrating any coordinate in U annihilates L_U; precisely those $U \subseteq B$ survive. The remaining latent coordinates are sampled independently from μ_x, as are all unused ambient coordinates. This is exactly the construction of ν_B^ι. □

The retained sum is a nonnegative normalized density with the prescribed label marginal. The ideal oracle therefore evaluates queries under an explicit probability law. The conditional truncation estimate controls its discrepancy from the target, and the online patch enforces tolerance validity on every history.

E.2 Conditional Counts for Random Embeddings

Fix the placement η of an old latent carrier A. Write $(m)_t = m(m-1)\cdots(m-t+1)$ for a falling factorial. For a particular latent support U, the $t = |U \setminus A|$ fresh coordinates are a uniform ordered injection into $n - |A|$ unused positions. Thus a prescribed ordered placement has probability $1/(n-|A|)_t$.

Lemma E.3 (Counting a prescribed ambient tensor)**.** *Let* $\mathbf{i}$ *be a fixed ambient tensor index. Conditional on* $\iota|_A = \eta$*, summing over all latent tensor indices* $\mathbf{j}$ *with support* U *gives*

$$\sum_{\mathrm{supp}(\mathbf{j})=U} \mathbb{P}\{\iota\mathbf{j} = \mathbf{i} \mid \eta\} \le \frac{t!}{(n-|A|)_t}.$$

If $t \ge k$ *and* $n > P$*, this is at most* $P!(n-P)^{-k}$.

Proof. If the old coordinate positions or the cardinality of the fresh ambient support disagree with $\mathbf{i}$, the sum is zero. Otherwise the t fresh latent positions can be matched to the t nonzero fresh ambient positions in at most $t!$ ways. Once such a matching is fixed, all basis indices in $\mathbf{j}$ are determined by $\mathbf{i}$. Each ordered placement has the probability stated above. Finally every factor of $(n-|A|)_t$ is at least $n - P$, $t! \le P!$, and $t \ge k$. □

The factorial counts latent-role assignments and contributes to the fixed-link constant. Appendix F instead exploits distinct Fourier levels to obtain explicit binomial counts for growing $P = bD$.

E.3 Conditional Energy and Truncation

Lemma E.4 (Conditional truncation)**.** *Fix* $A \subseteq B \subseteq [P]$*, an injection* $\eta : A \hookrightarrow [n]$*, and* $k \ge 1$*. Draw* ι *uniformly among injections extending* η*. Suppose*

$$U \in \mathcal{C}_{\mathrm{SQ}}(\rho), \quad U \not\subseteq B \quad \Longrightarrow \quad |U \setminus A| \ge k. \tag{22}$$

There is a constant C_ρ*, independent of* n*, such that every* $\phi \in L_2(\nu_0)$ *satisfies*

$$\mathbb{E}_{\iota|\iota|_A=\eta}\left[\left(\mathbb{E}_{\nu^\iota}\phi - \mathbb{E}_{\nu^\iota_B}\phi\right)^2\right] \le C_\rho (n-P)^{-k} \|\phi\|^2_{L_2(\nu_0)}. \tag{23}$$

Proof. Let $\{\psi_j\}_{j\in I}$ be a countable orthonormal basis of $L_2(\mu_x)$ with $\psi_0 = 1$. For a tensor index $\mathbf{j} \in I^P$, let $\psi_{\mathbf{j}}(z) = \prod_{u=1}^P \psi_{j_u}(z_u)$ and $\mathrm{supp}(\mathbf{j}) = \{u : j_u \ne 0\}$. Write

$$L(y,z) = \frac{d\rho}{d(\rho_y \otimes \mu_x^P)}(y,z),$$

and define latent and ambient coefficients

$$a_{\mathbf{j}}(y) = \mathbb{E}_Z[L(y,Z)\psi_{\mathbf{j}}(Z)], \qquad b_{\mathbf{i}}(y) = \mathbb{E}_X[\phi(y,X)\psi_{\mathbf{i}}(X)].$$

Parseval gives

$$\sum_{\mathbf{j}\in I^P} \|a_{\mathbf{j}}\|^2_{L_2(\rho_y)} = \|L\|^2_{L_2(\rho_y\otimes\mu_x^P)}, \qquad \sum_{\mathbf{i}\in I^n} \|b_{\mathbf{i}}\|^2_{L_2(\rho_y)} = \|\phi\|^2_{L_2(\nu_0)}. \tag{24}$$

Let $\iota\mathbf{j} \in I^n$ place j_u at ambient coordinate $\iota(u)$ and zeros elsewhere. Conditional projection onto $(Y, X_{\iota(B)})$ removes exactly the tensor indices not supported in B, hence

$$\begin{aligned}\Delta_\iota(\phi) &:= \mathbb{E}_{\nu^\iota}\phi - \mathbb{E}_{\nu^\iota_B}\phi \\ &= \sum_{\mathrm{supp}(\mathbf{j})\not\subseteq B} \langle a_{\mathbf{j}}, b_{\iota\mathbf{j}}\rangle_{L_2(\rho_y)}.\end{aligned} \tag{25}$$

Only detectable supports contribute. Indeed, if $a_{\mathbf{j}} \ne 0$, take $T = a_{\mathbf{j}}$ and $T_u = \psi_{j_u}$ in (4); their correlation is $\|a_{\mathbf{j}}\|_2^2 > 0$. Let $\mathcal{J}$ contain the tensor indices whose supports are detectable but not contained in B. Cauchy–Schwarz over $\mathcal{J}$ and (24) give

$$|\Delta_\iota(\phi)|^2 \le \|L\|_2^2 \sum_{\mathbf{j}\in\mathcal{J}} \|b_{\iota\mathbf{j}}\|_2^2. \tag{26}$$

Fix one eligible support U and put $t = |U \setminus A| \geq k$. Conditional on the anchor η, its t unanchored latent coordinates map as a uniform injection. For any fixed ambient tensor index, at most $t!$ assignments match it. Consequently,

$$\sum_{\mathrm{supp}(\mathbf{j})=U} \Pr\{\iota\mathbf{j} = \mathbf{i} \mid \eta\} \leq \frac{t!}{(n-|A|)_t} \leq P!(n-P)^{-k}.$$

Summing over at most 2^P supports, applying Tonelli, and using the ambient Parseval identity in (24) bounds the conditional expectation of the last sum in (26) by

$$2^P P!(n-P)^{-k} \|\phi\|^2_{L_2(\nu_0)}.$$

This proves the lemma with $C_\rho = 2^P P! \|L\|_2^2$. □

E.4 Upper and Lower Bounds for Frozen Batches

We now prove Theorem 2.2.

Proof. Write $\mathcal{C} = \mathcal{C}_{\mathrm{SQ}}(\rho)$ and $k = k_R$.

Upper bound. For each nonempty $U \in \mathcal{C}$, fix witnesses in (4) and normalize the product of their L_2 norms to one. Since P is fixed, the minimum absolute nonzero witness correlation β is positive. Use tolerance at most $\beta/4$ and threshold returned magnitudes at $\beta/2$.

Let $A_0 = \varnothing$ and $A_r = \Gamma_k(A_{r-1})$. Let K_r be the union of all ambient coordinates appearing in thresholded hits through batch r. We prove by induction that

$$\iota(A_r) \subseteq K_r \subseteq \iota([P]). \tag{27}$$

Given this invariant at $r-1$, the learner retains every injection $a : A_{r-1} \hookrightarrow K_{r-1}$ rather than choosing latent roles. There are at most P^P such maps because $K_{r-1} \subseteq \iota([P])$. For every U with $|U \setminus A_{r-1}| \leq k$, it extends each map by every injection of the fresh latent coordinates and submits the corresponding product witness. In particular, the true restriction $\iota|_{A_{r-1}}$ and its true extension are retained, so every element of $\iota(A_r)$ appears in a hit. A tuple containing an irrelevant coordinate has zero expectation because its centered factor is independent of the label and other factors; a wrong latent matching may reveal another relevant coordinate but cannot create a false one. This proves (27). Role ambiguity enlarges the constant-size state family but cannot remove the true state.

There are at most 2^P templates, P^P retained old-coordinate maps, and R batches; at most k fresh ambient coordinates are enumerated. The total is at most $R2^P P^P n^k$ queries. Since $A_R = [P]$, the final certified set is exactly the hidden support.

Lower bound. Run the failed closure one unit below k:

$$H_0 = \varnothing, \qquad H_r = \Gamma_{k-1}(H_{r-1}).$$

Minimality of k gives $H_R \neq [P]$, and

$$U \in \mathcal{C}, \quad U \not\subseteq H_r \quad \Longrightarrow \quad |U \setminus H_{r-1}| \geq k. \tag{28}$$

Draw ι uniformly from all injections. Strengthen the learner by revealing $\iota|_{H_{r-1}}$ before batch r and $\iota|_{H_r}$ after that batch is committed. In an ideal interaction, answer every batch-r query by its expectation under $\nu^{\iota}_{H_r}$. Conditional on the old grant, all queries in the batch are fixed and the remaining injection is uniform. Lemma E.4, (28), Markov's inequality, and a union bound give

$$\Pr\{\text{some ideal answer is not tolerance-valid}\} \leq \frac{C_\rho q}{\tau^2 (n-P)^k}. \tag{29}$$

For each fixed target, turn this ideal interaction into a legal online oracle. It returns a complete truncated batch whenever every coordinate is tolerance-valid; at the first failure it returns the exact expectation vector and continues exactly thereafter. This public-history rule is target-specific and stateful but does not observe private learner coins. Outside the event in (29), the legal and ideal transcripts coincide.

The ideal final transcript depends only on $\iota|_{H_R}$ and the learner's coins. If $m = P - |H_R| \geq 1$, the remaining unordered support is uniform, so its posterior guessing probability is at most

$$\binom{n - |H_R|}{m}^{-1}.$$

Average jointly over the random injection and learner coins. Success against the legal oracle is at most

$$\frac{C_\rho q}{\tau^2 (n-P)^k} + \binom{n - |H_R|}{m}^{-1}.$$

For a sufficiently small constant multiple of $(n-P)^k$, this is below $2/3$ for all large n. Hence some fixed injection and legal oracle defeat the learner. Since P is fixed, $(n-P)^k = \Theta(n^k)$, proving the claim. □

E.5 Conditional Uniformity of Adaptive Transcripts

The conditioning variables are the granted latent coordinates and independent learner coins. They leave the unrevealed injection uniform along every ideal transcript. Validity of the ideal answers is handled afterward by a union bound and pathwise coupling.

Here is the induction in detail. Let Ω denote the private random seed, independent of the target injection, and let

$$\mathcal{F}_r = \sigma\big(\Omega, \iota|_{H_r}\big).$$

The sets H_r are deterministic functions of the link and the failed width $k-1$. Before round one, every query is $\mathcal{F}_0$-measurable. Suppose the ideal transcript through round $r-1$ is $\mathcal{F}_{r-1}$-measurable. The next query batch is a function of that transcript, the old grant, and Ω, so it is also $\mathcal{F}_{r-1}$-measurable. Its truncated expectations depend only on the additional placement $\iota|_{H_r}$, making the enlarged transcript $\mathcal{F}_r$-measurable. Conditional on $\mathcal{F}_{r-1}$, the remaining injection is uniform because the seed is independent and no other target information enters the ideal transcript.

Normalize each nonzero query by its reference norm. A zero-norm query is zero almost everywhere under the reference and, by absolute continuity, under the target; it can be answered exactly and omitted. For a normalized batch-r query, Lemma E.4 and Markov's inequality give

$$\mathbb{P}\big\{|\Delta_\iota(\phi)| > \tau \mid \mathcal{F}_{r-1}\big\} \leq C_\rho \tau^{-2} (n-P)^{-k}.$$

The bound holds for every value of the conditioning variables for which the conditional law is defined. Integrating it does not require independence between different queries. Summing over all queries gives (29), even when later queries depend on earlier ideal answers. A learner that stops early can be padded by zero queries so that this union bound has a deterministic budget.

The oracle is legal on every path. The actual oracle checks every coordinate of the committed batch against its true tolerance interval. It returns the ideal vector when all coordinates are valid; at the first invalid batch, it returns exact expectations and remains exact thereafter. This rule is legal for every submitted history. The state bit and batch index are public-history functions, and private coins enter through submitted queries. On every valid ideal trajectory, the actual and ideal transcripts coincide by induction, yielding the success bound.

Recovering the support after the last grant. Let $h = |H_R|$ and $m = P - h > 0$. Conditional on the final grant, the image of the m ungranted latent coordinates is a uniform m-subset of the $n-h$ unoccupied ambient coordinates. Each such subset has exactly $m!$ ordered completions, so forgetting latent roles preserves uniformity. An output omitting a granted coordinate cannot be correct. Every other fixed output corresponds to one candidate remaining subset and succeeds with probability exactly $\binom{n-h}{m}^{-1}$. Randomizing the output averages these probabilities and cannot improve the maximum.

E.6 Witness Normalization and Link Constants

For each nonempty detectable U, choose one product witness and rescale it to reference norm one. There are at most $2^P - 1$ choices, and every chosen correlation is nonzero, so

$$\beta = \min_{U \in \mathcal{C} \setminus \{\varnothing\}} \left| \mathbb{E}_\rho \left[T_U(Y) \prod_{i \in U} T_{U,i}(Z_i) \right] \right| > 0.$$

The upper-bound constants can depend on these particular witnesses. A tolerance at most $\beta/4$ gives true returned magnitudes at least $3\beta/4$ and zero-mean false returned magnitudes at most $\beta/4$. The threshold $\beta/2$ is therefore separated from both. A false latent assignment containing only relevant coordinates is allowed to pass; the algorithm only certifies relevance and retains the true assignment among all candidate role maps. This is why no uniqueness assumption on the latent roles appears in the theorem.

The constants in the lower bound are also explicit enough to expose their dependence: $C_\rho = 2^P P! \, \|L\|_2^2$ is sufficient for the truncation estimate. For n large enough that the residual guessing probability is at most $1/3$, success $2/3$ implies

$$\frac{q}{\tau^2} \geq \frac{(n-P)^k}{3C_\rho}.$$

If additionally $n \geq 2P$, this is at least $n^k/(3 \cdot 2^k C_\rho)$. The resulting ambient exponent holds with (P, R, ρ) fixed. The rigid and critical calculations give the corresponding explicit dependence on structural size, signal, and witness margins.

F Rigid Staircase Round Complexity

F.1 Likelihood Decomposition and Detectable Prefixes

Let g_+ and g_- be Lebesgue densities of $N(a, \sigma^2)$ and $N(-a, \sigma^2)$. Their equal mixture has density $m = (g_+ + g_-)/2$. Direct division gives

$$\frac{g_+(y)}{m(y)} = 1 + w(y), \qquad \frac{g_-(y)}{m(y)} = 1 - w(y), \qquad w(y) = \tanh\Big(\frac{ay}{\sigma^2}\Big). \tag{30}$$

Indeed $g_+(y)/g_-(y) = \exp(2ay/\sigma^2)$, and $(e^{2z}-1)/(e^{2z}+1) = \tanh z$. The function w is odd, bounded by one in magnitude, and centered under the symmetric mixture μ_y.

For rigid covariates, each $\chi_{A_t}(X)$ is a fair sign, so the marginal label law is μ_y for every target. The joint likelihood ratio relative to $\nu_0 = \mathrm{Rad}^n \otimes \mu_y$ is

$$L_{\mathbf{B}}(x, y) = 1 + \frac{w(y)}{D} \sum_{t=1}^{D} \chi_{A_t}(x). \tag{31}$$

This density lies in $[0, 2]$. In particular, every reference-L_2 query has a finite expectation under the target. Orthogonality of distinct parity characters gives

$$\|L_{\mathbf{B}}\|^2_{L_2(\nu_0)} = 1 + \frac{\beta_{\mathrm{lr}}^2}{D}, \qquad \beta_{\mathrm{lr}}^2 = \mathbb{E}_{\mu_y} w(Y)^2. \tag{32}$$

Thus the analytic likelihood norm stays bounded even when the total degree grows.

Proposition F.1 (Only prefixes are detectable at the rigid endpoint)**.** *The nonempty detectable sets of the hidden-rung link on its bD latent coordinates are exactly $A_1, \ldots, A_D$.*

Proof. The centered L_2 space of a fair sign is one-dimensional, generated by $x \mapsto x$. Hence every nonzero centered product witness on a set U is a scalar multiple of χ_U. For any $h \in L_2(\mu_y)$ and nonempty U, (31) gives

$$\mathbb{E}[h(Y)\chi_U(X)] = \frac{\mathbb{E}_{\mu_y}[h(Y)w(Y)]}{D} \sum_{t=1}^{D} \mathbf{1}\{U = A_t\}.$$

This is zero unless U is a prefix. For a prefix, choosing $h = w$ gives $\beta_{\mathrm{lr}}^2/D > 0$, since $a \neq 0$ and $\sigma > 0$. Thus all and only the prefixes are detectable. This identifies the closure family for Lemma D.5 without restricting the class of full SQs. □

F.2 Optimal and Bounded Label Witnesses

The lower bound uses the entire antisymmetric label channel. The upper bound uses its bounded sign correlation, which admits a simple empirical estimator.

Lemma F.2 (Label-channel signal scales)**.** *Let* $\kappa = 2\Phi(a/\sigma) - 1$ *and* $\gamma_{\mathrm{sgn}} = |\kappa|$. *Then*

$$\mathbb{E}[\mathrm{sign}(Y) \mid \chi_{A_T}(X) = s] = s\kappa, \qquad 0 < \gamma_{\mathrm{sgn}} \le \beta_{\mathrm{lr}} < 1.$$

Moreover

$$\sup_{\|h\|_{L_2(\mu_y)} \le 1} |\mathbb{E}_{\mu_y}[h(Y)w(Y)]| = \beta_{\mathrm{lr}},$$

with equality for $h = w/\beta_{\mathrm{lr}}$.

Proof. For $Y \sim N(as, \sigma^2)$, symmetry and the normal distribution function give $\mathbb{E}\,\mathrm{sign}(Y) = s(2\Phi(a/\sigma) - 1)$. Averaging the two components shows

$$\mathbb{E}_{\mu_y}[\mathrm{sign}(Y)w(Y)] = \frac{1}{2}\big(\mathbb{E}_{Q_+}\,\mathrm{sign}(Y) - \mathbb{E}_{Q_-}\,\mathrm{sign}(Y)\big) = \kappa.$$

The sign function has reference norm one. Cauchy–Schwarz implies $|\kappa| \le \|w\|_2 = \beta_{\mathrm{lr}}$. Since the two Gaussian components are distinct but mutually absolutely continuous, w is nonzero and strictly between -1 and 1 almost everywhere, proving the strict bounds. The variational assertion is the equality case of Cauchy–Schwarz. □

Writing $s = |a|/\sigma$, these scales have the same order at low SNR:

$$\gamma_{\mathrm{sgn}}^2 = \frac{2}{\pi}s^2 + O(s^4), \qquad \beta_{\mathrm{lr}}^2 = s^2 + O(s^4) \quad (s \downarrow 0). \tag{33}$$

The first expansion follows from the Taylor expansion of Φ at zero. For the second, represent $Y/\sigma = sS + G$ with independent $S \sim \mathrm{Rad}$ and $G \sim N(0,1)$, absorbing the sign of a into S. Then the argument of $\tanh$ is $s^2S + sG$. Its second moment is $s^2 + s^4$, its fourth moment is $O(s^4)$ for $s \le 1$, and $|\tanh^2 z - z^2| \le Cz^4$ for all real z. Taking expectations proves (33). As $s \to \infty$, both scales tend to one: the sign formula is immediate, and $|s^2S + sG| \to \infty$ almost surely permits dominated convergence for $\tanh^2$.

F.3 Reduction of Coupled Statistical Queries

Let

$$w(y) = \tanh\Big(\frac{ay}{\sigma^2}\Big), \qquad d_\phi(x) = \int \phi(x,y)w(y)\mu_y(dy).$$

Lemma F.3 (Full-SQ antisymmetric slice)**.** *Every* $\phi \in L_2(\nu_0)$ *satisfies*

$$\mathbb{E}_P\phi - \mathbb{E}_{\nu_0}\phi = \frac{1}{D}\sum_{t=1}^{D}\langle d_\phi, \chi_{A_t}\rangle,$$
$$\|d_\phi\|_2 \le \beta_{\mathrm{lr}}\,\|\phi\|_{L_2(\nu_0)}.$$

Proof. Relative to μ_y, the two Gaussian label densities have likelihood ratios $1 + w$ and $1 - w$. Therefore

$$\mathbb{E}_{Q_s}\phi(x,Y) = \mathbb{E}_{\mu_y}\phi(x,Y) + sd_\phi(x), \qquad s \in \{-1,1\}.$$

Averaging over the hidden selector proves the first identity. Pointwise Cauchy–Schwarz followed by integration over x proves the norm bound. □

F.4 Hybrid Truncation and Fourier Energy

For $0 \le h \le D$, define the scalar truncation operator exactly by

$$M_h^{\mathbf{B}}(\phi) = \mathbb{E}_{\nu_0}\phi + \frac{1}{D}\sum_{t=1}^{h} \langle d_\phi, \chi_{A_t} \rangle . \tag{34}$$

This is the expectation under a genuine hybrid distribution: draw the same hidden selector T; if $T \le h$, use the target Gaussian label, and if $T > h$, draw Y independently from μ_y. Hence $M_h^{\mathbf{B}}$ depends on the target only through $(B_1, \dots, B_h)$.

Lemma F.4 (Conditional unseen-suffix energy)**.** *Assume $n \ge 2bD$. Condition on the first k blocks, where $0 \le k \le h < D$, and let ϕ be fixed under this conditioning. Then*

$$\mathbb{E}\left[\left| \mathbb{E}_P \phi - M_h^{\mathbf{B}}(\phi) \right|^2 \,\middle|\, B_{1:k} \right] \le \frac{\beta_{\mathrm{lr}}^2 \, \|\phi\|_{L_2(\nu_0)}^2}{D\binom{n-bk}{b(h+1-k)}}.$$

Proof. Put $K = A_k$ and expand $d_\phi = \sum_C \widehat{d}_\phi(C)\chi_C$. The discrepancy is

$$\Delta = \frac{1}{D}\sum_{t>h} \widehat{d}_\phi(A_t).$$

Cauchy–Schwarz across at most D unseen rungs gives

$$\Delta^2 \le \frac{1}{D}\sum_{t>h} \widehat{d}_\phi(A_t)^2.$$

Conditional on $A_k = K$, the extension $A_t \setminus K$ is a uniform $b(t-k)$-subset of the $n - bk$ unused coordinates. Hence

$$\mathbb{E}\left[\widehat{d}_\phi(A_t)^2 \,\middle|\, A_k = K \right] = \frac{1}{\binom{n-bk}{b(t-k)}} \sum_{\substack{U \subseteq [n]\setminus K \\ |U| = b(t-k)}} \widehat{d}_\phi(K \cup U)^2.$$

The assumption $n \ge 2bD$ gives $b(t-k) \le (n-bk)/2$, so these denominators are nondecreasing in t. The Fourier levels for distinct values of t are disjoint. Therefore

$$\begin{aligned}
\mathbb{E}[\Delta^2 \mid A_k = K] &\le \frac{1}{D\binom{n-bk}{b(h+1-k)}} \sum_{t>h} \sum_{\substack{U \subseteq [n]\setminus K \\ |U| = b(t-k)}} \widehat{d}_\phi(K \cup U)^2 \\
&\le \frac{\|d_\phi\|_2^2}{D\binom{n-bk}{b(h+1-k)}} \\
&\le \frac{\beta_{\mathrm{lr}}^2 \, \|\phi\|_{L_2(\nu_0)}^2}{D\binom{n-bk}{b(h+1-k)}}.
\end{aligned}$$

The second line is Parseval and the last is Lemma F.3. □

F.5 Recovery Schedules and Legal Lower-Bound Oracles

Theorem F.5 (Rigid round law, restatement)**.** *The upper and lower conclusions of Theorem 3.1 hold.*

Proof. **Upper bound.** Suppose A_{p_r} is known at the beginning of batch $r+1$. For every $1 \le s \le s_r$ and every bs-subset $U \subseteq [n] \setminus A_{p_r}$, query

$$\phi_U(x, y) = \mathrm{sign}(y)\chi_{A_{p_r} \cup U}(x).$$

This query is pointwise bounded and has reference norm one. Fourier orthogonality gives

$$|\mathbb{E}_P \phi_U| = \frac{\gamma_{\mathrm{sgn}}}{D} \mathbf{1}\{U = A_{p_r+s} \setminus A_{p_r}\}.$$

Every true returned magnitude is at least $3\gamma_{\mathrm{sgn}}/(4D)$, while every false magnitude is at most $\gamma_{\mathrm{sgn}}/(4D)$. Thresholding identifies the unique true extension at every requested size. It therefore identifies $A_{p_r+s_r}$ for the next batch. Summing the candidate counts gives (8); after $J \le R$ batches the learner knows A_D.

Lower bound. Draw the ordered blocks uniformly and strengthen the learner by revealing the first $(r-1)\ell$ blocks before it commits to batch r. Let Ω be the learner's private coins and define the filtration

$$\mathcal{G}_r = \sigma(\Omega, B_1, \dots, B_{r\ell}), \qquad r = 0, \dots, R.$$

In an ideal interaction, answer every query in batch r using $M^{\mathbf{B}}_{r\ell}$ from (34). Since $R\ell < D$, every truncation is strict. By induction, the transcript before round r and the newly submitted batch are $\mathcal{G}_{r-1}$-measurable. The ideal answers use only this batch and $(B_1, \dots, B_{r\ell})$, so the transcript after the round is $\mathcal{G}_r$-measurable. Conditional on $\mathcal{G}_{r-1}$, the ungranted ordered block suffix remains uniform because Ω is independent of the target prior.

Normalize a nonzero query to reference norm one. Apply Lemma F.4 with $k = (r-1)\ell$ and $h = r\ell$. Markov's inequality bounds the conditional probability of an invalid ideal answer by

$$\frac{\beta_{\mathrm{lr}}^2}{D\tau^2 \binom{n-b(r-1)\ell}{b(\ell+1)}} \le \frac{\beta_{\mathrm{lr}}^2}{D\tau^2 B_{n,b,D,R}}.$$

The ideal transcript is defined regardless of validity. The tower property and a union bound over all submitted queries give

$$\mathbb{P}\{\text{some ideal answer is invalid}\} \le \frac{\beta_{\mathrm{lr}}^2 q}{D\tau^2 B_{n,b,D,R}}.$$

For each fixed target, define a legal online oracle with one public state bit F, initialized to zero. Given a submitted finite batch, compute its ideal vector and exact target vector. If $F = 0$ and every ideal coordinate lies inside its own true tolerance interval, return the entire ideal vector. Otherwise set $F = 1$, return the entire exact vector, and return exact vectors forever after. Every returned coordinate is legal. The map is target-specific, measurable in the public history and submitted queries, and depends on private coins only through those public objects. Whenever all ideal answers are valid, F remains zero and the actual and ideal transcripts coincide inductively.

Condition on the learner's coins, the final ideal transcript, and every grant. These objects depend on the target only through the first $R\ell$ blocks. The remaining active coordinates are therefore a uniform subset of size

$$m = b(D - R\ell)$$

among the $n - bR\ell$ unused coordinates. Even an unbounded output rule identifies the active union with conditional probability at most $g_{n,b,D,R}$ from (10), which is at most $1/2$. Thus the prior-averaged success probability against the patched oracle is at most

$$g_{n,b,D,R} + \frac{\beta_{\mathrm{lr}}^2 q}{D\tau^2 B_{n,b,D,R}}.$$

If every fixed target and its patched oracle had larger coin-averaged success, their prior average would also be larger. Hence some fixed target–oracle pair obeys the same bound. A learner guaranteed success s against every such pair must therefore satisfy

$$\frac{q}{\tau^2} \ge \frac{D(s - g_{n,b,D,R})}{\beta_{\mathrm{lr}}^2} B_{n,b,D,R},$$

proving (11). The $s = 2/3$ simplification uses $g_{n,b,D,R} \le 1/2$. □

F.6 Exact Candidate Counts and Endpoint Regimes

The schedule uses $L = \lceil D/R \rceil$, $J = \lceil D/L \rceil$, and $p_r = rL$. Since $L \ge D/R$, one has $J \le R$. At a round starting at depth p_r, every candidate of fresh size bs lies in the $n - bp_r$ unused coordinates, giving exactly $\binom{n-bp_r}{bs}$ queries. Summing over the nonempty requested sizes yields (8). In particular,

$$K_{n,b,D,R} \ge \binom{n}{bL}, \qquad K_{n,b,D,R} \le D\binom{n}{bL} \le RL\binom{n}{bL}.$$

The upper inequality uses $\sum_r s_r = D$ and monotonicity of binomial coefficients up to the middle layer. The assumption $bL \le bD \le n/2$ is exactly what permits that monotonicity.

For the lower schedule, $\ell = L - 1$ satisfies $R\ell < D$, since $\lceil D/R \rceil - 1 < D/R$. Thus the final ungranted support is nonempty. If $N_0 = n - bR\ell$ and $m = b(D - R\ell)$, then

$$N_0 \ge 2bD - bR\ell \ge 2m,$$

so $1 \le m \le N_0/2$ and $\binom{N_0}{m} \ge 2$. This proves $g_{n,b,D,R} \le 1/2$, including the smallest admissible cases. The $2/3$-success simplification therefore holds at every admissible finite dimension.

At $R = 1$, the largest queried set has size bD and the lower binomial is $\binom{n}{bD}$. At $R = D$, each jump has width b and

$$K_{n,b,D,R} = \sum_{p=0}^{D-1} \binom{n - bp}{b}, \qquad B_{n,b,D,R} = \binom{n}{b}.$$

For $b = 1$, the former equals $Dn - D(D-1)/2$, explaining the $O(nD)$ evaluation count. For $D = 1$, the two query counts coincide at $\binom{n}{b}$; their explicit signal and success factors remain as stated in the theorem. These cases check the interpretation of batch endpoints directly.

Finite-parameter ratio. With $k = bL$ and $\Delta = b(R-1)(L-1)$, direct cancellation gives

$$\frac{\binom{n}{k}}{\binom{n-\Delta}{k}} = \prod_{i=0}^{k-1} \frac{n-i}{n-\Delta-i} \le \left(\frac{n}{n-\Delta-k+1}\right)^k.$$

Combining with the upper count proves Remark 3.2. If $(\Delta+k)/n$ tends to zero, $-\log(1-x) = O(x)$ gives an exponential correction $\exp(O(k(\Delta+k)/n))$. When density is constant, this correction can itself be exponential in the structural parameters. The two binomial bounds preserve these density-dependent factors throughout the admissible parameter range.

F.7 Query-Dependent Precision Budgets

The same proof allows a precision selected separately for each query. Suppose a nonzero query ϕ_i uses tolerance $\tau_i \|\phi_i\|_2$, with $\tau_i > 0$ committed before its answer. Assume the precision budget $\sum_i \tau_i^{-2}$ is bounded by a deterministic $\mathcal{W}$ on every interaction path. Then the conditional invalidity bound for that query is $\beta_{\mathrm{lr}}^2/(DB_{n,b,D,R}\tau_i^2)$. Summing along the ideal path and taking expectations gives

$$\mathbb{P}(\text{some ideal answer is invalid}) \le \frac{\beta_{\mathrm{lr}}^2 \mathcal{W}}{DB_{n,b,D,R}}.$$

Consequently success at least s for every target and legal oracle requires

$$\mathcal{W} \ge \frac{D(s - g_{n,b,D,R})}{\beta_{\mathrm{lr}}^2} B_{n,b,D,R}. \tag{35}$$

A common tolerance recovers $\mathcal{W} = q/\tau^2$. This extension confirms that the price is precision-weighted query effort. Its resource measure accounts directly for the tolerance assigned to each scalar query.

F.8 Finite Precision and Exact-Real Encodings

An exact scalar expectation can distinguish finitely many supports even when it is impractical to estimate. Enumerate all nonempty ambient subsets with sizes in $\{b, 2b, \dots, bD\}$ as $C_1, \dots, C_M$, and define the finite query

$$\phi(x, y) = w(y) \sum_{i=1}^{M} 2^{-i} \chi_{C_i}(x).$$

Its reference norm is finite because the characters are orthogonal. Equation (31) gives

$$\mathbb{E}_{\mathbf{B}} \phi = \frac{\beta_{\mathrm{lr}}^2}{D} \sum_{i:\, C_i \in \{A_1, \dots, A_D\}} 2^{-i}.$$

Different finite subsets of the indexed powers have different sums: after multiplying by 2^M, the result is an integer with a unique finite binary expansion. The exact answer therefore encodes the entire prefix family, and in particular its largest support. Decoding may require enormous computation and precision. The smallest spacing is on the scale of 2^{-M}, making tolerance an essential component of query cost. The currency q/τ^2 records this precision requirement explicitly.

G Unary Recovery at Critical Dispersion

The observed input magnitude provides a unary witness with an exact mean and second moment. These identities quantify the signal introduced by dispersion and the variance paid to estimate it. Independent sample blocks convert this signal into simultaneous support recovery.

G.1 Amplitude Decoding and Unary Signals

For $0 < u < 1$, let ξ_j, R_j be independent fair signs and set

$$X_j = \xi_j \sqrt{1 + \sqrt{u}\, R_j}.$$

Both radii are strictly positive and distinct. Thus the latent construction has an observable inverse:

$$\xi_j = \operatorname{sign}(X_j), \qquad R_j = \frac{X_j^2 - 1}{\sqrt{u}} \in \{-1, 1\}.$$

This inverse decodes the amplitude bit from X_j whenever $u > 0$; at $u = 0$, the two radii coincide and the amplitude channel disappears.

The coordinate law is symmetric and standardized. Directly,

$$\begin{aligned} &\mathbb{E}X_j = 0, && \mathbb{E}X_j^2 = 1, && \mathbb{E}X_j^4 = 1 + u, \\ &\mathbb{E}R_j = 0, && \mathbb{E}R_j^2 = 1, && \mathbb{E}[X_j^2 R_j] = \sqrt{u}. \end{aligned} \tag{36}$$

For any integer $s \geq 0$, its even moment is

$$\mathbb{E}X_j^{2s} = \frac{(1 + \sqrt{u})^s + (1 - \sqrt{u})^s}{2}.$$

Independence across coordinates makes monomial moments products of these scalar moments. These identities follow directly from the two-radius coordinate law.

G.2 Exact Means and Fourth Moments

Let $h_j = |\{t : j \in A_t\}|$. This count is zero outside the active union; for $j \in B_s$, it is $D - s + 1$. Define $Z_j = Y^2 R_j$ and $V_u = \mathbb{E}Y^4$.

Lemma G.1 (Response moments)**.** *For the two-radius hidden-rung model,*

$$\mathbb{E}Z_j = \frac{a^2\sqrt{u}}{D} h_j, \tag{37}$$

$$\mathbb{E}Y^2 = a^2 + \sigma^2, \tag{38}$$

$$V_u = \frac{a^4}{D} \sum_{t=1}^{D} (1 + u)^{bt} + 6a^2\sigma^2 + 3\sigma^4, \tag{39}$$

$$\operatorname{Var}(Z_j) = V_u - \left(\frac{a^2\sqrt{u}}{D} h_j \right)^2 \leq V_u. \tag{40}$$

In particular $V_u \leq a^4(1 + u)^{bD} + 6a^2\sigma^2 + 3\sigma^4$.

Proof. Condition on $T = t$ and write $F_t = \prod_{k \in A_t} X_k$. Expanding the square gives

$$Y^2 R_j = a^2 F_t^2 R_j + 2aF_t \epsilon R_j + \epsilon^2 R_j.$$

The middle term has mean zero because ϵ is centered and independent. The final term has mean $\sigma^2 \mathbb{E} R_j = 0$. If $j \notin A_t$, then R_j is independent of F_t^2 and centered. If $j \in A_t$, independence and (36) give

$$\mathbb{E}[F_t^2 R_j] = \mathbb{E}[X_j^2 R_j] \prod_{k \in A_t \setminus \{j\}} \mathbb{E} X_k^2 = \sqrt{u}.$$

Averaging over the D equally likely rungs proves (37).

Likewise $\mathbb{E} F_t^2 = 1$, so Gaussian noise expansion yields (38). Expanding the fourth power and integrating the independent Gaussian noise gives

$$\mathbb{E}[Y^4 \mid T = t] = a^4 \mathbb{E} F_t^4 + 6a^2\sigma^2 \mathbb{E} F_t^2 + 3\sigma^4.$$

Odd noise moments vanish. Independence gives $\mathbb{E} F_t^4 = (1+u)^{bt}$, proving (39). Finally $R_j^2 = 1$ pointwise, even though R_j and Y need not be independent. Consequently $\mathbb{E} Z_j^2 = \mathbb{E} Y^4 = V_u$ and subtracting the squared mean proves (40). The last inequality bounds each term in the average by its largest term. □

Selecting a single rung gives Y^2 the coordinate energy of that monomial. This removes cross-rung products $F_s F_t$ and yields the exact unary mean in (37). Proposition 2.3 identifies the distinct screening mechanism for deterministic sums.

G.3 Reference-Normalized SQ Recovery

Let $\mu_{y,u}$ be the marginal law of Y and let $\nu_{0,u} = \mu_u^n \otimes \mu_{y,u}$. Relabeling input coordinates does not change the law of any degree-bt monomial, hence this reference is common to all block embeddings. The known parameters (a, σ, b, D, u) determine the reference moments used to normalize each query.

Lemma G.2 (Norm and margin of a unary query). *For $\phi_j(x, y) = y^2(x_j^2 - 1)/\sqrt{u}$, one has*

$$\|\phi_j\|_{L_2(\nu_{0,u})} = \sqrt{V_u}, \qquad \frac{\mathbb{E}\phi_j}{\|\phi_j\|_{L_2(\nu_{0,u})}} = \frac{a^2\sqrt{u}}{D\sqrt{V_u}} h_j.$$

Every active coordinate therefore has normalized margin at least $\lambda_u = a^2\sqrt{u}/(D\sqrt{V_u})$.

Proof. Under the product reference, the squared norm factors into $\mathbb{E} Y^4$ and $\mathbb{E} R_j^2 = 1$. The mean under the target is (37). The product reference supplies the query norm; the true joint law supplies the correlated unary mean. □

If $\tau < \lambda_u/4$, every inactive returned value has magnitude below $a^2\sqrt{u}/(4D)$, while every active value is positive and above $3a^2\sqrt{u}/(4D)$. Thresholding at $a^2\sqrt{u}/(2D)$ proves Corollary 3.4. The n queries are fixed before any answer is observed. Their joint precision cost is $O(nD^2V_u/(a^4u))$ when tolerance is selected as a fixed fraction of the displayed margin. This expression retains the dependence on growing block width that a fixed-link exponent would hide.

G.4 Simultaneous Recovery by Block Voting

Set $\Delta_u = a^2\sqrt{u}/D$. Let a block contain m independent rows, and define

$$S_j = \sum_{i=1}^{m} Y_i^2 R_{i,j}, \qquad W_j = \mathbf{1}\{S_j > 0\}.$$

The rows are independent, whereas different coordinates within a row need not produce independent votes.

Lemma G.3 (Active and inactive vote probabilities). *If $m \geq 3V_u/\Delta_u^2$, then $\mathbb{P}(W_j = 1) \geq 3/4$ for every active j. For every inactive j, $\mathbb{P}(W_j = 1) \leq 1/2$ for every block size.*

Proof. For an active coordinate, write $\mu_j = \mathbb{E} Z_j \geq \Delta_u$. Independence of the rows gives $\mathbb{E} S_j = m\mu_j$ and $\mathrm{Var}(S_j) \leq mV_u$. The one-sided variance inequality in Lemma L.1 gives

$$\mathbb{P}(S_j \leq 0) \leq \frac{mV_u}{mV_u + m^2\mu_j^2} \leq \frac{1}{1 + m\Delta_u^2/V_u} \leq \frac{1}{4}.$$

This includes ties, consistent with defining a positive vote by strict inequality.

If j is inactive, the entire vector $(R_{1,j}, \dots, R_{m,j})$ consists of independent fair signs independent of $(Y_1, \dots, Y_m)$. Conditional on the latter vector, replacing every sign by its negative preserves the distribution and sends S_j to $-S_j$. Therefore

$$\mathbb{P}(S_j > 0 \mid Y_{1:m}) = \mathbb{P}(S_j < 0 \mid Y_{1:m}) \le \tfrac{1}{2}.$$

Averaging proves the inactive bound. The strict-vote definition includes ties in the nonpositive event, so the argument also covers atomic laws. □

Theorem G.4 (Critical recovery, restatement)**.** *With*

$$m \ge \frac{3D^2}{u}\left((1+u)^{bD} + 6\frac{\sigma^2}{a^2} + 3\frac{\sigma^4}{a^4}\right), \qquad K \ge 32 \log \frac{2n}{\delta},$$

Algorithm 4.1 recovers the active union with probability at least $1-\delta$ *using* mK *rows.*

Proof. For any fixed coordinate, its K votes come from disjoint blocks of independent rows and are independent Bernoulli variables. For an active coordinate, the mean vote fraction is at least $3/4$; for an inactive coordinate, it is at most $1/2$. The threshold $5/8$ is separated from both by $1/8$. Hoeffding's inequality (Hoeffding, 1963) bounds either error probability by

$$\exp\big(-2K(1/8)^2\big) = e^{-K/32}.$$

Union bounding over the n coordinates gives error at most $ne^{-K/32} \le \delta/2 \le \delta$. The stated m is sufficient for Lemma G.3 by the fourth-moment bound. Rounding both allocations up to positive integers preserves the guarantee. □

Critical scaling. For $u = c/(bD)$ with $0 < c < bD$,

$$(1+u)^{bD} \le e^c, \qquad \frac{D^2}{u} = \frac{bD^3}{c}.$$

Substitution gives the sample bound in Theorem 3.3. In the phrase $O_D(b \log n)$, depth, c, the SNR, and the desired failure probability convention are held fixed. The displayed bound retains D^3 for comparisons across depths. Appendix I optimizes the criterion with its noise contribution and open-interval endpoint conditions.

Formalized identities. The amplitude decoding, unary expectation, response fourth moment, variance bound, and reference-norm identity have Lean proofs. Appendix B gives the corresponding declarations.

G.5 Streaming Estimation and Median of Means

A streaming implementation keeps a vector of n block sums and a vector of n vote counts. For each row it computes Y_i^2 once, decodes every $R_{i,j}$, and adds $Y_i^2 R_{i,j}$. After m rows it updates the vote counts and clears the sums. After K blocks it thresholds the counts. This requires $O(nN)$ arithmetic and $O(n)$ working memory in addition to the incoming row. Retaining the whole dataset is unnecessary for this nonadaptive algorithm.

The block vote is chosen because inactive coordinates have an exactly symmetric statistic and because $R_j^2 = 1$ makes its variance easy to control. A median-of-means estimator gives another implementation based on the same finite second moment (Lugosi & Mendelson, 2019). It estimates the actual levels $\Delta_u h_j$, rather than only their signs; Appendix K uses this observation for a more informative output. Both guarantees apply under the finite second moment established above.

H Passive Information Lower Bounds

We quantify passive recovery from IID observations under unrestricted computation. KL packing establishes the baseline dependence on support size; Hellinger affinity reveals the exponential obstruction already present in a binary test. Both reductions retain explicit constants (Yu, 1997; Tsybakov, 2009).

H.1 Monomial Orthogonality under Product Laws

Lemma H.1 (Monomial second moments)**.** *Suppose the coordinates are independent, centered, and have variance one. For distinct finite sets $C, C' \subseteq [n]$,*

$$\mathbb{E}\chi_C(X)^2 = 1, \qquad \mathbb{E}[\chi_C(X)\chi_{C'}(X)] = 0, \qquad \mathbb{E}(\chi_C(X) - \chi_{C'}(X))^2 = 2.$$

Proof. The square is a product of second powers and its expectation factors into unit second moments. In the cross product, coordinates in $C \cap C'$ occur twice and coordinates in $C \triangle C'$ occur once. The symmetric difference is nonempty, so at least one factor is centered and the product expectation is zero. All products are integrable by independence and the finite second moments. Expanding the squared difference proves the final identity. □

The calculation remains true for nonconstant amplitudes and for laws with atoms. The identity fixes the L_2 separation across degrees. The affinity calculation below quantifies how rare covariate realizations can carry that separation.

H.2 Mixture KL Bounds and Support Packing

Fix $B_2, \dots, B_D$ and put $m = n - b(D-1)$. Index hypotheses by $S \in \binom{[m]}{b}$, placing $B_1 = S$ on the unused coordinates. Let $P_{S,t}$ be the joint law of (X, Y) conditional on $T = t$, and let $P_S = D^{-1}\sum_t P_{S,t}$ be the observed one-row law.

Lemma H.2 (Pairwise mixture KL)**.** *For distinct S, S',*

$$\mathrm{KL}(P_{S,t} \| P_{S',t}) = \frac{a^2}{\sigma^2}, \qquad \mathrm{KL}(P_S \| P_{S'}) \le \frac{a^2}{\sigma^2}.$$

Proof. The covariate marginal is common to both hypotheses. Conditional on $X = x$, the labels are Gaussian with equal variance and means $a\chi_{A_t}(x)$ and $a\chi_{A'_t}(x)$. The KL divergence between $N(m, \sigma^2)$ and $N(m', \sigma^2)$ is $(m - m')^2/(2\sigma^2)$. Integrating and using Lemma H.1 gives the first identity.

Consider also the augmented observation (T, X, Y). The distribution of T is common, so its KL divergence is the average of the D component divergences. Forgetting T is a measurable map and cannot increase KL; equivalently, apply the log-sum inequality pointwise to the common-weight mixtures. This proves the second assertion. We use an inequality here, because observing T would generally supply additional information. □

Proposition H.3 (Critical information lower bound, restatement)**.** *Even with $B_2, \dots, B_D$ supplied, recovering B_1 with worst-case success at least $3/4$ requires*

$$N \ge \frac{\sigma^2}{4a^2} \log \binom{m}{b} \ge \frac{\sigma^2}{4a^2} b \log(m/b).$$

In particular this proves Proposition 3.5.

Proof. Let $M = \binom{m}{b}$ and choose a uniform hypothesis index V. For N IID rows, pairwise KL is at most Na^2/σ^2. Let $\overline{P}_N$ be the mixture of the M sample laws. Convexity in the second KL argument gives

$$I(V; \mathcal{D}_N) = \frac{1}{M}\sum_S \mathrm{KL}(P_S^{\otimes N} \| \overline{P}_N) \le \frac{1}{M^2}\sum_{S,S'} \mathrm{KL}(P_S^{\otimes N} \| P_{S'}^{\otimes N}) \le \frac{Na^2}{\sigma^2}.$$

Fano's inequality implies an average error at least $1 - (I + \log 2)/\log M$. If $M \ge 4$, error at most $1/4$ therefore requires

$$\frac{Na^2}{\sigma^2} \ge \frac{3}{4}\log M - \log 2 \ge \frac{1}{4}\log M.$$

For $M = 2$ or 3, restrict to two hypotheses. Worst-case success $3/4$ requires total variation at least $1/2$. Pinsker's inequality and Lemma H.2 imply $Na^2/\sigma^2 \ge 1/2 \ge (\log M)/4$. This handles the small packing sizes for which the basic Fano numerator is insufficient.

Finally $m \geq 2b$ under the assumed density condition, and

$$\binom{m}{b} = \prod_{i=0}^{b-1} \frac{m-i}{b-i} \geq (m/b)^b.$$

Taking logarithms gives the second inequality. Because the lower bound applies after granting all later blocks, it also applies to the original recovery problem. □

Proposition H.4 (A packing for the entire active union)**.** *Without the grant of later blocks, exact active-union recovery with worst-case success at least* $3/4$ *requires*

$$N \geq \frac{\sigma^2}{4a^2} \log \binom{n}{bD} \geq \frac{\sigma^2}{4a^2} bD \log \frac{n}{bD}.$$

Proof. For each union S of size bD, partition its sorted coordinates into D successive blocks of size b; this selects one valid target per distinct union. For any two targets and any common selector value t, their prefixes either coincide, giving component KL zero, or are distinct standardized monomials, giving component KL a^2/σ^2. Thus pairwise observed KL is still at most a^2/σ^2. Apply the preceding Fano and binary argument to the $\binom{n}{bD}$ distinct outputs. The binomial estimate is the same. This packing bound captures the full union size bD. Proposition 3.5 separately gives a bound that remains valid when all later blocks are revealed. □

H.3 Gaussian Affinity and Fractional-Moment Bounds

Now fix two targets differing only by exchanging an active coordinate j with an inactive coordinate k. For the moment, take a single monomial of degree q_0. The shared factor is $A = \prod_{v \in C} X_v$, with $|C| = q_0 - 1$ and C disjoint from $\{j, k\}$. The two regression means are aAX_j and aAX_k.

Lemma H.5 (Conditional affinity identity)**.** *For the two one-component joint laws,*

$$H^2(P_{q_0}, P'_{q_0}) = \mathbb{E}\left[1 - \exp\left\{-\frac{a^2 A^2 (X_j - X_k)^2}{8\sigma^2}\right\}\right]. \tag{41}$$

Proof. With respect to the common covariate law times Lebesgue measure on labels, the conditional square-root density product of $N(m, \sigma^2)$ and $N(m', \sigma^2)$ is

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-m)^2 + (y-m')^2}{4\sigma^2}\right\}.$$

Complete the square:

$$(y-m)^2 + (y-m')^2 = 2\left(y - \frac{m+m'}{2}\right)^2 + \frac{(m-m')^2}{2}.$$

Integration over y therefore gives $\exp(-(m-m')^2/(8\sigma^2))$. Substitute $m - m' = aA(X_j - X_k)$, then integrate over X and use $H^2 = 1 -$ affinity. □

Lemma H.6 (Adjacent component clipping)**.** *For* $0 < \alpha \leq 2$*,*

$$H^2(P_{q_0}, P'_{q_0}) \leq \left(\frac{|a|}{\sqrt{8}\sigma}\right)^\alpha D_\alpha(\mu) M_\alpha(\mu)^{q_0 - 1}.$$

Proof. For every real z, $1 - e^{-z^2} \leq \min\{1, z^2\} \leq |z|^\alpha$. Apply this inequality inside (41). The factor A is independent of $X_j - X_k$, and its absolute αth moment is the product of $q_0 - 1$ coordinate moments. This gives the stated bound. All moments exist: finite coordinate second moment implies finite αth moments for $\alpha \leq 2$, and $D_\alpha(\mu) \leq 2^{\alpha/2}$ for centered standardized inputs by the monotonicity of L_p norms. □

The clipping step preserves a feature that a purely quadratic bound would lose. A quadratic bound sees $\mathbb{E}A^2 = 1$ at every degree. A fractional power sees $M_\alpha^{q_0 - 1}$, which decays when amplitudes are not constant.

H.4 Mixture Contraction and IID Testing

Lemma H.7 (Common-weight mixture contraction)**.** *For probability weights π_t and paired probability laws P_t, Q_t,*

$$H^2\left(\sum_t \pi_t P_t, \sum_t \pi_t Q_t\right) \le \sum_t \pi_t H^2(P_t, Q_t).$$

Proof. Choose a common dominating measure for the finite collection of laws. Pointwise Cauchy–Schwarz gives

$$\sum_t \pi_t \sqrt{p_t q_t} \le \sqrt{\left(\sum_t \pi_t p_t\right)\left(\sum_t \pi_t q_t\right)}.$$

Integrate and subtract from one. The resulting inequality has the required direction: hiding a common selector can make the two models harder to distinguish. □

Theorem H.8 (Passive collapse, restatement)**.** *For a swap in B_1, the adjacent hidden-rung laws obey (14). For each fixed symmetric standardized law with nonconstant modulus, fixed depth, and fixed nonzero SNR, testing the pair with worst-case success at least $3/4$ requires exponentially many rows in b.*

Proof. The swap affects every prefix. Conditional on $T = t$, its monomial degree is bt. Apply Lemma H.6 to each component and then Lemma H.7 with weights $1/D$. This gives exactly (14).

For $0 < \alpha < 2$, strict concavity of $v \mapsto v^{\alpha/2}$ on $[0, \infty)$ and nonconstant X^2 give

$$0 < M_\alpha(\mu) = \mathbb{E}[(X^2)^{\alpha/2}] < (\mathbb{E}X^2)^{\alpha/2} = 1.$$

Fix such an α and write $r = M_\alpha(\mu)$. Bounding the mixture average by its first term yields $H^2(P, P') \le Cr^{b-1}$ with $C = (|a|/(\sqrt{8}\sigma))^\alpha D_\alpha(\mu)$.

Let $h = H^2(P, P')$. Independence of rows multiplies affinities, so

$$H^2(P^{\otimes N}, P'^{\otimes N}) = 1 - (1-h)^N \le Nh.$$

A test with success at least $3/4$ under both hypotheses has average error at most $1/4$ and hence needs total variation at least $1/2$. Since $\mathrm{TV} \le \sqrt{2H^2}$, it follows that $Nh \ge 1/8$. Therefore

$$N \ge \frac{1}{8Cr^{b-1}} = \frac{r}{8C}(r^{-1})^b.$$

The base $r^{-1} > 1$ and all prefactors are independent of b for a fixed law and fixed SNR. This is the claimed exponential lower bound. □

Formalized information tools. The Lean artifact verifies the Gaussian affinity identity, common-weight mixture contraction, IID Hellinger tensorization, and binary-testing reduction. These components correspond to the information tools listed in Appendix B.

H.5 Later-Rung Perturbations and Rare-Event Information

If the exchanged coordinate belongs to B_s instead of B_1, components $t < s$ are identical and contribute zero Hellinger distance. The same proof gives the refinement

$$H^2(P, P') \le \left(\frac{|a|}{\sqrt{8}\sigma}\right)^\alpha D_\alpha(\mu) \frac{1}{D} \sum_{t=s}^{D} M_\alpha(\mu)^{bt-1}. \tag{42}$$

For a last-block swap, only the degree-bD component changes. At a fixed nonconstant-modulus law, the displayed upper bound is Cr^{bD-1}/D. This yields an even harder pair when depth increases. The main theorem uses the first-block swap because it already establishes the fixed-depth exponential obstruction and affects all rungs uniformly.

KL and Hellinger measure different aspects of the same rare-event mechanism. The squared mean difference has constant expectation, which can be carried by rare, large covariate realizations. Equal-variance Gaussian KL weights this difference quadratically; affinity applies the bounded transformation $1-\exp(-z^2)$. The KL upper bound supports the baseline packing argument, while the affinity bound yields the exponential testing obstruction. Appendix I quantifies the preserved energy on rare events.

I Dispersion Tuning and Learning Regimes

Proposition 3.7 links an explicit unary sample guarantee to the dispersion scale. We determine the exact minimizer of its sufficient-bound objective, including noise and open-interval boundary cases, and quantify the rare-event mechanism behind passive information loss.

I.1 Uniform Fractional-Moment Decay

For the two-radius law, direct averaging gives

$$M_\alpha(\mu_u) = \frac{(1+\sqrt{u})^{\alpha/2} + (1-\sqrt{u})^{\alpha/2}}{2}. \tag{43}$$

In particular,

$$\begin{aligned} M_1(\mu_u)^2 &= \frac{1+\sqrt{1-u}}{2} \le 1 - \frac{u}{4}, \\ M_1(\mu_u) &\le \sqrt{1-u/4} \le e^{-u/8}. \end{aligned} \tag{44}$$

The first inequality follows by concavity of the square root, or by squaring $\sqrt{1-u} \le 1-u/2$. The second uses $1-z \le e^{-z}$. These estimates hold uniformly throughout $0<u<1$.

For independent centered standardized coordinates, $\mathbb{E}(X-X')^2 = 2$, so $D_1(\mu_u) \le \sqrt{2}$. Theorem 3.6 with $\alpha = 1$ therefore yields

$$H^2(P_u, P_u') \le \frac{|a|}{2\sigma D} \sum_{t=1}^{D} e^{-u(bt-1)/8} \le \frac{|a|}{2\sigma} e^{-u(b-1)/8}.$$

This proves (16) with an explicit constant. The expression may exceed one for some parameters, in which case it can be intersected with the trivial upper bound one.

For completeness, the difference moment itself has the exact formula

$$D_1(\mu_u) = \frac{3\sqrt{1+\sqrt{u}} + \sqrt{1-\sqrt{u}}}{4}.$$

Conditional on two nonnegative radii r, r', averaging independent fair signs gives $\mathbb{E}|\xi r - \xi' r'| = \max\{r, r'\}$. The larger of two independent two-radius draws is the large radius with probability $3/4$, proving the formula. The simpler $\sqrt{2}$ upper bound is sufficient for the theorem.

I.2 Signal-Only Optimization and Boundary Cases

Write $q_0 = bD$ for total degree. The signal contribution to the unary sample bound is proportional to

$$f_{q_0}(u) = \frac{(1+u)^{q_0}}{u}, \qquad 0<u<1.$$

Differentiation gives

$$f_{q_0}'(u) = \frac{(1+u)^{q_0-1}}{u^2}\big((q_0-1)u - 1\big). \tag{45}$$

If $q_0 > 2$, the derivative changes sign once inside the domain, at $u_* = 1/(q_0-1)$, so this is the unique minimizer. If $q_0 = 2$, the derivative is negative throughout the open interval and tends to zero at $u=1$; the infimum is 4 and is not attained. If $q_0 = 1$, the function is $1+u^{-1}$, with infimum 2 as $u \uparrow 1$. This establishes all boundary cases in Proposition 3.7.

For large total degree,

$$f_{q_0}(u_*) = (q_0 - 1)\left(1 + \frac{1}{q_0 - 1}\right)^{q_0} = e\, q_0(1 + o(1)).$$

The explicit choice $u = 1/q_0$ achieves the same leading order, giving linear dependence on block width at fixed depth.

I.3 Noise-Dependent Optimization and Critical Scaling

Put $C_{\text{noise}} = 6\sigma^2/a^2 + 3\sigma^4/a^4 \geq 0$. The sufficient bound actually optimizes

$$g_{q_0,C}(u) = \frac{(1+u)^{q_0} + C}{u}, \qquad C = C_{\text{noise}}.$$

Its derivative has the sign of

$$G_{q_0,C}(u) = (1+u)^{q_0-1}\big((q_0 - 1)u - 1\big) - C. \tag{46}$$

For $q_0 > 1$,

$$G'_{q_0,C}(u) = q_0(q_0 - 1)u(1+u)^{q_0-2} > 0.$$

Since $G(0) = -1 - C$, there is at most one interior optimizer. For $q_0 > 2$, it exists exactly when $C < 2^{q_0-1}(q_0 - 2)$, and it is the unique solution of $G(u) = 0$. If this condition fails, the objective decreases to its infimum at the boundary $u \uparrow 1$. For $q_0 \leq 2$, it is likewise decreasing throughout the allowed interval. These conditions completely characterize the optimizer of the fourth-moment sample guarantee.

At fixed C and $q_0 \to \infty$, the interior solution has

$$u_* = \frac{z_C}{q_0} + O(q_0^{-2}), \qquad e^{z_C}(z_C - 1) = C, \tag{47}$$

where $z_C \geq 1$ is unique. To see this, substitute $u = z/q_0$ into (46); on compact positive z intervals the expression converges uniformly to $e^z(z-1) - C$. That limiting function is strictly increasing for $z > 0$ and has a simple zero at z_C. A Taylor expansion of $(1 + z/q_0)^{q_0-1}$ has uniform error $O(q_0^{-1})$ near the root. The nonzero limiting derivative $z_C e^{z_C}$ then yields $q_0 u_* = z_C + O(q_0^{-1})$. Thus fixed SNR changes the tuning constant, while preserving inverse-total-degree scaling.

There is also a simple rate statement without solving an equation. For $q_0 \geq 2$, Bernoulli's inequality gives $g_{q_0,C}(u) \geq q_0$ for all admissible u, while $u = 1/q_0$ gives $g_{q_0,C}(u) \leq q_0(e + C)$. Therefore the best value of this sufficient-bound objective is $\Theta_C(q_0)$. This is an optimization fact about the certificate; the independent information lower bound is supplied by Appendix H.

Formalized optimization. The Lean development proves the exact optimizer, its uniqueness and existence threshold, and the open-interval boundary infimum. The declarations in Appendix B cover the objective $g_{q_0,C}$ and its zero-noise specialization.

I.4 Sample Bounds with the Exact Rung Average

The largest-rung bound in the main theorem can be sharpened by retaining (39). Define

$$F_{b,D}(u) = \frac{1}{D}\sum_{t=1}^{D}(1+u)^{bt}.$$

Then $m \geq 3D^2(F_{b,D}(u) + C_{\text{noise}})/u$ is sufficient. For $u > 0$, a geometric sum gives

$$F_{b,D}(u) = \frac{(1+u)^b\big((1+u)^{bD} - 1\big)}{D\big((1+u)^b - 1\big)}.$$

At fixed D and $u = c/(bD)$, this average converges as $b \to \infty$ to $D^{-1}\sum_{t=1}^{D} e^{ct/D}$. It is bounded above by e^c uniformly in b. The main theorem uses that simpler bound to display the resource scaling; an implementation with known parameters may use the exact average for a smaller certified block size.

I.5 Rare-Event Concentration of Signal Energy

Let $A_{q_0} = \prod_{i=1}^{q_0} X_i$ under a fixed symmetric standardized law with nonconstant modulus. Standardization always gives $\mathbb{E}A_{q_0}^2 = 1$. Yet for any $0 < \alpha < 2$ and any fixed $\varepsilon > 0$,

$$\mathbb{P}(|A_{q_0}| > \varepsilon) \le \varepsilon^{-\alpha} M_\alpha(\mu)^{q_0}.$$

The event is exponentially rare. On its complement $A_{q_0}^2 \le \varepsilon^2$, so

$$\mathbb{E}\big[A_{q_0}^2 \mathbf{1}\{|A_{q_0}| > \varepsilon\}\big] \ge 1 - \varepsilon^2.$$

For $\varepsilon < 1$, an exponentially small event therefore carries a nonvanishing fraction of the second moment. The fractional-moment argument applies to atomic and heavy-tailed standardized laws, including atoms at zero.

For the two-radius model there is an additional exact logarithmic description:

$$\begin{aligned} \log A_{q_0}^2 &= \frac{q_0}{2}\log(1-u) + \frac{1}{2}\log\frac{1+\sqrt{u}}{1-\sqrt{u}}\sum_{i=1}^{q_0} R_i, \\ \mathbb{E}\log A_{q_0}^2 &= \frac{q_0}{2}\log(1-u), \\ \mathrm{Var}(\log A_{q_0}^2) &= \frac{q_0}{4}\left(\log\frac{1+\sqrt{u}}{1-\sqrt{u}}\right)^2. \end{aligned} \tag{48}$$

As $u \downarrow 0$, the mean is $-q_0 u/2 + O(q_0 u^2)$ and the variance is $q_0 u(1 + O(u))$. Thus accumulated dispersion is governed by degree times u. At inverse-degree dispersion these quantities remain of constant order. If $q_0 u \to \infty$ while u stays bounded away from one, the negative mean dominates its standard deviation. This typical-energy calculation is explanatory; the testing lower bound itself uses the uniform fractional-moment estimate.

I.6 Coordinate Laws and the Three Learning Regimes

The following exact analytic moments locate standard coordinate laws within the three regimes.

Coordinate design	**Exact analytic moments**	
Standardized coordinate law	$\mathbb{E}\boldsymbol{X^4}$	$\boldsymbol{M_1 = \mathbb{E}\lvert X\rvert}$
Fixed-magnitude endpoint		
Rademacher	1	1
Variable-magnitude coordinate laws		
$N(0,1)$	3	$\sqrt{2/\pi}$
Uniform on $[-\sqrt{3}, \sqrt{3}]$	$9/5$	$\sqrt{3}/2$
0 with probability $1-p$; $\pm p^{-1/2}$ with probability $p/2$	$1/p$	$\sqrt{p}$

For the last three laws with $p < 1$, $M_1 < 1$ immediately gives an exponential adjacent-target obstruction. The Gaussian and uniform values follow from direct one-dimensional integration; the three-point values follow by summing its atoms.

At $u = 0$, the unary margin vanishes and the round hierarchy governs SQ search. At $u = \Theta(1/(bD))$, the explicit unary estimator certifies polynomial sample cost. At fixed $u > 0$, the Hellinger upper bound forces exponential passive sample complexity as b grows. The estimator and the testing reduction establish the positive and negative regimes, respectively.

J Sample Reuse and Computational Cost

A single dataset supports the entire adaptive carrier path through simultaneous concentration over a predetermined parity family. This uniform event controls every candidate before selection (Dwork et al., 2015; Russo & Zou, 2016), giving an explicit IID implementation of the rigid upper bound.

J.1 A Fixed Family of Candidate Correlations

Let $P = bD$ and define

$$\mathcal{F}_P = \{(x, y) \mapsto \operatorname{sign}(y)\chi_C(x) : C \subseteq [n],\ |C| \le P\}.$$

Every candidate query used by the rigid path algorithm lies in this family, regardless of which earlier prefixes were selected. A sharper family could retain only the rung-compatible sizes, but the larger family gives a simple uniform bound. Its size is

$$M_P = \sum_{s=0}^{P} \binom{n}{s} \le \left(\frac{en}{P}\right)^P, \qquad 1 \le P \le n/2.$$

A proof of this standard counting estimate is given in Lemma L.3.

For each C define the population and empirical correlations

$$c_C = \mathbb{E}[\operatorname{sign}(Y)\chi_C(X)], \qquad \widehat{c}_C = \frac{1}{N}\sum_{i=1}^{N} \operatorname{sign}(Y_i)\chi_C(X_i).$$

For nonempty C, the exact rigid likelihood calculation gives

$$c_C = \frac{\kappa}{D}\sum_{t=1}^{D} \mathbf{1}\{C = A_t\}, \qquad \kappa = 2\Phi(a/\sigma) - 1. \tag{49}$$

Hence every true prefix has absolute correlation γ_{sgn}/D, and every other nonempty candidate has zero correlation.

Lemma J.1 (Uniform accuracy on one dataset)**.** *If*

$$N \ge \frac{32D^2}{\gamma_{\mathrm{sgn}}^2}\left(P\log\frac{en}{P} + \log\frac{2}{\delta}\right), \tag{50}$$

then, with probability at least $1 - \delta$*,*

$$\sup_{|C|\le P} |\widehat{c}_C - c_C| \le \frac{\gamma_{\mathrm{sgn}}}{4D}.$$

Proof. For each fixed C, its summands are independent across rows and lie in $[-1, 1]$. Hoeffding's inequality gives

$$\mathbb{P}(|\widehat{c}_C - c_C| > \varepsilon) \le 2e^{-N\varepsilon^2/2}.$$

The functions in $\mathcal{F}_P$ are fixed independently of the data. Union bounding over all M_P functions yields $2M_P e^{-N\varepsilon^2/2}$. Substitute $\varepsilon = \gamma_{\mathrm{sgn}}/(4D)$ and the bound on $\log M_P$ to obtain failure probability at most δ. □

J.2 Empirical Frozen-Batch Recovery

Take $L = \lceil D/R \rceil$ and the schedule (p_r, s_r) from Theorem 3.1. The following procedure uses the same N rows throughout.

Algorithm J.2 (ReusedSamplePath)**.** Initialize $\widehat{A}_0 = \varnothing$. At stage $r = 0, \dots, J-1$, freeze every set $\widehat{A}_{p_r} \cup U$ with $U \subseteq [n] \setminus \widehat{A}_{p_r}$ and $|U| \in \{b, 2b, \dots, bs_r\}$. Evaluate its empirical correlation and declare a hit when its absolute value exceeds $\gamma_{\mathrm{sgn}}/(2D)$. At each requested size, retain the unique hit; if uniqueness fails, report failure. Use the hit of largest size as $\widehat{A}_{p_r+s_r}$ for the next stage. Return $\widehat{A}_D$.

Theorem J.3 (One-reused-sample path, restatement)**.** *Under* (50)*, Algorithm J.2 succeeds with probability at least* $1 - \delta$*, uses* $K_{n,b,D,R}$ *scalar correlation evaluations on its successful path, and satisfies the arithmetic and storage bounds in Theorem 3.8.*

Proof. Work on the single event in Lemma J.1. Initially the carrier is correct. Suppose $\widehat{A}_{p_r} = A_{p_r}$. For each requested size bs, the true extension $A_{p_r+s} \setminus A_{p_r}$ is among the candidates. By (49), it has empirical magnitude at least $3\gamma_{\mathrm{sgn}}/(4D)$. Every other candidate of that size has magnitude at most $\gamma_{\mathrm{sgn}}/(4D)$. Thus there is exactly one hit, it is correct, and the largest hit supplies the correct next carrier. Induction proves all stages succeed and the final union is exact.

The number of evaluated candidates in stage r is exactly $\sum_{s=1}^{s_r} \binom{n-bp_r}{bs}$. Summing yields $K_{n,b,D,R}$. Each candidate has at most bL fresh coordinates, so the cost calculation below gives $O(NbLK_{n,b,D,R})$ arithmetic. The uniformly accurate event has probability at least $1-\delta$, proving the theorem. An unsuccessful path can terminate early; the same count is a worst-case budget if the remaining scheduled evaluations are included before reporting failure. □

J.3 Uniform Accuracy under Adaptive Selection

The event controls every member of $\mathcal{F}_P$ simultaneously, including candidates outside the realized path. Every adaptively selected query is accurate on this event because it belongs to the fixed family. This argument permits selection to depend on all earlier empirical answers.

The sample bound in (50) uses the entropy of the complete candidate pool, with total degree $P = bD$. The evaluation count $K_{n,b,D,R}$ records the realized path, whose fresh extensions have size at most bL. These two quantities account for uniform statistical accuracy and computational search, respectively.

The empirical answers implement the bounded queries used by the algorithm on one high-probability event. Appendix H analyzes the information available from whole IID rows, yielding the separate passive lower bounds.

J.4 Sample, Arithmetic, and Storage Costs

Cache a length-N carrier vector $v_i = \chi_{A_{p_r}}(X_i)$ and the label signs. For a candidate U of size bs, form

$$\widehat{c}_{A_{p_r} \cup U} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{sign}(Y_i) v_i \prod_{j \in U} X_{i,j}.$$

Direct evaluation takes $O(N(bs+1)) \le O(NbL)$ operations. Updating the cached carrier at the end of a batch takes at most $O(NbL)$ more operations. Over the path, the update cost is dominated by the candidate evaluations. Reading the dataset also fits within the bound: the first batch already has at least $\binom{n}{b} \ge n$ candidates under $1 \le b \le n/2$.

If all covariates are retained explicitly, storage is $O(Nn)$, with an additional $O(N+n)$ for cached products, signs, and coordinate sets. The storage bound uses explicit covariate storage and cached carrier products. At $b = 1$ and $R = D$, the query count is $Dn - D(D-1)/2$ and direct arithmetic is $O(NnD)$.

The critical unary algorithm streams each row through n fixed accumulators, using $O(nN)$ arithmetic and $O(n)$ working memory. The rigid algorithm evaluates a combinatorial candidate family on a reused dataset. Reporting observations and evaluations separately exposes the computational benefit of the unary path.

K Extensions of the Hidden-Rung Model

The same identities yield ordered-block recovery, nonuniform-selector guarantees, binary-channel extensions, and critical recovery for general bounded amplitude perturbations. Each result below states the assumptions that determine its signal and resource costs.

K.1 Ordered-Block Recovery from Unary Levels

Under the uniform hidden selector, the exact response-energy mean encodes more than whether a coordinate is active. With $\Delta_u = a^2\sqrt{u}/D$,

$$\mathbb{E}[Y^2 R_j] = \begin{cases} \Delta_u(D-s+1), & j \in B_s, \\ 0, & j \notin A_D. \end{cases}$$

These $D+1$ levels are separated by the same gap Δ_u.

Proposition K.1 (Ordered-block recovery under the two-radius design)**.** *Suppose (a, b, D, u) are known. The same n one-batch relative-L_2 queries as Corollary 3.4, at tolerance $\tau < \Delta_u/(4\sqrt{V_u})$, recover every ordered block B_s by rounding their answers to the nearest multiple of Δ_u. An IID implementation using coordinatewise median-of-means estimates succeeds with probability at least $1-\delta$ when*

$$m \geq \frac{36V_u}{\Delta_u^2}, \qquad K \text{ is odd}, \qquad K \geq 8\log\frac{n}{\delta}.$$

Proof. Each SQ error is less than $\Delta_u/4$, whereas two possible means are separated by Δ_u. Nearest-level rounding therefore returns the exact integer $h_j \in \{0, \dots, D\}$. Set $\widehat{B}_s = \{j : h_j = D-s+1\}$.

For the IID statement, $\mathrm{Var}(Y^2 R_j) \leq V_u$ by Lemma G.1. Lemma L.2 with accuracy $\eta = \Delta_u/3$ gives a simultaneously accurate median-of-means estimate for all coordinates at the displayed allocations. Rounding again recovers every h_j. The required sample count has the same order in (b, D, n) as the unary support guarantee, with different constants. □

Algorithm 4.1 uses sign votes for union recovery. Retaining the mean levels recovers the ordered blocks with the same n coordinate queries. The known common coefficient and uniform selector weights determine the level spacing used by the decoder.

K.2 Known Nonuniform Selector Weights

Suppose $\mathbb{P}(T=t) = \pi_t > 0$ with $\sum_t \pi_t = 1$, and keep the common coefficient a. At the rigid endpoint,

$$L_{\mathbf{B}}(x,y) = 1 + w(y)\sum_{t=1}^{D} \pi_t \chi_{A_t}(x).$$

Every prefix is still detectable and no new nonempty set becomes detectable. Hence the fixed-link batch-leap profile is unchanged. A sign-query upper bound uses the smallest margin $\gamma_{\mathrm{sgn}} \min_t \pi_t$ rather than γ_{sgn}/D.

For the two-radius law, the exact identities become

$$\mathbb{E}[Y^2 R_j] = a^2\sqrt{u} \sum_{t:\, j \in A_t} \pi_t,$$
$$\mathbb{E}Y^4 = a^4 \sum_t \pi_t (1+u)^{bt} + 6a^2\sigma^2 + 3\sigma^4. \tag{51}$$

The proof is the same conditioning argument as Lemma G.1, with weighted rather than uniform averaging. For union recovery the weakest active mean is $a^2\sqrt{u}\,\pi_D$, attained in the last block. The block-vote proof works with this value as its minimum signal. For ordered recovery, adjacent block levels differ by $a^2\sqrt{u}\,\pi_s$, so the smallest level gap depends on $\min_s \pi_s$. These are different quantities if the selector is imbalanced.

The lower-bound energy calculation also retains an explicit weighted form. Conditional on the first k blocks, write $p_{>h} = \sum_{t>h} \pi_t$ and $M = \binom{n-bk}{b(h+1-k)}$. Weighted Cauchy–Schwarz gives

$$\left(\sum_{t>h} \pi_t \widehat{d}(A_t)\right)^2 \leq p_{>h} \sum_{t>h} \pi_t \widehat{d}(A_t)^2.$$

The same disjoint-level Parseval calculation therefore bounds its conditional expectation by

$$\frac{p_{>h}\max_{t>h}\pi_t}{M}\,\beta_{\mathrm{lr}}^2\,\|\phi\|_2^2\,.$$

Uniform weights recover the main bound because $p_{>h}\le 1$ and $\max_t \pi_t = 1/D$. Fixed positive weights change margins and prefactors rather than the ambient exponent. If weights themselves vanish with a growing parameter, those prefactors must be retained.

K.3 General Binary Label Channels

Gaussian noise gives convenient closed-form signal constants, but the rigid projection argument has a more general channel interpretation. Let Q_+ and Q_- be any two distinct known probability measures, put $\mu_y = (Q_+ + Q_-)/2$, and write $q_\pm = dQ_\pm/d\mu_y$. These derivatives always exist and satisfy $q_+ + q_- = 2$ almost everywhere. Set

$$w = \frac{q_+ - q_-}{2}, \qquad \beta^2 = \mathbb{E}_{\mu_y} w^2 \in (0,1].$$

Then $q_s = 1 + sw$, $|w| \le 1$, and $\mathbb{E}_{\mu_y} w = 0$. Draw Y from $Q_{\chi_{A_T}(X)}$ at the rigid endpoint. Its likelihood again has the form (31); the slice and conditional energy proofs apply with β in place of β_{lr}.

For the upper bound, use the bounded query $w(Y)\chi_C(X)$. Its reference norm is β and its target expectation is β^2/D exactly for true prefixes. Thus tolerance $\tau < \beta/(4D)$ gives the same thresholding guarantee and candidate counts. The SQ lower bound likewise retains the same form with β. The rigid round exponent therefore extends to every distinct known binary label channel through β. The Gaussian specialization also supplies the exact affinity formula used for the passive-information bound.

K.4 Unary Recovery beyond Gaussian Noise

Suppose instead that ϵ is independent and centered, with variance σ^2 and finite fourth moment m_4. The unary identity is unchanged. Under symmetric covariates, every nonempty monomial F_t is symmetric and hence $\mathbb{E}F_t = 0$. In the fourth-power expansion, the possibly nonzero term involving $\mathbb{E}\epsilon^3$ is multiplied by $\mathbb{E}F_t = 0$. The term involving $\mathbb{E}F_t^3$ is multiplied by $\mathbb{E}\epsilon = 0$. Consequently

$$\mathbb{E}Y^4 = \frac{a^4}{D}\sum_{t=1}^{D}(1+u)^{bt} + 6a^2\sigma^2 + m_4.$$

For an inactive coordinate, R_j remains a fair sign independent of Y, even if the noise itself is asymmetric. The entire block-vote proof therefore continues to hold with $3\sigma^4$ replaced by m_4. This extension concerns the unary upper bound; the passive Gaussian result uses the affinity identity in Appendix H.

K.5 Critical Recovery under Bounded Amplitude Perturbations

The inverse-degree recovery guarantee extends from two radii to an entire class of amplitude distributions. Let Z have any law satisfying

$$\mathbb{E}Z = 0, \qquad \mathbb{E}Z^2 = 1, \qquad |Z| \le M \quad \text{almost surely}, \qquad 0 < u < M^{-2}. \tag{52}$$

Necessarily $M \ge 1$. Independently across coordinates, draw copies Z_j and independent fair signs ξ_j, and set

$$X_j = \xi_j\sqrt{1+\sqrt{u}\,Z_j}, \qquad G_j = \frac{X_j^2 - 1}{\sqrt{u}} = Z_j. \tag{53}$$

The signs make X_j symmetric even when Z_j is asymmetric. The law of Z may be discrete or continuous. Its first two moments give $\mathbb{E}X_j^2 = 1$, $\mathbb{E}X_j^4 = 1+u$, and $\mathrm{Var}(X_j^2) = u$ exactly.

Theorem K.2 (Critical recovery for bounded amplitudes)**.** *Consider* (6) *under* (52)–(53)*, with known* (a, b, D, u, σ) *and* $0 < \delta < 1$*. Define*

$$\Delta_u = \frac{a^2\sqrt{u}}{D}, \quad h_j = |\{t : j \in A_t\}|, \quad r_u = 1 + M\sqrt{u} < 2,$$

$$V_u = \frac{a^4}{D}\sum_{t=1}^{D}(1+u)^{bt} + 6a^2\sigma^2 + 3\sigma^4.$$

Then $\mathbb{E}[Y^2 G_j] = \Delta_u h_j$, $\mathbb{E}Y^4 = V_u$, *and*

$$\mathrm{Var}(Y^2 G_j) \le r_u^2 V_u < 4V_u. \tag{54}$$

Coordinatewise median-of-means estimates from $N = mK$ *shared IID rows recover all ordered blocks, and hence their union, with probability at least* $1-\delta$ *if*

$$m \ge \frac{36 r_u^2 V_u}{\Delta_u^2}, \qquad K \text{ is odd}, \qquad K \ge 8\log\frac{n}{\delta}. \tag{55}$$

One may replace $36r_u^2$ *by* 144, *independently of the shape of the amplitude law. At* $u = c/(bD)$, *where* $c > 0$ *and* $bD > cM^2$, *suitable integer allocations give*

$$N = O\left(\frac{bD^3}{c}\left(e^c + 6\frac{\sigma^2}{a^2} + 3\frac{\sigma^4}{a^4}\right)\log\frac{2n}{\delta}\right), \qquad \text{work} = O(nN). \tag{56}$$

The constants are uniform over all amplitude laws satisfying the assumptions. In full SQ, the n *fixed queries* $\phi_j(x,y) = y^2 G_j(x)$ *recover the ordered blocks in one batch at relative product-reference tolerance* $\tau < \Delta_u/(4\sqrt{V_u})$.

Proof. Write $s = \sqrt{u}$ and $F_t = \prod_{i\in A_t} X_i$. Since $\mathbb{E}G_j = 0$ and

$$\mathbb{E}[X_j^2 G_j] = \mathbb{E}[(1+sZ_j)Z_j] = s,$$

independence gives $\mathbb{E}[F_t^2 G_j] = s$ for $j \in A_t$, and zero otherwise. Conditioning on the centered independent Gaussian noise and averaging the selector proves $\mathbb{E}[Y^2 G_j] = a^2 s h_j/D$. Likewise,

$$\mathbb{E}F_t^2 = 1, \qquad \mathbb{E}F_t^4 = (1+u)^{bt},$$

so the Gaussian fourth-power expansion yields $\mathbb{E}Y^4 = V_u$.

The IID variance requires a mixed moment under the actual joint law. For an active coordinate $j \in A_t$, the bounds $0 < 1 + sZ_j \le r_u$ and $\mathbb{E}Z_j^2 = 1$ imply

$$\mathbb{E}[F_t^4 G_j^2] = \mathbb{E}[(1+sZ_j)^2 Z_j^2]\,(1+u)^{bt-1} \le r_u^2(1+u)^{bt-1},$$
$$\mathbb{E}[F_t^2 G_j^2] = \mathbb{E}[(1+sZ_j)Z_j^2] \le r_u.$$

For $j \notin A_t$, independence gives the respective values $(1+u)^{bt}$ and 1. Since $r_u \ge 1$, averaging the Gaussian fourth-power expansion thus gives

$$\begin{aligned}\mathbb{E}[Y^4 G_j^2] &\le \frac{r_u^2 a^4}{D}\sum_{t=1}^{D}(1+u)^{bt} + 6r_u a^2\sigma^2 + 3\sigma^4 \\ &\le r_u^2 V_u.\end{aligned}$$

Variance is at most the second moment of $Y^2 G_j$, proving (54). This calculation uses the normalized second moment of Z, so the uniform bound $r_u < 2$ controls the constant throughout the admissible parameter range.

Apply Lemma L.2 with variance bound $r_u^2 V_u$ and accuracy $\Delta_u/3$. The allocations in (55) make every coordinate estimate accurate simultaneously with probability at least $1-\delta$. Rounding each estimate to the nearest level in $\{0, \Delta_u, \dots, D\Delta_u\}$ recovers h_j. Set $\widehat{B}_t = \{j : h_j = D - t + 1\}$. Independence across coordinates of the estimates is unnecessary: each uses independent row blocks, followed by a union bound over coordinates. Computing all block means, medians, and labels costs $O(nN)$ arithmetic using a linear-time median selection algorithm.

At $u = c/(bD)$, $(1+u)^{bt} \le e^c$ and $r_u^2 < 4$. Substitution into (55), with integer rounding, proves (56). For the SQ claim, the common reference $\nu_{0,u}$ is the product of the input law and the label marginal of this amplitude design. Symmetry and identical coordinate laws make the label marginal independent of the embedding. Therefore

$$\|\phi_j\|^2_{L_2(\nu_{0,u})} = \mathbb{E}Y^4\,\mathbb{E}G_j^2 = V_u.$$

The stated tolerance makes each answer's error less than $\Delta_u/4$, so the same level decoder succeeds. □

Discrete, continuous, and asymmetric designs. The theorem includes $Z \sim \mathrm{Rad}$ with $M = 1$, $Z \sim \mathrm{Unif}[-\sqrt{3}, \sqrt{3}]$ with $M = \sqrt{3}$, and the asymmetric law $\mathbb{P}(Z = -1/2) = 4/5$, $\mathbb{P}(Z = 2) = 1/5$ with $M = 2$. All three have mean zero and variance one. Thus inverse-total-degree dispersion produces a unary path for both continuously distributed magnitudes and asymmetric amplitude fluctuations. The IID decoder needs the displayed moments and signal scale, without estimating the shape of the amplitude distribution.

Moment interface for general coordinate laws. For any symmetric standardized law with finite $\kappa_4 := \mathbb{E}X^4 > 1$, set $v = \kappa_4 - 1$ and $G_j = (X_j^2 - 1)/\sqrt{v}$. The same calculation gives

$$\mathbb{E}[Y^2 G_j] = \frac{a^2\sqrt{v}}{D} h_j, \qquad \|y^2 G_j(x)\|^2_{L_2(\mu^n \otimes \mu_y)} = \mathbb{E}Y^4. \tag{57}$$

When $\mu_6 = \mathbb{E}X^6$ and $\mu_8 = \mathbb{E}X^8$ are finite, the active-coordinate factors in the IID variance are

$$\mathbb{E}[F_t^4 G_j^2] = \frac{\mu_8 - 2\mu_6 + \kappa_4}{v} \kappa_4^{bt-1}, \qquad \mathbb{E}[F_t^2 G_j^2] = \frac{\mu_6 - 2\kappa_4 + 1}{v}.$$

Theorem K.2 controls these factors uniformly along a degree-dependent design, turning the general magnitude witness into a critical recovery guarantee.

L Auxiliary Probability and Product-Space Results

We collect the elementary tools used above, with the normalizations needed for the constants in the paper. These are standard results; their inclusion makes the proof chain self-contained at the level of concentration and testing reductions. Background references are O'Donnell (2014); Hoeffding (1963); Lugosi & Mendelson (2019); Yu (1997); Tsybakov (2009).

L.1 Boolean Orthogonality and Parseval

For $X \sim \mathrm{Rad}^n$, define $\chi_C(X) = \prod_{j \in C} X_j$. Then

$$\chi_C \chi_{C'} = \chi_{C \triangle C'}, \qquad \mathbb{E}\chi_C \chi_{C'} = \mathbf{1}\{C = C'\}.$$

The second identity follows because a nonempty symmetric difference contains an independent centered sign. There are 2^n such orthonormal functions in the 2^n-dimensional function space, so they form a basis. For every real-valued function f on the cube,

$$f = \sum_{C \subseteq [n]} \widehat{f}(C)\chi_C, \qquad \widehat{f}(C) = \mathbb{E}[f(X)\chi_C(X)], \qquad \sum_C \widehat{f}(C)^2 = \mathbb{E}f(X)^2.$$

Multiplication by a carrier χ_K permutes this basis: the coefficient at U of $f\chi_K$ is the coefficient at $U \triangle K$ of f. This is the analytic version of carrier cancellation. The conditional counting argument in Lemma F.4 uses only this orthogonality and uniform placement of an unseen subset.

L.2 Cantelli's One-Sided Variance Bound

Lemma L.1 (Cantelli's inequality)**.** *If Z has mean zero and variance $v < \infty$, then for $t > 0$,*

$$\mathbb{P}(Z \le -t) \le \frac{v}{v + t^2}.$$

Proof. For any $s \ge 0$, the event $Z \le -t$ implies $(Z - s)^2 \ge (t + s)^2$. Markov's inequality gives

$$\mathbb{P}(Z \le -t) \le \frac{\mathbb{E}(Z - s)^2}{(t + s)^2} = \frac{v + s^2}{(t + s)^2}.$$

Choose $s = v/t$ and simplify. The case $v = 0$ is immediate. □

This bound applies to $Y^2 R_j$ through its finite variance, without requiring the statistic itself to be bounded. The bounded variables in the next step are the block votes.

L.3 Concentration and Median-of-Means Estimation

If independent variables U_i take values in intervals $[a_i, b_i]$, then

$$\mathbb{P}\left(\sum_i (U_i - \mathbb{E}U_i) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right). \tag{58}$$

Here is a short derivation. For a bounded variable U, the centered log moment-generating function $\psi(\lambda) = \log \mathbb{E}e^{\lambda(U-\mathbb{E}U)}$ has $\psi(0) = \psi'(0) = 0$. Its second derivative is the variance under an exponentially tilted distribution supported on the same interval. This variance is at most $(b-a)^2/4$, because squared distance from the interval midpoint is bounded by that quantity and variance minimizes squared distance over centers. Integrating twice gives $\psi(\lambda) \leq \lambda^2(b-a)^2/8$. Independence adds these bounds. Exponential Markov and optimization over $\lambda > 0$ give (58). Applying the same argument to $-U_i$ gives the lower tail; adding the two tails gives the factor two used for empirical parity means.

Lemma L.2 (Coordinatewise median of means)**.** *Suppose $Z^{(1)}, \ldots, Z^{(n)}$ are coordinates of a random vector, each with variance at most V, and rows are IID. Split mK rows into K blocks, with K odd, and estimate each coordinate mean by the median of its block means. If $m \geq 4V/\eta^2$, then*

$$\mathbb{P}\left(\max_{j \leq n} |\widehat{\mu}_j - \mu_j| > \eta\right) \leq ne^{-K/8}.$$

Proof. A block mean has variance at most V/m, so Chebyshev's inequality bounds the probability that it deviates by more than η by $1/4$. For a fixed coordinate the K bad-block indicators are independent. If its median deviates by more than η, at least half the blocks are bad. Their mean is at most $1/4$, and (58) bounds this probability by $e^{-2K(1/4)^2} = e^{-K/8}$. Union bound over coordinates. No independence between vector coordinates is needed. □

L.4 Binomial Counts for Sparse Candidate Families

Lemma L.3 (Binomial entropy bound)**.** *For integers $1 \leq P \leq n/2$,*

$$\sum_{s=0}^{P} \binom{n}{s} \leq e^{nh(P/n)} \leq (en/P)^P,$$

where $h(p) = -p \log p - (1-p)\log(1-p)$.

Proof. Put $p = P/n$ and $x = p/(1-p) \leq 1$. Since $x^s \geq x^P$ for $s \leq P$,

$$x^P \sum_{s=0}^{P} \binom{n}{s} \leq \sum_{s=0}^{P} \binom{n}{s} x^s \leq (1+x)^n.$$

The logarithm of the resulting upper bound $x^{-P}(1+x)^n$ is $nh(p)$. Finally $-(1-p)\log(1-p) \leq p$, obtained by integrating $1/(1-t)$ from 0 to p and multiplying by $1-p$. Hence $nh(p) \leq P\log(n/P) + P$. □

L.5 Hellinger Distance and Product Observations

For densities p, q relative to any common dominating measure, write $h = H^2(P,Q) = 1 - \int \sqrt{pq}$. Cauchy–Schwarz gives

$$\begin{aligned} \mathrm{TV}(P,Q) &= \frac{1}{2}\int |\sqrt{p} - \sqrt{q}|(\sqrt{p} + \sqrt{q}) \\ &\leq \frac{1}{2}\sqrt{2h(4-2h)} = \sqrt{h(2-h)} \leq \sqrt{2h}. \end{aligned}$$

Affinity multiplies under independent products by Fubini's theorem, proving $H^2(P^{\otimes N}, Q^{\otimes N}) = 1 - (1-h)^N \leq Nh$.

For an equal-prior binary test that chooses P on a measurable set A, its average error is

$$\frac{1}{2}\big(P(A^c) + Q(A)\big) = \frac{1}{2}\big(1 - (P(A) - Q(A))\big).$$

Maximizing $P(A) - Q(A)$ over A gives $\mathrm{TV}(P, Q)$; the maximizing set can be chosen as $\{p \geq q\}$. Thus the best average error is $(1 - \mathrm{TV}(P, Q))/2$. Randomized tests cannot improve it because they average deterministic decision probabilities pointwise. These facts justify the testing threshold used in Appendix H.

L.6 KL Contraction, Pinsker, and Fano

For joint laws of (T, Z), the chain rule for KL gives

$$\mathrm{KL}(P_{T,Z} \| Q_{T,Z}) = \mathrm{KL}(P_Z \| Q_Z) + \mathbb{E}_{P_Z} \mathrm{KL}(P_{T|Z} \| Q_{T|Z}).$$

The conditional term is nonnegative, so forgetting T contracts KL. With a common distribution of T, expanding the other order of the chain rule gives the average conditional KL. These two observations prove the mixture bound used earlier. The identities follow by writing a joint density as marginal density times conditional density and splitting the logarithm; infinite KL cases follow by the usual extended-value convention.

For completeness, Pinsker's inequality $\mathrm{TV}(P, Q) \leq \sqrt{\mathrm{KL}(P\|Q)/2}$ follows by projecting onto the event $A = \{p \geq q\}$. This gives Bernoulli probabilities $r = P(A)$ and $s = Q(A)$ with $r - s = \mathrm{TV}(P, Q)$. For fixed $s \in (0, 1)$, the second derivative in r of Bernoulli KL is $1/(r(1-r)) \geq 4$. Since its value and first derivative vanish at $r = s$, integration gives $\mathrm{KL}(\mathrm{Ber}(r)\|\mathrm{Ber}(s)) \geq 2(r-s)^2$. Contraction yields the result; boundary cases follow by limits or are trivial when KL is infinite.

Finally let V be uniform on M hypotheses and let $\widehat{V}$ be any data-based estimate. If $p_e = \mathbb{P}(\widehat{V} \neq V)$, then introducing the error indicator gives

$$H(V \mid \mathcal{D}) \leq \log 2 + p_e \log M.$$

Indeed, conditional on no error, V is determined by the estimate; conditional on error it has at most $M - 1$ possible values. For a randomized estimate its independent random seed can be included in $\mathcal{D}$ without changing $I(V; \mathcal{D})$. Since $H(V) = \log M$, rearranging $I(V; \mathcal{D}) = H(V) - H(V \mid \mathcal{D})$ gives the version of Fano's inequality used in the proof.

M Model Assumptions and Resource Regimes

Fixed links versus growing latent structure. Theorem 2.2 fixes (P, R, ρ) while n grows; witness margins and likelihood constants are link-dependent. The hidden-rung theorem retains explicit binomial counts and Gaussian channel constants for growing $P = bD$. These statements provide fixed-link exponents and finite-parameter resource bounds, respectively.

Frozen batches versus within-batch feedback. Each batch applies one closure wave. The learner may compute arbitrarily between batches and commits every current-batch query before receiving its answers. Immediate within-batch feedback would increase the query dependency depth and belongs to a different round budget. The lower-bound oracle is stateful and target-specific, legal on every submitted query, and accesses private randomness through public queries.

Finite precision and the reference norm. Tolerance is measured in L_2 of the product reference, whose label marginal is common across embeddings. The precision-weighted currency q/τ^2 distinguishes finite-tolerance access from the exact-real encoding in Appendix F. The IID upper bound implements the specified query family of the algorithm on a high-probability event.

Product structure and disjoint blocks. Independent product coordinates provide centered-factor cancellation, tensor orthogonality, and uniform placement counts. Disjoint blocks give prefixes of distinct degrees and uniform-subset descriptions of unseen suffixes. These are the structural conditions used in the displayed rates.

A fresh hidden selector. Independent selector draws across rows give the IID mixture and its likelihood identity. Selecting one rung per row removes cross-rung products and supports the comparison of unrestricted SQ rounds with unary distributional witnesses. A selector shared across rows or a simultaneous sum defines a different observation law.

Recovery task and known parameters. The main task is active-union recovery with known structural parameters and signal scales. Proposition K.1 additionally recovers ordered blocks under known uniform selector weights and a common coefficient. Its decoder uses the resulting known separation between unary levels.

Information bounds and dispersion. The passive obstruction fixes the coordinate law and nonzero SNR as width grows. The critical design instead sets $u \asymp 1/(bD)$ and is certified by the constructive unary estimator. Appendix I optimizes the sufficient fourth-moment sample bound of that estimator over $0 < u < 1$.

Resource comparisons across scales. At fixed depth, SNR, and failure probability, the critical upper bound is $O(b \log n)$ and the information lower bound is $\Omega(b \log(n/b))$. They match up to constants when $b \le n^{1-\varepsilon}$ for fixed $\varepsilon > 0$. At constant active density, the stated logarithms remain distinct. Dispersion optimization in Appendix I minimizes the sufficient-bound objective of the specified unary estimator. The explicit formulas retain the separate depth and SNR factors. For growing active density, Remark 3.2 gives the ratio between the rigid binomial bounds. Frozen SQ rounds quantify information dependencies, with architecture and optimization dynamics treated as separate modeling choices.