SHARP RATES FOR INFORMATION COMPOSITION UNDER SOURCE UNCERTAINTY

Anonymous authors
Author review draft

ABSTRACT

Several observations are compressed separately and then used together for prediction. We ask how their individual information losses constrain the joint loss for a fixed finite source and deterministic encoders. The comparison ranges over all finite task alphabets. At positive finite whole-cell kernel depth d, we determine the sharp exponent to be $1 - 2^{-d}$. Infinite depth gives a linear bound. At depth zero, exact local preservation can fail to compose. The depth is computable in polynomial time for rational inputs. We construct binary witnesses, with two views sufficient for every positive finite depth. We also allow the source law to change and the relative posterior floor to approach zero. For an eight-atom, depth-two source, the sharp joint modulus is $\min\{u^{3/4}, \tau^{-1/3}u\} + e\log(1/e)$, where u is the local information budget, τ the floor, and e the source radius. Its source-error coefficient stays bounded as the floor vanishes. Exact depth can change under small mass perturbations on the same support, preventing uniform finite-sample identification. A prediction calculation and fixed-encoder experiments give a sample-size crossover whose coefficient depends on smoothing. Generic power existence and the conic restriction operation follow from established theory.

1 INTRODUCTION

Suppose we compress several observations separately and then use the representations to predict a label. The raw tuple can contain task information that none of its views reveals alone. An individual encoder's loss does not measure that joint information. We ask which fixed source laws and encoders make zero loss in every view imply zero joint loss for every task.

Once this implication holds, we still need to know how large the joint loss can be when the local losses are small. Task posteriors can approach zero, where a linear bound may fail. We follow those boundary cases through nonnegative functions whose local conditional-mean contrasts vanish. Each zero set gives a smaller state. We retain an original coarse cell only when the entire cell lies in that state. The shortest chain reaching a failed implication gives the sharp power $1 - 2^{-d}$. Binary witnesses attain this power. For rational inputs, we can also compute the depth exactly in polynomial time.

A positive posterior floor restores a linear bound. Its coefficient grows as the floor approaches zero, so transferring the bound to an approximately known source can amplify the source-error term. Our noisy-floor induction carries that error separately. Matching witnesses on an eight-atom system give the joint rate as local loss, floor, and source error vary together.

In this example, an arbitrarily small change in a positive source mass can change the exact depth even with the support fixed.

The prediction interpretation comes from log loss: joint conditional information is exactly the additional Bayes risk of using the compressed tuple. We calculate a sample-dependent risk crossover for smoothed decoders, then compare raw, compressed, and partially pooled predictors with the same tuning choices. The sample-size scale involves the structural exponent and a coefficient that depends on smoothing. These comparisons use fixed encoders.

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.

2 RELATED WORK AND PRECISE COMPARISON

Multi-view information bottleneck methods choose representations through prediction–compression tradeoffs, including consensus and incremental formulations (Huang et al., 2022). Here we take the source and deterministic encoders as given. We ask for a bound that holds for every finite task kernel on that system.

When exact composition holds, general results already guarantee that some power law exists. The quadratic perspective losses are continuous and semialgebraic on the normalized nonnegative simplex. Effective Łojasiewicz theory (Basu & Mohammad-Nezhad, 2024), semialgebraic maximization, and Puiseux asymptotics give an attained rational optimal exponent. An elementary binary embedding gives the same exponent. Appendix K writes out these reductions. Our failure-depth theorem determines the exponent's value from the source and partitions.

Kenig & Suciu (2022); Kenig (2025) prove linear bounds for specified classes of universal conditional-independence implications. Applying them here requires checking their saturated, marginal, or recursive hypotheses; exactness of a fixed source alone does not supply those hypotheses. The compatible-marginal forward-reverse Brascamp–Lieb theorem (Liu et al., 2018) also gives a formulation of our scalar affine inequality $G \le CL+b$. The rate question concerns how its optimal slack behaves near zero.

Conic error bounds relate rates to degeneracy (Sturm, 2000). The anchored whole-cell restriction is already an instance of SOCP expansion (Lourenço et al., 2016), and maximal-rank facial reduction also produces shortest chains (Sremac et al., 2021). For our homogeneous feasible lift, the radial bounds of Nishijima et al. (2026) give a square-root estimate: its partial-polyhedral Slater degree is one at every information failure depth. The capped, contrast-normalized zero-loss lift is infeasible. Appendix K states the lower mass estimate, uniform in the cap, needed to recover the sharp homogeneous bound. One approximating sequence gives an upper mass estimate. The lower estimate still needs its own argument.

Conditional-entropy continuity bounds source error by $O(e\log(1/e))$ independently of the task alphabet (Alhejji & Smith, 2020). Direct substitution into a nominal floor bound gives $\Phi_\tau(u+E(e))+O(E(e))$, where $E(e) \asymp e\log(1/e)$ and $\Phi_\tau(u) = \min\{u^{3/4}, \tau^{-1/3}u\}$ for the eight-atom example. We obtain $\Phi_\tau(u)+O(E(e))$, with a source-error coefficient that stays bounded as the floor vanishes. Matching witnesses show that both terms are needed.

3 FIXED-SOURCE MODEL AND THE FAILURE-DEPTH INVARIANT

Let S be the finite actual support of $X = (X_1, \dots, X_v)$, with $N = |S|$ and $\mu_x > 0$ for $x \in S$. The raw tuple distinguishes every atom. Fix deterministic encoders $Z_i = f_i(X_i)$, write $Z = (Z_1, \dots, Z_v)$, and let $\mathcal{J}$ contain the nonempty fibers of Z. We allow dependence between views and non-Cartesian support. A task is any finite label alphabet with a kernel $P(Y \mid X)$. The source and encoders are fixed before we choose the task.

All information quantities use natural logarithms. Define

$$\delta_\mu(Y) = \sum_{i=1}^{v} I_\mu(Y; X_i \mid Z_i), \qquad \Delta_\mu(Y) = I_\mu(Y; X \mid Z). \tag{1}$$

A local coarse cell C is a fiber of Z_i. Its fine cells $D \subseteq C$ are the fibers of X_i. For a nonnegative scalar function f, write $\langle f\rangle_E = \sum_{x\in E} \mu_x f_x/\mu(E)$. We use states T that are nonempty unions of whole joint cells $J \in \mathcal{J}$. An original coarse cell survives precisely when $C \subseteq T$. If it crosses the boundary, we remove it entirely. Its intersection with T adds no constraint. Every retained cell keeps its original source weights.

For each retained pair $D \subseteq C$, the local map A_T has row $\langle f\rangle_D - \langle f\rangle_C$. The target map B_T has rows $f_x - \langle f\rangle_{J(x)}$, for $x \in T$. Call T **exact** when

$$\ker A_T \subseteq \ker B_T. \tag{2}$$

An exact state has an edge to U if U is the proper, nonempty zero set of some $q \geq 0$ satisfying $A_T q = 0$. Exactness ensures that U is a union of whole joint cells. Nonexact states are terminal. Define $d(T)$ to be the shortest number of edges from T to a nonexact state, with $d(T) = \infty$ when none is reachable; write $d = d(S)$. Every edge removes at least one joint cell, so a finite positive depth is at most $|\mathcal{J}| - 1$.

At the root, (2) says exactly that $\delta_\mu = 0 \Rightarrow \Delta_\mu = 0$ for all finite tasks. Expand the conditional mutual informations into nonnegative scalar Jensen gaps to obtain one direction. For the other, shift and scale a signed kernel vector into a binary posterior.

4 SHARP EXPONENTS, COMPUTATION, AND REALIZABILITY

Theorem 1 (fixed-source classification). Suppose at least one joint cell contains more than one atom. The largest exponent $\alpha > 0$ for which a finite constant C, depending only on the fixed source and encoders, satisfies

$$\Delta_\mu(Y) \leq C\delta_\mu(Y)^\alpha \tag{3}$$

for every finite task is characterized as follows.

Failure depth	Sharp behavior
$d = 0$	A binary task has $\delta_\mu = 0 < \Delta_\mu$; no positive exponent works.
$1 \leq d < \infty$	The endpoint exponent is $\alpha_d = 1 - 2^{-d}$.
$d = \infty$	A finite linear coefficient exists, and the largest exponent is 1.

For finite positive depth, a bounded polynomial family of binary task kernels satisfies

$$\delta_\mu(t) = \Theta(t^{2^d}), \qquad \Delta_\mu(t) = \Theta(t^{2^d-1}), \qquad t \downarrow 0. \tag{4}$$

When all joint cells are singletons, $\Delta_\mu \equiv 0$; every exponent is vacuously valid and there is no largest one.

The proof begins by comparing each scalar entropy defect with a quadratic perspective. We choose a nonnegative local-kernel function of maximal mass among those lying below the profile and subtract it. The remaining mass is controlled by the square root of the local defect. A split relative to the chosen target cell then lets us continue on a smaller whole-cell state. This gives $\alpha_h = (1+\alpha_{h-1})/2$, starting at $1/2$. Infinite depth is handled by induction on the number of target cells. A shortest failed chain supplies the lower family (4). Appendix A gives these arguments and the summation over arbitrary finite label alphabets.

Theorem 2 (exact depth in polynomial time). For explicitly listed rational source masses and deterministic partitions of total binary input length L, the exact value of d, including 0 and ∞, is computable in time polynomial in L, using standard rational linear programming (Khachiyan, 1979) and rank tests. The algorithm uses $O(r^2)$ rank tests and $O(r^2 N)$ polynomial-size LP feasibility queries, where $r = |\mathcal{J}|$. A shortest failure certificate has polynomial bit size.

For each target cell J, the algorithm follows the canonical zero set

$$\mathcal{C}_J(T) = \{x \in T : q_x = 0 \text{ for every } q \geq 0,\ A_T q = 0,\ q|_J = 0\}. \tag{5}$$

At each state, test exactness and stop if the test fails or the chain reaches a fixed point. The minimum failure length over all J equals d. To prove this, Appendix B shows that the canonical chain stays inside any failed chain anchored at its terminal offending target cell. Following these canonical chains avoids enumerating the full state graph, which may be exponential. The theorem gives ordinary

polynomial bit complexity for explicit inputs. Strong polynomiality and estimating the source from data are outside its scope.

Proposition 3 (two views suffice). Every positive finite depth occurs with exactly two deterministic views. Four atoms suffice for $d = 1$, and $4d$ atoms suffice for $d \geq 2$. One view has identical local and joint loss and cannot exhibit positive finite depth. Appendix C constructs the two-view systems and identifies every reachable state. The support and alphabet sizes grow with d.

5 AN EXPLICIT EIGHT-ATOM SYSTEM

Use the following depth-two source. Each encoder keeps the letter prefix of its raw label.

Atom	Probability	X_1	X_2
a_1	$1/14$	L0	N0
a_2	$1/14$	L0	N1
b_1	$1/14$	M0	N1
b_2	$1/14$	M1	N0
c	$1/14$	M2	P0
z	$1/7$	L1	P1
e_1	$1/4$	L0	Q0
e_2	$1/4$	L1	Q1

Let $A = \{a_1, a_2\}$, $B = \{b_1, b_2\}$, $E = \{e_1, e_2\}$, $U = A \cup B \cup \{c, z\}$, and $K = B \cup \{c, z\}$. The joint cells are $A, B, \{c\}, \{z\}, E$. Direct conditional-mean equations give

$$\ker A_S = \operatorname{span}\{\mathbf{1}_U, \mathbf{1}_E\}, \qquad \ker A_U = \operatorname{span}\{\mathbf{1}_A, \mathbf{1}_K\}. \tag{6}$$

The root has two children, U and E. From U, we can move to A or K. The states E, K are exact leaves. At A, every original local coarse cell has disappeared, leaving the target comparison between a_1, a_2 unconstrained. This is the first possible failure, at depth $d = 2$.

The exponent can be attained with a balanced label. In the table's atom order, set

$$\eta_t = P(Y = 1 \mid X) = (t^3, 0, t^2, t^2, t^2, t^2, v_t, v_t), \qquad v_t = 1 - \frac{(5+t)t^2}{7}. \tag{7}$$

For $0 \leq t \leq 1$, this is a valid kernel with $P(Y = 1) = 1/2$. Writing h for binary entropy, direct conditioning and expansion yield

$$\delta(t) = \frac{29}{392}t^4 + O(t^5), \qquad \Delta(t) = \frac{\log 2}{14}t^3 + O(t^6). \tag{8}$$

Thus the nominal $3/4$ rate already occurs with a balanced marginal. We allow rare labels in the later source-perturbation construction.

Appendix D gives the conditional-mean equations, expansions, and conservative all-task bounds, including $\Delta_{\mu_0} \leq 114\delta^{3/4}$ and a floor-dependent linear bound. The coefficients have not been optimized.

6 POSTERIOR FLOORS AND PREDICTION RISK

Fix an exact system of finite positive depth. For $0 < \tau < 1$, require every task label to satisfy

$$P(Y = y \mid X = x) \geq \tau P(Y = y). \tag{10}$$

Labels with zero marginal contribute no information. The relative floor still allows positive marginals to be arbitrarily small. Let $F_\tau(u)$ be the supremum of Δ_μ over tasks satisfying (10) and $\delta_\mu \le u$. Write C_τ for the best linear comparison coefficient on this class.

Theorem 4 (relative-floor modulus). With $\alpha = 1 - 2^{-d}$ and $\beta = 1/(2^d - 1)$, there exist source-dependent positive constants and a common $u_0 > 0$ such that, uniformly for $0 < \tau \le 1/2$ and $0 < u \le u_0$,

$$C_\tau \asymp \tau^{-\beta}, \qquad F_\tau(u) \asymp \min\{u^\alpha, \tau^{-\beta} u\}. \tag{11}$$

Binary tasks attain the lower orders, and the two branches meet at a budget of order $\tau^{2^d/(2^d-1)}$. The result leaves the optimal numerical threshold undetermined. It also leaves open whether binary tasks attain the unrestricted optimum at each fixed budget.

The upper proof follows the floor through each child state and combines that estimate with Theorem 1. For the lower proof, begin with a positive shortest-chain witness whose smallest relative posterior has order t^{2^d-1}. A hidden one-sided channel thins its first label and preserves the floor. Local information falls by at least the thinning factor, and the first scalar target gap supplies the required lower bound. Appendix E gives the construction.

The same floor class can be described by keeping an underlying label with probability $1 - \tau$ and otherwise drawing independently from its marginal. This changes the task distribution. Whether applying such noise to training labels improves clean-label prediction requires a separate analysis.

For fixed encoders and n independent labeled compressed examples, the Laplace decoder

$$\widehat{q}(y \mid z) = \frac{N_{zy} + 1}{N_z + m} \tag{12}$$

on a known m-label alphabet obeys

$$\mathbb{E} R_{\log}(\widehat{q}) - H(Y \mid X) \le \Delta_\mu + \frac{s(m-1)}{n+1}, \qquad s = |\operatorname{supp}(Z)|. \tag{13}$$

Every decoder restricted to Z incurs structural excess at least Δ_μ. Substituting (3) or (11) in the standard estimation bound gives an achievable risk bound. Appendix F proves the estimation step; it makes no joint minimax claim.

Decoder experiment. We use the same eight-atom source and balanced binary priors for three comparisons: a depth-two boundary family, an independently certified depth-one encoder control, and an interior task on the depth-two system. The frozen grid contains three task scales and four training sizes from 10^3 to 10^6, with 512 independent training/test replicates in each cell. It gives 18,432 decoder pairs. Each test sample has 10^6 observations stored as exact multinomial counts. The seeds and counts are included in the supplement.

At depth two with $t = 0.1$, the population losses are $\delta = 6.76 \times 10^{-6}$ and $\Delta = 4.95 \times 10^{-5}$. The compressed-minus-raw held-out risk is $-1.426 \times 10^{-4} \pm 8.06 \times 10^{-6}$ at $n = 10^4$ and $+3.015 \times 10^{-5} \pm 1.01 \times 10^{-6}$ at $n = 10^5$. These are 95% Monte Carlo intervals for the means.

Proposition 5 (smoothing-dependent prediction crossover). Fix $a > 0$. For task (7), train binary add-a decoders $\widehat{q}^{(a)}_{W,n}(y \mid w) = (N_{wy} + a)/(N_w + 2a)$, $W = X, Z$, on the same n labeled samples. As $t \downarrow 0$ and $nt^3 \to c \in (0, \infty)$,

$$n\,\mathbb{E}[R(\widehat{q}^{(a)}_{Z,n}) - R(\widehat{q}^{(a)}_{X,n})] \longrightarrow c \log(2)/14 - (a+1). \tag{14}$$

The result gives the scale $n \asymp \delta^{-3/4}$ and a limiting sign change at $c = 14(a+1)/\log 2$. In the calculation, the rare positive-label numerator cancels between the two decoders. The boundary pair contributes $-a/n$ to their estimation difference. Each of the two equal-posterior pairs contributes $-1/(2n)$. Appendix P extends the add-one calculation in Appendix L to fixed $a > 0$; it does not cover a random, data-selected pseudocount or a minimax threshold.

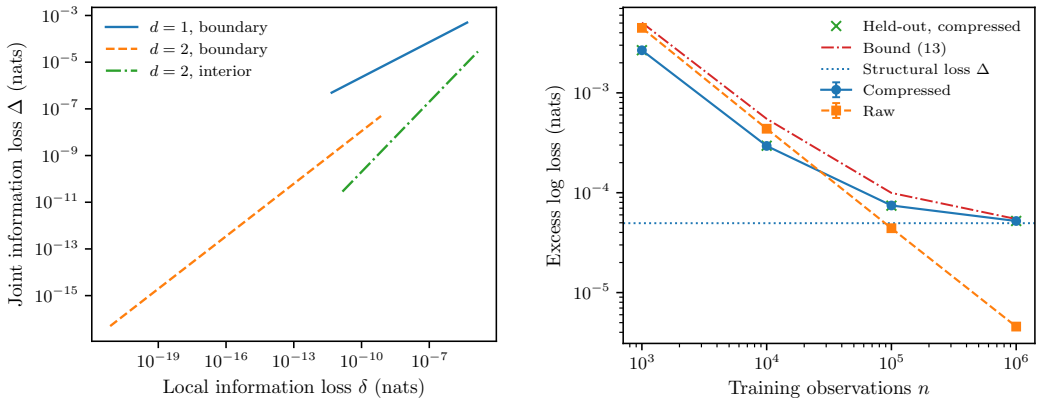

Figure 1: Left: population losses for the three specified task families. The interior family stays on the depth-two system. Right: raw and compressed Laplace decoders for the depth-two task at $t = 0.1$. Error bars give 95% Monte Carlo intervals for the mean population excess risk; crosses show the compressed decoder's mean held-out excess risk. The dotted line is the structural loss, and the dash-dot curve is bound (13).

Equal tuning and partial pooling. The follow-up protocol fixes $t = 0.1$ and five sample sizes, with 2,048 paired replicates at each size. Raw and compressed predictors receive the same five-fold smoothing grid. Their compressed-minus-raw population-risk difference is -7.386×10^{-5} at $n = 10^4$, with simultaneous 95% Monte Carlo interval $[-8.452, -6.319] \times 10^{-5}$. At $n = 10^5$, it is $+3.868 \times 10^{-5}$, with interval $[3.796, 3.940] \times 10^{-5}$. It is already positive near the original add-one sign-change bracket, at $n = 40{,}242$.

We also let the raw-access learner select a pooling weight. It selects no pooling in every replicate at 10^6. At each smaller size, its mean risk exceeds that of the better tuned endpoint. Tuning itself increases risk at 10^4 for both endpoints relative to fixed add-one smoothing. Appendix P includes these outcomes along with the full conditions.

7 Source perturbations and limits of sample certificates

For the eight-atom source μ_0, fix the full ambient raw alphabet

$$\Omega = \{\mathrm{L0}, \mathrm{L1}, \mathrm{M0}, \mathrm{M1}, \mathrm{M2}\} \times \{\mathrm{N0}, \mathrm{N1}, \mathrm{P0}, \mathrm{P1}, \mathrm{Q0}, \mathrm{Q1}\}, \qquad |\Omega| = 30.$$

The same prefix encoders apply to every ambient atom. Define

$$R_{\mu_0,\tau}(u, e) = \sup\{\Delta_\nu : \mathrm{TV}(\nu, \mu_0) \le e,\ \delta_\nu \le u,\ P(y \mid x) \ge \tau P_\nu(y)\}, \tag{15}$$

where both source and finite task may vary, and the floor is imposed at positive-source atoms.

Theorem 6 (sharp joint source/floor modulus). Put $\Phi_0(u) = u^{3/4}$, and $\Phi_\tau(u) = \min\{u^{3/4}, \tau^{-1/3}u\}$ for $\tau > 0$. There are source-dependent positive constants and common $u_0, e_0 > 0$ such that, uniformly for $0 \le \tau \le 1/2$, $0 \le u \le u_0$, and $0 \le e \le e_0$,

$$R_{\mu_0,\tau}(u, e) \asymp \Phi_\tau(u) + e\log(1/e). \tag{18}$$

The constants do not vary with task alphabet, label prior, or floor within the stated range. Taking $e\log(1/e) = 0$ at zero includes both axes and the no-floor case. The optimal linear coefficient grows as $\tau^{-1/3}$. For a fixed $1/2 < \tau < 1$, Appendix G gives order $u + e\log(1/e)$ with constants that may depend on τ. At $\tau = 1$, the modulus is zero. Appendix O proves the uniform statement.

For the lower construction, add mass $e/2$ to each missing cross-pair $(\mathrm{L0}, \mathrm{Q1})$, $(\mathrm{L1}, \mathrm{Q0})$, and multiply the nominal masses by $1 - e$. Let T indicate a new atom and use

$$P(Y = 1 \mid T) = \tau e + (1 - \tau)T. \tag{19}$$

The new-atom indicator carries no extra information within either local coarse cell, so $\delta_\nu = 0$. Both labels meet their relative floors and $P(Y = 1) = e$. Direct entropy expansion gives

$$\Delta_\nu = (1 - \tau)e \log(1/e) + O_\tau(e). \tag{20}$$

To obtain one lower bound for every $\tau \le 1/2$, use channel (19) with floor $1/2$. It meets each smaller requested floor and gives $\Delta_\nu \ge \frac{1}{2} e \log(1/e) - 2e$. This and the nominal witnesses give the sum in (18). The upper bound uses the noisy-floor estimate below because the fixed-floor continuity argument alone does not give uniform constants.

Now consider a certificate based on raw source observations. Let $F_\tau(\nu) = R_{\nu,\tau}(0, 0)$ for the same ambient alphabet and encoders. Suppose the nonnegative statistic C_n uses only n independent raw-source samples and has uniform coverage:

$$P_\nu\{C_n \ge F_\tau(\nu)\} \ge 1 - \beta \tag{21}$$

on a source class containing μ_0 and all sufficiently small contaminations above. Then, for sufficiently large n,

$$P_{\mu_0}\left\{C_n \ge \left[\frac{(1 - \tau)\log n - \log 2}{2n}\right]_+\right\} \ge 1 - 2\beta. \tag{22}$$

A contaminated sample containing no new atoms has exactly the nominal product law. This observation gives (22), as shown in Appendix H. The certifier receives only raw source samples. Giving it an exact-support promise, labeled observations, or the task kernel would change the problem.

Theorem 7 (general source stability with a vanishing floor). Let μ be a fixed finite exactly composing source of finite positive depth d, minimum positive nominal atom mass m, and fixed encoders on a finite ambient alphabet of size N_Ω. Put $\alpha = 1 - 2^{-d}$ and $\beta = 1/(2^d - 1)$. Uniformly over finite tasks with relative floor $0 < \tau \le 1$ under ν, and $\mathrm{TV}(\nu, \mu) \le e \le \min\{m/2, 1/2\}$,

$$\Delta_\nu \le C_\mu \min\{\delta_\nu^\alpha, \tau^{-\beta}\delta_\nu\} + C_\mu e + 2e \log N_\Omega + 2h(e). \tag{23}$$

The constant is independent of τ. At zero floor, retain the fractional branch. At infinite depth, replace the minimum by δ_ν. The ambient source need not compose exactly.

For each task column f_y, set $a_y = \sum_x |\nu_x - \mu_x| f_y(x) + eM_\mu(f_y)$. These allowances sum to $\sum_y a_y \le 3e$. Appendix J handles zero floor. Appendix O uses the inherited relative floor ρ to prove $Q_{G,T}(f) \le C\rho^{-1/(2^h-1)}\ell + Ca$ for local row loss ℓ and depth at least h. The important step is to remove children of mass $O(a)$ using the mass bound first. On the remaining children, the source allowance stays additive without a factor $\rho^{-1/(2^h-1)}$. Floor transport, label summation, and finite-source entropy continuity then give (23).

For a direct comparison with the earlier bounds, set $\eta = E(e) \asymp e \log(1/e)$, $u = \eta^{6/5}$, and $\tau = \eta^{3/5}$. Combining direct substitution with the no-floor robust estimate gives the better of orders $\eta^{4/5}$ and $\eta^{9/10}$. Equation (18) gives order η. This calculation compares those particular estimates; it does not settle every possible reduction to prior theory.

Proposition 8 (depth instability and identification). In the eight-atom source, multiply the mass of z by $1 + q$, renormalize, and retain the support and encoders. For $q > -1$, the exact depth is ∞ if $q < 0$, 2 if $q = 0$, and 1 if $q > 0$. Moreover $\mathrm{TV}(\mu_q, \mu_0) = 6|q|/[7(7 + q)]$. For any fixed sample size, every estimator of exact depth has worst-case error at least $2/3$ in every neighborhood of $q = 0$. Appendix M proves the cone classification and statistical consequence. The result limits exact-depth estimation around a known nominal model.

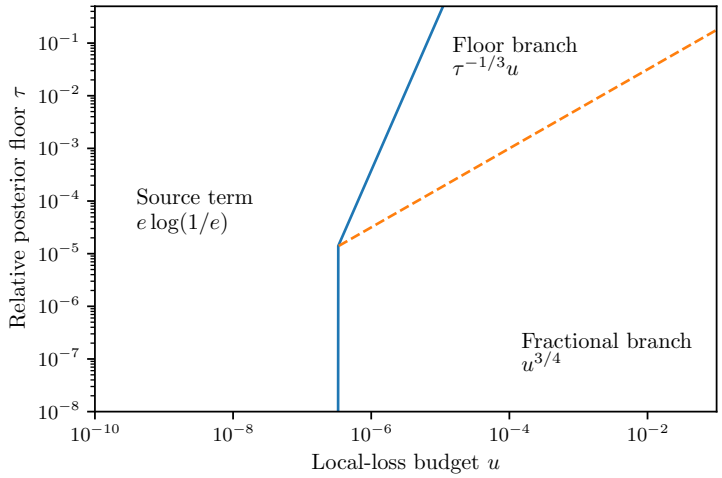

Figure 2: Terms controlling the depth-two rate at $e = 10^{-6}$, with multiplicative constants suppressed. The nominal branches meet at $u = \tau^{4/3}$. The source term dominates where $\Phi_\tau(u) \leq e\log(1/e)$. These boundaries belong to the asymptotic rate expression; optimal finite-budget thresholds remain unspecified.

8 Reproducibility and Limitations

The verifier reconstructs each matrix, edge, and terminal rank condition from the rational source. A floating-point LP proposes a certificate, then exact rational arithmetic checks it. Independent full cone-graph enumeration and canonical chains agree on 188 systems, including 180 fresh sources and structured depth-one/two cases. The original implementation agrees on 187 valid-model cases across 3,928 coordinate decisions. Another replay checks 50 archived certificates with 9,885 checks.

High-precision calculations check witness orders at depths one through four. The original 36-cell decoder experiment reproduces from its frozen seeds and counts, and the supplement retains all 10,240 follow-up training replicates. Appendix I lists the relevant code. The universal claims and rational bit-complexity result rest on the proofs.

The constants can depend on the source masses and partitions. Matching source-error lower bounds are established for the stated eight-atom source and ambient model. Our experiments use theorem-directed synthetic tasks with known groupings and fixed encoders. We have not evaluated natural data or trained encoders, and population local-information estimation and joint minimax analysis remain open.

Appendix K shows that finite-label attainment needs at most $N + 1$ labels for a fixed source, or 31 for the ambient robust problem. It does not give an efficient global optimizer. Estimating exact depth uniformly requires a separation promise.

REPRODUCIBILITY STATEMENT

The supplement contains the proofs, source tables, deterministic partitions, frozen protocols, seeds, paired counts, and numerical outputs. Its README gives quick checks, full experiment commands, dependency versions, and a claim-to-evidence map. Structural certificates are checked with exact arithmetic. High-precision numerical checks and Monte Carlo intervals are recorded separately. The independent critiques in the record were AI-assisted; external human certification remains outstanding.

AI ASSISTANCE STATEMENT

AI systems assisted with mathematical development, proof reconstruction and critique, literature comparison, code generation, computational checks, experiment design and analysis, and manuscript drafting and revision. The separate agent checks were also AI-assisted. Human verification and author approval remain to be completed. The submitting authors are responsible for checking the mathematics, code, citations, and final text before submission.

REFERENCES

Mohammad A. Alhejji and Graeme Smith. A tight uniform continuity bound for equivocation. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 2270–2274, 2020. doi: 10.1109/ISIT44484.2020.9174350. URL https://arxiv.org/abs/1909.00787v3.

Saugata Basu and Ali Mohammad-Nezhad. Improved effective łojasiewicz inequality and applications. *Forum of Mathematics, Sigma*, 12:e115, 2024. doi: 10.1017/fms.2024.66. URL https://arxiv.org/abs/2211.10034v2.

Teng-Hui Huang, Aly El Gamal, and Hesham El Gamal. On the multi-view information bottleneck representation. arXiv:2202.02684v4, 2022. URL https://arxiv.org/abs/2202.02684v4.

Batya Kenig. Approximate implication for probabilistic graphical models. *Journal of Artificial Intelligence Research*, 82:1–37, 2025. doi: 10.1613/jair.1.16467. URL https://arxiv.org/abs/2310.13942.

Batya Kenig and Dan Suciu. Integrity constraints revisited: From exact to approximate implication. *Logical Methods in Computer Science*, 18(1), 2022. doi: 10.46298/lmcs-18(1:5)2022. URL https://arxiv.org/abs/1812.09987v8.

L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademii Nauk SSSR*, 244(5):1093–1096, 1979. URL https://www.mathnet.ru/eng/dan42319.

Jingbo Liu, Thomas A. Courtade, Paul W. Cuff, and Sergio Verdú. A forward-reverse Brascamp–Lieb inequality: Entropic duality and Gaussian optimality. *Entropy*, 20(6):418, 2018. doi: 10.3390/e20060418. URL https://doi.org/10.3390/e20060418.

Bruno F. Lourenço, Masakazu Muramatsu, and Takashi Tsuchiya. Weak infeasibility in second order cone programming. *Optimization Letters*, 10(8):1743–1755, 2016. doi: 10.1007/s11590-015-0982-4. URL https://arxiv.org/abs/1509.05168.

Mitsuhiro Nishijima, Guoyin Li, and Bruno F. Lourenço. Radial-type error bounds for semidefinite feasibility problems without strict feasibility: Qualitative estimates and asymptotic tightness. arXiv:2608.09013v2, 2026. URL https://arxiv.org/abs/2608.09013v2.

Stefan Sremac, Hugo J. Woerdeman, and Henry Wolkowicz. Error bounds and singularity degree in semidefinite programming. *SIAM Journal on Optimization*, 31(1):812–836, 2021. doi: 10.1137/19M1289327. URL https://arxiv.org/abs/1908.04357.

Jos F. Sturm. Error bounds for linear matrix inequalities. *SIAM Journal on Optimization*, 10(4):1228–1248, 2000. doi: 10.1137/S1052623498338606. URL https://doi.org/10.1137/S1052623498338606.

A PROOF OF THE FIXED-SOURCE CLASSIFICATION

A.1 SCALARIZATION AND QUADRATIC PERSPECTIVES

For a state T, put $M_T(f) = \sum_{x\in T} \mu_x f_x$. With $\phi(a) = a\log a$, $\phi(0) = 0$, define

$$G_T(f) = \sum_{J\subseteq T} \left[\sum_{x\in J} \mu_x \phi(f_x) - \mu(J)\phi(\langle f\rangle_J)\right],$$

$$L_T(f) = \sum_{C\subseteq T} \left[\sum_{D\subseteq C} \mu(D)\phi(\langle f\rangle_D) - \mu(C)\phi(\langle f\rangle_C)\right]. \tag{A1}$$

Keep the view index in the sums over C, D, including when two views have identical cells. Both functions are nonnegative, continuous, and homogeneous of degree one. Their zero sets satisfy $G_T(f) = 0 \iff B_T f = 0$ and $L_T(f) = 0 \iff A_T f = 0$. Write $c_\ell = \langle f\rangle_{C_\ell}$, $a_\ell = A_\ell f$, and $b_J = \langle f\rangle_J$, and define

$$Q_{L,T}(f) = \sum_\ell \mu(D_\ell)\frac{a_\ell^2}{c_\ell}, \qquad Q_{G,T}(f) = \sum_{J\subseteq T}\sum_{x\in J} \mu_x \frac{(f_x - b_J)^2}{b_J}. \tag{A2}$$

When a denominator is zero, nonnegativity forces all values in that cell to be zero. We assign zero to that term. Fix $R \geq 1$ bounding the point-to-target and fine-to-coarse mean ratios. For $0 \leq a \leq Rb$,

$$\frac{(a-b)^2}{2Rb} \leq a\log(a/b) - a + b \leq \frac{(a-b)^2}{b}. \tag{A3}$$

The lower inequality follows from the curvature $1/a$; the upper one follows from $\log t \leq t - 1$. Continuity covers the endpoints. After the linear terms cancel within each cell, we obtain $Q/(2R) \leq H \leq Q$ for $H = L, G$.

Write $M = M_T(f)$ and $Q = Q_{L,T}(f)$. Finiteness and positive source weights imply

$$Q \leq CM, \qquad Q_{G,T} \leq CM, \qquad \|A_T f\| \leq C\sqrt{MQ}. \tag{A4}$$

If T is exact, each target row is a fixed linear combination of local rows. Weighted Cauchy–Schwarz then gives, for a target cell with $b_J > 0$,

$$Q_{G,J}(f) \leq C_J MQ/b_J. \tag{A5}$$

Also $Q_{G,J}(f) \leq C'_J b_J$, since $f_x \leq \mu(J)b_J/\mu_x$. Taking the minimum proves $Q_{G,T} \leq C\sqrt{MQ}$. This is the depth-one upper bound.

A.2 A MAXIMAL KERNEL MINORANT

Let $\mathcal{K}_T = \{q \geq 0 : A_T q = 0\}$. Choose $q \in \mathcal{K}_T$ maximizing $M_T(q)$ subject to $q \leq f$. Compactness gives a maximizer. Set $r = f - q \geq 0$. If its support contained a nonzero nonnegative local-kernel vector, we could add a small multiple of that vector to q and increase the objective. Thus no such vector exists.

On any support H containing no such vector, consider the compact mass-one simplex of nonnegative functions. The minimum of $\|A_T r\|$ on it is positive. Taking the smallest minimum over the finitely many eligible supports and rescaling gives

$$M_T(r) \leq C\|A_T r\| = C\|A_T f\| \leq C\sqrt{MQ}. \tag{A6}$$

The inequality also holds for a zero residual. Its constants depend only on the finite source and partitions.

The mass-one slice of $\mathcal{K}_T$ is a bounded polytope with finitely many vertices. Each vertex is constant on each target cell by exactness. Fix J. In $q = q_0 + q_+$, group vertices according to whether they vanish or are positive on J, and set $g = q_+ + r$. Write the vertex decomposition as $q = \sum_k \lambda_k p^k$, with a_k the value of p^k on J. If the positive group is nonempty, set $a_* = \min_{a_k>0} a_k$. Then

$$M_T(q_+) = \sum_{a_k>0} \lambda_k \le a_*^{-1} \sum_{a_k>0} \lambda_k a_k = a_*^{-1} \langle q \rangle_J \le a_*^{-1} b_J.$$

If every vertex vanishes on J, take $q_+ = 0$. Together with (A6), this gives

$$f = q_0 + g, \quad q_0|_J = 0, \quad g|_J = f|_J, \quad M_T(g) \le C\big(b_J + \sqrt{MQ}\big). \tag{A7}$$

If $q_0 \neq 0$, its zero set U is a legal proper child containing J. On that child,

$$f|_U = g|_U, \qquad Q_{L,U}(f|_U) \le Q, \qquad M_U(f|_U) \le M_T(g). \tag{A8}$$

The whole-cell rule leaves every child loss term unchanged from the parent. Since all terms are nonnegative, (A8) follows.

A.3 Quantitative induction

For each integer $h \ge 1$, we prove that $d(T) \ge h$ implies

$$Q_{G,T}(f) \le C_{T,h} M_T(f)^{2^{-h}} Q_{L,T}(f)^{1-2^{-h}}. \tag{A9}$$

The depth-one estimate gives the base case $h = 1$. For $h \ge 2$, put $n = 2^h$ and fix J. Nonnegativity and exactness give zero target loss whenever the mass, local loss, or target mean is zero. In the other cases we use (A7).

If $b_J \ge \sqrt{MQ}$, then $M_T(g) \le C b_J$. When $q_0 \neq 0$, the child is exact and (A5), applied in the child with (A8), gives $Q_{G,J} \le CQ$. When $q_0 = 0$, the same estimate holds at T, because $f = g$. Since $Q \le CM$, this is bounded by a constant times $M^{1/n} Q^{1-1/n}$.

If $b_J < \sqrt{MQ}$ and $q_0 \neq 0$, the child has depth at least $h - 1$ and mass at most $C\sqrt{MQ}$. The induction hypothesis gives

$$Q_{G,J} \le Q_{G,U} \le C(MQ)^{1/n} Q^{1-2/n} = C M^{1/n} Q^{1-1/n}. \tag{A10}$$

In the remaining case, $b_J < \sqrt{MQ}$ and $q_0 = 0$, equation (A7) gives $M \le C\sqrt{MQ}$, so $M \le C^2 Q$. The bound $Q_{G,J} \le CM$ now suffices. Choose a common constant over the finitely many children and target cells. This proves (A9), and (A3) converts it to the entropy bound.

If $d(T) = \infty$, each child also has infinite depth. Induct on the number of target cells using the same three cases. The first and third give linear bounds. In the second, apply the child's linear bound to obtain $Q_{G,J} \le CQ$. A state with one target cell has no nonzero q_0 vanishing on that cell, which starts the induction. This proves the infinite-depth case directly.

For a task, set $f_y(x) = P(Y = y \mid X = x)$. Expanding the entropies gives $\Delta = \sum_y G_S(f_y)$, $\delta = \sum_y L_S(f_y)$, and $\sum_y M_S(f_y) = 1$. Apply Hölder's inequality:

$$\sum_y M_S(f_y)^{1-\alpha} L_S(f_y)^{\alpha} \le \delta^{\alpha}, \tag{A11}$$

This makes the coefficient independent of label count. For $\alpha = 1$, direct summation of the linear estimates gives the result.

A.4 SHARPNESS FROM A SHORTEST FAILED CHAIN

Take a shortest path $T_0 = S \to T_1 \to \cdots \to T_d$. For each edge, choose $q_i \geq 0$ vanishing exactly on T_{i+1} and extend it by zero outside its state. At T_d, take a signed vector in $\ker A_{T_d} \setminus \ker B_{T_d}$. Adding a large enough constant produces $g > 0$ with $L_{T_d}(g) = 0 < G_{T_d}(g) = a$.

For an edge $T \to U$, its witness q, and bounded $h \geq 0$ supported on U,

$$G_T(q + \epsilon h) = \epsilon G_U(h), \qquad L_T(q + \epsilon h) = \epsilon L_U(h) + O(\epsilon^2). \tag{A12}$$

Since q is constant on each joint cell, the target identity follows. Cells retained inside U scale exactly. Every other retained coarse cell C has the same positive fine and coarse mean m_C at q. With $h_D = \langle h \rangle_D$ and $h_C = \langle h \rangle_C$, its contribution is

$$L_C(q + \epsilon h) = \frac{\epsilon^2}{2m_C} \sum_{D \subseteq C} \mu(D)(h_D - h_C)^2 + O(\epsilon^3).$$

The first derivative cancels. The contribution is nonnegative, and the remainder is uniform over bounded h. At the final parent, the quadratic coefficients cannot all vanish: that would make every crossing row annihilate the terminal witness. The child rows already annihilate it, so parent exactness would force its nonzero target contrast to vanish.

At the final parent, take $q_{d-1} + tg$. Its target loss is at. If the crossing quadratic coefficient vanished, every crossing row would annihilate g. The child rows already do so, contradicting exactness at the parent. The local loss is therefore $\Theta(t^2)$.

Proceed upward. For a child with target order t^r and local order t^s, use $q_i + t^s h_{i+1}(t)$. By (A12), the parent has target order t^{s+r} and local order t^{2s}; the scaled child term gives the local lower bound. Beginning with $(s, r) = (2, 1)$ gives root orders 2^d and $2^d - 1$. The coordinates are polynomials bounded uniformly for small t.

Fix $c > 0$ such that $\eta_t = ch_0(t) \in [0, 1/2]$ and take this as the first binary posterior. Its scalar losses have the required orders. The complementary posterior lies in $[1/2, 1]$, where its defects are bounded by squared contrasts. Exactness gives $\|Bh_0\| \leq C\|Ah_0\|$, and boundedness with (A3) gives $\|Ah_0\|^2 \leq CL(h_0)$. Its losses are $O(t^{2^d})$. This proves (4) and rules out every larger exponent.

At depth zero, the signed-vector construction gives a binary task with zero local loss and positive joint loss. At infinite depth with a nonvacuous target, take $1/2 + tv$ with $Bv \neq 0$. Exactness gives $Av \neq 0$, and strict convexity makes both losses order t^2. An exponent above one is impossible. This completes Theorem 1.

B POLYNOMIAL DEPTH COMPUTATION

For each exact T and anchor $J \subseteq T$, compute (5) by testing, for each $x \in T$, feasibility of

$$A_T q = 0, \quad q \geq 0, \quad q|_J = 0, \quad q_x = 1. \tag{B1}$$

Homogeneity shows that (B1) is feasible exactly when some vector in the constrained kernel is positive at x. For a proper canonical zero set, sum one feasible witness for each outside coordinate. This gives a nonnegative local-kernel vector whose zero set is exactly the canonical set, hence a legal edge. If the canonical set equals T, the constrained cone is zero and the chain stops at a fixed point.

Run the chain for each anchor, checking $\ker A_T \subseteq \ker B_T$ before each move. Recorded failures are genuine paths, so their minimum length is at least d. To prove the reverse inequality, take any failed path $T_0 \to \cdots \to T_k$ and a terminal offending row on J. Write V_i for the J-anchored chain, with virtual repetition at a fixed point. Before any earlier canonical failure, $V_i \subseteq T_i$. Indeed, restrict the original edge witness to V_i: retained whole cells keep their rows, so the restricted vector stays in the nonnegative local kernel and vanishes on J. Hence

$$V_{i+1} = \mathcal{C}_J(V_i) \subseteq V_i \cap T_{i+1}. \tag{B2}$$

At a canonical fixed point, the constrained cone is zero. The restricted edge witness must vanish, giving $V_i \subseteq T_{i+1}$ and allowing virtual repetition. If V_k were exact, extend the row representation of the offending target row by zeros to T_k. This contradicts the terminal failure. The canonical chain therefore fails in at most k proper steps. A shortest failed path then gives the computed minimum d.

We run r chains, each with at most $r-1$ edges and at most N LP tests per state. Exactness is checked by comparing the ranks of A_T and $[A_T; B_T]$. Clearing the rational source denominators once gives integer rows of polynomial bit length, reusable at every state. Rational LP feasibility (Khachiyan, 1979) and exact elimination therefore give polynomial runtime.

For a rational edge certificate, maximize ϵ subject to $A_T q = 0$, $q|_U = 0$, $\sum_{T\setminus U} q_x = 1$, and $q_x \geq \epsilon \geq 0$ outside U. At a vertex optimum, ϵ is positive. Cramer's rule and determinant bounds give polynomial-bit coordinates. Obtain a terminal witness by solving $A_T v = 0$ and $bv = 1$ for an offending row b. With at most $r-1$ edges, the full certificate has polynomial size. It verifies the path; shortestness follows from the canonical-chain argument above.

The binary witness can also be written sparsely. Set $E_i = 2^d - 2^{d-i}$, $0 \leq i \leq d$, and

$$h(t) = \sum_{i=0}^{d-1} t^{E_i} q_i + t^{2^d-1} g. \tag{B3}$$

The supports of the terms are disjoint: each lies on a removed shell or the terminal state. Each coordinate has one nonzero monomial. Binary storage of the exponents and polynomial-bit rational coefficients gives a polynomial-size description. Dense expansion or exact evaluation at a general rational parameter can require exponentially many output bits. The runtime claim excludes those operations.

C EVERY DEPTH IS REALIZABLE WITH TWO VIEWS

For this construction, use the canonical family before the balanced-prior reweighting μ_0 of Section 5. Fix $d \geq 2$ and $h = d-2$. Take pair cells $A = \{a_1, a_2\}$, $B = \{b_1, b_2\}$, $E = \{e_1, e_2\}$, and $O_r = \{o_{r,1}, o_{r,2}\}$ for $1 \leq r \leq h$, with singleton atoms c, z, g_r, k_r. Give z weight two and every other atom weight one. The $4d$ atoms then have total weight $4d+1$.

The first view has these coarse cells and their fine partitions:

1. $A \cup \{z\} \cup E$, split into $A \cup \{e_1\}$ and $\{z, e_2\}$.
2. $B \cup \{c\}$, fully split into singleton fine cells.
3. For each r, $\{g_r, k_r\} \cup O_r$, split into $\{g_r, o_{r,1}\}$ and $\{k_r, o_{r,2}\}$.

The second view has $A \cup B$, split into $\{a_1, b_2\}$ and $\{a_2, b_1\}$; the fully split cell $\{c, z, k_1, \ldots, k_h\}$; and the following fully split cells. If $h = 0$, the final cell is E. If $h \geq 1$, they are $E \cup \{g_1\}$, $O_r \cup \{g_{r+1}\}$ for $1 \leq r < h$, and O_h. These lists partition all atoms. Their joint fine labels distinguish every atom, and their joint coarse cells are the named pairs and singleton atoms.

Set $P = A \cup B \cup \{c, z, k_1, \ldots, k_h\}$. For $h = 0$, put $R_0 = E$; otherwise put $R_0 = E \cup \{g_1\}$, $R_r = O_r \cup \{g_{r+1}\}$ for $1 \leq r < h$, and $R_h = O_h$. Let $T_{-1} = P$ and $T_j = P \cup R_0 \cup \cdots \cup R_j$. The root is T_h.

For $0 \leq j \leq h$, at T_j the fully split cells overlapping at c force all entries on $P \setminus A$ to a common value y. The $A \cup B$ mean equation forces $a_1 = a_2 = x$. Each retained second-view cell R_r has a common value v_r. The first coarse cell gives

$$(2x + v_0)/3 = (2y + v_0)/3,$$

so $x = y$. For each retained indexed first-view cell $r \leq j$, its mean equation gives $(v_{r-1} + v_r)/2 = (y + v_r)/2$, hence $v_{r-1} = y$. The cell at $r = j+1$ is omitted whole. Consequently

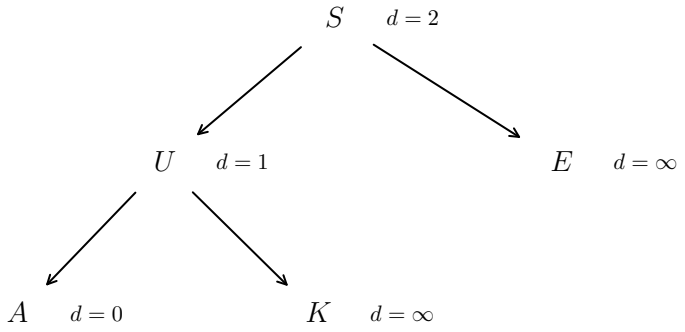

Figure 3: The five reachable states of the eight-atom system. The shortest failed path is $S \to U \to A$, so $d(S) = 2$. Each restriction retains original coarse cells contained in the child. A crossing cell is omitted in full.

$$\ker A_{T_j} = \operatorname{span}\{\mathbf{1}_{T_{j-1}}, \mathbf{1}_{R_j}\}. \tag{C1}$$

The two blocks are unions of target cells, so the state is exact. Its only proper nonempty zero-set children are T_{j-1} and R_j. Each R_j is fully split and is an exact leaf. At P, the first mean constraint involving E has been removed, giving

$$\ker A_P = \operatorname{span}\{\mathbf{1}_A, \mathbf{1}_{P\setminus A}\}. \tag{C2}$$

The fully split coarse cells cover $P \setminus A$ and overlap, making it an exact leaf. At A, all original coarse cells have been removed. Its two atoms are still compared by the target, so the state is nonexact. The failure path is $T_h \to \cdots \to T_0 \to P \to A$, with an exact leaf branching from each earlier state. Its length is $h + 2 = d$. The kernel calculations list every possible child and exclude a shorter path.

For $d = 1$, take four uniform atoms a, b, c, z. The first view has coarse cells $\{a, b\}$ and $\{c, z\}$, with fine cells $\{a, b\}, \{c\}, \{z\}$. The second has one coarse cell and fine cells $\{a, z\}, \{b, c\}$. In the root kernel, functions are constant on each target pair. The first pair is a failed child; the second is an exact leaf. With only one view, raw fine labels distinguish the atoms, so the local and joint gaps agree at every state. Proposition 3 follows.

D Details of the eight-atom example

D.1 Exactness and the balanced sharpness family

Using atom names for scalar values, the root local-kernel equations are

$$2a_1 + 2a_2 + 7e_1 = 4z + 7e_2, \quad b_1 = b_2 = c, \quad a_1 + b_2 = a_2 + b_1, \quad c = z, \quad e_1 = e_2. \tag{D1}$$

These equations give (6). At U, retain only M, N, and P. Their constraints give $b_1 = b_2 = c = z$ and $a_1 = a_2$, as used in Section 5.

For the posterior (7), only target cell A contributes. Hence

$$\Delta(t) = \frac{1}{7}\left[h(t^3/2) - \frac{1}{2}h(t^3)\right]. \tag{D2}$$

Define $J(p, q) = h((p + q)/2) - [h(p) + h(q)]/2$,

$$u_1 = (2t^3 + 7v_t)/11, \quad u_2 = (4t^2 + 7v_t)/11, \quad w_1 = (t^2 + t^3)/2, \quad w_2 = t^2/2.$$

Direct conditional means give

$$\delta(t) = \frac{11}{14} J(u_1, u_2) + \frac{2}{7} J(w_1, w_2). \tag{D3}$$

The first midpoint tends to $7/11$, where its Taylor expansion contributes $11/196$ at order t^4. The second midpoint has order t^2. Expanding in relative probabilities gives coefficient $1/56$, so the sum is $29/392$. At the boundary, (D2) expands to $(\log 2)t^3/14 + O(t^6)$, proving (8).

D.2 Explicit nominal coefficients

All local fine-to-coarse mean ratios are at most three, so (A3) gives $Q_L \le 6L$. Also $G \le Q_G$ and $Q_L \le 4M$. Exact rational row representations of target contrasts, followed by weighted Cauchy–Schwarz, give

$$Q_{G,J} \le D_J M_T Q_{L,T}/b_J, \qquad Q_{G,J} \le E_J b_J,$$

with the following coefficients on each listed child and also valid at the root:

Target	Child	D_J	$E_J = \mu(J)$
A	U	$35/3$	$1/7$
B	U	$14/3$	$1/7$
E	E	2	$1/2$

Summing the geometric-mean bounds gives $Q_G \le 4\sqrt{MQ_L}$, hence $G \le 10\sqrt{ML}$. The root kernel in (6) permits the explicit subtraction $q = (\min_U f)\mathbf{1}_U + (\min_E f)\mathbf{1}_E$. For $r = f - q$, exact row certificates give

$$M(r)^2 \le \frac{13733}{300} M Q_L \le 49 M Q_L. \tag{D4}$$

For each target cell in its child T, $\mu(T) = 1/2$, so $M_T(f) \le b_J/2 + 7\sqrt{MQ_L}$. Since $Q_{L,T} \le Q_L$,

$$Q_{G,J} \le \min\{E_J b_J,\ D_J Q_L/2 + 7 D_J \sqrt{MQ_L}\, Q_L/b_J\}. \tag{D5}$$

The elementary inequality $\min\{Eb, a + c/b\} \le a + \sqrt{Ec}$, together with $Q_L \le 4M$, yields

$$Q_G \le \frac{85}{3} M^{1/4} Q_L^{3/4}, \qquad G \le \frac{85}{3} 6^{3/4} M^{1/4} L^{3/4} < 114 M^{1/4} L^{3/4}. \tag{D6}$$

Use the conservative integer bounds (4,3,3) for the three square roots. Equation (A11) gives $\Delta_{\mu_0} \le \min\{11 \log 2/14, 10\sqrt{\delta}, 114\delta^{3/4}\}$ after scalarization. The entropy cap is $H_{\mu_0}(X \mid Z) = 11 \log 2/14$.

For the floor coefficient, define quadratic forms

$$L_0 = A^\top \operatorname{diag}(\mu(D_\ell))A, \qquad G_0 = B^\top \operatorname{diag}(\mu_x)B. \tag{D7}$$

The rational certificate gives $G_0 \preceq (7/2)L_0$: all 255 nonempty principal minors of the difference are nonnegative. For a normalized task profile $f_y = P(y \mid x)/P(y)$, we have $\tau \le f_y \le 14$. Curvature of $a \log a$ gives $L(f_y) \ge f_y^\top L_0 f_y/28$ and $G(f_y) \le f_y^\top G_0 f_y/(2\tau)$. Combine these inequalities and sum with label-prior weights to obtain $\Delta \le 49\delta/\tau$. The supplied row and principal-minor certificates check the constants exactly, without sampling tasks.

E PROOF OF THE RELATIVE-FLOOR MODULUS

E.1 UNIFORM UPPER BOUND

At a state T, put $w_T = \mu(T)$ and define the relative floor of a nonzero scalar profile by

$$\rho_T(f) = w_T \min_{x \in T} f_x / M_T(f). \tag{E1}$$

We prove for every $h \geq 1$ and state with $d(T) \geq h$ that

$$Q_{G,T}(f) \leq C_{T,h} a^{-\beta_h} Q_{L,T}(f), \qquad \beta_h = (2^h - 1)^{-1}, \qquad \rho_T(f) \geq a, \tag{E2}$$

uniformly for $0 < a \leq 1$. Normalize $M_T(f) = 1$. At $h = 1$, (A5) and $b_J \geq a/w_T$ give the result. At $h \geq 2$, the first and third cases of the induction in A.3 already give linear bounds. In the remaining case the child U has mass $m \leq C\sqrt{Q}$, and

$$\rho_U(f|_U) \geq (w_U/w_T)a/m. \tag{E3}$$

This is an admissible floor, since $m \geq (w_U/w_T)a$. With $\beta = \beta_{h-1}$, child induction and (A8) give

$$Q_{G,J}/Q \leq C a^{-\beta} Q^{\beta/2}. \tag{E4}$$

Equation (A9), with $n = 2^h$, gives the second estimate $Q_{G,J}/Q \leq C Q^{-1/n}$. Split at $Q = a^{n/(n-1)}$. Below this value, the increasing bound (E4) is at most $C a^{-1/(n-1)}$; above it, the decreasing bound has the same upper estimate. Take finite maxima over children and target cells. Entropy comparison and scalarization give $\Delta \leq C\tau^{-\beta_d}\delta$. Combining with Theorem 1 proves the upper part of (11).

E.2 BINARY LOWER WITNESSES AND THINNING

Choose a strictly positive terminal profile in the construction of A.4. Each atom has one positive monomial, with degree between zero and $r = 2^d - 1$, and the degree-zero shell is nonempty. After a fixed rescaling to posterior $\eta_t \leq 1/2$, choose constants $A, b, k, t_0 > 0$ such that, for $0 < t \leq t_0$,

$$\delta(\eta_t) \leq A t^{r+1}, \qquad G(\eta_t) \geq b t^r, \qquad \min_{x,y} \frac{P_t(y \mid x)}{P_t(y)} \geq k t^r. \tag{E5}$$

The first-label marginal stays away from both endpoints, and the complementary posterior is at least $1/2$. We fix these constants before choosing τ, u.

For $0 < \rho \leq 1$, replace η with $\rho\eta$. The first posterior-to-prior ratio stays unchanged. The second becomes a convex combination of its old value and one. Hence thinning preserves the floor. Convexity and homogeneity of the local scalar defect, with nonnegativity of the target defect, give

$$\delta(\rho\eta) \leq \rho\delta(\eta), \qquad \Delta(\rho\eta) \geq \rho G(\eta). \tag{E6}$$

The learner does not observe the thinning coin. The two total information losses therefore need not scale by the same factor.

Choose fixed $a_0, b_0 > 0$ with $k a_0^r \geq 1$, let $t = \max\{a_0\tau^{1/r}, b_0 u^{1/(r+1)}\}$, and let $\rho = \min\{1, u/\delta(\eta_t)\}$. For sufficiently small common thresholds on τ, u, this task satisfies the floor and local budget. Equations (E5)–(E6) imply

$$\Delta(\rho\eta_t) \geq b \min\{t^r, u/(At)\} \geq c \min\{u^{r/(r+1)}, \tau^{-1/r} u\}. \tag{E7}$$

For τ in a fixed interval away from zero and up to $1/2$, use a small fixed interior binary perturbation with positive local and target defects and floor at least $1/2$. Thin this task and reduce the common

u_0 as needed. This gives the remaining linear lower order uniformly. Binary tasks therefore attain the budget orders. Dividing by the budget on the linear branch gives the lower bound for C_τ; E.1 gives its upper bound.

F THE FINITE-SAMPLE LOG-LOSS COROLLARY

Let p be a multinomial law on m labels. From k samples with counts A_y, form $q_y = (A_y + 1)/(k + m)$. Using $\log t \le t - 1$ gives $D(p\|q) \le \sum_y p_y^2/q_y - 1$. For p_y positive, use the binomial identity

$$\mathbb{E}\frac{1}{A_y + 1} = \frac{1 - (1 - p_y)^{k+1}}{(k + 1)p_y} \le \frac{1}{(k + 1)p_y}. \tag{F1}$$

Zero-probability coordinates contribute zero, so $\mathbb{E}D(p\|q) \le (m - 1)/(k + 1)$. Condition on the feature assignments and apply this estimate inside each compressed cell z. For $w_z = P(Z = z) > 0$, we have $N_z \sim \mathrm{Binomial}(n, w_z)$, and (F1) gives

$$w_z\mathbb{E}\frac{1}{N_z + 1} = \frac{1 - (1 - w_z)^{n+1}}{n + 1}. \tag{F2}$$

Using the decomposition $R_{\log}(q) = H(Y \mid Z) + \sum_z w_z D(P(Y \mid z)\|q(\cdot \mid z))$ and combining the two bounds gives the slightly sharper form

$$\mathbb{E}R_{\log}(\widehat{q}) - H(Y \mid X) \le \Delta + \frac{m - 1}{n + 1}\sum_{z:w_z>0}[1 - (1 - w_z)^{n+1}], \tag{F3}$$

which implies (13). Nonnegativity of the KL terms gives the structural lower bound for every decoder based only on Z. The argument allows random training data and $n = 0$. The encoders must be fixed independently of the data.

G SHARP SOURCE-PERTURBATION RESULTS

G.1 ALPHABET-UNIFORM CONDITIONAL-ENTROPY CONTINUITY

If two joint laws of finite A and arbitrary finite B have total variation at most $e \le 1/2$, then

$$|H_P(A \mid B) - H_Q(A \mid B)| \le e \log |A| + h(e). \tag{G1}$$

Write the laws as $(1-e)R+eP'$ and $(1-e)R+eQ'$, and introduce a mixture-component indicator. Adding this indicator to the conditioning changes each conditional entropy by a conditional mutual information in $[0, h(e)]$. The residual conditional entropies differ by at most $\log |A|$. Subtraction gives the bound. A common stochastic kernel can only decrease total variation.

To avoid a label-count factor, apply (G1) to the raw inputs: use $I(Y; X_i \mid Z_i) = H(X_i \mid Z_i) - H(X_i \mid Y, Z_i)$ and the analogous identity for X. Evaluating one task kernel under μ_0 and ν on the fixed ambient alphabet gives

$$|\delta_\nu - \delta_{\mu_0}| \le E(e) := 2e \log 30 + 4h(e), \qquad |\Delta_\nu - \Delta_{\mu_0}| \le B(e) := 2e \log 30 + 2h(e). \tag{G2}$$

The first expression uses raw alphabet sizes five and six, whose log sizes sum to $\log 30$.

G.2 THE ZERO-LOCAL-LOSS UPPER BOUNDARY

Fix an exactly composing source μ with support S in a finite ambient alphabet of size N. Exactness gives a finite matrix H satisfying $B_\mu = HA_\mu$. For a nominal-positive conditioning cell C with $\nu(C) > 0$,

$$\|\nu(\cdot \mid C) - \mu(\cdot \mid C)\|_1 \le 2e/\mu(C). \tag{G3}$$

First normalize both restricted measures by $\mu(C)$. The numerator is bounded by their restricted ℓ_1 difference plus the absolute difference of their total masses. Together these equal twice the larger positive or negative discrepancy in C, at most $2e$.

Set $f_y(x) = P(Y = y \mid X = x)$. At $\delta_\nu = 0$, fine and coarse label posteriors coincide under ν. Apply the common kernel to (G3) on D_ℓ, C_ℓ to get

$$\sum_y |(A_\mu f_y)_\ell| \le 2e \left[\mu(D_\ell)^{-1} + \mu(C_\ell)^{-1}\right]. \tag{G4}$$

The row identity bounds the total variation between $Q(x, y) = \mu_x f_y(x)$ and $\bar{Q}(x, y) = \mu_x \langle f_y \rangle_{J(x)}$ by $t \le K_\mu e$, where

$$K_\mu = \sum_x \mu_x \sum_\ell |H_{x\ell}|[\mu(D_\ell)^{-1} + \mu(C_\ell)^{-1}]. \tag{G5}$$

The two laws have identical XZ and YZ marginals. Under $\bar{Q}$, X and Y are conditionally independent given Z. Thus $\Delta_\mu = H_{\bar{Q}}(X \mid Y, Z) - H_Q(X \mid Y, Z)$. For small enough e with $K_\mu e \le 1/2$, use (G1) and source transfer:

$$\Delta_\nu \le K_\mu e \log |S| + h(K_\mu e) + 2e \log N + 2h(e). \tag{G6}$$

The right side is increasing on $[0, 1/2]$, which justifies substituting the upper bound for t. If K_μ is zero, so is its entropy term. For the eight-atom source, the rational certificate gives $K_{\mu_0} = 34/3$. Taking $e \le 1/28$ keeps the nominal local cells positive and the entropy argument in range. The resulting bound is $O(e \log(1/e))$ without a posterior floor.

G.3 The matching rare-label construction

Put mass $1/2$ on each missing cross-pair (L0, Q1) and (L1, Q0) to define q, and set $\nu_e = (1-e)\mu_0 + eq$. Its TV distance from μ_0 is e. Let T indicate a new atom. Inside first-view cell L, the nominal masses of L0 and L1 are both $11/28$, and q splits them equally. Inside second-view cell Q, Q0 and Q1 have nominal mass $1/4$ each and receive equal mass under q. Therefore $I(T; X_i \mid Z_i) = 0$ for each view.

Apply channel (19) to T. The label marginal is $P(Y = 1) = e$. On old atoms the label-one posterior is exactly τe; on new atoms the label-zero posterior is exactly $\tau(1 - e)$. The other two floor inequalities follow directly. Conditional data processing gives $\delta_{\nu_e} = 0$.

All joint loss lies in the L,Q coarse cell. Put

$$a = (1 + e)/2, \qquad r = 2e/(1 + e), \qquad p_J = \tau e + (1 - \tau)r.$$

The exact joint loss is

$$D_\tau(e) = a\, h(p_J) - \frac{1 - e}{2} h(\tau e) - e\, h(1 - \tau + \tau e). \tag{G7}$$

For fixed $0 \le \tau < 1$, expand $h(ce)$ to obtain $D_\tau(e) = (1 - \tau)e \log(1/e) + O_\tau(e)$. At $\tau = 0$, use $h(0) = 0$ directly. With (G6), this proves the zero-budget boundary. We also need the finite lower bound

$$D_\tau(e) \ge e \left[(1 - \tau) \log \frac{1}{2e} - \log 2\right], \qquad 0 < e \le 1/2. \tag{G8}$$

Keep only the nonnegative new-atom term $eD_{\mathrm{Ber}}(1-\tau+\tau e\|p_J)$. For Bernoulli parameters v, p, use $D_{\mathrm{Ber}}(v\|p) \geq v\log(1/p) - \log 2$. In this construction $p_J \leq 2e$ and $v \geq 1-\tau$.

G.4 Fixed-positive-floor joint upper bound

Transport the floor to the nominal source. For $e \leq 1/28$, set $c = 1 - 14e \geq 1/2$. The TV bound on each singleton gives

$$\nu_x \geq \mu_{0,x} - e \geq c\mu_{0,x}. \tag{G9}$$

Hence $\nu \geq c\mu_0$ as measures on the ambient alphabet. Applying any nonnegative task profile gives $p_y^\nu \geq cp_y^{\mu_0}$. Each nominal atom stays ν-positive. A floor τ under ν gives floor $\sigma = \tau c$ under μ_0, including arbitrarily small label priors. Use the coefficient from Appendix D to obtain

$$\Delta_\nu \leq \frac{49}{\tau(1-14e)}[u + E(e)] + B(e). \tag{G10}$$

For each fixed positive floor, this is $O_\tau(u + e\log(1/e))$ uniformly over finite task alphabets. The argument uses nominal exactness, its finite quadratic-form coefficient, and the transported floor.

For a lower bound proportional to u, keep $\nu = \mu_0$ and set the first binary posterior to $1/2 + t$ at a_1, $1/2 - t$ at a_2, and $1/2$ elsewhere. The prior is balanced, and every fixed floor $\tau < 1$ holds for sufficiently small t. Direct conditioning gives

$$\delta(t) = \frac{2}{7}[\log 2 - h(1/2 + t/2)] \sim t^2/7, \qquad \Delta(t) = \frac{1}{7}[\log 2 - h(1/2+t)] \sim 2t^2/7. \tag{G11}$$

Continuity and local monotonicity let us choose a witness at every small positive budget. The contamination family supplies the source-error term. Each witness lies in the same budget rectangle, so the supremum bounds their maximum and hence half their sum. This also covers the axes, with $e\log(1/e) = 0$ at $e = 0$.

G.5 Optional weighted upper certificate

We can also retain an $O(e)$ nominal perturbation term before applying source continuity. Put $p_y = P_\nu(Y = y) > 0$ and $\|v\|_p^2 = \sum_y v_y^2/p_y$. For a ν-positive cell C, the posterior satisfies $P_\nu(y \mid C) \leq p_y/\nu(C)$. KL curvature then gives, for a fine/coarse pair,

$$\|P_\nu(Y \mid D) - P_\nu(Y \mid C)\|_p \leq \frac{\sqrt{2u}}{\nu(D)} \leq \frac{\sqrt{8u}}{\mu_0(D)} \tag{G12}$$

when $e \leq 1/28$. The conditional-source difference $d_x = \mu_0(x \mid C) - \nu(x \mid C)$ satisfies $|d_x| \leq 2\nu_x/\mu_0(C)$ by nominal support retention. Weighted Cauchy–Schwarz and (G3) give

$$\|P_{\mu_0}(Y \mid C) - P_\nu(Y \mid C)\|_p^2 \leq \sum_{\nu_x > 0} \frac{d_x^2}{\nu_x} \leq \frac{4e}{\mu_0(C)^2}. \tag{G13}$$

Omit zero ν_x, for which $d_x = 0$. Set $T_x = \sum_\ell |H_{x\ell}|/\mu_0(D_\ell)$ and $W_x = \sum_\ell |H_{x\ell}|[\mu_0(D_\ell)^{-1} + \mu_0(C_\ell)^{-1}]$. Equations (G12)–(G13), together with $B = HA$, give

$$\|f(x) - \langle f\rangle_{J(x)}\|_p \leq \sqrt{8u}\, T_x + 2\sqrt{e}\, W_x. \tag{G14}$$

The pointwise ν-floor applies on all nominal atoms, so each nominal target posterior is at least τp_y. Apply KL$\leq \chi^2$ and sum:

$$\Delta_\nu \le \frac{16\mathsf{A}}{\tau}u + \frac{8\mathsf{B}}{\tau}e + B(e), \quad \mathsf{A} = \sum_x \mu_{0,x}T_x^2, \quad \mathsf{B} = \sum_x \mu_{0,x}W_x^2. \tag{G15}$$

In source-table order, the stored certificate gives $T = (28, 28, 14, 14, 0, 0, 4, 4)$ and $W = (119/3, 119/3, 56/3, 56/3, 0, 0, 6, 6)$. Thus $\mathsf{A} = 148$ and $\mathsf{B} = 2633/9$, giving $2368u/\tau + 21064e/(9\tau) + B(e)$. These conservative constants do not establish the optimal dependence on τ.

H SOURCE-ONLY CERTIFICATE OBSTRUCTION

Under ν_e, a sample of n raw observations contains no new atom with probability $(1-e)^n$. Given that event, its law is $\mu_0^{\otimes n}$. Independent algorithmic randomness preserves the identity. Since $F_\tau(\nu_e) \ge D_\tau(e)$, uniform validity in (21) requires

$$\beta \ge P_{\nu_e}\{C_n < D_\tau(e)\} \ge (1-e)^n P_{\mu_0}\{C_n < D_\tau(e)\}. \tag{H1}$$

Rearranging gives $P_{\mu_0}\{C_n \ge D_\tau(e)\} \ge 1 - \beta/(1-e)^n$. Take $e = 1/(2n)$. Then Bernoulli's inequality gives $(1-e)^n \ge 1/2$, and (G8) gives the threshold in (22). Since C_n is nonnegative, we can take the positive part of that threshold. If the source class is a small neighborhood, choose n large enough for the contamination to stay inside it. This proves the source-only lower bound. A matching achievable sample complexity remains open.

I PROOF AND REPRODUCIBILITY MAP

The following paths are relative to the supplement root. They separate proofs, exact structural checks, and numerical tests of the stated formulas.

Claim	Proof	Executable evidence
Classification and shortest failure depth	A–C	`audit_central/independent_depth_audit.py`; `algorithm/`
Rational source certificates	B, M	`audit_central/replay_source_certificates.py`; `audit_central/compare_new_inputs.py`
Sharp witness orders	A, D	`audit_central/independent_witness_orders.py`
Posterior-floor modulus	E	`validation/`; `robust_extension/check_vanishing_floor.py`
Finite-sample decoder bound	F	Analytical proof
Source continuity and certificate lower bound	G–H	`robustness/`; `validation/check_source_stability_independent.py`
General source stability	J, O	`robust_extension/check_vanishing_floor.py`; `robustness/check_nofloor_robust.py`
Finite-label and prior-theorem reductions	K	Analytical reductions and verified bibliography
Original decoder study	L	`operational/`
Smoothing and learned pooling	P	`learning_extension/`

Commands for quick verification and full reproduction are in `README.md`. The audit files identify symbolic arguments, exact finite checks, deterministic numerical sums, and Monte Carlo intervals separately. Personal correspondence and identifying development records are excluded from the anonymous supplement.

J SOURCE STABILITY WITHOUT A POSTERIOR FLOOR

We use the sharp nominal exponent to bound source perturbations. Throughout the proof the nominal weights stay fixed. Logarithms are natural, and $h(e) = -e \log e - (1-e)\log(1-e)$.

Theorem J.1 (source stability). Let μ be a fixed finite source with support S, minimum positive mass m, and fixed deterministic encoders satisfying exact composition. Let d be its whole-cell failure depth, and put $\alpha = 1 - 2^{-d}$ for $1 \le d < \infty$, and $\alpha = 1$ for $d = \infty$. Extend the raw variables and encoders to a fixed finite ambient alphabet $\Omega \supseteq S$, with $N_{\mathrm{amb}} = |\Omega|$. There is a finite constant C_μ such that every source ν and every finite task kernel satisfy

$$\Delta_\nu \le C_\mu \delta_\nu^\alpha + C_\mu e + 2e \log N_{\mathrm{amb}} + 2h(e), \qquad \mathrm{TV}(\nu, \mu) \le e \le \min\{m/2, 1/2\}. \tag{J1}$$

The constant depends only on the fixed nominal system. It is uniform over task alphabets and label priors, with no posterior floor. It may grow as positive nominal masses approach zero.

J.1 A DETERMINISTIC EXTENSION OF THE WHOLE-CELL INDUCTION

A nominal state T is a union of target cells. It retains precisely the original local coarse cells wholly contained in T, with their original fine cells and weights. For $f \ge 0$ on T, write

$$M_T(f) = \sum_{x \in T} \mu_x f_x, \qquad A_\ell f = \langle f \rangle_{D_\ell} - \langle f \rangle_{C_\ell}, \qquad c_\ell(f) = \langle f \rangle_{C_\ell}.$$

For a target cell $J \subseteq T$, let $b_J = \langle f \rangle_J$ and

$$Q_{G,J}(f) = \sum_{x \in J} \mu_x \frac{(f_x - b_J)^2}{b_J}, \qquad Q_{G,T} = \sum_{J \subseteq T} Q_{G,J}. \tag{J2}$$

If $b_J = 0$, all values on J vanish by nonnegativity, and we set the contribution to zero. At an exact state T, the target rows lie in the row span of A_T. Every nonnegative vector in $\ker A_T$ is therefore constant on each target cell. Its proper nonempty zero sets are legal children under the same whole-cell rule.

The depth $d(T)$ counts the shortest child path to a failed exactness test and is infinite if no such path exists. If $d(T) \ge k \ge 1$, the state T is exact and each child has depth at least $k-1$. Every step removes a target cell. There are therefore finitely many possible states, rows, and cells over which to choose constants, even when failure is unreachable.

Lemma J.2 (noisy nominal rows). Fix $K < \infty$. Suppose $f \ge 0$ and $a, \ell \ge 0$ obey

$$|A_j f| \le K \left[\sqrt{(c_j(f) + a)\ell} + a \right] \tag{J3}$$

for every retained row. If $d(T) \ge k \ge 1$, then

$$Q_{G,T}(f) \le C_{T,k,K} (M_T(f) + a)^{2^{-k}} \ell^{1 - 2^{-k}} + C_{T,k,K} a. \tag{J4}$$

If $d(T) = \infty$, the bound is $Q_{G,T}(f) \le C_{T,K}(\ell + a)$.

Proof. Write $M = M_T(f)$ and $V = M + a$. Positive finite source weights give $c_j \le C_j M$, $Q_{G,J} \le C_J b_J$, and $Q_{G,T} \le CM$. Constants below depend on these fixed systems. By (J3),

$$\|A_T f\| \le Cs, \qquad s = \sqrt{V\ell} + a.$$

At an exact state, the target row representations therefore yield

$$Q_{G,J}(f) \le \min\left\{Cb_J, \frac{C(V\ell + a^2)}{b_J}\right\}. \tag{J5}$$

The minimum is bounded by $C\sqrt{V\ell} + Ca$, giving (J4) for $k = 1$. Zero target means contribute nothing, and exactness handles $a = \ell = 0$. When $\ell > V$, the mass bound already gives every fractional estimate. We can restrict attention to $\ell \le V$, where

$$\ell \le V^{1-\gamma}\ell^{\gamma} \quad \text{for } 0 < \gamma \le 1. \tag{J6}$$

Maximize $M_T(q)$ over $0 \le q \le f$ and $A_T q = 0$, choose a maximizer q, and set $r = f - q$. The feasible set is nonempty and compact. A nonzero nonnegative kernel vector on the support of r would allow a small multiple to be added to q, violating maximality. On any such support, the nonnegative mass-one simplex has positive distance from $\ker A_T$. Take the largest reciprocal bound over the finitely many supports:

$$M_T(r) \le C\|A_T r\| = C\|A_T f\| \le Cs.$$

Normalize the nonnegative kernel cone by nominal mass to obtain a polytope with finitely many vertices. Fix J and decompose q into q_0 from vertices vanishing on J and q_+ from vertices positive there. Each vertex is constant on J by exactness. The finite set of positive values has a positive minimum. Hence $M_T(q_+) \le Cb_J$ and

$$f = q_0 + g, \qquad g = q_+ + r \ge 0, \qquad g|_J = f|_J, \qquad M_T(g) \le C(b_J + s). \tag{J7}$$

If $q_0 \ne 0$, its zero set U is a proper nonempty legal child containing J, with $f|_U = g|_U$. A retained row in U has exactly the same weights as its original row, so (J3) continues to hold with the same a, ℓ. The normalization is unchanged.

Suppose now $k \ge 2$, and write $b = b_J > 0$. If $b \ge s$, the applicable state is U when $q_0 \ne 0$, and T otherwise. It is exact and has mass plus a at most Cb by (J7). Applying (J5) there gives $Q_{G,J} \le C(\ell + a^2/b) \le C(\ell + a)$. Equation (J6) gives the desired power.

If $b < s$ and $q_0 \ne 0$, then $d(U) \ge k - 1$ and $M_U + a \le Cs$. With $\beta = 2^{-(k-1)}$, induction gives

$$\begin{aligned} Q_{G,J}(f) &\le Cs^{\beta}\ell^{1-\beta} + Ca \\ &\le CV^{\beta/2}\ell^{1-\beta/2} + Ca^{\beta}\ell^{1-\beta} + Ca \\ &\le CV^{2^{-k}}\ell^{1-2^{-k}} + C(\ell + a). \end{aligned} \tag{J8}$$

Use subadditivity for $\beta \le 1$ and weighted arithmetic–geometric mean. Equation (J6) absorbs the ℓ term.

For $b < s$ and $q_0 = 0$, (J7) gives $M \le Cs$ and hence $V \le C'(\sqrt{V\ell} + a)$. Apply Young's inequality and absorb half of V to get $V \le C''(\ell + a)$. The target mass bound and (J6) finish the estimate. Sum over target cells and maximize over the finite collection of children.

For infinite depth, induct on target-cell count. Each proper child has fewer cells and still has infinite depth. The first and third cases give $C(\ell + a)$ directly. The second uses the child's linear estimate. A one-cell state has no nonzero q_0 vanishing on that cell, so the first and third cases form the base. Zero target rows contribute zero. The induction uses finitely many states. □

J.2 Perturbed sources satisfy the noisy-row hypothesis

Extend μ by zero outside S. For one nonnegative profile f on Ω, define

$$M = \sum_S \mu_x f_x, \qquad D_f = \sum_{\Omega} |\nu_x - \mu_x| f_x, \qquad a = D_f + eM, \qquad \ell = L_\nu(f). \tag{J9}$$

Here L_ν sums the local scalar entropy gaps under ν and retains the view index. A fine/coarse pair contributes $\nu(D)[v\log(v/c) - v + c]$, with $v = \langle f\rangle_{\nu,D}$ and $c = \langle f\rangle_{\nu,C}$. Each term is nonnegative. The linear terms cancel inside each coarse cell.

The condition $e \le m/2$ gives $\nu_x \ge \mu_x - e \ge \mu_x/2$ on each original atom. Every nominal-positive fine or coarse cell E then satisfies $\nu(E) \ge \mu(E)/2 \ge m/2$. For ν, take E to be the full ambient fiber, including new atoms. Its nominal mass and mean are the same as on the original support. Subtracting the conditional means gives

$$\begin{aligned} |\langle f\rangle_{\nu,E} - \langle f\rangle_{\mu,E}| &\le \frac{D_f}{\nu(E)} + \frac{e\langle f\rangle_{\mu,E}}{\nu(E)} \\ &\le \frac{2D_f}{m} + \frac{2eM}{m^2} \le \frac{2a}{m^2}. \end{aligned} \tag{J10}$$

We used $|\nu(E) - \mu(E)| \le e$, $\langle f\rangle_{\mu,E} \le M/\mu(E)$, and $m \le 1$.

For a retained fine/coarse pair, $v \le Rc$ with $R = \nu(C)/\nu(D) \le 2/m$. The curvature of $t\log t$ gives

$$v\log(v/c) - v + c \ge \frac{(v-c)^2}{2Rc}.$$

Multiplying by $\nu(D) \ge m/2$ and bounding this contribution by ℓ yields

$$|v - c| \le \sqrt{8c\ell/m^2}. \tag{J11}$$

If $c = 0$, then $v = 0$ and the row is zero. Combining (J10) for both means, (J11), and $c \le c_\mu + 2a/m^2$ proves (J3) with a fixed constant depending only on m.

For every retained child row, repeat the argument with its unchanged nominal means. Keep a equal to the root budget in (J9). Extra fine cells under ν add nonnegative terms to $L_\nu(f)$. We use the nominal row-span identities throughout; the perturbed source ν may fail exact composition.

Nominal entropy is bounded above by its quadratic perspective, $G_\mu(f) \le Q_{G,S}(f)$, using $\log t \le t - 1$. Lemma J.2 therefore gives

$$G_\mu(f) \le C(M + a)^{1-\alpha} L_\nu(f)^\alpha + Ca \tag{J12}$$

when $\alpha < 1$. At infinite depth it gives the direct linear bound $C[L_\nu(f) + a]$.

J.3 Summing labels and transferring the target loss

Apply (J12) to the actual unnormalized profiles $f_y(x) = P(Y = y \mid x)$. Stochastic normalization implies

$$\sum_y M_y = 1, \qquad \sum_y a_y = \|\nu - \mu\|_1 + e \le 3e, \qquad \sum_y L_\nu(f_y) = \delta_\nu.$$

For $\alpha < 1$, Hölder's inequality gives

$$\Delta_\mu \le C(1 + 3e)^{1-\alpha}\delta_\nu^\alpha + 3Ce. \tag{J13}$$

For $\alpha = 1$, direct summation gives $C\delta_\nu + 3Ce$. This also covers zero label priors without a zero-power convention.

Expand each conditional mutual information into one scalar entropy gap per label, using the same kernel under both sources. This avoids division by label probabilities. A label appearing only on new atoms has zero nominal mass and may have a positive error allowance. Even then, all allowances sum to at most $3e$. The coefficient therefore stays independent of label count.

For two laws of finite X and arbitrary finite W at TV distance t, the common-part argument gives $|H_P(X \mid W) - H_Q(X \mid W)| \le t \log N_{\mathrm{amb}} + h(t)$. Decompose each law into the shared component of mass $1-t$ and a residual of mass t. Adding the component indicator to the conditioning changes conditional entropy by a number in $[0, h(t)]$. Residual conditional entropies lie in $[0, \log N_{\mathrm{amb}}]$. Subtract to obtain the bound. The same kernel and deterministic encoders preserve TV. Apply the estimate twice to $\Delta = H(X \mid Z) - H(X \mid Y, Z)$ with $t \le e \le 1/2$. Combining it with (J13) and absorbing the bounded prefactor proves (J1).

J.4 Sharp joint order for the eight-atom source

For the eight-atom source μ_0 and its fixed thirty-atom ambient alphabet, define

$$R_{\mu_0,0}(u, e) = \sup\{\Delta_\nu : \mathrm{TV}(\nu, \mu_0) \le e,\ \delta_\nu \le u\},$$

where the supremum includes all finite task alphabets. Theorem J.1, $d = 2$, and $m = 1/14$ give the upper order $u^{3/4} + e \log(1/e)$ for sufficiently small budgets.

For the nominal lower term, use the balanced binary posterior $\eta_t = (t^3, 0, t^2, t^2, t^2, t^2, v_t, v_t)$ with $v_t = 1 - (5+t)t^2/7$ in the Section 5 atom order. Appendix D gives $\delta(t) \asymp t^4$ and $\Delta(t) \asymp t^3$. Choose t using an upper bound for $\delta(t)/t^4$. This gives $\Delta \ge c u^{3/4}$ for every small u and requires no monotonicity assumption.

For the source term, let q place mass $1/2$ on each missing cross-pair $(\mathrm{L0}, \mathrm{Q1})$ and $(\mathrm{L1}, \mathrm{Q0})$, take $\nu_e = (1-e)\mu_0 + eq$, and let Y indicate a new atom. The equal marginal splits within the L and Q coarse cells give $\delta_{\nu_e} = 0$. Only the joint L,Q cell is ambiguous, so

$$\Delta_{\nu_e} = \frac{1+e}{2} h\left(\frac{2e}{1+e}\right) = e \log(1/e) + O(e).$$

Both witnesses lie in the same budget rectangle. Their maximum is a lower bound for the supremum and is at least a fixed multiple of their sum. Hence

$$R_{\mu_0,0}(u, e) \asymp u^{3/4} + e \log(1/e) \tag{J14}$$

jointly near $(0, 0)$ on both axes, using the zero-value convention. The general theorem supplies the upper bound. For another source, a matching source-error lower bound requires its own contamination construction. Optimal constants are not determined here. Appendix O treats uniformity in a vanishing floor separately.

J.5 An explicit certificate for the eight-atom source

For the source in Section 5, the source-stability coefficient can be made explicit. For $0 \le e \le 1/28$, every task with $\delta_\nu \le u$ and $\mathrm{TV}(\nu, \mu_0) \le e$ satisfies

$$\Delta_\nu \le \min\left\{\log 6,\ 741696 u^{3/4} + 23525 e + 2e \log 30 + 2h(e)\right\}. \tag{J15}$$

The constants in this certificate are conservative. The cap uses $\Delta_\nu \le H_\nu(X \mid Z)$ and the ambient joint-cell sizes: there are $2 \times 2 = 4$ raw pairs in each L-prefix fiber and $3 \times 2 = 6$ in each M-prefix fiber.

For the coefficient calculation, take a scalar profile with $M = M_{\mu_0}(f)$, $a = \sum_x |\nu_x - \mu_{0,x}| f_x + eM$, $V = M + a$, and $Q = Q_\nu(f)$. Root row identities bound within-U and within-E pair differences by coefficients with ℓ_1-norm at most $79/5 < 16$. Child target-row coefficients are at most 3. The conditional-mean perturbation estimates then bound the residual mass by $448\sqrt{VQ} + 12544a$. For a nontrivial target J with mean b, the child mass is $1/2$, giving

$$Q_{G,J} \le \mu_0(J) \min\left\{b,\ 7056Q + \frac{14112(448\sqrt{V}Q^{3/2} + 12545aQ + 784a^2)}{b}\right\}.$$

Terms with $b = 0$ vanish. Using $\min\{b, v + w/b\} \le v + \sqrt{w}$, the integer square-root upper bounds $2515, 13306, 3327$ for the three terms, $\sqrt{aQ} \le (a + Q)/2$, and $\sum_J \mu_0(J) = 11/14$, gives

$$G_{\mu_0}(f) \le \frac{11}{14}\left(13709Q + 2515V^{1/4}Q^{3/4} + 9980a\right).$$

Since $Q \le 54V$, $Q \le 56L_\nu(f)$, $54^{1/4} \le 3$, and $56^{3/4} \le 21$, the coefficient of $V^{1/4}L_\nu(f)^{3/4}$ is at most $(11/14)43642 \cdot 21$. Summing label profiles uses $\sum_y V_y \le 31/28$, $(31/28)^{1/4} \le 103/100$, and $\sum_y a_y \le 3e$. This gives the exact rational coefficients $74169579/100$ and $164670/7$ before upward rounding. Target-information continuity completes the proof.

The evaluator takes a supplied population local-loss budget and a supplied valid source-TV radius. It returns a floating-point approximation to the bound. The evaluator does not estimate its inputs or certify roundoff.

K PRIOR-THEOREM REDUCTIONS AND FINITE TASK ALPHABETS

This appendix works through the links to prior theory. Semialgebraic error bounds supply an attained rational exponent, a binary embedding realizes it, and entropy duality gives a functional-inequality formulation. We state the required hypotheses before comparing these conclusions with the exact classification and source-stability bounds.

K.1 GENERIC SEMIALGEBRAIC EXISTENCE AND ENDPOINT ATTAINMENT

Fix the positive source on S, let $N = |S|$ and $m = \min_x \mu_x$, and normalize profiles on

$$\mathcal{K} = \{f \ge 0 : M(f) = \textstyle\sum_x \mu_x f_x = 1\}. \tag{K1}$$

This compact simplex satisfies $f_x \le 1/m$. Each local perspective summand in Q_L has form $w_\ell(a_\ell f)^2/c_\ell(f)$, with coarse-cell mean c_ℓ. Fixed positive cell weights give $|a_\ell f| \le C_\ell c_\ell(f)$. The quotient therefore tends to zero when its denominator tends to zero. The target terms have the same property. The zero extensions of Q_L, Q_G are consequently continuous, piecewise rational, and semialgebraic.

At an exact root, $Q_L^{-1}(0) \subseteq Q_G^{-1}(0)$ on $\mathcal{K}$. Apply Basu & Mohammad-Nezhad (2024, Theorem 1, arXiv version 2) to obtain $r > 0, C < \infty$ with $Q_G^r \le CQ_L$. Let c, j count distinct local coarse-cell and target-cell denominators. On each stratum, multiply by the active denominators to describe the graph with degree at most their number plus one. Thus degree $D = \max\{2, c + 1, j + 1\}$ suffices, and the theorem allows $r \le (8D)^{2(N+7)}$.

For a nontrivial target, define

$$\Psi(u) = \max\{Q_G(f) : f \in \mathcal{K},\ Q_L(f) \le u\}. \tag{K2}$$

Compactness gives the maximum, and quantifier elimination gives a semialgebraic graph. Exactness implies $\Psi(u) \to 0$ as $u \downarrow 0$: otherwise, a convergent subsequence of maximizers would have zero local loss and positive target loss. Choose a mean-zero v with $Bv \ne 0$. Along $1 + tv$, both perspective losses are positive and quadratic, since exactness forces $Av \ne 0$. Hence $\Psi(u) > 0$ for small positive u.

The one-variable Puiseux expansion gives

$$\Psi(u) = au^\alpha + o(u^\alpha), \qquad a > 0, \quad \alpha \in \mathbb{Q}_{>0}. \tag{K3}$$

The interior perturbation implies $\Psi(u) \ge c_0 u$. The row-span and perspective bound gives $\Psi(u) \le C_0\sqrt{u}$, so $1/2 \le \alpha \le 1$. Equation (K3) attains the endpoint and rules out larger exponents. Compactness extends the upper bound away from zero. If $Q_G \equiv 0$, every exponent is valid and there is no largest one. The whole-cell classification determines the exact exponent in the nonvacuous case.

K.2 BINARY TASKS REALIZE THE GENERIC EXPONENT

Set $s = m/2$. For $f \in \mathcal{K}$, use

$$P_f(Y = 1 \mid x) = sf_x, \qquad P_f(Y = 0 \mid x) = 1 - sf_x. \tag{K4}$$

The channel has fixed label-one marginal s; its complementary profile is in $[1/2, 1]$. From Appendix A, $Q_L/(2R_L) \le L \le Q_L$ and $Q_G/(2R_G) \le G \le Q_G$ with source-dependent constants. The first label contributes $sL(f), sG(f)$. On the complementary interval, the curvature of $z \log z$ is at most 2. Its local gap is at most $s^2 \sum_\ell w_\ell (a_\ell f)^2 \le (s^2/m) Q_L(f)$, with the same argument for the target. Thus

$$\begin{aligned} \frac{s}{2R_L} Q_L(f) \le \delta(P_f) \le (s + s^2/m) Q_L(f), \\ \frac{s}{2R_G} Q_G(f) \le \Delta(P_f) \le (s + s^2/m) Q_G(f). \end{aligned} \tag{K5}$$

Rescale the budget by the upper constant and use maximizers in (K2) to realize α with binary tasks. The prior can be unbalanced. This argument does not provide efficient maximizing formulas or coefficient bounds.

For an arbitrary finite task, set $p_y = P(Y = y) > 0$ and $f_y = P(y \mid x)/p_y \in \mathcal{K}$. Homogeneity, perspective comparison, and concavity give

$$\Delta \le C \sum_y p_y Q_L(f_y)^\alpha \le C \left(\sum_y p_y Q_L(f_y) \right)^\alpha \le C' \delta^\alpha. \tag{K6}$$

Hence the optimal exponents for binary and arbitrary finite tasks coincide. Exact binary optimality at a fixed budget and balanced-prior sharpness need separate arguments.

K.3 AN EXACT COMPATIBLE-MARGINAL ENTROPY-DUALITY FORMULATION

Let F_i, C_i, T denote the fine, local coarse, and joint compressed partitions. For $f \in \mathcal{K}$, put $P_x = \mu_x f_x$. Then

$$G(f) = D(P\|\mu) - D(P_T\|\mu_T), \qquad L(f) = \sum_i [D(P_{F_i}\|\mu_{F_i}) - D(P_{C_i}\|\mu_{C_i})]. \tag{K7}$$

Apply Liu et al. (2018, Theorem 6, equations (55)–(56)). For $C > 0$, use reverse maps $(T, F_1, \dots, F_v)$ with weights $(1, C, \dots, C)$ and nominal marginal references. Use forward maps $(\mathrm{id}, C_1, \dots, C_v)$ with the same weights and corresponding references. Finite alphabets and positive image reference measures give the compatible-marginal hypothesis. The arbitrary-marginal hypothesis of their Theorem 1 is stronger and may fail here.

Exactness makes the compatible input unique. Write P_T, P_{F_i}, P_{C_i} for nominal conditional-expectation operators. If P, P' share the reverse marginals, then $r = (P - P')/\mu$ satisfies $P_T r = 0$ and $P_{F_i} r = 0$. Refinement gives $P_{C_i} r = 0$ and therefore $Ar = 0$. By exactness, $Br = 0$, so $r = P_T r = 0$. The compatible-input minimum reduces to

$$G(f) \le CL(f) + b \qquad (f \in \mathcal{K}). \tag{K8}$$

The optimal scalar slack $b(C) = \max_{\mathcal{K}}(G - CL)$ is finite for $C \ge 0$. Under exactness, compactness makes it tend to zero as $C \to \infty$. The classification determines whether zero slack is attained at finite C and, otherwise, its decay. Optimizing the finite-depth modulus gives $b(C) \asymp C^{-(2^d - 1)}$. Equality of scalar and all-task additive slacks remains unproved.

K.4 FINITE-LABEL ATTAINMENT, INCLUDING NEARBY SOURCES

For a label with positive marginal, put $q_y = P(X = \cdot \mid Y = y)$. Tasks are mixtures satisfying $\sum_y p_y q_y = \mu$; any such mixture gives $P(y \mid x) = p_y q_y(x)/\mu_x$. A relative floor is equivalent to $q_y(x) \ge \tau \mu_x$. Define

$$\begin{aligned} Q_\tau &= \{q \ge \tau\mu : \textstyle\sum_x q_x = 1\}, \\ \ell(q) &= \sum_i [D(q_{X_i}\|\mu_{X_i}) - D(q_{Z_i}\|\mu_{Z_i})], \\ g(q) &= D(q\|\mu) - D(q_Z\|\mu_Z). \end{aligned} \tag{K9}$$

Both functions are continuous on compact Q_τ, and $\delta = \sum_y p_y \ell(q_y)$ and $\Delta = \sum_y p_y g(q_y)$. Carathéodory's theorem applied to $(q_1, \dots, q_{N-1}, \ell(q), g(q)) \in \mathbb{R}^{N+1}$ preserves each attainable loss pair using at most $N+2$ labels. The convex hull is compact and convex, as is its loss region at barycenter μ.

A maximum at fixed budget exists. If a maximizing mixture has $N+2$ positive weights, the vectors $(1, q_1, \dots, q_{N-1}, \ell(q))$ are linearly dependent. Perturb the weights along a dependence. This preserves total weight, barycenter, and local loss. Both signs are initially feasible, so optimality forces zero global-loss derivative. Continue until one weight vanishes. Repeating gives at most $N+1$ labels, including under a floor; nine suffice for the nominal example. Compactness makes the budget modulus vanish under exactness. Its rate still needs a quantitative argument.

For the robust problem on 30 ambient atoms, 31 labels suffice for attainment. At each source the preceding argument uses its actual support size, at most 30. Fix 31 labels and optimize over p_{xy}. Source-TV and local-loss constraints are closed, as is

$$p_{xy} \geq \tau \left(\sum_{y'} p_{xy'} \right) \left(\sum_{x'} p_{x'y} \right). \tag{K10}$$

The feasible set is compact and conditional mutual informations are continuous at zero probabilities. This reduces the problem to finite optimization. An efficient global solver or a certificate for a sampled grid requires further work.

At fixed source, averaging a universal inequality

$$g(q) \leq \lambda \ell(q) + a \cdot q \quad (q \in Q_\tau), \qquad \lambda \geq 0,\ a \in \mathbb{R}^N, \tag{K11}$$

gives $F_\tau(u) \leq \lambda u + a \cdot \mu$. This weak-duality certificate does not require dual attainment. A posterior grid can suggest coefficients; (K11) requires verification over the whole domain. Mixing tasks and retaining the mixture indicator in the label preserves the floor and mixes both losses. The budget envelope is therefore nondecreasing and concave. These properties can be checked against numerical output.

K.5 Relation to conventional multiview assumptions

Define total correlation as $\mathrm{TC}(W) = \sum_i H(W_i) - H(W)$ and its conditional form through conditional entropies. Expanding gives

$$\Delta - \delta = [\mathrm{TC}(X \mid Y) - \mathrm{TC}(Z \mid Y)] - [\mathrm{TC}(X) - \mathrm{TC}(Z)]. \tag{K12}$$

Deterministic local maps decrease total correlation, so $\Delta \leq \delta + \mathrm{TC}(X \mid Y)$. For conditionally independent views, $\Delta \leq \delta$. A fractional worst-case regime therefore needs task-conditional dependence. This gives a baseline under the assumption in Huang et al. (2022, Section II).

Encoder functional dependencies do not by themselves turn fixed-source exactness into a universal independence implication. The linear-relaxation results of Kenig & Suciu (2022, Sections 4–6) and Kenig (2025, Theorems 4.2–4.4, arXiv version 1) need further universal hypotheses. If their added antecedents all vanished on a sharp fractional witness, the resulting finite linear coefficient would contradict that witness. This excludes that direct application. Other nonlinear and conic reductions remain unresolved.

K.6 Conic expansion and the cap-uniform quantitative statement

Conic error bounds relate rates to singularity (Sturm, 2000). In the quadratic-perspective lift, an anchored recession direction satisfies $q \geq 0$, $Aq = 0$, and $q|_J = 0$. Under Lourenço et al. (2016, Theorem 2 and Definition 3), a local constraint is removed exactly when its coarse mean on q is positive. This is the whole-cell restriction. The homogeneous feasible lift has partial-polyhedral Slater degree one; Nishijima et al. (2026, Corollary 3.4 and Theorem 3.19) therefore give the square-root baseline. Their feasibility hypothesis fails for the contrast-one zero-loss slice, which is infeasible.

To state the remaining rate requirement, fix finite depth $d \geq 1$, let $n = 2^d$, and take β as either sign of target row $f_x - b_J(f)$. Assign $+\infty$ to an empty infimum and define

$$m_{J,\beta,K}(\varepsilon) = \inf\{M(f) : f \geq 0,\ \beta(f) = 1,\ b_J(f) \leq K,\ Q_L(f) \leq \varepsilon\}. \tag{K13}$$

Up to source-dependent constants, $Q_G \le CM^{1/n}Q_L^{1-1/n}$ is equivalent to

$$m_{J,\beta,K}(\varepsilon) \ge cK^{-n}\varepsilon^{-(n-1)} \quad (K > 0,\ 0 < \varepsilon \le \varepsilon_0), \tag{K14}$$

uniformly over the finitely many rows and signs. Both $c, \varepsilon_0 > 0$ must be independent of K.

For the forward implication, use the target summand μ_x/K. Conversely, for $a = \beta(f) > 0$ apply (K14) to f/a, taking $K = b_J(f)/a$ and $\varepsilon = Q_L(f)/a$. If $\varepsilon \le \varepsilon_0$, obtain $a^{2n} \le c^{-1}M(f)b_J(f)^n Q_L(f)^{n-1}$ and the required bound on $a^2/b_J(f)$. Otherwise use positivity: $a \le Rb_J(f)$, hence $a^2/b_J(f) \le RQ_L(f)/\varepsilon_0$. Apply $Q_L \le C_0 M$ and sum the rows. Zero contrasts contribute zero; exactness covers $Q_L = 0$. A fixed-cap asymptotic estimate leaves the cap dependence of its constants and threshold unresolved. Constructing one approximate point gives an upper bound for (K13). The whole-cell proof gives the required lower bound (K14).

K.7 Continuity comparison for the joint floor/source modulus

The conditional-entropy result in Alhejji & Smith (2020, Main Result, equation (7)) is $e\log(k-1)+h(e)$, where k is the alphabet size of the variable whose entropy is measured. Under a common task kernel, source/task TV equals source TV. Write $I(Y;X \mid Z) = H(X \mid Z) - H(X \mid Y,Z)$ and the corresponding local identities to bound both loss changes by $O(E(e))$ independently of task-label count. Here $E(e) = e\log(1/e)$ for small positive e and $E(0) = 0$.

For $e \le m/2$, domination $\nu \ge (1 - e/m)\mu$ transports the floor to at least $\tau/2$ under μ. When $0 < \tau \le 1/2$, substitution in the nominal depth-two modulus gives

$$\Delta_\nu \le C\Phi_\tau(u + C'E(e)) + C''E(e), \qquad \Phi_\tau(u) = \min\{u^{3/4}, \tau^{-1/3}u\}. \tag{K15}$$

The joint theorem gives $C\Phi_\tau(u) + C'E(e)$. To compare it with the minimum of (K15) and the no-floor robust bound, set $\eta = E(e)$, $u = \eta^{6/5}$, and $\tau = \eta^{3/5}$. The joint expression has order η; the previous minimum has order $\eta^{9/10}$. This establishes the improvement over that envelope of direct substitutions. It leaves other possible uses of the prior machinery unresolved.

L Decoder comparison and controlled prediction experiment

L.1 A sample-dependent reversal for the balanced depth-two family

Train binary add-one Laplace decoders $\widehat{q}_{Z,n}$ and $\widehat{q}_{X,n}$ on the same n independent labeled observations. They use the five compressed cells and eight raw atoms from Section 5, respectively. Let $R(q) = \mathbb{E}[-\log q(Y \mid \cdot)]$ be population log loss for the balanced task η_t.

Proposition L.1. For any fixed $c \in (0,\infty)$, as $t \downarrow 0$ and integer $n \to \infty$ with $nt^3 \to c$,

$$n\,\mathbb{E}[R(\widehat{q}_{Z,n}) - R(\widehat{q}_{X,n})] \longrightarrow \frac{c\log 2}{14} - 2. \tag{L1}$$

For $c < 28/\log 2$, the compressed decoder eventually has smaller expected risk. For $c > 28/\log 2$, the raw decoder does. Since $\delta(t) \sim (29/392)t^4$, these two estimators have comparison scale proportional to $\delta^{-3/4}$. The proposition does not identify a minimax threshold or prove that the finite-sample crossing is unique.

For $B \sim \mathrm{Bin}(n,q)$, put $g_n(q,b) = \mathbb{E}\log(B+b)$. A cell with source mass w and binary posterior p contributes

$$L_n(w,p) = wg_n(w,2) - wpg_n(wp,1) - w(1-p)g_n(w(1-p),1) \tag{L2}$$

to expected population risk. We take expectations of the numerator and denominator logarithms separately, so their counts need not be independent. The singleton terms c, z cancel. For B and E, pooling changes the risk by $L_n(2w,p) - 2L_n(w,p)$, with $(w,p) = (1/14, t^2)$ and $(1/4, v_t)$, respectively.

We use the binomial logarithm expansion

$$g_n(q,b) = \log(nq) + \frac{b-(1-q)/2}{nq} + O((nq)^{-2}), \qquad b \in \{1,2\}, \tag{L3}$$

uniformly for $0 < q \le 1$ and $nq \ge 1$. Put $\lambda = nq$ and expand to third order on $|B-\lambda| \le \lambda/2$. The centered third moment is $O(\lambda)$ and the fourth absolute moment is $O(\lambda^2+\lambda)$, which give the remainder. On the lower tail, bound the centered logarithm by $\log((\lambda+b)/b)$ and apply the Chernoff probability bound. This term is smaller. On the upper tail,

$$\log\frac{B+b}{\lambda+b} \le \frac{B-\lambda}{\lambda+b}. \tag{L4}$$

Cauchy–Schwarz, the variance bound, and Chernoff concentration make this expectation $O(\lambda^{-1/2}e^{-\kappa\lambda})$ for some $\kappa > 0$. This avoids an uncontrolled $\log n$ factor. The same bounds handle the Taylor terms omitted on the tails.

For w bounded away from zero, substitution into (L2) yields

$$L_n(w,p) = wh(p) + \frac{1}{2n} + O\!\left(\frac{1}{n^2\min\{p,1-p\}}\right). \tag{L5}$$

The coefficient is $1/2$: the denominator contributes $3/2 + w/2$, and the numerators sum to $1 + w/2$. In B and E, the smaller label probabilities have order t^2. Since $nt^2 \to \infty$, pooling either pair changes estimation risk by $-1/(2n) + o(1/n)$.

For cell A, the positive count has bounded expectation in this limit and atom a_2 has zero posterior. We therefore handle it directly. Put $a = 1/14$ and $p = t^3$. The positive-label count in a_1 is the same count as in A, with law $\mathrm{Bin}(n, ap)$ and weight ap. Its logarithmic numerator cancels. The remaining difference is

$$\begin{aligned} D_{A,n} = {} & 2a\{g_n(2a,2) - g_n(a,2)\} - (2a-ap)g_n(2a-ap,1) \\ & + a(1-p)g_n(a(1-p),1) + ag_n(a,1). \end{aligned} \tag{L6}$$

All binomial probabilities in (L6) stay away from zero. Their leading logarithms give $\Delta(t) = 2ah(p/2) - ah(p)$. The $1/n$ coefficients are $-3/2$ from the denominators and $+1/2$ from the negative-label numerators. Uniformly for small t,

$$D_{A,n} = \Delta(t) - \frac{1}{n} + O(n^{-2}). \tag{L7}$$

Combining the three pairs gives

$$\mathbb{E}[R(\widehat{q}_{Z,n}) - R(\widehat{q}_{X,n})] = \Delta(t) - \frac{2}{n} + O\!\left(\frac{1}{n^2t^2}\right). \tag{L8}$$

Substitute $\Delta(t) = (\log 2)t^3/14 + O(t^6)$ and $nt^2 \to \infty$ to obtain (L1).

L.2 Fixed protocol and structural controls

The simulation protocol was fixed and hashed before execution. Primary conditions use the same eight source masses, raw observations, and balanced binary prior. The depth-one control keeps $\mathrm{L0}, \mathrm{L1}$ separate in the first encoder and merges the three M values. Its second encoder is unchanged. Its task is

$$\eta_t^{(1)} = (t, 0, 1/2, 1/2, 1/2, 1/2, v, v), \qquad v = 1 - (t+5/2)/7. \tag{L9}$$

The local kernel is constant on each of A, $K = \{b_1, b_2, c, z\}$, and E. The root is exact and has the failed child A, giving depth one. For the interior control, retain both original encoders and use

$$\eta_t^{\mathrm{int}} = (1/2 + t, 1/2 - t, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2). \tag{L10}$$

For the auxiliary infinite-depth control, keep the original first encoder and make the second constant. Equality of all second-view fine-label means, together with first-view M constraints, forces every scalar entry to agree. Its kernel has no proper nonempty zero-set child. This nonvacuous system has the same losses for (L10) as the original interior control.

We computed population entropies at 100-decimal precision on 25 logarithmically spaced values $10^{-5} \le t \le 10^{-2}$. Fits used the smallest ten values, selected in advance.

Condition	Depth	Compressed cells	Fitted $\log \Delta$ versus $\log \delta$ slope
Original boundary family	2	5	0.7500073
Refined-encoder boundary family	1	6	0.5000154
Original encoders, interior task	2	5	1.0000000
Auxiliary interior control	∞	2	1.0000000

The coinciding auxiliary values give no additional independent evidence. We constructed these families to check the theorem's rates. They do not estimate the worst-case exponent from random tasks, and the interior family has a different rate within the same depth-two system.

For each primary condition with $t \in \{0.05, 0.1, 0.2\}$ and $n \in \{10^3, 10^4, 10^5, 10^6\}$, we generate 512 independent training samples, each with a separate test sample of one million observations. The 36 settings produce 18,432 decoder pairs. We store exact multinomial counts of the observations. Independent streams use seed 20260912, and all prescribed settings and replicates are retained.

L.3 Observed risks and deterministic expectation checks

At depth two and $t = 0.1$, $\delta = 6.7593 \times 10^{-6}$ and $\Delta = 4.95284 \times 10^{-5}$ nats. Population risk evaluates each fitted decoder against the known distribution; test risk uses independent observations.

Training n	Compressed population excess	Raw population excess	Test compressed-minus-raw risk
10^3	2.6692×10^{-3}	4.4753×10^{-3}	$(-1.8053 \pm 0.0529) \times 10^{-3}$
10^4	2.9409×10^{-4}	4.3731×10^{-4}	$(-1.4263 \pm 0.0806) \times 10^{-4}$
10^5	7.4511×10^{-5}	4.4109×10^{-5}	$(3.0150 \pm 0.1009) \times 10^{-5}$
10^6	5.2090×10^{-5}	4.5667×10^{-6}	$(4.7387 \pm 0.0563) \times 10^{-5}$

Excess risk is measured against the raw Bayes oracle. The intervals quantify 95% Monte Carlo uncertainty for means. At $n = 10^6$, compressed estimation cost is 5.2% of Δ. Across the settings, the largest mean estimation cost is 0.637 times the proved $s/(n+1)$ bound. The maximum error in the structural-loss-plus-decoder-divergence identity is 1.96×10^{-16}.

After the simulation, we derived Proposition L.1 and calculated finite-binomial expectations from (L2). Centered logarithms reduce numerical cancellation. The risk differences have opposite signs at the endpoints of these brackets:

t	Asymptotic scale $(28/\log 2)t^{-3}$	Numerical sign-change bracket for n
0.2	5,049	4,986–4,991
0.1	40,395	40,236–40,247
0.05	323,164	322,605–322,672
0.02	5,049,433	5,045,818–5,046,829
0.01	40,395,461	40,379,106–40,387,185

Multiplying the bracket midpoint by t^3 gives values approaching $28/\log 2 \approx 40.3955$. We widened brackets where adjacent-integer differences approached the summation-tail error. Roundoff remains outside the certificate. Four deterministic expectations agreed with independent simulation means within 1.82 Monte Carlo standard errors. This calculation was developed after the preregistered simulation.

The supplement includes the protocol, code, counts, 36 result rows, replicate metrics, high-precision population losses, and numerical-tail details. All experiments use fixed representations on this controlled finite source. Empirical local-information certification, learned encoders, and natural-data validation remain untested. The risk ordering can also change with the estimators.

M EXACT DEPTH INSTABILITY AND COMPUTATIONAL VERIFICATION

M.1 A SOURCE-MASS TRICHOTOMY WITH FIXED SUPPORT AND ENCODERS

The support alone does not determine failure depth. In the following family, we change one positive atom mass and keep the support, encoders, and joint target partition fixed.

Proposition M.1. Let μ_0 be the eight-atom source of Section 5, with atom order $a_1, a_2, b_1, b_2, c, z, e_1, e_2$ and masses $(1, 1, 1, 1, 1, 2, 7/2, 7/2)/14$. For $q > -1$, multiply the unnormalized mass of z by $1 + q$ and renormalize. Keep the raw variables and prefix encoders fixed, and denote the resulting source by μ_q. Then

$$d(\mu_q) = \begin{cases} \infty, & -1 < q < 0, \\ 2, & q = 0, \\ 1, & q > 0, \end{cases} \qquad \mathrm{TV}(\mu_q, \mu_0) = \frac{6|q|}{7(7+q)}. \tag{M1}$$

The corresponding sharp all-finite-task exponents are 1, $3/4$, and $1/2$ on identical support and deterministic partitions.

Proof. Put $A = \{a_1, a_2\}$, $K = \{b_1, b_2, c, z\}$, and $E = \{e_1, e_2\}$. Every source mass is positive for $q > -1$. The fully split first-view cell M equates the entries on b_1, b_2, c. The fully split second-view cell P equates those on c, z. Equal masses in the two cross-pairs of N then give equality on a_1, a_2. Finally, the fully split Q cell gives equality on E. Thus each local-kernel vector is constant on A, K, E; denote its values by x, y, v.

The only remaining condition is equality of the two fine-cell means inside the first-view coarse cell L:

$$\frac{2x + (7/2)v}{11/2} = \frac{(2 + 2q)y + (7/2)v}{(11/2) + 2q}. \tag{M2}$$

Equivalently,

$$(11 + 4q)x - (11 + 11q)y + 7qv = 0. \tag{M3}$$

Conversely, every block-constant vector satisfying (M3) satisfies all original local equations. Each joint target cell lies within one of these blocks, so the root is exact for every q.

For $q > 0$, solve (M3) for y:

$$y = \frac{11 + 4q}{11 + 11q} x + \frac{7q}{11 + 11q} v. \tag{M4}$$

Both coefficients are positive, so $x, v \geq 0$ parametrizes the nonnegative cone. Its only proper nonempty zero sets are A and E. At A, both original coarse cells containing A extend outside A. Removing them leaves a nonzero two-atom target contrast, so this child is nonexact. The shortest failed path has length one.

For $-1 < q < 0$, solve for x:

$$x = \frac{11 + 11q}{11 + 4q} y - \frac{7q}{11 + 4q} v. \tag{M5}$$

Both coefficients are again positive. The only proper nonempty zero sets are K and E. Each is an exact leaf. On K, fully split cells M and P overlap at c and cover the state, forcing a constant kernel. The fully split Q cell does the same on E. Both leaves have no proper nonempty zero set, so failure is unreachable.

At $q = 0$, (M3) becomes $x = y$. The root children are $U = A \cup K$ and E, the latter an exact leaf. On U, remove the original L and Q cells in full. The retained M, P, N equations force constants on A and K separately. Hence U is exact, with children A and K. The only nonexact child is A. The depth is two.

Finally, $\mu_q(z) = (1 + q)/(7 + q)$, and every other mass is multiplied by $7/(7 + q)$. The change at z is $6q/[7(7 + q)]$ and has the opposite sign from every other change. Summing absolute differences proves (M1).

M.2 A FINITE-SAMPLE LIMITATION FOR EXACT DEPTH ESTIMATION

Corollary M.2. Fix $n \geq 1$ and $0 < \eta < 1$. If an estimator receives n independent raw samples and may use arbitrary additional randomness, then

$$\inf_{\widehat{d}} \sup_{|q| < \eta} \mathbb{P}_{\mu_q^{\otimes n}} \big(\widehat{d} \neq d(\mu_q) \big) = \frac{2}{3}. \tag{M6}$$

Proof. The sample laws converge in total variation to $\mu_0^{\otimes n}$ as q approaches zero from either side. This follows, for example, from

$$\mathrm{TV}(\mu_q^{\otimes n}, \mu_0^{\otimes n}) \leq n \, \mathrm{TV}(\mu_q, \mu_0) \longrightarrow 0. \tag{M7}$$

For a fixed randomized estimator, let p_1, p_2, p_∞ be the probabilities of outputs $1, 2, \infty$ under $\mu_0^{\otimes n}$. As positive q tends to zero, success tends to p_1. At zero it is p_2, and along negative q it tends to p_∞. Their sum is at most one, so the infimum success probability is at most $1/3$. This proves the lower bound. Uniformly choosing one of the three depths independently of the data attains error $2/3$ across the family.

This is a limit on uniform exact-depth estimation near the critical source law. It allows pointwise consistency away from that law and estimation under a separation condition. The robust modulus still bounds loss near the critical law through its additive source-error terms.

M.3 SOURCE-BOUND CERTIFICATES AND FINITE COMPUTATIONAL CHECKS

The implementation reads a positive rational source and deterministic partitions. Its standalone verifier rebuilds all retained equations using the whole-original-coarse-cell rule, then checks exactness, coordinate witnesses, anchor traces, and minimum failure depth. Positive-coordinate decisions

include a nonnegative rational kernel vector; forced-zero decisions include a rational separator. Numerical LPs propose the certificates, and exact rational arithmetic checks them. Replay needs no LP solver. Supplying the source separately also lets the verifier reject a certificate for a different input.

Inputs, seed, and per-case time limits were fixed before the full run. An independent exponential reference enumerates one-dimensional positive kernel supports, takes unions of extreme-ray supports, and traverses the complete zero-face graph. This reference is independent of the LP solver and canonical-anchor recursion.

Validation	Prespecified cases	Outcome
Known two-view family	Depths 1 through 8, with $4d$ atoms	All expected depths recovered
Independent complete graphs	44 tiny systems	All depths agree
Full rational-source suite	50 systems	All certificates verify
Invalid source inputs	8 malformed or model-invalid inputs	All rejected
Tampered certificates	7 source, equation, trace or conclusion changes	All rejected

The suite includes 32 seeded sources with four to six atoms, adversarial multi-ray systems, the operational controls, and four source reweightings from Proposition M.1. All 50 cases completed without timeout. They required $9{,}885$ coordinate decisions and $2{,}115$ numerical candidate calls; no exact-solver fallback was needed. The largest recorded certificate numerator or denominator used 17 bits.

For the depth-eight case with 32 atoms, computation plus immediate verification took 5.14 seconds. A separate solver-free replay took 0.73 seconds on the recorded host. These are single-run timings. Caching exact geometry avoids repeated rank and pseudoinverse calculations at each coordinate. The completed runs return exact certificates; the numerical backend has no established polynomial worst-case runtime.

N Earlier analytical certificate illustration

The following calculation evaluates Appendix D's conservative certificates using known population quantities.

Certificate calculation. Consider binary labels, five compressed cells, and $n = 10{,}000$. Bound (13) contributes the same estimation term, $5/10001$, to either structural certificate. For a risk target of 0.01 nats, the square-root certificate permits $u \le 9.03 \times 10^{-7}$, and the depth-two certificate permits $u \le 3.64 \times 10^{-6}$, a factor of 4.03 larger. At target 0.1 nats, the square-root certificate permits the larger budget. Figure 4 shows the bounds and their inverses. The calculation uses known source and population inputs, with no training or empirical information estimate. Its conservative coefficients require very small local-loss budgets.

O Source robustness with a vanishing posterior floor

We now let the posterior floor vanish in Appendix J's noisy-row argument. The nominal source, its positive masses, the encoders, and the finite ambient alphabet stay fixed. Constants may depend on nominal states, row representations, residual-support separation margins, and normalized kernel vertices. They must remain uniform over task alphabets, label priors, floors, local-loss budgets, and source radii. We retain the sharp floor coefficient and keep source error additive.

O.1 A noisy relative-floor estimate

Keep the states, masses, retained rows, and quadratic target defects from Appendix J. A child contains only whole original coarse cells with their original weights and normalization. For a nonzero nonnegative profile on T, define

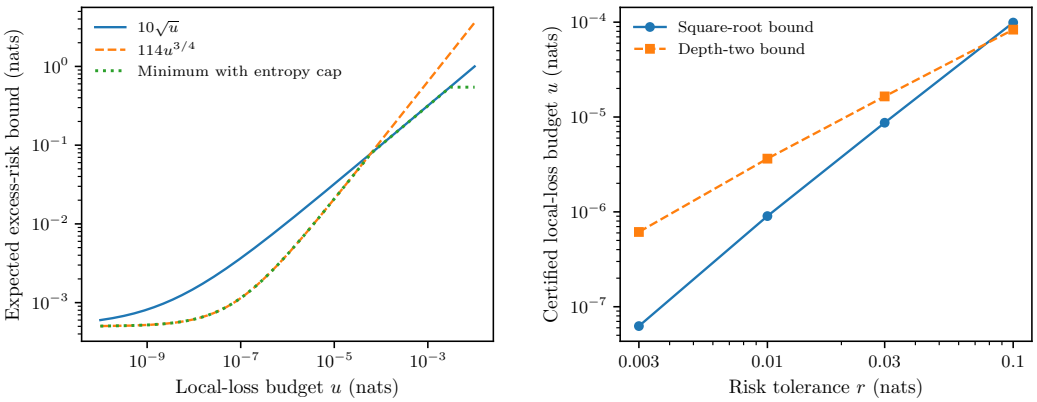

Figure 4: Left: the two structural certificates and their minimum with the entropy cap. Every curve includes the common estimation allowance $5/10001$. Right: sufficient local-loss budgets obtained by inverting the two certificates. The sample size is $n = 10{,}000$ throughout. The calculation uses the stated bounds and known population inputs. Optimal constants remain uncomputed.

$$\rho_T(f) = \frac{\mu(T)\min_{x\in T} f_x}{M_T(f)}. \tag{VF1}$$

Lemma O.1 (noisy relative-floor bound). Fix $K < \infty$. Let $a, \ell \geq 0$ and suppose that every retained local row satisfies

$$|A_j f| \leq K\left[\sqrt{(c_j(f)+a)\ell} + a\right]. \tag{VF2}$$

If $d(T) \geq h \geq 1$ and $\rho_T(f) \geq \rho$ for $0 < \rho \leq 1$, then

$$Q_{G,T}(f) \leq C_{T,h,K}\rho^{-\beta_h}\ell + C_{T,h,K}a, \qquad \beta_h = \frac{1}{2^h - 1}. \tag{VF3}$$

The constant is uniform in f, a, ℓ, ρ. If $M_T(f) = 0$, the target defect is zero and we do not form a floor ratio. At infinite depth, Lemma J.2 already gives $Q_{G,T}(f) \leq C_{T,K}(\ell + a)$ independently of the floor.

Proof. Multiplying f, a, ℓ by a common positive factor scales both hypothesis and conclusion equally. Normalize $M = M_T(f) = 1$. The mass bound $Q_{G,T} \leq CM$ handles $a \geq 1$ and $\ell \geq 1$. Assume $0 \leq a < 1$ and $0 \leq \ell < 1$ and put $V = 1 + a \leq 2$.

For $h = 1$, fix a target cell of mean $b > 0$. The row-span and mass estimates in (J5) give

$$Q_{G,J} \leq \min\left\{Cb, \frac{C(V\ell + a^2)}{b}\right\}.$$

For $b < a$, the first estimate is at most Ca. For $b \geq a$, use $a^2/b \leq a$ and $b \geq \min_T f \geq \rho/\mu(T)$ in the second estimate to obtain $C\rho^{-1}\ell + Ca$. Sum over target cells, assigning zero when the target mean vanishes.

Suppose $h \geq 2$. For a target cell J of mean $b > 0$, use the kernel-minorant decomposition (J7):

$$s = \sqrt{V\ell} + a, \qquad f = q_0 + g, \qquad g|_J = f|_J, \qquad M_T(g) \leq C(b + s). \tag{VF4}$$

For $q_0 \neq 0$, its zero set U is a proper legal child containing J, with $f|_U = g|_U$ and $d(U) \geq h - 1$. The rows retained in U are unchanged, so they satisfy (VF2) with the same a, ℓ.

As in Lemma J.2, $b \geq s$ and $q_0 = 0$ each give $Q_{G,J} \leq C(\ell + a)$. For $b \geq s$, apply (J5) to $f|_U$ if $q_0 \neq 0$ and otherwise to f on T. The state is exact, and its mass plus a is at most Cb, yielding $C(\ell + a^2/b) \leq C(\ell + a)$. For $b < s$ and $q_0 = 0$, we have $g = f$ and $M \leq C(\sqrt{V\ell} + a)$. Young's inequality gives $M \leq C(\ell + a)$, which completes the case through the mass bound. Since $\rho \leq 1$, these estimates imply (VF3).

It remains to consider $b < s$ and $q_0 \neq 0$. Write $m_U = M_U(f|_U)$. By (VF4),

$$m_U \leq Cs \leq C(\sqrt{\ell} + a). \tag{VF5}$$

When $a \geq \sqrt{\ell}$, use the child's mass bound to get $Q_{G,J} \leq Cm_U \leq Ca$. Otherwise $a < \sqrt{\ell}$ and $m_U \leq C\sqrt{\ell}$. The inherited floor then satisfies

$$\rho_U(f|_U) \geq \frac{\mu(U)}{\mu(T)} \frac{\rho}{m_U}. \tag{VF6}$$

This is an admissible floor because $m_U \geq \mu(U) \min_T f \geq \mu(U)\rho/\mu(T)$. Apply child induction with that floor and $\beta = \beta_{h-1}$. The state-weight ratio is a fixed positive constant, so

$$Q_{G,J} \leq C\rho^{-\beta} m_U^{\beta} \ell + Ca \leq C\rho^{-\beta} \ell^{1+\beta/2} + Ca. \tag{VF7}$$

If the root allowance a exceeds the child mass, the trivial case already included in the induction hypothesis still applies. The no-floor bound (J4) at T also gives

$$Q_{G,J} \leq C\ell^{1-2^{-h}} + Ca. \tag{VF8}$$

Set $n = 2^h$, so $\beta = 2/(n-2)$. The elementary inequality

$$\min\left\{\ell^{1-1/n}, \rho^{-2/(n-2)} \ell^{1+1/(n-2)}\right\} \leq \rho^{-1/(n-1)} \ell \tag{VF9}$$

follows from the second term for $\ell \leq \rho^{n/(n-1)}$ and the first term above that threshold. Choose one fixed constant covering (VF7) and (VF8) and take their minimum. Sum over targets to get (VF3). Rescaling f, a, ℓ restores arbitrary $M > 0$. □

O.2 General uniform source-and-floor bound

Theorem O.2 (uniform source and floor stability). Let μ be a fixed finite exactly composing source with minimum positive nominal atom mass m, and extend its encoders to a fixed finite ambient alphabet Ω. At finite positive depth d, put $\alpha = 1 - 2^{-d}$ and $\beta = (2^d - 1)^{-1}$. There is a finite source-dependent constant C_μ such that every finite task with relative posterior floor $0 < \tau \leq 1$ under ν satisfies

$$\Delta_\nu \leq C_\mu \min\{\delta_\nu^\alpha, \tau^{-\beta}\delta_\nu\} + C_\mu e + 2e \log|\Omega| + 2h(e), \qquad \mathrm{TV}(\nu, \mu) \leq e \leq \min\{m/2, 1/2\}. \tag{VF10}$$

At $\tau = 0$, use δ_ν^α in the braces. At infinite depth, use δ_ν. The coefficient and radius are independent of τ, task-label count, and label priors.

Proof. The singleton TV bound gives $\nu_x \geq \mu_x - e \geq (1 - e/m)\mu_x \geq \mu_x/2$ on the nominal support. Thus $\nu \geq \mu/2$ as measures on Ω, and $p_y^\nu \geq p_y^\mu/2$ for every label. Transporting the task floor gives

$$f_y(x) \geq \tau p_y^\nu \geq (\tau/2) p_y^\mu$$

on every nominal atom. For $M_\mu(f_y) = p_y^\mu > 0$ the nominal relative floor is at least $\tau/2$. A zero nominal mass contributes zero target defect.

Use precisely the allowances verified in Appendix J,

$$a_y = \sum_{x\in\Omega} |\nu_x - \mu_x| f_y(x) + eM_\mu(f_y), \qquad \ell_y = L_\nu(f_y).$$

These profiles satisfy (VF2) with source-dependent K and unchanged child allowances. Apply Lemma O.1 and $G_\mu \le Q_G$, then sum using stochastic label normalization:

$$\Delta_\mu \le C_\mu \tau^{-\beta} \sum_y \ell_y + C_\mu \sum_y a_y \le C_\mu \tau^{-\beta} \delta_\nu + 3C_\mu e, \tag{VF11}$$

because $\sum_y a_y = \|\nu-\mu\|_1 + e \le 3e$. Independently, (J13) gives $\Delta_\mu \le C_\mu \delta_\nu^\alpha + C_\mu e$ after absorbing its bounded prefactor. Take the smaller of these upper bounds and enlarge the fixed constants to obtain the minimum in (VF10) for Δ_μ. Apply target conditional-entropy continuity once, as in J.3. This gives the additive source allowance. Lemma J.2 covers zero floor and infinite depth. □

O.3 Sharp three-parameter law for the eight-atom source

Define, for $u \ge 0$,

$$\Phi_0(u) = u^{3/4}, \qquad \Phi_\tau(u) = \min\{u^{3/4}, \tau^{-1/3}u\} \quad (\tau > 0), \qquad \Phi_\tau(0) = 0. \tag{VF12}$$

Theorem O.3 (uniform three-parameter modulus). For the eight-atom source μ_0, the prefix encoders, and the fixed thirty-atom ambient alphabet in Section 7, there are constants $c, C, u_0, e_0 > 0$ such that

$$c\{\Phi_\tau(u) + e\log(1/e)\} \le R_{\mu_0,\tau}(u,e) \le C\{\Phi_\tau(u) + e\log(1/e)\} \tag{VF13}$$

simultaneously for $0 \le \tau \le 1/2$, $0 \le u \le u_0$, and $0 \le e \le e_0$. Use $e\log(1/e) = 0$ at $e = 0$. Both coordinate axes and all vanishing-floor paths are included. Uniformity as the floor approaches $\tau = 1$ remains outside the statement.

Proof. Theorem O.2 with $h(e) = O(e\log(1/e))$ gives the upper bound. At $\nu = \mu_0$, Theorem 4 gives the nominal lower order $c\Phi_\tau(u)$ uniformly for $0 < \tau \le 1/2$. Its no-floor endpoint gives $\tau = 0$.

For the source-error term, use the Section 7 cross-pair contamination with channel floor fixed at $1/2$ for every requested τ. Put mass $e/2$ on each missing pair $(\mathrm{L0}, \mathrm{Q1})$ and $(\mathrm{L1}, \mathrm{Q0})$ and scale the nominal masses by $1 - e$. With T indicating a new atom, set

$$P(Y = 1 \mid T) = e/2 + T/2.$$

The local loss is zero, $P(Y = 1) = e$, and both posterior-to-prior ratios are at least $1/2$. The task is admissible for every $0 \le \tau \le 1/2$.

Only the joint coarse cell LQ contributes target information. Its old-atom mass is $(1 - e)/2$, its new-atom mass is e, and its total mass is $(1 + e)/2$. Direct conditioning gives the exact formula

$$\Delta_\nu = \frac{1+e}{2} h\left(\frac{e(3+e)}{2(1+e)}\right) - \frac{1-e}{2} h(e/2) - eh((1+e)/2). \tag{VF14}$$

For a uniform remainder bound, set $q = e(3 + e)/[2(1 + e)]$ and $w = (1 + e)/2$. For $0 < e \le 1/2$, use

$$z\log(1/z) \le h(z) \le z\log(1/z) + z, \qquad q \le 3e/2, \qquad h((1+e)/2) \le \log 2.$$

The coefficient of $\log(1/e)$ in the resulting lower bound is $e(1+e)/2$. Moreover, $wq = e(3+e)/4 \le 7e/8$ and $(1 - e)e/4 \le e/4$. Consequently

$$\begin{aligned}\Delta_\nu &\geq \frac{1}{2}e\log(1/e) - e\left\{\frac{7}{8}\log(3/2) + \frac{5}{4}\log 2 + \frac{1}{4}\right\} \\ &\geq \frac{1}{2}e\log(1/e) - 2e.\end{aligned} \tag{VF15}$$

The last bound uses $\log(3/2) < 1/2$ and $\log 2 < 1$. Hence $\Delta_\nu \geq \frac{1}{4}e\log(1/e)$ for $0 < e \leq \exp(-8)$, with both coefficient and radius independent of the requested floor.

Both witnesses fit the budgets (u, e). The supremum is at least their maximum, hence at least half their sum after adjusting constants. This proves (VF13). We may choose the binary witness for whichever term is larger. □

The nominal branches cross at $u = \tau^{4/3}$. Source error dominates when $e\log(1/e)$ exceeds $\Phi_\tau(u)$. The upper theorem holds for general exactly composing nominal sources. A matching source-error lower bound for another ambient system needs a corresponding contamination witness.

P SMOOTHING-DEPENDENT CROSSOVER AND LEARNED POOLING

Proposition (smoothing-dependent prediction crossover). Fix a pseudocount $a > 0$ for the balanced depth-two task. Let $\widehat{q}^{(a)}_{W,n}(y \mid w) = (N_{wy} + a)/(N_w + 2a)$, for $W = X, Z$, be binary symmetric add-a decoders trained on the same n independent labeled observations. As $t \downarrow 0$ and $nt^3 \to c \in (0, \infty)$,

$$n\,\mathbb{E}\Big[R(\widehat{q}^{(a)}_{Z,n}) - R(\widehat{q}^{(a)}_{X,n})\Big] \longrightarrow \frac{c\log 2}{14} - (a+1).$$

The limiting sign changes at $c = 14(a+1)/\log 2$, giving scale $n \asymp \delta^{-3/4}$ with a smoothing-dependent coefficient. Taking $a = 1$ recovers add-one smoothing. The result assumes fixed positive a. A data-selected pseudocount requires separate analysis.

Proof. Write $g_n(q, b) = \mathbb{E}\log(B + b)$ for $B \sim \mathrm{Bin}(n, q)$. Fix $b > 0$ and apply the centered Taylor and binomial-tail estimates from the add-one case to get

$$g_n(q, b) = \log(nq) + \frac{b - (1-q)/2}{nq} + O_b((nq)^{-2})$$

uniformly for $q > 0$ with $nq \geq 1$. Constants may depend on b and need not stay bounded as $b \downarrow 0$. A cell with mass w and posterior p contributes

$$L_{n,a}(w, p) = wg_n(w, 2a) - wpg_n(wp, a) - w(1-p)g_n(w(1-p), a).$$

For cell masses away from zero, $0 < p < 1$ and $nw\min\{p, 1-p\} \geq 1$, substitution gives

$$L_{n,a}(w, p) = wh(p) + \frac{1}{2n} + O_a\left(\frac{1}{n^2\min\{p, 1-p\}}\right).$$

The denominator coefficient is $2a - 1/2 + w/2$. The numerator coefficients sum to $2a - 1 + w/2$, so the difference is $1/2$ independently of a. Each equal-posterior pair B, E therefore contributes $-1/(2n) + O_a(n^{-2}t^{-2})$ under pooling.

For A, set $w_0 = 1/14$ and $p = t^3$. The positive-label count in a_1 equals the pooled count, and the other atom has zero posterior. Equal pseudocounts therefore cancel the positive-label numerator exactly, including when its expected count stays bounded. The remaining difference is

$$\begin{aligned}D_{A,n,a} = {} & 2w_0\{g_n(2w_0, 2a) - g_n(w_0, 2a)\} \\ & - (2w_0 - w_0p)g_n(2w_0 - w_0p, a) \\ & + w_0(1-p)g_n(w_0(1-p), a) + w_0g_n(w_0, a).\end{aligned}$$

All remaining binomial probabilities stay away from zero. Leading logarithms give $\Delta(t) = 2w_0h(p/2) - w_0h(p)$. The $1/n$ coefficient is $-(2a - 1/2)$ for the denominator difference and $a - 1/2$ for the negative-label numerators, yielding

$$D_{A,n,a} = \Delta(t) - \frac{a}{n} + O_a(n^{-2}).$$

The singleton terms cancel. Adding the three pair differences gives

$$\mathbb{E}\Big[R(\widehat{q}^{(a)}_{Z,n}) - R(\widehat{q}^{(a)}_{X,n})\Big] = \Delta(t) - \frac{a+1}{n} + O_a(n^{-2}t^{-2}).$$

Now use $\Delta(t) = (\log 2)t^3/14 + O(t^6)$ and $nt^2 \to \infty$ to obtain the limit. □

Interpretation. The two equal-posterior pairs contribute one unit to the estimation gain for every fixed pseudocount. The pair with an impossible label supplies the additional a. These terms explain why the crossing coefficient depends on the decoder even though the exponent fixes the asymptotic sample scale for this task family.

P.1 Regularization sensitivity and learned partial pooling

The follow-up protocol was frozen after inspection of the original experiment and before any new simulation. It fixes $t = 0.1$, sample sizes $10^3, 10^4, 40{,}242, 10^5, 10^6$, and 2,048 independent training replicates per size. Each replicate has an independent test set of one million observations. Partition the n i.i.d. training observations into five multinomial-count folds. Raw and compressed predictors each select $a \in \{0.01, 0.03, 0.1, 0.3, 0.5, 1, 3\}$ by total five-fold validation log loss, then refit on all n observations. A raw-access learner jointly selects the pseudocount and $\lambda \in \{0, 1/4, 1/2, 3/4, 1\}$ for

$$(1-\lambda)\widehat{q}^{(a)}_{X,n}(\cdot \mid X) + \lambda \widehat{q}^{(a)}_{Z,n}(\cdot \mid Z).$$

The compressed grouping is given to this learner. Source probabilities, population risks, and test data are used only for evaluation.

We specified the three primary paired contrasts before the follow-up run. Fixed add-one and add-one-half predictors remain as secondary controls for each representation. The supplement retains every fold, selection, and outcome. This experiment assumes known source groupings and fixed encoders.

n	Comp. CV − raw CV	Pool CV − raw CV	Pool CV − comp. CV
1,000	-71.508 ± 11.146	-2.400 ± 8.590	69.109 ± 7.011
10,000	-7.386 ± 0.712	-5.200 ± 0.363	2.185 ± 0.564
40,242	2.190 ± 0.116	0.721 ± 0.085	-1.469 ± 0.091
100,000	3.868 ± 0.048	0.273 ± 0.034	-3.594 ± 0.057
1,000,000	4.858 ± 0.004	0.000 ± 0.000	-4.858 ± 0.004

The table reports paired mean population-risk differences in 10^{-5} nats with nominal 95% Student-t Monte Carlo intervals. A negative value favors the first predictor. The supplement also gives independent test risks, oracle excesses, training-sample quantiles, and Bonferroni simultaneous 95% intervals for all 15 primary population comparisons. These intervals measure simulation uncertainty for this source.

With equal smoothing selection, compression has lower risk at $n = 10^3, 10^4$. The difference becomes positive before the original add-one bracket. All five compressed-versus-raw signs survive simultaneous intervals, placing a reversal between 10^4 and $40{,}242$ on the tested grid. The grid does not identify a unique crossing.

At $n = 10^6$, the raw-access learner selects no pooling in all 2,048 replicates. Its mean risk exceeds that of the better CV endpoint at every smaller size. Its difference from tuned raw is insignificant at $n = 10^3$; every other nonzero primary sign survives the simultaneous intervals. The observed zero difference at 10^6 comes from identical predictions in those sampled replicates and need not hold on other training samples.

Tuning also has costs here. At $n = 10^3$, fixed add-one-half compression has mean excess 1.983×10^{-3}, below compressed CV at 3.180×10^{-3}. At $n = 10^4$, both tuned endpoints have higher mean risk than their add-one versions. The tables retain these outcomes. Equal tuning gives each method the same opportunity, without establishing a general benefit from selection. Proposition 5 assumes fixed $a > 0$. The adaptive predictor also observes X, so its ability to avoid structural loss is consistent with the lower bound for predictors restricted to Z.

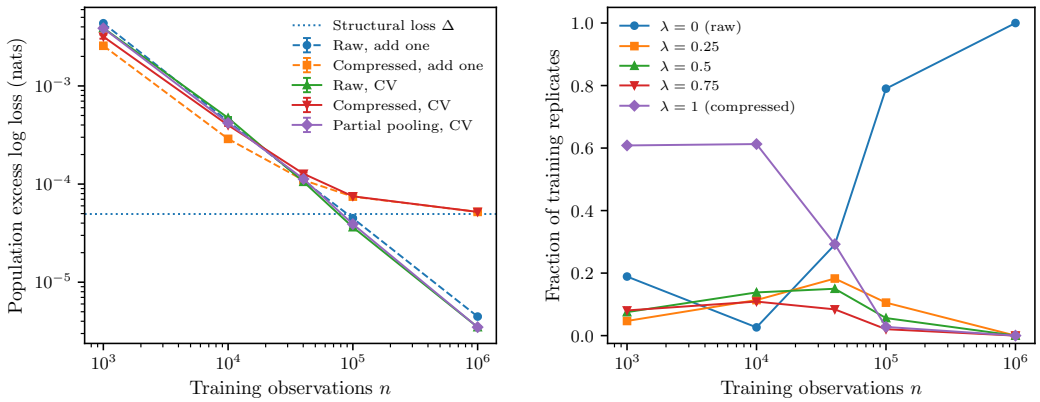

Figure 5: Smoothing selection and partial pooling on task (7) at $t = 0.1$. Left: mean population excess over the raw oracle, with 95% Monte Carlo intervals from 2,048 training replicates at each sample size. Dashed curves show the fixed add-one controls. Right: the fraction of replicates selecting each pooling weight, summed over pseudocounts. Here $\lambda = 0$ uses the raw predictor and $\lambda = 1$ uses the compressed predictor. Lines connect the tested sample sizes.

Deterministic smoothing controls. At $t = 0.1$, the add-a risk difference changes sign across the brackets $[23{,}132, 23{,}139]$ for $a = 0.1$, $[30{,}373, 30{,}382]$ for $a = 0.5$, $[40{,}236, 40{,}247]$ for $a = 1$, and $[81{,}998, 82{,}017]$ for $a = 3$. An independent long-double recurrence over the full binomial support agrees with the centered calculation to 5.52×10^{-17} nats at all twelve prescribed checks. The tail bounds exclude roundoff. The brackets therefore remain numerical sign checks without interval certification or a uniqueness proof.