# ORDER STABILITY IN MASKED GENERATION: SWAP DEPTH AND SELECTIVE REPAIR

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

We develop a quantitative theory of order dependence for positive masked conditional tables. Two structural effects govern reliability. First, along a route with $D$ layers of disjoint adjacent swaps and total source-weighted local Hellinger energy $E$, the endpoint laws satisfy $\mathrm{TV}^2 \le 15DE$. A matching product-distribution construction makes the depth dependence necessary, although the upper bound allows arbitrary statistical dependence. Second, every fixed order can be exponentially close to an anchor while a content-adaptive order remains a constant distance away, even with all token probabilities in $[1/4, 3/4]$. The same construction gives an exact exponential coverage penalty for independent-mask auditing. We connect local swap audits to anchor-conditional errors and selective repair. For a fixed deterministic policy, replacing selected rows has exactly additive forward-KL benefits. Under a known probability floor, independent anchor replay learns a near-optimal repair budget and certifies its residual error; bounded unresolved-prefix width makes the required conditional evaluations tractable. Bounded-score Chernoff certificates expose the finite audit cost. Exhaustive checks and sample-selected repair experiments demonstrate the structural scalings and nonvacuous certificates. The guarantees measure reliability relative to an anchor; external accuracy requires a separate anchor-quality bound.

## 1 INTRODUCTION

A masked generator may fill the same missing positions in different orders. Even when every query returns a normalized distribution, the resulting path products need not agree. Conditional specifications and masked language models have long motivated this compatibility question (Heckerman et al., 2000; Young et al., 2024); masked diffusion makes it directly relevant to decoding (Sahoo et al., 2024). Exact compatibility is characterized by local two-position exchanges: Kim (2026) identifies the corresponding log-ratio, or curl, and decomposes global order differences through adjacent swaps. We ask the quantitative question: *when do small local defects certify a final distribution, and when can selected errors be repaired at finite cost?*

**Depth, not just the number of swaps.** Our first structural result bounds squared total variation by $15DE$, where $D$ is the number of layers of disjoint adjacent order edits and $E$ is their source-weighted Hellinger energy (Theorem 2). Within a layer, errors aggregate as energy even under arbitrary dependence; between layers, distances accumulate. A matching lower construction has product endpoint distributions (Theorem 3). Thus the depth penalty is inherent to repeatedly moving the same coordinate, rather than an artifact of correlated outputs. This depth measures order editing, not parallel generation rounds.

**Adaptive selection can defeat every fixed-order guarantee.** Our second structural result has worst fixed-order error $2^{-m}/4$ but adaptive error $1/4$, with all conditional probabilities in $[1/4, 3/4]$ (Theorem 5). It also quantifies the exact coverage cost of an independent-mask audit: eliminating its missing mass requires density amplification $2^m$. The failure is therefore selection of conditional interfaces, not vanishing token probabilities. Replaying the decoder on complete anchor samples measures the appropriate visitation law.



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.

**From diagnosis to a budgeted, certified intervention.** A local-to-conditional bridge transfers swap energy to mean anchor-conditional error without inverse context-probability factors (Theorem 6). Exact anchor replacements then remove identifiable terms from forward KL for a frozen deterministic policy. We turn this additive identity into a finite-sample procedure: estimate group benefits on independent replay trajectories, select a repair budget, and certify the residual on held-out replay (Theorem 10). The KL selection guarantee assumes a known probability floor; a Hellinger version needs no such floor. Unresolved-prefix width controls the cost of the conditional evaluations, yielding an explicit polynomial-query regime without enumerating a mask library.

**Position relative to prior work.** Zhang et al. (2026) separate order error from factorized parallel-decoding bias, and Wen et al. (2026) analyze adaptive parallel sampling through conditional total correlation. Here both orientations of each swap remain sequential. Amin et al. (2026) train multi-path consistency, whereas our intervention is post-hoc conditional replacement. A concurrent manuscript also studies discrete curl and compatibility (Anonymous, 2026). Reveal-graph curl, exact compatibility, KL chain rules, and one-sided Hellinger subadditivity are background tools. Our focus is the matching depth law, the fixed–adaptive separation and its coverage price, and their connection to statistically certified selective repair.

## 2 A BOUNDED LOCAL AUDIT

Let $X = (X_1, \dots, X_n)$ have a finite product alphabet, each coordinate of size at most $K$. A context $h = (S, x_S)$ specifies revealed coordinates. Assume every $q_i(\cdot \mid h)$, $i \notin S$, is strictly positive and normalized, and depends on the revealed set and values, not on their earlier reveal order. A fixed prompt is absorbed into the table. For a permutation $\pi$,

$$P_\pi(x) = \prod_{t=1}^{n} q_{\pi_t}(x_{\pi_t} \mid x_{\pi_{<t}}). \tag{1}$$

These are genuine normalized joint distributions, although different permutations may give different ones. We use $H^2(P, Q) = \frac{1}{2}\sum_x (\sqrt{P(x)} - \sqrt{Q(x)})^2$, so $\mathrm{TV}^2 \le 2H^2$.

For $f = (S; i, j)$ and $h = x_S$, define normalized two-position kernels

$$\begin{aligned} A_f(a, b \mid h) &= q_i(a \mid h) q_j(b \mid h, i = a), \\ B_f(a, b \mid h) &= q_j(b \mid h) q_i(a \mid h, j = b), \qquad c_f = \log(A_f / B_f). \end{aligned} \tag{2}$$

The local energy and its bounded audit score are

$$\begin{gathered} e_f(h) = H^2(A_f, B_f), \qquad z_f = \frac{(\sqrt{A_f} - \sqrt{B_f})^2}{A_f + B_f} = 1 - \operatorname{sech}(c_f/2), \\ \mathbb{E}_{(A_f + B_f)/2} z_f = e_f(h). \end{gathered} \tag{3}$$

Thus $0 \le z_f \le 1$ and $z_f = c_f^2/8 + O(c_f^4)$ near zero. Unlike an unbounded squared log-ratio, this score admits a distribution-free concentration bound. It still requires the correct audit distribution.

**Lemma 1** (Exact single-swap identity)**.** *If $\pi'$ exchanges adjacent $i, j$ after a common prefix in $\pi$, and $R$ is the prefix law under either order, then*

$$H^2(P_\pi, P_{\pi'}) = \mathbb{E}_{h \sim R} e_{S;i,j}(h). \tag{4}$$

*Proof.* The two joint laws factor as $R(h)A_f(a, b \mid h)T(y \mid h, a, b)$ and $R(h)B_f(a, b \mid h)T(y \mid h, a, b)$. The suffix kernel is identical because the revealed state after the swap is identical. Sum over $y$ in the square-root distance. □

The same cancellation gives the familiar log-ratio identity: along an adjacent-swap route, $\log(P_\pi / P_\sigma)$ is the sum of the *oriented* local curls (Kim, 2026). Section A records the orientation and exact compatibility proof for completeness. A two-position swap is local in operations, not necessarily in its dependence on $x_S$.

## 3 ORDER-EDIT DEPTH CONTROLS ACCUMULATION

A *swap layer* reverses a set of disjoint adjacent pairs in the current permutation. Let $\pi^{(0)}, \dots, \pi^{(D)}$ be a route of $D$ such layers. For each flipped pair $f$ in layer $\ell$, let $R_{\ell f}$ be its prefix law under the *source order* $\pi^{(\ell-1)}$, and write

$$E_\ell = \sum_{f \text{ in layer } \ell} \mathbb{E}_{R_{\ell f}} e_f, \qquad E = \sum_{\ell=1}^{D} E_\ell. \tag{5}$$

The relevant probabilistic tool is one-sided Hellinger subadditivity: for two sequential processes $P, Q$,

$$H^2(P, Q) \le C_H \mathbb{E}_P \sum_t H^2(P_t(\cdot \mid X_{<t}), Q_t(\cdot \mid X_{<t})). \tag{6}$$

Foster et al. (2024, Lemma D.2) give $C_H = 7$. We use $C_H = 15/2$, for which Section B provides a self-contained stopping proof. The structural dependence, rather than the universal constant, is the issue.

**Theorem 2** (Depth–energy stability)**.** *For any positive masked conditional table and any swap-layer route,*

$$\mathrm{TV}(P_{\pi^{(0)}}, P_{\pi^{(D)}}) \le \min\left\{1, \sqrt{15} \sum_{\ell=1}^{D} \sqrt{E_\ell}\right\}, \qquad \mathrm{TV}^2(P_{\pi^{(0)}}, P_{\pi^{(D)}}) \le 15DE. \tag{7}$$

*No weak-dependence, mixing, or uniform lower bound on the conditional probabilities is required.*

*Proof.* Treat each flipped pair as one block of a sequential process. Its conditional kernel is $A_f$ or $B_f$; unchanged singleton blocks have zero discrepancy. Equation (6) gives $H^2(P_{\pi^{(\ell-1)}}, P_{\pi^{(\ell)}}) \le (15/2)E_\ell$. Apply the triangle inequality for $H$, then $\mathrm{TV} \le \sqrt{2}H$, and finally Cauchy–Schwarz across layers. □

For $m$ swaps with average local energy at most $\varepsilon^2$, the result scales as $\sqrt{Dm}\,\varepsilon$, rather than $m\varepsilon$. The same-layer aggregation is valid with arbitrarily dependent blocks; it is not a product-measure argument. The source-prefix weights change between layers and cannot be replaced by a single empirical data law without a coverage argument. The sharper route cost $\sum_\ell \sqrt{E_\ell}$ can be minimized over any prespecified candidate set; an optimal-route algorithm is not needed for the guarantee.

The proof also covers two decoders sharing a content-adaptive rule that selects the next singleton or unordered pair from completed blocks, while differing only in pair orientation. Arbitrary adaptive policies do not in general share such a block decomposition (Section C).

**Theorem 3** (Matching depth dependence)**.** *For integers $g, D \ge 1$ and $0 < |\theta| D\sqrt{g} \le 1/(2\sqrt{3})$, there is a positive binary masked table with $m = gD$ swaps and minimum swap-layer depth $D$, whose endpoint laws are product distributions, such that*

$$\frac{DE}{6} \le \mathrm{TV}^2(P_0, P_1) \le 15DE, \qquad E = \Theta(m\theta^2). \tag{8}$$

*The constants are universal on the stated parameter range.*

**Construction and mechanism.** Each of $g$ groups contains a main spin $X \in \{-1, 1\}$ and $D$ fair auxiliary positions. After $k$ auxiliaries have been revealed, the main conditional is $q(X = x \mid k) = e^{k\theta x}/(2\cosh(k\theta))$, independently of their values. Move each main position from the beginning to the end of its group, one slot per layer. Each swap has energy $\Theta(\theta^2)$, while each endpoint main spin has tilt $D\theta$. Thus $g$ independent endpoint changes have squared TV of order $gD^2\theta^2 = DE$. Moving a coordinate by $D$ slots requires at least $D$ layers. Section D gives the nonasymptotic calculation.

# 4 COVERAGE AND ADAPTIVE SELECTION

## 4.1 WHAT A FINITE LOCAL AUDIT CERTIFIES

Fix a route containing $m > 0$ marked swaps. Let $\mu$ choose a swap uniformly, sample its actual source prefix, then sample the pair from $(A_f + B_f)/2$. Then $m\mathbb{E}_\mu z = E$. For an audit law $\nu$ on the complete tuple (swap, mask, context, pair), assume $\kappa \geq 1$, $0 \leq \tau \leq 1$, and

$$\sum_\omega [\mu(\omega) - \kappa\nu(\omega)]_+ \leq \tau. \tag{9}$$

Since $z \in [0,1]$, $\mathbb{E}_\mu z \leq \kappa\mathbb{E}_\nu z + \tau$.

For Bernoulli relative entropy $d_{\mathrm{B}}$, define the upper Chernoff endpoint

$$\mathcal{U}(s,a) = \sup\{v \in [s,1] : d_{\mathrm{B}}(s\|v) \leq a\}, \qquad \mathcal{U}(0,a) = 1 - e^{-a}. \tag{10}$$

**Corollary 4** (Finite-sample, coverage-aware certificate)**.** *For $N \geq 1$ independent samples from $\nu$ and $0 < \delta < 1$, let $\bar{z}$ be their average score. With probability at least $1 - \delta$,*

$$\mathrm{TV}^2(P_{\pi^{(0)}}, P_{\pi^{(D)}}) \leq 15Dm \left[\kappa\mathcal{U}\left(\bar{z}, \frac{\log(1/\delta)}{N}\right) + \tau\right]. \tag{11}$$

A bounded-variable Chernoff argument proves the endpoint inversion (Section E). A simpler alternative replaces $\mathcal{U}$ by $2\bar{z} + 8\log(1/\delta)/N$. For $B$ prespecified certificates, use $\log(B/\delta)$. This also permits data-dependent selection among those certificates; unrestricted route selection requires new control or a held-out audit.

For $0 < \varepsilon < 1$, even if $\bar{z} = 0$ and $(\kappa, \tau) = (1, 0)$, certifying $\mathrm{TV} \leq \varepsilon$ by (11) requires

$$N \geq \frac{\log(1/\delta)}{-\log(1 - \varepsilon^2/(15Dm))} \sim \frac{15Dm\log(1/\delta)}{\varepsilon^2}. \tag{12}$$

This is the budget of this certificate, not a minimax sample lower bound. Coverage must be justified by design: $\nu = (1-\lambda)\nu_{\mathrm{data}} + \lambda\mu$ guarantees $(\kappa, \tau) = (1/\lambda, 0)$, but sampling $\mu$ costs source-prefix rollouts. Conversely, a positive two-position example has endpoint TV $3/8$ while its data-weighted score tends to zero (Section E).

## 4.2 A FIXED-ORDER GUARANTEE DOES NOT CERTIFY ADAPTIVE DECODING

**Theorem 5** (Exponential fixed–adaptive gap)**.** *For every $m \geq 1$, there is a binary table on $n = 3m+1$ positions, with every conditional probability in $[1/4, 3/4]$, an anchor $p = P_\sigma$, and a deterministic content-adaptive policy $\mathcal{A}$ satisfying*

$$\max_{\pi \text{ fixed}} \mathrm{TV}(p, P_\pi) = 2^{-m}/4, \qquad \mathrm{TV}(p, P^q_{\mathcal{A}}) = 1/4. \tag{13}$$

**Construction.** Take fair selector bits $U_1, \ldots, U_m$, paired markers $B^0_j, B^1_j$, and a target bit $X$. Only $X$ can be biased: it has probability $3/4$ of being one precisely when all selectors and exactly one marker per pair have been revealed, with each revealed marker label equal to $U_j$. Otherwise it is fair. The anchor reveals $X$ first and is uniform. For a fixed order, the required label match occurs with probability at most $2^{-m}$. An adaptive order first reveals the selectors, then the matching markers, then $X$; the error is always triggered. All remaining conditionals are fair. Section F proves the exact distances.

**An exact coverage price.** In this construction, let $\nu_{\mathrm{mask}}$ choose the target's marker-label vector uniformly, independently of an anchor context, and let $\mu_{\mathrm{sel}}$ be the adaptive target-row replay law. Then

$$\sum_\omega [\mu_{\mathrm{sel}}(\omega) - \kappa\nu_{\mathrm{mask}}(\omega)]_+ = (1 - \kappa 2^{-m})_+. \tag{14}$$

An independent-mask draw hits a biased row with probability $2^{-m}$, whereas replay always does. Section F derives both laws and the resulting exponential detection budget.

To measure policy-specific errors, fix an anchor $p = P_\sigma$ and *replay* a policy on complete samples $X \sim p$, revealing the requested entries without resampling them. Define

$$k_i(h) = D_{\mathrm{KL}}(p_i(\cdot \mid h) \| q_i(\cdot \mid h)), \qquad r_i(h) = H^2(p_i(\cdot \mid h), q_i(\cdot \mid h)). \tag{15}$$

For a fixed deterministic nonanticipating state-based policy, the standard chain rule gives

$$D_{\mathrm{KL}}(p \| P^q_{\mathcal{A}}) = \mathbb{E}_{p,\mathcal{A}} \sum_t k_{I_t}(H_t), \qquad \mathrm{TV}^2(p, P^q_{\mathcal{A}}) \le 15 \mathbb{E}_{p,\mathcal{A}} \sum_t r_{I_t}(H_t). \tag{16}$$

The Hellinger statement also holds for randomized policies. For those policies, the KL equality is for the action/seed transcript, not necessarily for the final output marginal. This distinction is also important in adaptive sampling analyses (Wen et al., 2026). The normalized replay occupancy $d^p_{\mathcal{A}}(h,i) = n^{-1}\mathbb{E}\sum_t \mathbf{1}\{(H_t, I_t) = (h,i)\}$ gives the analogous certificate $\mathrm{TV}^2 \le 15n(\kappa \mathbb{E}_\nu r + \tau)$ when $d^p_{\mathcal{A}}$ satisfies (9) relative to $\nu$. Replay visitation cannot be replaced by independent random masks for free.

## 5 From swap audits to selective conditional repair

Relabel coordinates so the anchor order is $(1, \dots, n)$. For a mask/action type $\eta = (S, i)$, put $T = [\max(S \cup \{i\})]$ and $U = T \setminus (S \cup \{i\})$. Let $Q_\eta$ be the path law on $T$ using the order $\tau = (S \text{ increasing}, i, U \text{ increasing})$. Then

$$Q_\eta(h, a, u) = A_S(h) q_i(a \mid h) R_\eta(u \mid h, a). \tag{17}$$

Here $R_\eta$ is a normalized sequential completion kernel. These objects involve only an anchor prefix; the remaining suffix is marginalized automatically.

**Theorem 6** (Local-to-conditional bridge)**.** *Choose a swap-layer route from the anchor order on $T$ to $\tau$, with energies $E_{\eta\ell}$, and let $\Gamma_\eta = (\sum_\ell \sqrt{E_{\eta\ell}})^2$. Then*

$$\mathbb{E}_{h \sim p_S} r_i(h) \le 2H^2(p_T, Q_\eta) \le 15\Gamma_\eta. \tag{18}$$

*If every action of an adaptive policy belongs to a finite library $\mathcal{L}$ of mask/action types, and all rows of the types in $\mathcal{R} \subseteq \mathcal{L}$ are replaced by their anchor conditionals, then, for deterministic or randomized policies,*

$$\mathrm{TV}^2(p, P^{q^{\mathcal{R}}}_{\mathcal{A}}) \le 225 \sum_{\eta \in \mathcal{L} \setminus \mathcal{R}} \Gamma_\eta. \tag{19}$$

*Proof.* For any joint $P, Q$ on $(S, A)$, conditional Hellinger error obeys $\mathbb{E}_{P_S} H^2(P_{A|S}, Q_{A|S}) \le 2H^2(P, Q)$; Section G proves the inequality directly. Apply marginal data processing to $p_T, Q_\eta$, whose latter conditional on $(S, i)$ is $q_i$. This gives (18). Under anchor replay, a fixed type is visited at most once and its visitation indicator is at most one. Thus the expected residual conditional energy is bounded by the sum of its $p_S$-averages over the unrepaired library. Apply (16). □

The library sum is an explicit cost: simultaneous local audits support reuse across policies and later selection of repair types. Policy-specific replay avoids summing over unused types and is often sharper; Section 7 gives a budgeted alternative.

**Proposition 7** (Normalized defect marginalization)**.** *Let $C_\eta = \log(Q_\eta / p_T)$, equivalently the sum of oriented swap defects along any route to $\tau$. Set*

$$Z_i(a \mid h) = \mathbb{E}_{u \sim R_\eta(\cdot \mid h, a)} e^{-C_\eta(h,a,u)}. \qquad p_i(a \mid h) = \frac{q_i(a \mid h) Z_i(a \mid h)}{\sum_b q_i(b \mid h) Z_i(b \mid h)}. \tag{20}$$

*Proof.* Use $p_T = Q_\eta e^{-C_\eta}$ and sum over $u$. The factor $A_S(h)$ in (17) cancels when normalizing over $a$. □

Equation (20) expresses the required conditional marginalization in swap defects. Normalizing a single unaveraged defect is generally insufficient; the unresolved completion must be integrated out.

**Proposition 8** (Safe selective replacement)**.** *Fix a deterministic policy $\mathcal{A}$ as a function of revealed values. Let $q^{\mathcal{R}}$ replace an arbitrary collection of conditional rows by* (20)*, leaving every other row unchanged. Then*

$$D_{\mathrm{KL}}(p\|P_{\mathcal{A}}^{q^{\mathcal{R}}}) = \mathbb{E}_{p,\mathcal{A}} \sum_t \mathbf{1}\{(H_t, I_t) \notin \mathcal{R}\} k_{I_t}(H_t). \tag{21}$$

*The improvement relative to $q$ is exactly the complementary nonnegative sum over repaired rows.*

*Proof.* The policy and the anchor replay law do not change. Apply the KL identity in (16); each replaced conditional KL is zero. □

The anchor order is itself preserved: at its prefix rows, $q_i$ already equals $p_i$, so replacing such a row does not change $P_\sigma$. For disjoint equal-cost groups of rows, define each benefit as its replay-weighted sum of $k_i(h)$. Selecting the $b$ largest *known* benefits exactly minimizes residual forward KL. Section 7 learns these benefits under bounded log loss. A per-trajectory repair budget instead has the Bellman characterization in Section J.

Two restrictions matter. A confidence-based policy must keep its original state-to-action map; recomputing that map after repair changes the theorem's object. Also, random mixtures of orders may cancel errors. There is a three-bit example with initial mixture equal to $p$ but one conditional repair producing $D_{\mathrm{KL}}(p\|P_{\mathrm{repaired}}) = \frac{1}{2}\log(16/15) > 0$ (Section K). Randomized transcript KL improves, but final-marginal KL need not.

## 6 Computational and accuracy boundaries

Define the *unresolved-prefix width* of a query by

$$w(S, i) = \max(S \cup \{i\}) - |S|. \tag{22}$$

It counts the unrevealed coordinates in the necessary anchor prefix, including $i$. This is distinct from order-edit depth.

**Proposition 9** (Width-bounded repair and its oracle boundary)**.** *An anchor conditional at width $w$ can be computed by enumerating at most $K^w$ prefix completions, with $O(nK^w)$ conditional-row evaluations. Along a reveal trajectory whose width never exceeds $w$, an incremental prefix-posterior filter uses $O(nK^w)$ row evaluations and at most $K^w$ posterior table entries. In the binary black-box model, exact deterministic evaluation of an arbitrary width-$w$ anchor conditional can require $\Omega(2^{w-1})$ row queries.*

The filter maintains the anchor posterior on unresolved prefix positions. It extends that prefix in anchor order, multiplying each table entry by the appropriate conditional, and conditions the table after every actual reveal. It remains valid when unrepaired reveals were sampled from $q$, because it computes the anchor posterior *given the observed evidence*. Context bookkeeping and the internal cost of a neural-network call are not included in the row-query bound.

For the lower bound, let $w - 1$ fair hidden bits index arbitrary probabilities $v_y \in [1/4, 3/4]$ of a final anchor bit. Querying this final bit first requires the mean $2^{-(w-1)} \sum_y v_y$. Any unqueried $v_y$ can change the exact answer while leaving the query transcript unchanged. Non-anchor conditional interfaces can be fixed to reveal no additional information. Section L gives full details.

The lower bound concerns exact conditional *evaluation*. Unconditional generation is easier: sample a complete anchor sequence in $n$ calls, then reveal it in any order. Selective repair serves a different interface—conditioning on observed evidence while retaining unselected $q$ decisions. Randomized approximation is outside the exact-query lower bound.

**Approximate repair.** If an approximate correction $\widehat{Z}_i$ satisfies $\mathrm{osc}_a(\log \widehat{Z}_i - \log Z_i) \le b_i(h)$, then the normalized repaired row obeys

$$D_{\mathrm{KL}}(p_i(\cdot \mid h)\|\widehat{p}_i(\cdot \mid h)) \le b_i(h)^2/8. \tag{23}$$

This is the bounded log-tilt inequality proved in Section M. These residuals replace the zeros in (21). A repair is certified safe when its residual bound does not exceed the original conditional KL. A sampled maximum is not a valid global oscillation certificate.

**Accuracy requires an anchor.** If $\mathrm{TV}(P_\star, p) \le \alpha$, then any deterministic-policy repair satisfies

$$\mathrm{TV}(P_\star, P_{\mathcal{A}}^{q^{\mathcal{R}}}) \le \alpha + \min\left\{ \sqrt{\tfrac{1}{2}\mathbb{E}_{p,\mathcal{A}} \sum_{\mathrm{unrep}} k_i(h)},\ \sqrt{15\mathbb{E}_{p,\mathcal{A}} \sum_{\mathrm{unrep}} r_i(h)} \right\}. \tag{24}$$

The $\alpha$ term is external: it is not implied by curl consistency, finite audit success, or the repair procedure.

## 7 Finite-budget repair from independent replay

Fix the deterministic policy $\mathcal{A}$ before sampling. Partition its reachable conditional rows into $M$ disjoint, equal-cost groups $G_j$, each visited at most $c$ times per trajectory. Assume a known floor $q_i(a \mid h) \ge \beta$, $0 < \beta < 1$, on reachable rows and put $L = \log(1/\beta)$. For an anchor sample $X$, replay gives

$$Y_j(X) = \sum_t \mathbf{1}\{(H_t, I_t) \in G_j\} k_{I_t}(H_t) \in [0, cL], \qquad B_j = \mathbb{E}_p Y_j. \tag{25}$$

Write $F(J) = D_{\mathrm{KL}}(p \| P_{\mathcal{A}}^{q^J}) = \sum_{j \notin J} B_j$. Only local anchor conditionals, not the full joint or population benefits, are needed to evaluate a sampled $Y_j$.

**Theorem 10** (Replay-selected repair and held-out certification)**.** *From $N$ independent anchor samples, compute $\widehat{B}_j = N^{-1}\sum_s Y_j(X^{(s)})$ and let $\widehat{J}_b$ contain the $b$ largest estimates. For $0 < \delta_{\mathrm{sel}} < 1$, set*

$$\rho_N = cL\sqrt{\frac{\log(2M/\delta_{\mathrm{sel}})}{2N}}. \tag{26}$$

*With probability at least $1 - \delta_{\mathrm{sel}}$, simultaneously for $b = 0, \dots, M$,*

$$0 \le F(\widehat{J}_b) - \min_{|J|=b} F(J) \le 2\min(b, M-b)\rho_N. \tag{27}$$

*For $0 < \delta_{\mathrm{val}} < 1$ and an independent validation batch of size $N_{\mathrm{val}}$, let $C_b = (M-b)cL$ and $\bar{V}_b$ be the empirical mean of $\sum_{j \notin \widehat{J}_b} Y_j / C_b$. With probability at least $1 - \delta_{\mathrm{val}}$, simultaneously for $b < M$,*

$$\mathrm{TV}^2(p, P_{\mathcal{A}}^{q^{\widehat{J}_b}}) \le \tfrac{1}{2}F(\widehat{J}_b) \le \frac{C_b}{2}\mathcal{U}\left(\bar{V}_b, \frac{\log(M/\delta_{\mathrm{val}})}{N_{\mathrm{val}}}\right). \tag{28}$$

*For $b = M$, the error is exactly zero. Both conclusions hold jointly with probability at least $1 - \delta_{\mathrm{sel}} - \delta_{\mathrm{val}}$.*

*Proof.* A union bound over $M$ bounded-variable estimates controls every $|\widehat{B}_j - B_j|$ by $\rho_N$. The optimal and selected size-$b$ sets differ in at most $\min(b, M-b)$ groups on each side; empirical optimality then gives (27). Conditional on the selection data, the validation losses are independent across trajectories and bounded by $C_b$. Apply (10), union over $b < M$, and use Pinsker. Section I supplies all details. □

**An explicit polynomial-query regime.** For position groups, $M = n$ and $c = 1$. An excess-KL target $\varepsilon > 0$ at budget $b$ is guaranteed by

$$N \ge \frac{2\min(b, n-b)^2 \log^2(1/\beta)\log(2n/\delta_{\mathrm{sel}})}{\varepsilon^2}. \tag{29}$$

If every policy query has width at most $w$, anchor sampling and conditional evaluation use $O((N + N_{\mathrm{val}})nK^w)$ row calls, plus the cost of evaluating the frozen policy. Thus fixed $K, w$ and polynomially bounded $\log(1/\beta)$, $\log(1/\delta_{\mathrm{sel}})$, and $1/\varepsilon$ give a polynomial-query procedure without enumerating a mask library. A complementary fast-rate bound is

$$F(\widehat{J}_b) \le 4F_b^\star + \frac{4C_b}{N}\log\left(\frac{2M\binom{M}{b}}{\delta}\right), \qquad F_b^\star = \min_{|J|=b} F(J), \tag{30}$$

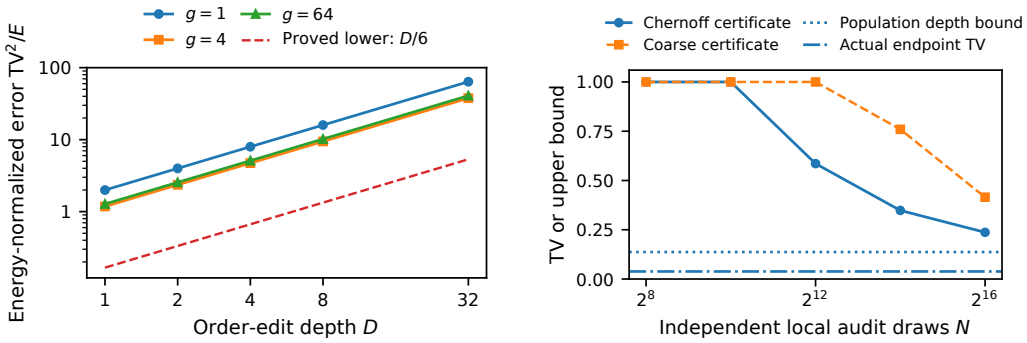



Figure 1: **Left:** exact depth scaling for the lower construction, with $\theta = 0.1/(D\sqrt{g})$; dividing by energy isolates the linear depth penalty. The dashed line is the proved lower bound; the full grid checks both bounds. **Right:** one finite local-audit run ($g = 4, D = 2, \theta = 0.025$), with simultaneous $95\%$ certificates over five sample budgets. The population depth bound is the limit of the certificate, not the actual endpoint TV.

with probability at least $1 - \delta$, simultaneously for $b < M$. In the small-residual regime, Section I gives explicit selection and held-out budgets proportional to $1/\varepsilon^2$ for a valid TV-$\varepsilon$ certificate. The combinatorial factor enters only through its logarithm; selection still sorts $M$ group scores.

The floor is used only to control KL estimation, not by the depth or adaptive-separation theorems. Replacing $k_i$ by $r_i$ and $L$ by 1 yields a floor-free selection guarantee for residual Hellinger energy and a TV certificate with factor 15 instead of $1/2$ (Section I). That version optimizes the additive certificate, not actual TV. Both procedures require anchor-conditional evaluation beyond a swap-score-only audit.

## 8 Finite-state validation and finite audit budgets

We distinguish algebraic checks, structural scaling, and sample-selected interventions. The original suite uses 16 positive five-bit tables (seeds $13000 + s$), all 120 fixed orders per table, and 15 random deterministic policies per table. All 28,842 checks pass at tolerance $3 \times 10^{-10}$, with largest identity residual $2.00 \times 10^{-15}$. The full identity and oracle-repair tables are in Section N; they are separate from the sample-based procedure below.

**Depth and the audit budget.** A 60-configuration grid checks the depth lower and upper bounds; 16 perturbed strongly correlated six-bit tables additionally check the dependence-free upper bound. Figure 1 makes the depth scaling and finite-sample overhead explicit. Its local audit samples actual swap occurrences, independent source-prefix rollouts, and pair mixtures with $(\kappa, \tau) = (1, 0)$. It records all row calls; the generic upper bound is $N(n + 2)$ for $N$ independent audit draws. The reported certificates retain the full sampling penalty rather than substituting population energy.

**Learning repairs rather than supplying oracle benefits.** For the same 16 tables, freeze the maximum-confidence state-to-position map. Select position groups using $N \in \{32, 128, 512, 2048\}$ independent ancestral anchor samples, then certify all six repair budgets using a separate 4096-sample replay batch. Anchor conditionals are computed by the incremental filter; full-joint enumeration is performed only after the chosen groups are frozen. Table 1 reports population KL for the learned sets, the oracle comparator, and held-out TV certificates. Confidence failure budgets are split across all 16 tables, four training budgets, and the nontrivial repair budgets; their joint coverage is at least $95\%$. At $N = 2048$, the largest excess KL over the oracle across all tables and repair budgets is 0.00776. After repairing three positions, every table has a nonvacuous TV certificate (largest bound 0.954). For the independent local audit, 4096 draws certify TV at most 0.586, using 37,156 row calls; the actual TV is 0.03837. All 5,143 additional recorded checks pass. Numerical checks use exhaustive floating-point sums or the disclosed samples, not formal proof verification.

Table 1: Sample-selected position repair: means over 16 five-bit tables. KL columns use independent population enumeration after selection. The last row averages per-table held-out TV bounds for $N = 2048$; confidence is simultaneous across the full reported collection. The oracle row is an evaluation comparator, not a selection input.

| Repaired positions $b$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Oracle residual KL | 3.227 | 1.880 | 0.971 | 0.468 | 0.122 | 0.000 |
| Replay-selected KL ($N = 128$) | 3.227 | 1.880 | 0.972 | 0.478 | 0.127 | 0.000 |
| Replay-selected KL ($N = 2048$) | 3.227 | 1.880 | 0.971 | 0.469 | 0.122 | 0.000 |
| Held-out TV certificate | 0.984 | 0.905 | 0.719 | 0.507 | 0.240 | 0.000 |

For the fixed–adaptive gap at $m = 2$, exhaustive enumeration of all $5040$ fixed orders gives maximum TV $0.0625$, whereas the adaptive decoder has TV $0.25$. Additional checks enumerate the exact coverage deficit (14) and the bounded-score Chernoff tails, and verify the finite-sample repair inequalities against independent population calculations. Section N gives the seeds, raw records, query accounting, and complete regeneration commands.

## 9 Discussion and limitations

The results identify distinct structural costs. Edit depth controls repeated order changes, replay coverage controls content-selected interfaces, and unresolved-prefix width controls exact conditional inference. The matching depth law and exponential adaptive separation make the first two costs unavoidable in the stated settings; the finite-budget theorem specifies when repair is both statistically controlled and computationally accessible.

The guarantees concern irreversible single-token revelation and a frozen policy. They exclude independent parallel draws, remasking, and recomputing a confidence policy after repair. Exact replacement is monotone in deterministic-policy forward KL, not randomized final-marginal KL or TV. KL selection requires a known probability floor, and low-width inference or well-covered local routes may be unavailable in large neural models. The synthetic experiments establish finite-state behavior, not language quality or neural inference speed. Accuracy relative to an external target retains the separate anchor-error term in (24).

## AI use statement

Generative AI (OpenAI ChatGPT) was used substantively for candidate statements and proofs, proof checking and revision, literature comparison, synthetic constructions and experiments, implementation and execution of verification code, analysis and figures, and manuscript/LaTeX drafting and editing. Synthetic tables and replay samples use disclosed seeded NumPy procedures. Executable checks are floating-point tests, not proof-assistant certification. No AI system is listed as an author.

## Ethics statement

The study uses mathematical models and synthetic distributions, with no human subjects, personal records, or language corpus. Consistency is not an accuracy or safety certificate: an inaccurate anchor can remain consistently inaccurate, and insufficient coverage can hide failures.

## Reproducibility statement

The appendices prove all formal results and specify seeds, grids, tolerances, and experimental interpretation. The anonymous supplement includes conditional-table, compiler, and frontier-filter implementations; both verification runners; sample-selection and held-out certification code; and raw records. Tables and figures regenerate from CSV/JSON. No GPU or external dataset is required. Selection, validation, and model-generation seeds are distinct.

## REFERENCES

Hasan Amin, Yuan Gao, Yaser Souri, Subhojit Som, Ming Yin, Rajiv Khanna, and Xia Song. Consistent diffusion language models. *arXiv preprint arXiv:2605.00161*, 2026. https://arxiv.org/abs/2605.00161.

Anonymous. Order-asymmetry in diffusion language models: Discrete curl and conditional compatibility. *ACL ARR 2026 May submission*, 2026. https://openreview.net/forum?id=dPmN9nom2I.

Dylan J. Foster, Yanjun Han, Jian Qian, and Alexander Rakhlin. Online estimation via offline estimation: An information-theoretic framework. *arXiv preprint arXiv:2404.10122*, 2024. https://arxiv.org/abs/2404.10122.

David Heckerman, David Maxwell Chickering, Christopher Meek, Robert Rounthwaite, and Carl Kadie. Dependency networks for inference, collaborative filtering, and data visualization. *Journal of Machine Learning Research*, 1:49–75, 2000. https://www.jmlr.org/papers/v1/heckerman00a.html.

Jeonseong Kim. Path-dependent denoising: A non-conservative field perspective on order collapse in diffusion language models. *arXiv preprint arXiv:2605.09303*, 2026. https://arxiv.org/abs/2605.09303.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. *arXiv preprint arXiv:2406.07524*, 2024. https://arxiv.org/abs/2406.07524.

Chuling Wen, Weijie Liang, and Jian Lu. Conditional total correlation and the serial depth of adaptive parallel sampling. *arXiv preprint arXiv:2608.25505*, 2026. https://arxiv.org/abs/2608.25505.

Tom Young, Yunan Chen, and Yang You. Inconsistencies in masked language models. *arXiv preprint arXiv:2301.00068v3*, 2024. https://arxiv.org/abs/2301.00068v3.

Shaorong Zhang, Longxuan Yu, Rob Brekelmans, Luhan Tang, Salman Asif, and Greg Ver Steeg. Generation order and parallel decoding in masked diffusion models: An information-theoretic perspective. *arXiv preprint arXiv:2602.00286*, 2026. https://arxiv.org/abs/2602.00286.

# A Conventions and exact path identities

All distributions in the proofs are on finite sets. Unless a counterexample states otherwise, every conditional row and every induced joint distribution is strictly positive. This makes all likelihood ratios, conditional distributions, and KL divergences well defined. Constants in the Hellinger results do not depend on the smallest probability. The symbol $h$ denotes both the revealed values and their labelled coordinate set; $h + (i = a)$ is the enlarged state. Every policy terminates after exactly $n$ one-position reveals. An initial prompt, when present, is fixed before all comparisons and absorbed into the conditional table.

The convention $H^2(P, Q) = 1 - \sum_x \sqrt{P(x)Q(x)}$ agrees with the square-root definition for normalized laws. The quantity $H$ is a metric because it is the Euclidean distance between square roots divided by $\sqrt{2}$. Cauchy–Schwarz gives

$$\begin{aligned} \mathrm{TV}(P, Q) &= \frac{1}{2} \sum_x |\sqrt{P} - \sqrt{Q}|(\sqrt{P} + \sqrt{Q}) \\ &\le \frac{1}{2} \left( \sum_x (\sqrt{P} - \sqrt{Q})^2 \right)^{1/2} \left( \sum_x (\sqrt{P} + \sqrt{Q})^2 \right)^{1/2} \le \sqrt{2} H(P, Q). \end{aligned} \tag{31}$$

For a marginalization map, the affinity cannot decrease: for each fiber $F$, $\sum_{x \in F} \sqrt{P(x)Q(x)} \le \sqrt{P(F)Q(F)}$. Thus $H^2$ contracts under marginalization. The same statement for a stochastic kernel follows by considering its joint extension and then marginalizing.

**Oriented curl sums.** Suppose the old order contains $(i, j)$ after a prefix $S$ and the new order contains $(j, i)$ there. Cancelling the common prefix and suffix gives

$$\log \frac{P_{\mathrm{new}}(x)}{P_{\mathrm{old}}(x)} = -c_{S;i,j}(x_S, x_i, x_j). \tag{32}$$

Consequently, if $\rho_0 = \sigma, \rho_1, \ldots, \rho_L = \tau$ is a route of individual adjacent swaps,

$$\log \frac{P_\tau(x)}{P_\sigma(x)} = \sum_{s=1}^{L} \log \frac{P_{\rho_s}(x)}{P_{\rho_{s-1}}(x)}. \tag{33}$$

Each summand is the oriented version in (32). This is the reveal-graph identity established by Kim (2026); it is used here as background, not as a novelty claim. It applies equally to a fixed prefix subset when generation starts from the empty state on that subset. Different routes give the same total even when individual curls are nonzero.

**Exact compatibility.** If all curls vanish, every adjacent swap leaves every path law unchanged, hence all fixed orders define a common law $p$. Use an order that reveals $S$, then $i$, then its complement. Summing over the complement gives $p(x_S, x_i) = p_S(x_S)q_i(x_i \mid x_S)$, so $q_i = p_i$ for every row. Conversely, common joint conditionals satisfy $q_i(a \mid h)q_j(b \mid h, i = a) = p(a, b \mid h) = q_j(b \mid h)q_i(a \mid h, j = b)$. Strict positivity is needed only to make all conditional rows identifiable and logarithms defined. This standard exact argument does not quantify approximate compatibility.

**The bounded score.** Writing $A/B = e^c$ yields $2\sqrt{AB}/(A + B) = \mathrm{sech}(c/2)$. Therefore $z = 1 - \mathrm{sech}(c/2)$ and

$$\sum_{a,b} \frac{A(a, b) + B(a, b)}{2} z(a, b) = \frac{1}{2} \sum_{a,b} (\sqrt{A} - \sqrt{B})^2 = H^2(A, B). \tag{34}$$

The Taylor expansion follows from $\cosh t = 1 + t^2/2 + O(t^4)$. Neither taking the data expectation of $c^2$ nor replacing the pair-mixture law by the data pair law is implicit in this identity.

# B A self-contained one-sided Hellinger chain bound

The bound in this section is a standard probability tool. The known constant 7 in Foster et al. (2024, Lemma D.2) is sharper than the constant used below. We include the stopping argument so no unproved probabilistic lemma is needed for the depth bound.

**Lemma 11** (One-sided chain inequality, constant $15/2$)**.** *Let $P, Q$ be positive sequential laws on $X_1, \dots, X_N$, with potentially different conditional distributions at every history. Put $d_t(X_{<t}) = H^2(P_t(\cdot \mid X_{<t}), Q_t(\cdot \mid X_{<t}))$. Then*

$$H^2(P, Q) \le \frac{15}{2} \mathbb{E}_P \sum_{t=1}^{N} d_t(X_{<t}). \tag{35}$$

*Proof.* Define the prefix likelihood ratio $Z_t = Q(X_{\le t})/P(X_{\le t})$, with $Z_0 = 1$. Conditional expectation gives

$$\mathbb{E}_P[\sqrt{Z_t} \mid X_{<t}] = \sqrt{Z_{t-1}} \sum_a \sqrt{P_t(a \mid X_{<t}) Q_t(a \mid X_{<t})} = \sqrt{Z_{t-1}}(1 - d_t). \tag{36}$$

Let $\tau$ be the first time $t$ for which $Z_t \ge 9$, or $N$ if the threshold has not been hit earlier. The stopped observation is the pair $(\tau, X_{\le \tau})$. The stopping rule is the same measurable function of prefixes for $P$ and $Q$. Its likelihood ratio is $Z_\tau$, and both stopped laws are normalized.

Because $\{t \le \tau\}$ is determined by $X_{<t}$, telescoping (36) gives

$$H^2(P^\tau, Q^\tau) = 1 - \mathbb{E}_P \sqrt{Z_\tau} = \mathbb{E}_P \sum_{t=1}^{N} \mathbf{1}\{t \le \tau\} \sqrt{Z_{t-1}} d_t \le 3 \mathbb{E}_P \sum_{t=1}^{N} d_t. \tag{37}$$

The last step uses $Z_{t-1} < 9$ on $\{t \le \tau\}$, including the final possible threshold crossing.

Factor the full affinity through the stopped record. For a record $s = (\tau, x_{\le \tau})$ with $\tau < N$, denote by $a_s \in [0, 1]$ the conditional affinity of the remaining suffix. For records with $\tau = N$, set $a_s = 1$. Then

$$H^2(P, Q) - H^2(P^\tau, Q^\tau) = \mathbb{E}_P\big[\sqrt{Z_\tau}(1 - a_s)\big] \le \mathbb{E}_P\big[\sqrt{Z_\tau} \mathbf{1}\{\tau < N\}\big]. \tag{38}$$

On $\{\tau < N\}$, let $r = \sqrt{Z_\tau} \ge 3$. The elementary inequality $r \le \frac{3}{4}(r-1)^2$ follows by rearranging $3r^2 - 10r + 3 \ge 0$, whose larger root is 3. Hence

$$\mathbb{E}_P\big[\sqrt{Z_\tau} \mathbf{1}\{\tau < N\}\big] \le \frac{3}{4} \mathbb{E}_P(\sqrt{Z_\tau} - 1)^2 = \frac{3}{2} H^2(P^\tau, Q^\tau). \tag{39}$$

Combining with (37) and (38) proves the result. No lower bound on a row probability, independence condition, or likelihood-ratio moment bound is used. □

The proof works on a finite tree whose next alphabet depends on its observed history. It also works for a fixed upper bound on the number of steps, by padding terminated paths with a common deterministic symbol. Structural zero probabilities for impossible histories can be omitted from the tree. All positive token kernels in the paper produce the same feasible policy tree under the two compared processes.

## C SWAP LAYERS AND THEIR ADAPTIVE EXTENSION

We give the detailed blocking argument behind Theorem 2. A layer in a permutation is specified by starting positions $J \subseteq \{1, \dots, n-1\}$ such that $|j - j'| \ge 2$ for distinct $j, j' \in J$. Partition the permutation into its selected two-position blocks and its unselected singleton blocks, preserving order. The target permutation has the same ordered blocks, with each selected block reversed.

For any given sequence of completed block values, both processes have exactly the same revealed state at the start of the next block. In the source process, a selected block has joint kernel $A_f$; in the target, it has joint kernel $B_f$, with the two variables recorded in a common labelled order. The prefix distribution of this block under the source process is exactly the corresponding source-prefix distribution in the original single-token decoder. This remains true even when an earlier flipped block changes the distribution of later contexts. Those changes are already present in the source law used by Lemma 11.

It follows that

$$H^2(P_{\pi^{(\ell-1)}}, P_{\pi^{(\ell)}}) \le \frac{15}{2} \sum_{f \text{ in layer } \ell} \mathbb{E}_{R_{\ell f}} e_f = \frac{15}{2} E_\ell. \tag{40}$$

The block transcript and the final labelled vector are in bijection for a fixed layer, so this is also the endpoint Hellinger distance. Applying the metric triangle inequality and then Cauchy–Schwarz gives (7). The $D = 0$ case is interpreted separately: the endpoints are identical, with zero distance.

**Shared adaptive block rules.** At each stage, let a common nonanticipating rule choose a singleton or an unordered pair from the remaining positions, using only completed blocks and their values. The two processes use this same selection rule and may differ only in the internal orientation of a selected pair. The rule must not inspect an internal reveal before deciding which positions belong to that pair, nor distinguish the two internal histories once the pair is complete. On the completed-block transcript, the action-selection kernel is identical, and the pair kernels are again $A_f, B_f$. Lemma 11 and final-output data processing give the same one-layer bound. A sequence of such comparisons gives the same route bound.

This extension is not an unrestricted representation theorem for adaptive decoders. If one decoder reveals $i$, inspects its value, and then decides which $j$ to reveal, while the other must choose $j$ before seeing $i$, a common unordered two-position block need not exist. The adaptive counterexample exploits precisely the ability to select the revealed set based on earlier content.

# D A MATCHING LOWER CONSTRUCTION

We prove Theorem 3 with explicit estimates. Each of $g$ disjoint groups has a main spin and $D$ auxiliary spins. Every auxiliary conditional is fair for every mask and context. If exactly $k$ auxiliaries of its own group have been observed, the main conditional has law

$$p_{k\theta}(x) = \frac{e^{k\theta x}}{2\cosh(k\theta)}, \qquad x \in \{-1, 1\}. \tag{41}$$

It does not depend on any observed value, or on observed coordinates in other groups. Thus every fixed-order output law is a product measure. The anchor order places the main first in each group; the target places it last. Inside each group, move the main one position to the right at every layer. Swaps from different groups are disjoint, yielding $D$ layers and $gD$ swaps. A single coordinate moves at most one slot per layer, so no standard swap-layer route between these particular endpoint permutations can have fewer than $D$ layers.

## D.1 LOCAL ENERGIES

For the swap crossing the $(k+1)$-st auxiliary, the auxiliary is fair in both directions and independent of the main value. Therefore its local Hellinger energy is

$$e_k = H^2(p_{k\theta}, p_{(k+1)\theta}) = 1 - \frac{\cosh((k+\frac{1}{2})\theta)}{\sqrt{\cosh(k\theta)\cosh((k+1)\theta)}}. \tag{42}$$

Let $F(t) = \log\cosh t$. The logarithm of the reciprocal affinity is the midpoint Jensen gap

$$\Delta_k = \frac{F(k\theta) + F((k+1)\theta)}{2} - F((k+\tfrac{1}{2})\theta). \tag{43}$$

Since $F''(t) = \operatorname{sech}^2 t$, Taylor's theorem with integral remainder bounds this gap by

$$\frac{\theta^2}{8} \inf_{|t| \le D|\theta|} \operatorname{sech}^2 t \le \Delta_k \le \frac{\theta^2}{8}. \tag{44}$$

On the theorem's parameter range, $D|\theta| \le 1/(2\sqrt{3}) < 1$ and $|\theta| < 1$, so $0 \le \Delta_k < 1$. The inequality $u/2 \le 1 - e^{-u} \le u$ for $u \in [0, 1]$ gives

$$\frac{\operatorname{sech}^2(1)}{16}\theta^2 \le e_k \le \frac{\theta^2}{8}. \tag{45}$$

The energy is independent of the prefix values. Thus $E = g\sum_{k=0}^{D-1} e_k = \Theta(gD\theta^2)$.

## D.2 ENDPOINT TOTAL VARIATION

Auxiliaries have the same law at both endpoints and can be suppressed. The $g$ main spins are fair under $P_0$ and independently tilted by $a = D\theta$ under $P_1$. Under $P_0$, let $S = \sum_{r=1}^{g} X_r$ and

$$R_a = \frac{P_1(X)}{P_0(X)} = \exp(aS - g\log\cosh a). \tag{46}$$

Its first derivative at zero is $S$, and

$$R_a'' = R_a\big[(S - g\tanh a)^2 - g\,\mathrm{sech}^2\, a\big]. \tag{47}$$

Under the tilted law, $\mathbb{E}S = g\tanh a$ and $\mathrm{Var}\, S = g\,\mathrm{sech}^2\, a$. Hence $\mathbb{E}_0|R_a''| \le 2g\,\mathrm{sech}^2\, a \le 2g$ for every real $a$. Integrating the second derivative twice, also for negative $a$, gives

$$\mathbb{E}_0|R_a - 1 - aS| \le ga^2. \tag{48}$$

For the fair sum, $\mathbb{E}S^2 = g$ and $\mathbb{E}S^4 = 3g^2 - 2g \le 3g^2$. Hölder's inequality in the form $\mathbb{E}|S|^2 \le (\mathbb{E}|S|)^{2/3}(\mathbb{E}|S|^4)^{1/3}$ therefore implies $\mathbb{E}|S| \ge \sqrt{g/3}$. It follows that

$$\mathrm{TV}(P_0, P_1) = \frac{1}{2}\mathbb{E}_0|R_a - 1| \ge \frac{|a|\sqrt{g}}{2\sqrt{3}} - \frac{ga^2}{2}. \tag{49}$$

For $|a|\sqrt{g} \le 1/(2\sqrt{3})$, this is at least $|a|\sqrt{g}/(4\sqrt{3})$, and hence $\mathrm{TV}^2 \ge ga^2/48$. On the other hand, (45) gives $DE \le gD^2\theta^2/8 = ga^2/8$. Consequently

$$\mathrm{TV}^2(P_0, P_1) \ge DE/6. \tag{50}$$

The upper bound is Theorem 2. This proves the claimed parameter-order matching without asserting optimal universal constants.

# E Coverage-aware concentration and its necessity

## E.1 Transfer of a bounded score

For arbitrary laws $\mu, \nu$ and any measurable $0 \le f \le 1$,

$$\mathbb{E}_\mu f - \kappa\mathbb{E}_\nu f = \sum_\omega f(\omega)(\mu(\omega) - \kappa\nu(\omega)) \le \sum_\omega [\mu(\omega) - \kappa\nu(\omega)]_+. \tag{51}$$

This proves the coverage transfer used throughout the paper. The deficit is an absolute missing-mass cost, not a multiplicative quantity. A support point with very small data mass may still have a large contribution to the deficit under a decoder's visitation law.

For the route audit, each draw first chooses one of the $m$ swap occurrences, not merely a pair of position labels. Repeated appearances of a pair with different masks or source layers are distinct occurrences. Sampling that occurrence's source prefix and local pair mixture yields

$$\mathbb{E}_\mu z = \frac{1}{m}\sum_{\ell,f} \mathbb{E}_{R_{\ell f}}\mathbb{E}_{(A_f+B_f)/2} z_f = \frac{E}{m}. \tag{52}$$

No assumption is made that swaps from different occurrences share the same context distribution.

## E.2 Chernoff inversion for bounded scores

For $x, v \in [0, 1]$, write

$$d_{\mathrm{B}}(x\|v) = x\log(x/v) + (1-x)\log((1-x)/(1-v)), \tag{53}$$

with the continuous boundary conventions and infinite value for a positive mass divided by zero. Let $Y_1, \dots, Y_N$ be independent $[0, 1]$ variables with common mean $u$, and put $\bar{Y} = N^{-1}\sum_j Y_j$. Convexity of the exponential gives, for any real $t$,

$$\mathbb{E}e^{tY_j} \le 1 - u + ue^t. \tag{54}$$

For $0 \le x < u$, use Markov's inequality with $t < 0$ and optimize:

$$\Pr(\bar{Y} \le x) \le \inf_{t<0} \exp\{N[\log(1 - u + ue^t) - tx]\} = e^{-Nd_{\mathrm{B}}(x\|u)}. \tag{55}$$

For $0 < x < u < 1$ the optimizer is $e^t = x(1-u)/(u(1-x))$; the endpoint $x = 0$ follows by taking a limit.

The event $u > \mathcal{U}(\bar{Y}, a)$ implies $\bar{Y} < u$ and $d_{\mathrm{B}}(\bar{Y}\|u) > a$. On $[0, u]$, $d_{\mathrm{B}}(x\|u)$ decreases with $x$. Thus this event is a lower tail ending at the unique root of $d_{\mathrm{B}}(x\|u) = a$, if such a root exists; otherwise it is empty. Equation (55) and a one-sided limit give

$$\Pr\{u > \mathcal{U}(\bar{Y}, a)\} \le e^{-Na}. \tag{56}$$

If $u = 0$ or $u = 1$, every observation equals $u$ almost surely and the statement is immediate. Taking $a = \log(1/\delta)/N$, then applying the coverage transfer and Theorem 2, proves Corollary 4. No Bernoulli-valued assumption on the observed score is used.

At $\bar{Y} = 0$, solving $d_{\mathrm{B}}(0\|v) = -\log(1-v) \le a$ gives $\mathcal{U}(0, a) = 1 - e^{-a}$. Substituting this into (11) with $(\kappa, \tau) = (1, 0)$ gives (12); its asymptotic expression uses $-\log(1-t) \sim t$. With coverage deficit $\tau$, even infinitely many zero-score observations leave the term $15Dm\tau$ in this certificate.

**A useful deterministic relaxation.** The variational identity for binary relative entropy, which also follows by differentiating its two-point log moment generating function, gives

$$d_{\mathrm{B}}(s\|v) \ge -s\log 2 - \log(1 - v/2) \ge v/2 - s\log 2. \tag{57}$$

Consequently every feasible $v$ in (10) is at most $2a + 2s\log 2$, and

$$\mathcal{U}(s, a) \le \min\{1, 2s + 2a\}. \tag{58}$$

The still simpler $2\bar{Y} + 8\log(1/\delta)/N$ expression in the main text follows as a conservative relaxation. We retain its numerical comparison to distinguish the gain from endpoint inversion from the structural depth improvement.

For later use, (54) with $t = 1$ yields $\mathbb{E}e^{Y_j} \le e^{(e-1)u}$, so

$$\Pr\{\bar{Y} > 2u + a\} \le \exp\{Nu(e-3) - Na\} \le e^{-Na}. \tag{59}$$

Combining this with (58), with probability at least $1 - e^{-Na}$,

$$\mathcal{U}(\bar{Y}, a) \le 4u + 4a. \tag{60}$$

This last event bounds the *size* of the certificate; (56) separately guarantees its validity.

If several statistics use the same independent samples, the statistics can be dependent across certificates; finite union bounds remain valid. The independent observations are complete trajectories or independently initiated audit draws, not the reveals inside one trajectory. A route chosen outside the controlled candidate collection requires a separate analysis or held-out audit.

### E.3 Two-position coverage counterexample

On the four states $(0,0), (0,1), (1,0), (1,1)$, define

$$P_0 = (1/4, 1/2, 1/8, 1/8), \qquad P_1 = (1/4, 1/8, 1/2, 1/8). \tag{61}$$

Set the two rows used by order $(1, 2)$ to the marginal/conditional factorization of $P_0$, and the two rows used by order $(2, 1)$ to those of $P_1$. They are disjoint interfaces, so this defines a valid positive masked table. The two local kernels are precisely $A = P_0, B = P_1$.

The endpoint distance is $\mathrm{TV}(P_0, P_1) = 3/8$. The score in (3) is zero at the first and fourth states. At either middle state its value is

$$\frac{(\sqrt{1/2} - \sqrt{1/8})^2}{1/2 + 1/8} = \frac{1}{5}. \tag{62}$$

For the strictly positive audit law $\nu_\eta = (1 - \eta, \eta/3, \eta/3, \eta/3)$ with $0 < \eta < 1$, we get $\mathbb{E}_{\nu_\eta} z = 2\eta/15 \to 0$ while endpoint TV stays $3/8$. The local pair-mixture law $(P_0 + P_1)/2$ assigns positive constant mass to the erroneous states. Any valid coverage transfer must therefore incur growing density amplification or a nonvanishing deficit.

# F ADAPTIVE REPLAY AND THE EXPONENTIAL GAP

## F.1 DETERMINISTIC ADAPTIVE FACTORIZATION

Fix a deterministic nonanticipating policy $\mathcal{A}(h)$ choosing an unrevealed coordinate from the labelled observed values. Given a complete vector $x$, it induces a unique sequence of positions $i_1(x), \dots, i_n(x)$. The event that a particular partial path has been followed is determined by the revealed values along that path. It imposes no additional restriction on the unrevealed coordinates. Conditional on a feasible revealed history $h = x_S$ under $X \sim p$, the next requested coordinate therefore has distribution $p_i(\cdot \mid h)$.

Along that same path, multiplication of the anchor conditionals telescopes:

$$\prod_{t=1}^{n} p_{i_t(x)}(x_{i_t(x)} \mid h_t(x)) = \prod_{t=1}^{n} \frac{p(x_{S_t})}{p(x_{S_{t-1}})} = p(x). \tag{63}$$

The model's output probability is the analogous product of $q$ rows. It is normalized, since it is the leaf law of a finite stochastic decision tree. Taking the logarithm of their ratio, then the expectation under $p$, gives

$$\begin{aligned} D_{\mathrm{KL}}(p \| P^q_{\mathcal{A}}) &= \mathbb{E}_{p,\mathcal{A}} \sum_t \log \frac{p_{I_t}(X_{I_t} \mid H_t)}{q_{I_t}(X_{I_t} \mid H_t)} \\ &= \mathbb{E}_{p,\mathcal{A}} \sum_t k_{I_t}(H_t). \end{aligned} \tag{64}$$

This is a chain-rule application, not a new entropy identity. Applying Lemma 11 to the labelled reveal transcript and using (31) proves the Hellinger inequality in (16).

## F.2 RANDOMIZED POLICIES AND OCCUPANCY

A randomized policy can be represented by an independent seed or by common action-selection kernels on the history tree. Compare two transcript laws: one uses anchor token conditionals and the other uses model token conditionals, but both use the same action rule. The action terms cancel in the log-ratio. The transcript KL is therefore the replay sum of conditional KLs. Summing over the seed/action transcript is a marginalization, so the final-output KL is only bounded above by that sum. The same construction and Hellinger contraction give the final-output TV bound in (16).

Every trajectory contains $n$ actions, so $d^p_{\mathcal{A}}(h,i) = n^{-1}\mathbb{E}\sum_t \mathbf{1}\{(H_t, I_t) = (h,i)\}$ is a probability law on rows. The coverage transfer for $r_i(h) \in [0,1]$ immediately gives

$$\mathrm{TV}^2(p, P^q_{\mathcal{A}}) \le 15n \sum_{h,i} d^p_{\mathcal{A}}(h,i) r_i(h) \le 15n(\kappa \mathbb{E}_\nu r + \tau). \tag{65}$$

The same bound holds with residual row errors after repair. An independent seed must be replayed with the same distribution for this interpretation.

## F.3 PROOF OF THEOREM 5

The selector construction in the main text defines all conditional rows, including masks that no selected policy will encounter. All selectors and markers are always fair. The target $X$ is fair unless the exact trigger condition holds; its only other row is Bernoulli$(3/4)$. Positivity and the claimed probability range follow immediately. The anchor revealing $X$ first is uniform.

Fix a permutation. Its revealed set immediately before $X$ is deterministic. Unless this set consists of all selectors and exactly one marker from each pair, the trigger never occurs and the entire output is uniform. If it has this form, let $v \in \{0,1\}^m$ record the revealed marker labels. The trigger event is $U = v$, which has probability $2^{-m}$. Given this event, only $X$ differs from the uniform anchor, by TV $1/4$; given its complement, all coordinates remain uniform. Since the two cases are disjoint in the observed selector values, the joint TV is $2^{-m}/4$. Such a fixed permutation exists, so the maximum is exactly this value.

The adaptive policy first reveals every selector, then marker $B_j^{U_j}$ for each $j$, then $X$, and finally all remaining markers. The trigger is certain. Every selector and marker is still a fair independent bit in the final labelled output; the target is Bernoulli$(3/4)$ independently of their values. Hence the final joint TV from the uniform anchor is $1/4$.

The lower example does not violate the fixed-order depth theorem. It shows instead that a uniform small error bound over fixed orders does not control a content-selected mixture of their context-specific path decisions. A content-adaptive policy is not generally a convex mixture of fixed-order output laws with weights independent of $x$.

### F.4 Exact coverage price of the adaptive selector

Let $V \in \{0,1\}^m$ denote the marker labels included in the target's context; $U$ are the observed selector values, and $W \in \{0,1\}^m$ are the observed values of those markers. Both audit laws query the target after all selectors and exactly one marker per pair have been revealed. Under the independent-mask audit, $V$ is uniform and the context is drawn from the uniform anchor marginal on that mask, giving

$$\nu_{\mathrm{mask}}(v,u,w) = 2^{-3m}. \tag{66}$$

Under adaptive anchor replay, $V = U$, and $U, W$ are independent uniform vectors, giving

$$\mu_{\mathrm{sel}}(v,u,w) = 2^{-2m}\mathbf{1}\{v = u\}. \tag{67}$$

The likelihood ratio is $2^m$ on the matching set $T = \{V = U\}$ and zero elsewhere. This set has masses $\mu_{\mathrm{sel}}(T) = 1$ and $\nu_{\mathrm{mask}}(T) = 2^{-m}$. Summing the positive part of their difference proves (14) for every $\kappa \geq 0$.

The target's conditional Hellinger error equals $r_\star = H^2(\mathrm{Bern}(1/2), \mathrm{Bern}(3/4))$ on $T$ and is zero off $T$. Its means are therefore $r_\star$ under replay and $2^{-m}r_\star$ under the independent-mask audit. Independent draws from the latter all miss a biased row with probability $(1-2^{-m})^N$. Seeing at least one such row with probability at least $1-\delta$ requires

$$N \geq \frac{\log(1/\delta)}{-\log(1-2^{-m})} = \Theta(2^m \log(1/\delta)). \tag{68}$$

This is the detection cost of the specified independent-mask sampling design, not a lower bound for an auditor allowed to choose target masks after reading $U$. Such an auditor can implement replay and always hit $T$. The exponential loss in (14) is thus an exact property of the audit law, not an unspecified distribution-shift constant.

## G The local-to-conditional bridge

**Lemma 12** (Conditional Hellinger contraction)**.** *For positive joint laws $P, Q$ of $(S, A)$,*

$$\mathbb{E}_{s\sim P_S} H^2(P_{A|s}, Q_{A|s}) \leq 2H^2(P,Q). \tag{69}$$

*Proof.* Let $d_s = H^2(P_{A|s}, Q_{A|s}) \in [0,1]$ and $z_s = \sqrt{Q_S(s)/P_S(s)}$. Decomposing the joint affinity gives

$$H^2(P,Q) = \mathbb{E}_{P_S}\left[\frac{1}{2}(1-z_s)^2 + z_s d_s\right]. \tag{70}$$

For any $z \geq 0$ and $d \in [0,1]$,

$$2\left[\frac{1}{2}(1-z)^2 + zd\right] - d = (z-(1-d))^2 + d(1-d) \geq 0. \tag{71}$$

Average this pointwise inequality under $P_S$. □

The left side uses the reference marginal $P_S$, not the other marginal and not a supremum over contexts. It is this weighting that avoids an inverse-probability factor for rare individual contexts. A pointwise conditional guarantee need not hold.

**Applying the lemma to a mask/action type.** For $\eta = (S, i)$, marginalize $p_T, Q_\eta$ to $(X_S, X_i)$. Because $R_\eta$ in (17) is normalized,

$$(Q_\eta)_{S\cup\{i\}}(h, a) = A_S(h)q_i(a \mid h). \tag{72}$$

Its conditional at $i$ given $S$ is exactly $q_i$. Lemma 12 and marginal Hellinger contraction imply

$$\mathbb{E}_{p_S} r_i \le 2H^2\big(p_{S\cup\{i\}}, (Q_\eta)_{S\cup\{i\}}\big) \le 2H^2(p_T, Q_\eta). \tag{73}$$

The layer proof, before converting to TV, gives $H^2(p_T, Q_\eta) \le (15/2)(\sum_\ell \sqrt{E_{\eta\ell}})^2$. This proves (18).

**Uniformity over a mask/action library.** Let $\mathcal{L}$ be a finite set of types $(S, i)$, not a list of individual sampled contexts. A policy is allowed to use only these types, but can choose between them based on all revealed content. Under anchor replay, each type can occur at most once: the revealed set grows strictly at each step. For a type not repaired, the expected contribution is

$$\mathbb{E}_{X\sim p,\,\mathcal{A}}\big[\mathbf{1}\{\mathcal{A} \text{ visits } (S, i)\}\, r_i(X_S)\big] \le \mathbb{E}_{X\sim p} r_i(X_S) = \mathbb{E}_{p_S} r_i. \tag{74}$$

The inequality holds after averaging any random policy seed as well. For repaired types, every possible value-row for that type uses $p_i$, so the contribution is zero. Sum over the unrepaired library, use (18), and then the replay Hellinger bound. This gives the constant $15 \cdot 15 = 225$ in (19).

This corollary can be numerically loose because different types may never occur on the same path. It provides a uniform structural certificate, not an exact characterization of adaptive error. Its cost cannot be omitted merely because each local diagnostic is easy. Audit all needed $E_{\eta\ell}$ under appropriate source-prefix coverage, with simultaneous confidence, before selecting $\mathcal{R}$ from their bounds. The number of such certificates is part of the sampling budget.

# H Conditional compilation and selective repair

## H.1 Full derivation of the normalized formula

For the identity-order anchor, its marginal on $T = [r]$ is

$$p_T(x_T) = \prod_{j=1}^{r} q_j(x_j \mid x_{<j}), \tag{75}$$

because summing the original path product over $r+1, \dots, n$ leaves one. For fixed $S \subseteq T$, $i \in T \setminus S$, define $U = T \setminus (S \cup \{i\})$ and the order $\tau$ in the main text. The corresponding prefix product is (17), with

$$A_S(h) = \prod_{j\in S \text{ increasing}} q_j(x_j \mid x_{S\cap[j-1]}), \tag{76}$$

and $R_\eta$ the product of the remaining $q$ rows in increasing $U$ order, conditional on $h, i = a$. Both $A_S$ and $R_\eta$ are normalized in their respective variables. Define $C_\eta = \log Q_\eta - \log p_T$. Then

$$\begin{aligned} p_{S\cup\{i\}}(h, a) &= \sum_u Q_\eta(h, a, u) e^{-C_\eta(h,a,u)} \\ &= A_S(h)q_i(a \mid h) \sum_u R_\eta(u \mid h, a) e^{-C_\eta(h,a,u)} \\ &= A_S(h)q_i(a \mid h) Z_i(a \mid h). \end{aligned} \tag{77}$$

Dividing by the sum over $a$ proves (20). The curl route computes the same $C_\eta$ for every completion, by (33). In particular, no arbitrary choice of a global normalization for raw curls is being made.

The formula is a version of change of measure followed by conditional normalization. The marginalization operation is essential: replacing $Z_i(a \mid h)$ by $e^{-C_\eta(h,a,u_0)}$ for a single completion $u_0$ need not yield the anchor conditional. Approximate completion integration needs the certified error discussed in Section M.

### H.2 Exact repair gain and position groups

For a fixed deterministic policy, the path chosen for each $x$ is the same function of $x$ before and after replacing conditionals. The anchor replay law is therefore also unchanged. Applying the KL chain identity to $q^{\mathcal{R}}$ gives

$$D_{\mathrm{KL}}(p\|P_{\mathcal{A}}^{q^{\mathcal{R}}}) = \mathbb{E}_{p,\mathcal{A}} \sum_t D_{\mathrm{KL}}\big(p_{I_t}(\cdot \mid H_t)\|q_{I_t}^{\mathcal{R}}(\cdot \mid H_t)\big). \tag{78}$$

For selected rows the local divergence is zero; otherwise it is $k_{I_t}(H_t)$. This proves Proposition 8. Subtracting from the same identity for $q$ gives the exact improvement. Nested row sets consequently give monotone forward KL, but neither monotone TV nor monotone reverse KL is asserted.

Let disjoint groups $G_1, \dots, G_M$ partition the available repair rows, with benefits

$$B_j = \mathbb{E}_{p,\mathcal{A}} \sum_t \mathbf{1}\{(H_t, I_t) \in G_j\} k_{I_t}(H_t). \tag{79}$$

The gain from repairing groups $J$ is exactly $\sum_{j\in J} B_j$. For a cardinality budget $|J| \le b$, replacing any chosen smaller benefit by an unchosen larger one cannot decrease the gain. Hence the $b$ largest benefits are optimal. The deterministic optimization identity assumes equal group costs; costs of different groups need not coincide in a neural implementation. The next section supplies statistical guarantees when the benefits are estimated from replay.

## I Finite-sample selection, certification, and explicit budgets

### I.1 Bounded replay losses and the sampling interface

The policy, row groups, and probability floor in Theorem 10 are fixed before the selection samples are drawn. The floor is needed only on rows that this fixed policy can query. Since $\sum_a p_i(a \mid h) \log p_i(a \mid h) \le 0$ and $-\log q_i(a \mid h) \le L$,

$$0 \le k_i(h) \le L = \log(1/\beta). \tag{80}$$

At most $c$ visits to one group imply $0 \le Y_j(X) \le cL$. Distinct groups on a trajectory may be arbitrarily dependent. Independence is used only between the complete anchor samples $X^{(1)}, \dots, X^{(N)}$.

Each sample is obtained by $n$ anchor-order calls. Replay chooses $i = \mathcal{A}(h)$, computes the entire anchor row $p_i(\cdot \mid h)$, evaluates the row divergence to $q_i$, and reveals the already sampled $X_i$. In particular, it does not estimate $k_i$ from the single realized log ratio; that realized log ratio can be negative or unbounded even when the row KL has a useful upper bound. The incremental filter supplies each exact anchor row without constructing the full joint.

### I.2 Uniform group estimates and the additive oracle inequality

For a random variable $Z$ in an interval of length $a$, let $g(t) = \log \mathbb{E}e^{tZ}$. Its second derivative is the variance under the exponential tilt, and is at most $a^2/4$. Twice integrating gives $g(t) - g(0) - tg'(0) \le t^2a^2/8$ for positive or negative $t$. Applied independently to $Y_j(X^{(s)})$, exponential Markov bounds and optimization yield

$$\Pr\{|\widehat{B}_j - B_j| > r\} \le 2\exp\left(-\frac{2Nr^2}{c^2L^2}\right). \tag{81}$$

The radius (26) and a union bound control every group with probability at least $1 - \delta_{\mathrm{sel}}$.

Work on this event. Fix $b$ and an optimal size-$b$ group set $J_b^\star$. The identity of Proposition 8 gives

$$F(\widehat{J}_b) - F(J_b^\star) = \sum_{j \in J_b^\star \setminus \widehat{J}_b} B_j - \sum_{j \in \widehat{J}_b \setminus J_b^\star} B_j. \tag{82}$$

The corresponding difference of empirical sums is nonpositive because $\widehat{J}_b$ maximizes the size-$b$ empirical sum. The two difference sets have the same size $s \le \min(b, M-b)$. Replacing each of their $2s$ terms by its estimate adds at most $2s\rho_N$, proving (27) for every $b$ on the same event. The bound vanishes at $b = 0, M$ as it should. The lower inequality follows from the definition of the optimum. Substituting $M = n, c = 1$ and solving $2\min(b, n-b)\rho_N \le \varepsilon$ proves (29).

### I.3 Independent validation and adaptive choice of the final budget

Condition on the entire selection batch, so all $\widehat{J}_b$ are fixed. For $b < M$, define the validation loss

$$W_b(X) = \frac{\sum_{j \notin \widehat{J}_b} Y_j(X)}{C_b}, \qquad C_b = (M-b)cL. \tag{83}$$

It lies in $[0,1]$ and has mean $F(\widehat{J}_b)/C_b$. Equation (56) with $a = \log(M/\delta_{\mathrm{val}})/N_{\mathrm{val}}$, followed by a union bound over $b = 0, \dots, M-1$, gives all the KL bounds in (28). Pinsker gives their TV form. Removing the conditioning preserves the failure probability, and combining with the selection event gives the stated joint coverage. The full-repair case has exactly zero error and does not need a validation sample.

Because the validation event covers every budget in the independently selected nested family, the final $b$ may be chosen after viewing these bounds, for example as the smallest budget certified below a specified tolerance. The family of repair sets itself must remain fixed independently of the validation batch; arbitrary reselection of groups on that batch is not justified by this argument. The same distinction applies when several models or selection sample sizes share a validation batch: their finite union must enter the confidence allocation, as it does in the reported experiment.

### I.4 A complementary fast-rate bound and a nonvacuous regime

The additive oracle inequality is useful at arbitrary residual levels. A relative bound gives a sharper sampling dependence when the best residual is small. Put $F_b^\star = \min_{|J|=b} F(J)$, $C_b = (M-b)cL$, and, for $b < M$,

$$a_b = \frac{\log\Big(2M\binom{M}{b}/\delta\Big)}{N}. \tag{84}$$

For each fixed size-$b$ set $J$, its normalized residual loss is in $[0,1]$. Equations (56), (58), and (59) show that, except on a set of probability at most $2e^{-Na_b}$,

$$F(J) \le 2\widehat{F}(J) + 2C_b a_b, \qquad \widehat{F}(J) \le 2F(J) + C_b a_b, \tag{85}$$

where $\widehat{F}(J) = \sum_{j \notin J} \widehat{B}_j$. Union over all $\binom{M}{b}$ sets and all $b < M$. With probability at least $1-\delta$, empirical minimization gives, simultaneously,

$$F(\widehat{J}_b) \le 4F_b^\star + 4C_b a_b. \tag{86}$$

This is (30). The $\binom{M}{b}$ factor enters only logarithmically in this analysis; implementation still sorts the $M$ empirical benefits, rather than searching over subsets.

Let $a_{\mathrm{val}} = \log(M/\delta)/N_{\mathrm{val}}$ and denote the held-out KL bound by $U_b^{\mathrm{KL}} = C_b \mathcal{U}(\bar{V}_b, a_{\mathrm{val}})$. Conditional on selection, (60) and another union bound give $U_b^{\mathrm{KL}} \le 4F(\widehat{J}_b) + 4C_b a_{\mathrm{val}}$ with probability at least $1-\delta$. Its validity $F(\widehat{J}_b) \le U_b^{\mathrm{KL}}$ has failure probability at most $\delta$ by the previous subsection. Therefore with probability at least $1-3\delta$, simultaneously for $b < M$,

$$F(\widehat{J}_b) \le U_b^{\mathrm{KL}} \le 16F_b^\star + 16C_b a_b + 4C_b a_{\mathrm{val}}. \tag{87}$$

For example, if $F_b^\star \le \varepsilon^2/32$, $0 < \varepsilon < 1$, the sufficient budgets

$$N \ge \frac{32C_b \log\Big(2M\binom{M}{b}/\delta\Big)}{\varepsilon^2}, \qquad N_{\mathrm{val}} \ge \frac{4C_b \log(M/\delta)}{\varepsilon^2} \tag{88}$$

make the right side of (87) at most $2\varepsilon^2$. Thus the reported TV bound is at most $\varepsilon$. The constants are conservative; the numerically inverted endpoint in the actual certificate is typically smaller. These sufficient conditions explicitly require a repairable residual, rather than claiming that a small repair budget succeeds for every table.

For position groups $M = n, c = 1$, $\log\binom{n}{b} \le n \log 2$ (and, for $b > 0$, also at most $b\log(en/b)$). At fixed $K, w$, the bounds above give a polynomial-query route from a small oracle residual to a nonvacuous, independently validated certificate, with sample dependence $1/\varepsilon^2$ in the TV tolerance. This is complementary to the additive excess-KL guarantee (29).

### I.5 Query cost and what is being preserved

For each independent trajectory, ancestral sampling uses $n$ calls, computing the $q$ rows for losses uses $n$ calls, and the width-$w$ anchor filter uses $O(nK^w)$ calls. Maintaining the $M$ group sums and sorting them requires ordinary arithmetic; evaluating the fixed policy and identifying a row's group are additional costs. Processing all selection and validation trajectories therefore uses $O((N+N_{\mathrm{val}})nK^w)$ conditional-evaluation calls plus those policy/group costs. Only $O(K^w)$ numerical frontier entries, an observed context, and $M$ accumulated group sums are necessary for a streaming implementation. The policy's state-to-action function is not recomputed after repair.

This procedure computes conditionals under the anchor even on unrepaired trajectories. Proposition 9 guarantees that its posterior updates remain valid after reveals actually drawn from $q$. It is thus a selective intervention, retaining every unselected original row, rather than unconditional anchor sampling disguised as a repair algorithm.

### I.6 Floor-free Hellinger selection and clipped KL

For arbitrary positive tables, define $Y_j^H(X) = \sum_t \mathbf{1}\{(H_t, I_t) \in G_j\} r_{I_t}(H_t)$ and $B_j^H = \mathbb{E}Y_j^H$. Then $Y_j^H \in [0, c]$. The same selection proof with $L = 1$ controls the excess residual energy $\sum_{j \notin \widehat{J}_b^H} B_j^H$ over its best size-$b$ counterpart. Its independent validation certificate is

$$\mathrm{TV}^2(p, P_{\mathcal{A}}^{q^{\widehat{J}_b^H}}) \le 15(M-b)c\mathcal{U}\left(\bar{V}_b^H, \frac{\log(M/\delta_{\mathrm{val}})}{N_{\mathrm{val}}}\right). \tag{89}$$

This uses the replay Hellinger bound, not a KL identity. Selecting by $B_j^H$ optimizes that additive energy, not actual TV or forward KL. No probability floor or tail assumption is used, but exact anchor-conditional evaluation is still required.

A second option is to clip KL at $L_0 > 0$. Define $\widetilde{k}_i = \min(k_i, L_0)$ and let $T_{L_0} = \mathbb{E}_{p,\mathcal{A}} \sum_t (k_{I_t}(H_t) - L_0)_+$. Repeating the empirical selection proof with $L_0$ gives

$$F(\widehat{J}_b^{\mathrm{clip}}) - F_b^\star \le 2\min(b, M-b)cL_0\sqrt{\frac{\log(2M/\delta)}{2N}} + T_{L_0}. \tag{90}$$

Indeed, the difference of true repaired benefits is bounded by the corresponding difference of clipped benefits plus the tail mass of the oracle-selected groups, which is at most $T_{L_0}$. Clipping alone does not certify $T_{L_0}$ or the true residual KL; a separate tail bound is required.

Strict positivity alone supplies no uniform bound on KL replay losses. For example, let a two-bit anchor have independent $U \sim \mathrm{Bern}(\eta)$ and $X \sim \mathrm{Bern}(1/2)$, realized by anchor order $(X, U)$. Let the fixed policy reveal $U$ first, keeping $q_X(1 \mid U = 0) = 1/2$ but setting $q_X(1 \mid U = 1) = e^{-c_0/\eta}$ for fixed $c_0 > 0$. Its rare-row KL is asymptotic to $c_0/(2\eta)$, so the mean replay KL tends to $c_0/2$ while $N$ sampled trajectories miss the row with probability $(1-\eta)^N$. All table entries remain positive. This explains the need for a known floor or tail control in KL-based replay estimation; the example is not an oracle-query lower bound for an algorithm allowed to inspect unsampled rows.

## J Budget allocation on a fixed adaptive policy

The per-trajectory formulation allows at most $b$ repairs, as opposed to selecting a globally fixed set of $b$ row groups. The state includes the revealed context $h$ and the remaining repair budget. Fix the deterministic position policy $i = \mathcal{A}(h)$. Let $V_b(h)$ be the minimum expected remaining forward-KL contribution under anchor continuation from $h$.

At a terminal state set $V_b(h) = 0$. When $b = 0$, the only available action is not to repair. For $b \ge 1$,

$$V_b(h) = \min\left\{k_i(h) + \sum_a p_i(a \mid h) V_b(h + (i = a)),\ \sum_a p_i(a \mid h) V_{b-1}(h + (i = a))\right\}. \tag{91}$$

The first branch retains the model row and incurs local anchor KL $k_i(h)$; the second uses the exact anchor row and incurs zero. In both branches, the continuation in the forward-KL chain identity is

evaluated under the *anchor* conditional $p_i$, not under the generated model conditional. Backward induction on the number of unrevealed positions proves the optimality of (91). The selected repair/no-repair policy, once fixed, is a deterministic function of the augmented history and thus has the same chain-rule interpretation.

There can be exponentially many revealed contexts. The Bellman equation is an exact characterization, not an efficient planner for arbitrary large masked models. Bounded unresolved-prefix width makes each needed anchor conditional tractable, but does not by itself bound the number of reachable planning states. Unequal repair costs can replace the decrement by an appropriate budget transition if the costs are known in advance.

# K WHY RANDOMIZED MARGINAL REPAIR CAN FAIL

Consider three binary coordinates. Let all rows for positions 2 and 3 be fair. For position 1, define

$$q_1(1 \mid x_{\{2\}}) = 3/4, \qquad q_1(1 \mid x_{\{3\}}) = 1/4, \tag{92}$$

independently of the observed values, and set all its other rows to $1/2$. The identity-order anchor $p$ is uniform. A fair independent coin chooses the fixed order $(2, 1, 3)$ or $(3, 1, 2)$. Their output laws are products with first-coordinate parameters $3/4$ and $1/4$ respectively, so their mixture equals $p$ exactly.

Replace only the $S = \{2\}$ rows of position 1 by the anchor value $1/2$. In the same random-order mixture, the first-coordinate parameter is now $(1/2 + 1/4)/2 = 3/8$. Consequently

$$D_{\mathrm{KL}}(p \| P_{\mathrm{after}}) = \frac{1}{2} \log \frac{1/2}{3/8} + \frac{1}{2} \log \frac{1/2}{5/8} = \frac{1}{2} \log \frac{16}{15} > 0. \tag{93}$$

The original marginal KL was zero. Conditional on each coin outcome, the selected-row repair has the nonnegative forward-KL gain of Proposition 8. The loss arises solely from marginalizing a previously cancelling mixture. This is why the theorem is stated for a fixed deterministic policy and why a randomized transcript guarantee must not be relabelled a randomized marginal monotonicity guarantee.

# L WIDTH-BOUNDED INFERENCE AND A BLACK-BOX LOWER BOUND

## L.1 ONE QUERY

A query $(S, i)$ needs the anchor prefix $T = [r]$, $r = \max(S \cup \{i\})$. There are $w = r - |S|$ unresolved positions in $T$, one of which is $i$. Enumerating their assignments requires at most $K^w$ configurations. Each anchor-prefix weight is a product of at most $n$ row entries. Summing those weights for each $x_i$ and normalizing gives $p_i(\cdot \mid x_S)$, using at most $O(nK^w)$ row evaluations. This is an alternative direct implementation of the same conditional as the curl compiler. Evaluating a long explicit curl route separately may use additional arithmetic; the row-query upper bound does not require doing so.

## L.2 INCREMENTAL PREFIX-POSTERIOR FILTER

After each completed reveal, let $r = \max S$ (or zero for $S = \varnothing$), and let $F = [r] \setminus S$ be the unresolved frontier. Maintain a table

$$W(y_F) = p(X_F = y_F \mid X_S = x_S). \tag{94}$$

Initially $F = \varnothing$ and its sole entry is one. To answer query $i$:

1. If $i > r$, extend the prefix successively through $j = r + 1, \ldots, i$. For each current frontier assignment, all coordinates before $j$ are either observed or specified in that assignment. Multiply its weight by each value of the anchor row $q_j(\cdot \mid x_{<j})$. Add $j$ to the frontier.
2. Marginalize the resulting table to $X_i$ and return that normalized row.
3. After the actual value $a$ at $i$ is revealed, retain the entries with $y_i = a$, renormalize, and remove $i$ from the frontier.

The invariant holds at initialization. An extension uses the actual factor of the identity-order anchor; future coordinates sum to one, so it produces the correct joint conditional over the enlarged frontier. The final step is ordinary conditioning. Induction proves (94) at every stage, irrespective of which distribution generated the observed evidence. Positivity guarantees that conditioning on an actual symbol is defined.

If every query has width at most $w$, there are at most $K^w$ entries at any point, including the table just before a reveal. Each anchor coordinate is extended at most once, and each such extension requires at most $K^{w-1}$ conditional-row queries and $K^w$ elementary weight updates. There are $n$ reveals, each requiring at most $O(K^w)$ marginal/conditioning updates. The claimed $O(nK^w)$ row-query bound is conservative and follows immediately. At most $K^w$ numerical posterior entries are stored. An implementation that explicitly stores a tuple of at most $w$ labels per entry uses $O(n + wK^w)$ words including the observed context; the table-entry claim does not erase this bookkeeping.

### L.3 Exact conditional evaluation lower bound

Take binary coordinates $H_1, \dots, H_{w-1}, X$ and let the anchor order place $X$ last. Every anchor conditional for $H_j$ is fair. For the last row let

$$q_X(1 \mid H = y) = v_y, \qquad v_y \in [1/4, 3/4], \tag{95}$$

with the $2^{w-1}$ entries otherwise arbitrary. All remaining masked-condition interfaces, which do not lie on this last fully revealed row, can be fixed independently to fair distributions. This is permitted because the input is an arbitrary masked specification, not a promise of consistency.

The requested width-$w$ anchor conditional with empty observed set is

$$p_X(1) = 2^{-(w-1)} \sum_{y \in \{0,1\}^{w-1}} v_y. \tag{96}$$

Consider a deterministic algorithm returning this value exactly from row queries. Run it on a table with every $v_y = 1/2$. If some full-hidden-state row $y_0$ is never queried, change only $v_{y_0}$ to $3/4$. All replies observed by the algorithm are unchanged, so adaptivity produces the same entire query transcript and the same output, but the correct mean has changed. This contradiction proves that at least $2^{w-1}$ informative row queries are required in the worst case. Counting both entries in a returned binary row still yields one informative query per $y$.

This is an exact deterministic query lower bound, not a computational-complexity assumption such as $\mathrm{P} \neq \mathrm{NP}$. It does not forbid estimating the mean by random sampling to a specified error. It also does not forbid drawing the hidden vector and then $X$, which generates a sample of $p$ in anchor order. Exact conditional evaluation under arbitrary evidence, and preservation of unselected $q$ decisions, are different interfaces from that unconditional sampling task.

## M Approximate repair and external accuracy

Let $p_i$ be the exact row from (20). If $\widehat{Z}_i > 0$, write $\delta_a = \log \widehat{Z}_i(a \mid h) - \log Z_i(a \mid h)$. The approximate normalized row is

$$\widehat{p}_i(a \mid h) = \frac{p_i(a \mid h) e^{\delta_a}}{\mathbb{E}_{p_i} e^{\delta}}. \tag{97}$$

Put $F(t) = \log \mathbb{E}_{p_i} e^{t\delta}$. Direct differentiation on the finite alphabet gives $F''(t) = \mathrm{Var}_{p_{i,t}}(\delta)$, where $p_{i,t}$ is the tilted law. If $\mathrm{osc}(\delta) \le b$, the elementary variance bound on an interval yields $F''(t) \le b^2/4$. Therefore

$$\begin{aligned} D_{\mathrm{KL}}(p_i \| \widehat{p}_i) &= F(1) - F(0) - F'(0) \\ &= \int_0^1 (1-t) F''(t)\, dt \le b^2/8. \end{aligned} \tag{98}$$

This proves (23). A bound on a common additive offset in $\delta$ is unnecessary because normalization cancels it; the relevant quantity is oscillation across possible output values.

For a deterministic policy with approximate replacements on selected rows, the exact chain formula becomes

$$D_{\mathrm{KL}}(p\|P^{\widehat{q}}_{\mathcal{A}}) = \mathbb{E}_{p,\mathcal{A}}\left[\sum_{\text{unrep}} k_i(h) + \sum_{\text{rep}} D_{\mathrm{KL}}(p_i(\cdot \mid h)\|\widehat{p}_i(\cdot \mid h))\right]. \tag{99}$$

The residual bounds $b_i(h)^2/8$ can replace the second sum. A local sufficient condition for safety is $b_i(h)^2/8 \le k_i(h)$ on each selected row. This is a condition on certified quantities, not a guarantee for an arbitrary Monte Carlo approximation to the completion integral.

Finally, if an external target law $P_\star$ obeys $\mathrm{TV}(P_\star, p) \le \alpha$, the triangle inequality for TV combines with either Pinsker's inequality or (16), proving (24). No triangle inequality for KL is used. For completeness, Pinsker follows by taking $A = \{x : P(x) \ge Q(x)\}$, applying log-sum to reduce $D_{\mathrm{KL}}(P\|Q)$ to binary relative entropy with mass difference $\mathrm{TV}(P, Q)$, and using the second derivative $1/(u(1-u)) \ge 4$ of binary relative entropy with respect to its first probability argument to lower-bound that divergence by $2\,\mathrm{TV}(P, Q)^2$.

A perfectly coherent but incorrect anchor is a simple counterexample to any accuracy claim based only on consistency: set every masked row equal to the conditionals of a positive $p \ne P_\star$. All curls and order distances vanish, while the external accuracy error is $\mathrm{TV}(P_\star, p)$. This error remains after any anchor-based repair.

# N Verification design and reproducibility

## N.1 Implementation independence and numerical interpretation

The implementation uses zero-based coordinate indices and the integer $-1$ for an unrevealed coordinate. The core class stores a separate normalized row for every valid pair $(h, i)$. The reference joint oracle independently enumerates full path products and conditions the resulting full joint. The curl compiler instead enumerates only the necessary anchor prefix, constructs an actual adjacent-swap route, sums its signed local curls, computes $\log Z$ by log-sum-exp, and normalizes. The frontier filter maintains its own incremental posterior table. Agreement between these three routes is a test of their distinct implementations, not a proof of the mathematical claims.

The suite records a row for every equality or inequality it checks, including its input identifiers and observed discrepancy. The tolerance is recorded in the runner and described below. The master tolerance is $3 \times 10^{-10}$ in absolute units. Tiny negative KL values or tiny upper-bound overruns from floating-point cancellation are not interpreted as mathematical violations. The largest recorded identity residual is below $2.00 \times 10^{-15}$. All full-state summations are floating-point; "exhaustive" refers to enumerating the state space, not to rational arithmetic or proof-assistant verification.

The tests were executed with Python 3.13.5 and NumPy 2.3.5 on a Linux CPU environment. No GPU, network access, pretrained weights, or language corpus is used by the runner. All reported numerical results are regenerated by the supplied executable suite.

**Original exact checks.** Table 2 records the principal identity and inequality checks. Table 3 reports exact-benefit repair as an oracle reference; it is not the sample-selection experiment in the main text.

Table 2: Selected exhaustive finite-state checks. "Residual" is the maximum absolute discrepancy for an identity; bounds are checked with the stated tolerance.

| Quantity | Checks | Maximum residual/result |
|---|---|---|
| Swap-to-conditional compiler | 6,480 | $6.66 \times 10^{-16}$ |
| Single-swap Hellinger identity | 7,680 | $3.33 \times 10^{-16}$ |
| One-layer / multi-layer upper bounds | 1,920 / 480 | All pass |
| Conditional contraction / curl bridge | 1,280 / 1,280 | All pass |
| Adaptive KL / repair gain identities | 240 / 240 | $2.00 \times 10^{-15}$ |
| Incremental anchor posterior | 1,600 | $2.22 \times 10^{-16}$ |

Table 3: Mean forward KL to the anchor for position-group repair. All quantities are exactly enumerated before floating-point evaluation; no test-set estimation is involved.

| Repaired positions $b$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Largest-benefit groups | 3.227 | 1.880 | 0.971 | 0.468 | 0.122 | 0.000 |
| Uniform-subset expectation | 3.227 | 2.582 | 1.936 | 1.291 | 0.645 | 0.000 |

### N.2 Random tables

Use $n = 5$, $K = 2$, and 16 seeds $13000 + s$, $s = 0, \ldots, 15$. For every context in lexicographic order and every unrevealed coordinate in increasing order, independently draw $v \sim \mathrm{Dirichlet}(0.7, 0.7)$ and set the row to $0.96v + 0.02(1, 1)$. This gives $5 \cdot 3^4 = 405$ rows per table, or $6{,}480$ compiler comparisons. No row is trained or fitted from a dataset.

For each table, enumerate all 120 fixed permutations. All four adjacent swaps per permutation yield $7{,}680$ single-swap identities over the 16 tables. The disjoint slot pairs $(0, 1), (2, 3)$ or $(1, 2), (3, 4)$, alternating with permutation index in zero-based notation, give one multi-swap layer per permutation, for $1{,}920$ one-layer bounds. Thirty source/target comparisons per table test odd–even swap-layer routes, for 480 multi-layer bounds. There are $5 \cdot 2^4 = 80$ mask/action types per table, giving $1{,}280$ conditional-contraction and $1{,}280$ curl-bridge tests. Every value-row for the type is averaged under the anchor marginal, not under an independently chosen mask-data distribution.

Audit randomness uses seeds $23000 + s$. Fifteen random state-based deterministic policies per table give 240 adaptive KL identities, 240 adaptive Hellinger bounds, and 240 repair identities and monotonicity checks. Each policy is drawn once, then kept fixed as its conditionals are repaired. Twenty full reveal trajectories per table compare the frontier filter against full-joint conditioning at each of five steps: 320 trajectories and $1{,}600$ conditional queries. For every compiler row, a bounded perturbation of its log correction is tested against (23), yielding $6{,}480$ approximate-repair inequalities.

### N.3 Sharpness and strong-dependence grids

The depth grid uses $g \in \{1, 4, 16, 64\}$, $D \in \{1, 2, 4, 8, 32\}$ and $s \in \{0.03, 0.10, 0.25\}$. Set $\theta = s/(D\sqrt{g})$, so the theorem's small-tilt condition holds. This gives 60 settings. The endpoint TV is evaluated by summing over the binomial count of positive main spins, an exact symmetry reduction of the full finite-state distribution. The local energies use (42). Both inequalities in (8) are checked. The CSV contains the ratio $\mathrm{TV}^2/(DE)$ and all parameters.

The dependence stress test starts from the positive six-bit law

$$p_{\mathrm{mix}} = \tfrac{1}{2}\mathrm{Bernoulli}(0.08)^{\otimes 6} + \tfrac{1}{2}\mathrm{Bernoulli}(0.92)^{\otimes 6}. \tag{100}$$

Enumerate all of its masked conditionals to obtain a coherent table. Perturb each binary logit by $\eta u$, with $u$ independently uniform on $[-1, 1]$, seeds $61000 + s$ for $s = 0, 1, 2, 3$, and $\eta \in \{0.01, 0.03, 0.10, 0.30\}$. For each perturbed table, compare identity and reverse orders using the actual source-prefix energies of an odd–even layer route. No Dobrushin coefficient, spectral gap, or other mixing diagnostic is supplied to the bound. This checks that the proof's dependence-free assumptions match the implementation; it does not establish practical robustness of a trained language model.

### N.4 Budgeted repair and exact boundary cases

For every five-bit random table, freeze the policy selecting the row with largest maximum token probability; ties are broken by coordinate index. Position groups contain all rows querying that position. Compute each group's exact anchor-replay KL benefit by full enumeration. For each budget $b = 0, \ldots, 5$, repair the $b$ largest-benefit positions and independently enumerate the repaired output law. The uniform-subset comparator is the exact expected residual over uniformly selected position subsets: each position is unselected with probability $1 - b/5$, so its expectation is $(1 - b/5)D_{\mathrm{KL}}(p\|P^q_{\mathcal{A}})$. The 96 configurations validate both the gain identity and the ordering relative to this comparator. This is an oracle-benefit comparison, not a sample-efficient repair-learning experiment.

For the adaptive gap at $m = 2$, enumerate all $7! = 5{,}040$ fixed orders over $2^7$ output vectors. The maximum is $0.0625$. Direct enumeration of the adaptive policy gives $0.25$. Smaller $m = 1$ is also checked. The random-mixture example is enumerated over eight states before and after one row-family replacement. The coverage counterexample uses $\eta \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ and checks its score expectation $2\eta/15$. The concentration check enumerates binomial tails for $N \in \{10, 50, 200\}$, means $\{0.01, 0.1, 0.4, 0.9\}$, and failure levels $\{0.05, 0.2\}$; this is a finite sanity check of the general proof, not an empirical derivation of the inequality.

### N.5 Sample-selected repair and held-out certificates

The additional runner is `code/run_finite_budget.py`. It uses the same 16 model tables, with selection RNG seeds $73000 + s$ and independent validation RNG seeds $83000 + s$, $s = 0, \ldots, 15$. Each selection batch contains $2048$ independent ancestral anchor samples; its prefixes of sizes $32, 128, 512, 2048$ define four training budgets. A separate batch contains $4096$ independent anchor samples. The model's minimum entry is at least $\beta = 0.02$ by construction, and position groups have $M = 5, c = 1$.

The maximum-confidence policy uses the original model table throughout and breaks ties by coordinate index. For each sampled vector, replay computes anchor rows through `FrontierFilter`, then evaluates their KL to the original row. Repeated identical sampled vectors may reuse deterministic filter computations; their occurrence counts still come from the independent random samples. This cache does not contain population expectations. All selected sets are fixed before constructing the full anchor joint for population evaluation. The latter is used only to report KL, TV, oracle benefits, and discrepancies in the new checks.

For each of the $16 \cdot 4$ model/training-size pairs, set $\delta_{\mathrm{sel}} = \delta_{\mathrm{val}} = 0.025/(16 \cdot 4)$. The selection radius unions over five group estimates, and the held-out bound unions over the five nontrivial budgets $b = 0, \ldots, 4$; full repair has zero error. The resulting selection and validation statements have joint failure probability at most $0.05$ over the entire reported collection. No independence is claimed between training prefixes, budgets, or certificates sharing a validation batch.

The CSV records retain every learned group set, empirical group benefit, population KL, oracle KL, regret, theoretical regret bound, actual TV, and held-out certificate. Query records report the ancestral sampling calls and filter calls incurred during the sampled phase, separately from later exhaustive evaluation. The largest observed frontier width is five, with at most 32 entries. The computational guarantee concerns row queries; memoization and neural batching can change wall-clock costs and are not used to claim a learned-model speedup.

### N.6 Finite local audit and new probability checks

The local-audit example has $g = 4$, $D = 2$, $n = 12$, $\theta = 0.025$, and $m = gD = 8$ swap occurrences. With seed $93000$, each independent draw selects one occurrence uniformly, samples its source prefix in the actual source order, and samples the two-position kernel from its fair mixture of orientations. Four conditional-row evaluations suffice for the pair score, in addition to the prefix draws. The source prefix contains at most $n - 2 = 10$ reveals. The runner records actual row-query counts as well as the bound $N(n + 2)$.

Audit budgets are the prespecified nested values $256, 1024, 4096, 16384, 65536$. Every certificate uses $\log(5/0.05)/N$, so the five displayed bounds hold simultaneously with probability at least $0.95$. The population energy and endpoint TV are evaluated independently from the construction's formulas, not substituted for the sampled mean. The comparison labeled "coarse" uses $2\bar{z} + 8\log(5/0.05)/N$ and the same samples and confidence allocation.

The new probability checks enumerate exact binomial tails for $N \in \{1, 5, 20, 100\}$, means $\{0, 0.001, 0.01, 0.1, 0.5, 0.9, 1\}$ and failure levels $\{0.001, 0.05, 0.2\}$. They also enumerate all iid score sequences on $\{0, 1/4, 1\}$ with probabilities $(0.5, 0.3, 0.2)$, at sample sizes one through six and failure levels $\{0.01, 0.05, 0.2\}$. The adaptive coverage identity is independently summed over selector and marker-label pairs for $m = 1, \ldots, 7$ and $\kappa \in \{1, 2, 3, 2^m, 2^m + 1\}$. Marker values have already summed to one in that check. Finally, all 60 lower-bound configurations are tested against the strengthened constant $1/6$ proved in Section D.

### N.7 Learning curves and regeneration commands

Figure 2 shows actual mean excess KL over the oracle separately for each nontrivial repair budget. An increased selection sample size does not enforce a monotone realized learning curve: small finite-sample changes can exchange near-tied groups. Monotonicity in Proposition 8 instead concerns nested repair sets at a fixed learned ranking.

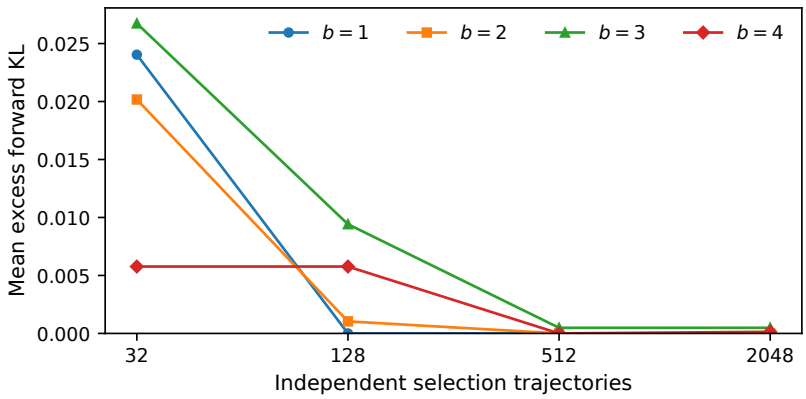



Figure 2: Actual sample-selection regret over the exact-benefit oracle, averaged over the same 16 model tables. Full-joint enumeration is only an evaluation device. The lines show means across synthetic tables, not confidence intervals or held-out certificates.

From the package root, run:

```
python code/run_checks.py --models 16
python code/run_finite_budget.py
python -m pytest code -q
python build.py
```

The build regenerates all tables and vector figures from CSV/JSON records, compiles three LaTeX passes, rejects unresolved references and overfull boxes, and checks the nine-page main-text limit. The figures use only the disclosed synthetic tables, exact finite sums, and independent samples. No trained-model performance is synthesized or inferred.