

IECZ–III: Hardcore Condensation Lift

Distributional and parameterized hardness with size-aware invariants

Marko Lela (ORCID: [0009-0008-0768-5184](#))

2025-10-10

Abstract

This paper develops a compact, size-aware blueprint for transferring structure through gadget lifts. Two low-order invariants—cumulative mod- q Fourier mass up to degree k and noise stability Stab_ρ —are treated as a reusable “profile”, tied to the gadget’s affine interface. Under coordinate permutations ($\Delta = 1$) the profile is preserved exactly; under bounded fan-in the degree budget relaxes by at most a factor Δ (cf. Theorem B.5), with all overheads tracked explicitly. In a balanced window $m = (1 + \gamma)n$ the framework yields a distributional lower bound for at least one monotone cost (focus: EC) and an “echo” to correlation against size-aware $\text{AC}^0 + \log$ and to logarithmic degree in the polynomial-calculus setting. The accounting keeps total-variation contraction and a single $O(\log N)$ prefix-free header visible end to end, avoiding hidden slack.

Scope and guiding principle. We work within a balanced window $m = (1 + \gamma)n$ and keep all overheads explicit: deterministic push-forwards are TV-contractive, the $3\text{XOR} \rightarrow 3\text{SAT}$ step has size factor $\times 4$, and meta descriptions are prefix-free with $O(\log N)$ bits. The guiding rule is to retain reproducibility while avoiding hidden slack: we expose the affine interface, preserve the low-order profile exactly at $\Delta = 1$, and account for bounded fan-in losses only through stated degree budgets and a single header amortized across the pipeline.

Contents

1	Introduction	3
1.1	Problem framing: balanced window without worst-case claims	3
1.2	Contributions	4
1.3	Conventions and notation	4
1.4	Motivation: structure profiles and contextual compression	5
1.5	Roadmap	5
1.6	Limitations and scope	5
2	Invariants and the size-aware API	6
2.1	Windowed API (balanced $3\text{XOR} \rightarrow 3\text{SAT}$)	6
2.2	Constants and loss table (size/meta exposed)	7
2.3	Controlled overheads and amortization	7
2.4	How the API is used downstream	8
2.5	Auxiliary notes and cross-references	8
2.6	Normalization conventions	8
2.7	API notes (practical checklist)	9
2.8	Summary	9

3	Window lower bounds (EC focus)	9
3.1	Setup and parameter path	10
3.2	Surviving parity along the switching path	11
3.3	From a live parity to erasures	12
3.4	Main theorem (EC in the window)	13
3.5	Robustness under XOR/Index gadgets	13
3.6	Parameter summary	14
3.7	Empirical summary (EC proxy) and cross-reference	15
4	Echo: correlation and degree	15
4.1	Information to correlation (bridge to Section 3)	16
4.2	Correlation echo against size-aware AC0+log	16
4.3	Degree echo in the polynomial-calculus framework	17
4.4	Empirical cross-references	18
4.5	Scope and constants	18
4.6	Formal bridge to Sections 2-3 and the appendices	19
5	Related work	19
5.1	Gadget reductions and affine interfaces	20
5.2	Fourier analysis of Boolean structure	20
5.3	Noise sensitivity and stability	20
5.4	Switching lemmas and decision-tree viewpoints	20
5.5	Sum of Squares and proof complexity for XOR and SAT	21
5.6	Encoding parity into CNF and window-aware lifts	21
5.7	Information theory, MDL, and compression as structure probes	21
5.8	Erasur complexity and reversible viewpoints on search	21
5.9	Relation to empirical work on spectral diagnostics	22
5.10	Summary of distinctions	22
6	Conclusion and outlook	22
6.1	Main contributions	22
6.2	Limitations	23
6.3	Open technical directions	23
6.4	Further directions	23
6.5	Empirical program and reproducibility	24
6.6	Open problems	24
A	TV contraction and prefix-free headers	25
A.1	TV contraction and prefix-free meta overhead	25
A.2	Notes on compression surrogates used for PCG	27
B	Hypergeometric concentration and affine pullbacks	27
B.1	Concentration for sampling without replacement	27
B.2	Affine pullbacks preserve low-degree structure	28
B.3	Affine pullbacks: degree, mass, and stability	28
B.4	Gadget pullbacks: XOR and Index	32

C	Switching path parameters (balanced window)	33
C.1	Parameter choices	33
C.2	Derived scales and routine consequences	33
C.3	Reuse across gadgets	34
C.4	Notation shorthands	34
D	Minimal reproducibility pack	35
D.1	Runtime and resources	35
D.2	Outputs	35
D.3	Scripts (brief)	35
D.4	Reproduction command	36
D.5	EC proxy — run summary	37
D.6	SPR proxy — k-gram on parity histories	38
D.7	PCG proxy — k-gram (label-block)	39
D.8	PCG proxy — LZ family (label-block)	39
D.9	PCG scaling with n (label-block, strong signal)	40
D.10	PCG sensitivity in side-context (k-gram)	40
D.11	PCG vs. label separation (k-gram, grid)	40
D.12	LZ family comparison (label-block)	41
D.13	Fine k -profile (label-block)	42
D.14	Multi-mode structure profiles (label-block & side)	42
D.15	Correlation with PCG (label-block)	43
D.16	Extended experiments (D.9–D.15)	44
E	Reproducibility and packaging (final + RC2)	46

1 Introduction

Series context and scope. This paper is the third part of a sequence. Part I develops size-aware lifts from 3XOR to 3SAT (see also [26, 27]); Part II formulates the minimax gadget view compatible with Sum-of-Squares [28]. This paper is standalone: all theorems and proofs needed for the stated results are contained here (or in the appendices). References to “IECZ I/II/...” are for context only; no results in this paper depend on unpublished parts of the series.

We study distributional and parameterized hardness via a small pair of invariants that behave predictably under size budgets. The two invariants are the cumulative mod- q Fourier mass up to degree k and the noise stability Stab_ρ [35]. We refer to them jointly as the *profile*. The profile is tied to the affine interface of a gadget. When the interface is a coordinate permutation ($\Delta = 1$), the profile is preserved exactly. For bounded fan-in $\Delta > 1$, the degree budget expands by at most a factor Δ , and we keep the associated overhead visible throughout.

Logarithm convention. Unless stated otherwise, $\log$ denotes base-2 and $\ln$ denotes the natural logarithm. All asymptotic statements (e.g., $O(\log N)$) are insensitive to constant base changes.

1.1 Problem framing: balanced window without worst-case claims

Fix a constant $\gamma \in (0, 1)$. Let μ be the balanced 3XOR distribution on the window and let R be the canonical four-clause map $3\text{XOR} \rightarrow 3\text{SAT}$. We set the working distribution to $\nu := R_\# \mu$. This

push-forward introduces *no additional* TV slack inside the window, and for any alternative source μ' the deterministic map is 1-Lipschitz:

$$\text{TV}(R_{\#}\mu, R_{\#}\mu') \leq \text{TV}(\mu, \mu').$$

We continue to track the explicit size factor $\times 4$ and a single prefix-free meta header of length $O(\log N)$. The clause count scales as

$$m' = 4m \quad \text{for} \quad m = (1 + \gamma)n.$$

We then apply constant XOR/Index gadgets that are affine over $\mathbb{F}_2$ after output mixing and analyze what the profile does under these steps. All statements are distributional with explicit parameter paths. We make no worst-case assertions and no P/poly claims.

1.2 Contributions

1. **API for invariants.** For $T \in \{\text{CMDL/PCG, SPR, EC, PROF}\}$ the data-processing monotonicity and the gadget stability properties hold with at most $O(\log N)$ meta overhead. Block aggregation is available for $T = \Theta(\log n)$.
2. **Window lower bound (EC focus).** Under the row-distance hypothesis $H_{\text{row-dist}}$ (Assumption 3.1), in the balanced window there is a distribution μ arising from the 3XOR mapping such that $\text{EC}_{\mu}(\text{SAT}) \geq m^{\beta}$ for some $\beta > 0$, via a switching-path argument with one surviving parity [19].
3. **Echo to correlation and degree.** Any gap in EC/SPR/PCG yields a size-aware correlation lower bound against $\text{AC}^0 + \log$ [26] and a polynomial degree lower bound $\deg \geq \zeta \log n$ in the polynomial-calculus framework (SoS optional) [40].
4. **Reproducibility artifacts.** Compact scripts generate CSV and PNG outputs. Seeds and run budgets are fixed and documented. All code and artifacts (software package only) are archived on Zenodo [29].

1.3 Conventions and notation

- **Balanced window.** $m = (1 + \gamma)n$ for fixed $\gamma > 0$. After $3\text{XOR} \rightarrow 3\text{SAT}$ the count is $m' = 4m$.
- **Size and meta.** Deterministic push-forward is TV-contractive. We track size factor $\times 4$ and a single prefix-free header of length $O(\log N)$ with total symbol budget $N = \Theta(m \log n)$.
- **Gadgets.** XOR and Index are constant and affine over $\mathbb{F}_2$ after output mixing. The Index gadget uses a constant address length (default 6 bits).
- **Switching path.** Depth $d \in \mathbb{N}$ (default $d = O(1)$; examples $d = 3$) and restriction parameter

$$p = m^{-\alpha/d} \quad (\text{default } \alpha = \tfrac{1}{3}).$$

When a subpolynomial bottom width is required we also use $d = \Theta(\log n)$; see Appendix C.

- **EC model.** Word RAM with $w = 64$. We count bit-level erasures for logically non-injective steps [23]. Algorithms may be randomized. Expectations are over internal randomness and $x \sim \mu$.

- **SPR predictors.** Finite-state, probabilistic models are allowed. In theory $k \leq c \log n$. Experiments may use $k \leq 8$.

On our synthetic families these profiles rise quickly and remain robust, and they correlate positively with PCG measured by k -gram and VLC surrogates.

1.4 Motivation: structure profiles and contextual compression

Low-order regularities visible in simple structure profiles tend to predict contextual compression gains. We use cumulative mod- q mass at small k and stability Stab_ρ at a short grid of ρ . Along synthetic families these profiles increase quickly and remain robust. They correlate with PCG measured by k -gram and by a variable-length coding surrogate. This supports the echo viewpoint used later: surviving low-order structure aids context models and also forces correlation and degree barriers. We refer to this multi-mode profile as PROF. By default we instantiate the profile on a small grid: $q \in \{2, 3, 5\}$, degree cap $k \leq 6$, and noise levels $\rho \in \{0.1, 0.2\}$.

1.5 Roadmap

Section 2 defines the invariants and states the API (Theorems A, A', A''). Section 3 proves the window lower bound for EC. Section 4 derives the echo to correlation and degree. The appendices provide the total-variation and prefix-free facts, the hypergeometric concentration tools, the affine pullback lemmas, and the reproducibility notes.

1.6 Limitations and scope

Our results are distributional and parameterized. We remain in a balanced window with explicit TV/size/meta accounting and constant XOR/Index gadgets. We do not claim worst-case results. Empirically, PCG and multi-mode profiles are proxies; short runs can show thin-context artifacts. All constants (γ, Δ , address length, k, ρ) are fixed and documented. Scaling beyond this window, alternate gadgets, and stronger codecs are left for follow-up work.

Glossary of symbols

m, n, N

Clause and variable counts; instance bitlength $N = \text{size}(x)$.

γ, d Window parameter and switching depth. Default $d = O(1)$ (examples $d = 3$); when a subpolynomial width bound is needed we also use $d = \Theta(\log n)$ (see Appendix C).

p, W

Restriction parameter $p = m^{-\alpha/d}$ and bottom width $W \leq m^{O(\alpha/d)}$ after switching.

ρ, θ Noise and mixing rates used in the profile and in gadget analyses.

q, k Modulus for mod- q mass and degree budget with $k = O(\log n)$.

TV Total-variation distance measured on the window.

$V_{\mathcal{C}}(\mu, \varepsilon)$

Size-aware MDL cost at error or advantage ε for class $\mathcal{C}$.

EC/SPR/PCG/PROF

Erasure count, short-path reversibility, pseudo-compressibility gain, and the multi-mode profile.

2 Invariants and the size-aware API

We formalize four invariants used throughout the pipeline: total variation with a single prefix-free header TV/Meta , short-path reversibility (SPR_k), erasure complexity (EC), and a multimode low-order structure profile (PROF). Each comes with a *size-aware API*: on windowed product measures and under the fixed prefix-coding scheme, every map we use preserves the invariant up to vanishing slack and a logarithmic meta overhead. The API is *composable*: it underwrites every step from the balanced 3XOR window through affine gadgets and switching restrictions to the final echo bounds.

A *window* is any product $\Omega \subseteq \{0,1\}^N$ with fixed Hamming marginals on the blocks we condition on; all API statements are either invariances or controlled dilations on such windows. The proofs are immediate from Appendices A/B ($\text{TV}/\text{headers}$ and affine pullbacks and hypergeometric concentration), together with standard amortization of the $O(\log N)$ prefix-free header across the pipeline.

2.1 Windowed API (balanced 3XOR $\rightarrow$ 3SAT)

Let $\text{XOR}_{n,m}$ denote the balanced 3XOR family in the window $m = (1 + \gamma)n$, and let $\text{SAT}_{n,4m}$ be its auxiliary-free 3CNF image under the canonical four-clause block map. For a product window Ω on XOR and a distribution μ supported on Ω , write $R_{\#}\mu$ for the push-forward of μ by a measurable map R . We write $N \asymp m \log n$ for the ambient bit length at the current stage.

Theorem 2.1 (A — API, windowed form). *There is a window-preserving map $R : \text{XOR} \rightarrow \text{SAT}$ with a prefix-free header of length $H = O(\log N)$ such that for $\nu = R_{\#}\mu$ the following hold.*

- **TV/Meta.** $\text{TV}(R_{\#}\mu, R_{\#}\mu') \leq \text{TV}(\mu, \mu')$. A single prefix-free header of length $H = O(\log N)$ decodes block structure and the gadget choice; headers compose additively and are charged once for the pipeline. See Theorem A.1 for TV contraction; for coding background cf. [12].
- **SPR_k.** For any $k = O(\log n)$,

$$\text{SPR}_k(R_{\#}\mu) \geq \text{SPR}_k(\mu) - O(\log N),$$

where the loss accounts for encoding/decoding the header inside a reversible wrapper [8].

- **EC.** For any randomized SAT solver $\mathcal{A}$ there exists a reversible wrapper $\mathcal{A}^*$ for XOR such that

$$\text{EC}_{\nu}(\text{SAT}; \mathcal{A}) \geq \text{EC}_{\mu}(\text{XOR}; \mathcal{A}^*) - O(\log N).$$

This uses a reversible simulation wrapper; the logical erasure bookkeeping follows [8, 10, 23]. Thus EC lower bounds transfer from XOR to SAT up to logarithmic slack.

PROF (multimode profile). *Deterministic push-forwards are TV-1-Lipschitz, so empirical fluctuations are preserved up to $o(1)$ on the window. Under output-mixing normalization ($\Delta = 1$) the affine relabeling is a coordinate permutation; the cumulative mod- q mass up to degree k and Stab_p are preserved exactly by pullback. For bounded fan-in Δ , Theorem B.5 yields the degree expansion $k \mapsto \Delta k$. We do not claim a general noise-stability inequality for $\Delta > 1$; all stability uses are under $\Delta = 1$. Under output-mixing normalization ($\Delta = 1$), degree is preserved exactly; see Theorem B.5.*

Theorem 2.2 (A' — Block product stability). *Let R be the blockwise product of $T = \Theta(\log n)$ independent copies of the 4-clause map, with a single shared header of length $H = O(\log N)$. Then TV contraction holds blockwise and globally; SPR_k drops by at most $O(\log N)$; EC adds across blocks with the header charged once; and the profile averages with variance $O(1/T)$ by concentration, hence equals the block-wise average up to $o(1)$.*

Arrow	TV loss on Win	Size blow-up	Meta (prefix-free)	Note
E1 Direct-product	$o(1)$	$\times T$ with $T = \lceil \varepsilon^{-2} \log N \rceil$	none	Chernoff exponent explicit; majority aggregation [11, 22, 33].
E2 Hardcore set	$o(1)$	$\times (1+o(1))$	$O(\log N)$ (selector tag)	Uniform selector; advantage N^{-c} on mass $\theta \geq c_2$.
E3 Seeded condensation	$o(1)$	$\times \text{polylog}(N)$	$O(\log N)$ (seed)	Preserves low-degree statistics up to $k = O(\log n)$.
E4 Index fixing	0	+meta only	$O(\log N)$ (indices)	Deterministic seed/schedule selection.

Table 1: TV/size/meta bookkeeping per arrow. The header is charged once and amortized.

Theorem 2.3 (A'' — Gadget stability under small branching). *The branch index is chosen once globally and encoded in the $O(\log N)$ header; branches do not vary position-wise. Allowing per-position branch choices would require $\Theta(m)$ meta bits and is outside our pipeline. Let R be a constant-branching mixture of injective affine maps $\{T_b\}_b$ selected by that header [14, 32]. Then TV contracts in each branch and hence in the mixture [41, 42]; a reversible wrapper stores the branch label and uncomputes it, so the total SPR_k drop is $O(\log N)$; EC lower bounds are preserved up to the header slack. Under $\Delta = 1$ each branch preserves the cumulative mod- q mass up to degree k and Stab_ρ exactly. For bounded fan-in Δ , degrees expand by a factor $\leq \Delta$; averaging across the $O(1)$ branches preserves the low-degree mass profile up to $o(1)$.*

Remark 2.4 (Normalization for EC). We evaluate EC on a $w=64$ word-RAM at bit granularity: #ERASE counts logically non-injective, irreversible bit operations; randomized solvers are measured in expectation over internal randomness and $x \sim \mu$. Reversible preprocessing/simulation follows Bennett-style checkpointing; the bookkeeping fits into a self-delimiting meta header of length $O(\log N)$ (prefix-free coding of indices/parameters), so the simulator overhead is absorbed by the meta budget; see [8, 31]. Optionally, #ERASE is motivated by Landauer’s principle [23].

2.2 Constants and loss table (size/meta exposed)

One-paragraph bound. In the calibrated window and under the fixed prefix-free coding scheme, every deterministic polynomial-time map R used in the blueprint satisfies

$$V_{\mathcal{C}}(R_{\#}\mu, \varepsilon) \leq V_{\mathcal{C}}(\mu, \varepsilon) + O(\log N),$$

and TV contracts under push-forward [42]. The single meta header has length $H = O(\log N)$ [14] and is charged once; it composes additively across constant-branching steps.

For later references we fix global placeholders

$$c_1 = 1, \quad c_2 = \frac{1}{4}, \quad c = \frac{1}{8}, \quad \eta = \frac{1}{8}, \quad \zeta = \frac{1}{16}, \quad k = \lfloor \frac{1}{4} \log n \rfloor, \quad T = \lceil \varepsilon^{-2} \log N \rceil,$$

which can be replaced by any comparable set without affecting downstream statements.

2.3 Controlled overheads and amortization

The $O(\log N)$ meta allowance covers all headers and constant-branching choices in the entire pipeline. Each header is prefix-free and charged once, so the total additive overhead stays $O(\log N)$. A

reversible wrapper loads the header once, keeps it in a persistent reversible register, and uncomputes it at the end. Along any computation of length T with $T = \text{poly}(n)$ the amortized per-step charge is $O(\log N)/T = o(1)$. Consequently the drop in SPR_k is $O(\log N)$ in total and $o(1)$ in the lim sup of the normalized units used downstream; EC loses at most $O(\log N)$ in the lower bound; PROF is unit-scale and unchanged up to $o(1)$.

2.4 How the API is used downstream

- **Window lower bounds (Section 3).** Reduce $\text{XOR} \rightarrow \text{SAT}$ via Theorems 2.1-2.2, then apply switching restrictions with $p = m^{-\alpha/d}$ to obtain $W \leq m^{O(\alpha/d)}$ (for the instantiation used in Section 3 we fix a switching calibration with exponent $\sigma = \frac{1}{2}$, yielding $W \leq \tilde{C} m^{\alpha/(2d)}$; cf. Remark 3.4); with decision depth $t = \Theta(\log n)$ the switching failure $(CpW)^t = o(1)$. In particular, $d = \Theta(\log n)$ yields $W = n^{o(1)}$, while constant d gives a polynomial W . EC then yields information lower bounds that survive the window. The meta overhead is absorbed by the $O(\log N)$ slack.
- **Echo (Section 4).** The correlation bound against $\mathcal{C}_{s,d}^{\log}$ and the degree echo $\deg_{\text{PC}} \geq \zeta \log n$ rely on (i) the switching-path width bound $W \leq m^{O(\alpha/d)}$ (instantiate $d = \Theta(\log n)$ when a subpolynomial width is required) and (ii) a parity surviving with probability $c_0 > 0$. With $W \leq m^{O(\alpha/d)}$ and decision depth $t = \Theta(\log n)$ the switching failure $(CpW)^t = o(1)$. PROF is preserved exactly under $\Delta = 1$; see Theorem B.5. For $\Delta > 1$ the degree budget relaxes to $k \mapsto \Delta k$, which is harmless in the $\Theta(\log n)$ regime.

Cross-reference. The full seed lists, grids, and exact commands for the final runs are collected in Appendix D; see §§ D.9–D.15. All resulting figures/CSVs are synced under `appendices/{assets,data}/...`

2.5 Auxiliary notes and cross-references

TV contraction and prefix-free headers (Appendix A). Deterministic push-forwards are 1-Lipschitz in TV (Lemma A.1; see also [42]). Headers cost $O(\log N)$ [14] and compose additively (Lemma A.2, Corollary A.3).

Affine pullbacks and concentration (Appendix B). Hypergeometric without-replacement bounds (B.1-B.2) control sampling fluctuations. Affine pullbacks under bounded fan-in Δ preserve degree/low-order mass with level expansion $k \mapsto \Delta k$ (Theorem B.5). For noise stability we only use the exact $\Delta = 1$ case.

Headers and reversible wrappers. The $O(\log N)$ header is kept in a dedicated register by the reversible wrapper and uncomputed at the end; this explains the additive $O(\log N)$ slack in SPR_k and EC [8, 12]. Output emission (decision bit or witness) is not charged as an erasure; a Bennett-style reversible simulation uncomputes all work tapes and ancillae, so that only the self-delimiting meta header h must eventually be disposed of, accounting for the additive $O(\log N)$ overhead.

2.6 Normalization conventions

Output-mixing normalization. Unless stated otherwise we work under output-mixing ($\Delta = 1$): every gadget induces an affine relabeling that is a coordinate permutation of the underlying bit

vector. This yields exact preservation of PROF, and it preserves the number of constraints m (wirewise bijection / coordinate renaming only).

Bounded fan-in. When a gadget uses output wiring with fan-in $\Delta > 1$, we apply Theorem B.5 for degree-level expansion $k \mapsto \Delta k$. For stability we rely only on the exact $\Delta = 1$ case. The $O(\log n)$ meta budget absorbs constant Δ factors.

2.7 API notes (practical checklist)

- **Headers.** Use a single prefix-free header of length $O(\log N)$ for the entire pipeline; decode once, reuse, uncompute.
- **SPR wrappers.** Implement a reversible state machine that loads the header once, keeps it in an ancilla register, and uncomputes it; the per-step charge amortizes over $T = \text{poly}(n)$.
- **Window discipline.** Perform each push-forward inside the product window; empirical fluctuations are preserved up to $o(1)$.

Technical note (switching path). Unless stated otherwise we take $d = O(1)$ (examples $d = 3$). Along the calibrated switching path with restriction parameter $p = m^{-\alpha/d}$ and depth $d \geq 1$, standard switching-lemma bounds imply that a bottom CNF/DNF of width w reduces, with failure probability at most $(Cpw)^t$, to decision-tree depth $\leq t$ [16, 19, 21].

Choosing $t = \Theta(\log n)$ and tracking widths across d rounds yields, with high probability, a bottom-width bound

$$W \leq \text{poly}(1/p) = m^{O(\alpha/d)}.$$

For the concrete calibration used later we take $\sigma = \frac{1}{2}$, hence $W \leq \tilde{C} m^{\alpha/(2d)}$; see Remark 3.4. With $W \leq m^{O(\alpha/d)}$ and $t = \Theta(\log n)$, the failure probability $(CpW)^t = o(1)$. When we require subpolynomial width we instantiate $d = \Theta(\log n)$, which gives $W = n^{o(1)}$; for constant d we retain a polynomial width bound. We will use this regime-specific width control; no claim is made that the bound is independent of the incoming width in a single round.

2.8 Summary

The size-aware API guarantees that TV/Meta , SPR_k , and EC are preserved up to a single $O(\log N)$ overhead, and that PROF is preserved exactly under output mixing ($\Delta = 1$) while the degree budget relaxes by at most a factor Δ under bounded fan-in (cf. Theorem B.5). This backbone lets windowed lower bounds propagate from the 3XOR source to 3SAT, and then echo as correlation and degree bounds against $\mathcal{C}_{s,d}^{\log}$ in the window $m = (1 + \gamma)n$.

3 Window lower bounds (EC focus)

We prove a distributional lower bound for *erasure complexity* (EC) in the balanced window. All parameters and overheads follow the size-aware API from Section 2. The backbone is a switching-path argument that leaves one *surviving parity* [19, 21, 36], combined with a reversible-simulation view that witnesses *irreversible* bit erasures under an explicit erasure condition (see Theorem 3.6) [8, 18, 23].

3.1 Setup and parameter path

Balanced window. We work with $m = (1 + \gamma)n$ for fixed $\gamma \in (0, 1)$; after the 3XOR→3SAT mapping the clause count is $m' = 4m$. Deterministic push-forwards are TV-contractive and have explicit size and meta accounting by Section 2. This regime aligns with the sharp 3XORSAT threshold at $m/n = 1$ (and its k -ary generalizations) [13, 37].

Restriction path. We use depth $d \in \mathbb{N}$ and a parameter $\alpha \in (0, 1)$, setting

$$p = m^{-\alpha/d} \quad \Rightarrow \quad m \cdot p^d = m^{1-\alpha}.$$

Gadgets. XOR and Index gadgets are constant-size and affine over $\mathbb{F}_2$ after output mixing; the Index gadget uses a constant address length (e.g., 6 bits).

EC model. We count bit-level erasures on a $w=64$ word-RAM. For a randomized algorithm A , $\#ERASE(A, I)$ is the number of logically non-injective steps; expectations are over internal randomness and $I \sim \mu$. We do not count the emission of the final decision bit (SAT/UNSAT) nor the return of a witness as erasures; a Bennett-style reversible simulation uncomputes all work tapes and ancillae, so that only the self-delimiting meta header h must be eventually disposed of, contributing the $O(\log N)$ term.

Let μ be the balanced 3XOR distribution, and let ν be its image under the TV-contractive, size- $\times 4$ push-forward to 3SAT. By the API (notably Theorems 2.1 and 2.3) we may analyze μ and transfer the bounds to ν and gadgetized variants at a meta cost of $O(\log N)$.

Notation. An instance is $I = (A, b)$, where A is the 3XOR clause-variable incidence over $\mathbb{F}_2$ and $b \in \{0, 1\}^m$ the right-hand side. An assignment is $y \in \{0, 1\}^n$. We use two parity types: (i) assignment parities $\chi_S(y) = \bigoplus_{i \in S} y_i \oplus b'$ with support $S \subseteq [n]$; (ii) RHS (instance) parities indexed by $\mu \in \{0, 1\}^m$, writing $T = \text{supp}(\mu)$ and $b_T := b|_T$. A *live RHS parity* means a nonzero μ with $\mu^\top A = 0$ (i.e. a row-dependency) whose support intersects the unfixed clause set after restriction.

Auxiliary references. We use facts fixed in the appendices: (i) TV contraction and a single prefix-free header of length $O(\log N)$, with a reversible wrapper that loads the header once and uncomputes it at the end (Appendix A); (ii) affine pullbacks and without-replacement concentration to control degree and sampling fluctuations (Appendix B). Under output mixing ($\Delta = 1$) the low-degree mass profile and Stab_ρ are preserved exactly; for bounded fan-in ($\Delta > 1$) we rely on the collision-aware upper bounds with factor M_k from the gadget pullback lemmas (B.5–B.6). The simple degree dilation $k \mapsto \Delta k$ serves only as a proxy where it is sound to do so.

Additionally, we use the ℓ_∞ (supremum) norm preservation under pullback, and we invoke the gadget pullback lemmas (B.5–B.6) to transfer bounds across gadget blocks.

Assumption (H_row-dist). Let $C \subseteq [m]$ denote the set of unfixed clauses after the restriction path (so $|C| = \Theta(m^{1-\alpha})$ w.h.p.). There exists a constant $\beta_{\text{row}} > 0$ such that, with probability $1 - o(1)$ over the instance and the restriction, the induced row-set A_C satisfies the minimum dependency weight bound

$$\min\{|\text{supp}(\mu)| : \mu \in \ker(A_C^\top) \setminus \{0\}\} \geq |C|^{\beta_{\text{row}}}.$$

We treat β_{row} as fixed throughout; it depends only on the window parameters (e.g., γ, d, α) and the sparsity regime. This is the standard row-distance hypothesis for sparse random 3-uniform systems, supported by LDPC distance/expansion analyses up to polynomial scales [1, 4, 5, 15].

Remark 3.1 (Scope of Section 3.1). The hypothesis holds for several standard sparse random ensembles. In LDPC-style parity-check matrices (left degree 3 and bounded right degrees), expansion at polynomial scales holds with high probability, which implies a polynomial lower bound on the minimum dependency weight in A_C , i.e. $\text{dist}_{\text{row}}(A_C^\top) \geq |C|^{\beta_{\text{row}}}$ for some constant $\beta_{\text{row}} > 0$ [1, 15]. For expander-code type constructions one even has linear distance (hence $\beta_{\text{row}} = 1$) under standard vertex/edge-expansion parameters [4, 5]. Our arguments only use that $\beta_{\text{row}} > 0$ (no optimization of the exponent is required).

3.2 Surviving parity along the switching path

Lemma 3.2 (Surviving parity). *There exists a restriction path ρ of depth d with parameter $p = m^{-\alpha/d}$ such that, with probability $\Omega(1)$ over ρ , the restricted instance retains a parity constraint $\chi(y) = b'$ (a nontrivial $\mathbb{F}_2$ -linear form over assignments with fixed bit b') that is not fixed by ρ .*

Proof. After d rounds each variable remains free with probability $p^d = m^{-\alpha}$. A 3-XOR clause is unfixed with probability $1 - (1 - p^d)^3 = 1 - (1 - m^{-\alpha})^3 \geq 3m^{-\alpha} - O(m^{-2\alpha})$. Over m clauses the expected number of unfixed clauses is $\Theta(m^{1-\alpha})$. A standard concentration bound (e.g., Azuma-Hoeffding along an exposure martingale) yields that with constant probability at least one clause remains unfixed. Any unfixed 3-XOR clause induces a nontrivial $\mathbb{F}_2$ -linear form χ over the assignment variables y that is not fixed. $\square$

Lemma 3.3 (Width reduction). *Under the same (d, p) , any DNF/CNF representation of the restricted instance has bottom width at most $\tilde{C} m^{\alpha/(2d)}$ (cf. Remark 3.4) with probability $1 - o(1)$.*

Proof. By Håstad's switching lemma, in each restriction round with parameter $p = m^{-\alpha/d}$ the bottom width drops to $W' \leq \tilde{C} m^{\alpha/(2d)}$ (cf. Remark 3.4) with failure probability at most $n^{-\Omega(1)}$ when the decision depth is chosen as $t = \Theta(\log n)$ [19, 21]. A union bound over $d \leq \Theta(\log n)$ rounds keeps the total failure probability $o(1)$. Hence after d rounds we have $W \leq \tilde{C} m^{\alpha/(2d)}$ (cf. Remark 3.4). $\square$

Remark 3.4 (Switching calibration and explicit $c_\star$). We use the following explicit calibration of the switching lemma (cf. [6, 19, 20]). There is an absolute constant $C_{\text{sw}} \geq 1$ such that for any bottom width W and survival rate p , the failure probability that the restricted formula has decision depth $> t$ is at most $(C_{\text{sw}} p W)^t$. We fix a concrete exponent $\sigma \in (0, 1)$, say $\sigma = \frac{1}{2}$, and define

$$W_\star(p) := \left\lfloor \left(\frac{1}{C_{\text{sw}} p} \right)^\sigma \right\rfloor.$$

Then $C_{\text{sw}} p W_\star(p) \leq (C_{\text{sw}} p)^{1-\sigma}$. For $p = m^{-\alpha/d}$ this gives (up to absolute constants)

$$W_\star(p) \leq C_{\text{sw}}^{-\sigma} m^{\sigma \alpha/d},$$

and for all sufficiently large n we have $C_{\text{sw}} p W_\star(p) \leq c_\star$ with the ****fixed constant**** $c_\star := \frac{1}{2}$. Consequently, choosing

$$t \geq \frac{10}{\log(1/c_\star)} \log n = \left\lceil \frac{10}{\log 2} \log n \right\rceil$$

makes the failure probability at most n^{-10} . Any constant $\sigma \in (0, 1)$ and any $c_\star \in (0, 1)$ would suffice; we fix $\sigma = \frac{1}{2}$ and $c_\star = \frac{1}{2}$ for definiteness. This matches the shorthand $W = m^{O(\alpha/d)}$ with the explicit exponent $W \leq \tilde{C} m^{\alpha/(2d)}$.

Lemma 3.5 (RHS product measure on a fixed leaf). *Let μ be the balanced 3XOR distribution where the RHS vector $b \in \{0,1\}^m$ is uniform and independent of the clause-variable incidence A . Let ρ be the switching-path restriction distribution of depth d and parameter $p = m^{-\alpha/d}$, and let L be any leaf event of the switching analysis that is measurable with respect to (A, ρ) (hence independent of b). Write $C = C(\rho) \subseteq [m]$ for the set of unfixed clauses on that leaf.*

Then, conditioned on L , the coordinates $b|_C$ remain mutually independent and each is $\text{Ber}(\frac{1}{2})$. In particular, for any nonzero $\mu_C \in \ker(A_C^\top)$ with support $T = \text{supp}(\mu_C) \subseteq C$, the parity $\mu_C^\top b = \bigoplus_{j \in T} b_j$ is uniform on $\{0,1\}$ (and remains independent of (A, ρ) given L).

Proof. By definition of the balanced 3XOR model, $b \sim \text{Ber}(\frac{1}{2})^m$ and is independent of A . The leaf event L is a function of (A, ρ) only and therefore independent of b . Hence the conditional law of b given L is still product $\text{Ber}(\frac{1}{2})^m$; restricting to the unfixed index set C preserves product form and marginals $\text{Ber}(\frac{1}{2})$. A parity of independent unbiased bits is unbiased, yielding the final claim. $\square$

3.3 From a live parity to erasures

Erasure-forcing event. For a (possibly randomized) SAT algorithm A on a restricted instance with a live RHS parity indexed by μ and support $T = \text{supp}(\mu)$ of size $L := |T|$, write $\text{EF}_\rho(A)$ for the event that over its run A discards $\Omega(L)$ bits of (Shannon) information about the RHS subvector $b_T := b|_T$; concretely, A performs logical merges that *merge many microstates that differ only on b_T* —hence encode $\Omega(L)$ bits about b_T —into a single state. This captures an unavoidable *loss* of instance information (about b_T), as opposed to merely processing it reversibly.

Lemma 3.6 (Information bottleneck implies erasures). *Consider any (possibly randomized) RAM algorithm A for SAT on the restricted instances. If A discards over its run $\Omega(L)$ bits of information about the RHS subvector b_T with $|T| = L$ (for a live RHS parity indexed by μ), then*

$$\mathbb{E}[\#\text{ERASE}(A, I)] \geq \Omega(L),$$

where the expectation is over the internal randomness and I drawn from the restricted distribution.

Proof. Model the execution as a stochastic map over machine configurations. Whenever the algorithm merges at least 2^Δ distinguishable microstates that encode Δ bits of b_T , the map is logically non-injective on a set of measure $\Omega(2^{-\Delta})$. By Landauer counting together with Bennett-style reversible simulation, any such merge forces $\Omega(\Delta)$ bit erasures in a w -RAM. The simulator overhead is $O(\log N)$ and is handled by the meta budget in Section 2 and appendix A. Summing over phases that discard $\Omega(L)$ bits about b_T gives the claim. In our w -RAM with a Bennett-style reversible wrapper, any single merge step that collapses at least 2^Δ distinct predecessor configurations forces the simulator to log Δ bits to remain injective; otherwise uncomputation would be impossible. When this history tape is later cleared to restore the working memory to a reference state, Landauer’s principle implies at least Δ physical bit-erasures, hence $\Omega(\Delta)$ erasures per such merge step in this model [8, 23]. $\square$

Lemma 3.7 (Length of a live RHS parity). *Let $C \subseteq [m]$ be the set of unfixed clauses after the restriction path, so $|C| = \Theta(m^{1-\alpha})$ with high probability. Assume the induced row-set on C has a minimum dependency weight $\text{dist}_{\text{row}}(A_C^\top) \geq |C|^{\beta_{\text{row}}}$ for some $\beta_{\text{row}} > 0$ (a standard “row-distance” hypothesis for sparse random 3-uniform systems). Then there exists a nonzero $\mu \in \ker(A_C^\top)$ with $|\text{supp}(\mu)| \geq |C|^{\beta_{\text{row}}}$. In particular, any live RHS parity supported on $T = \text{supp}(\mu)$ has effective length*

$$L = |T| \geq \Omega(|C|^{\beta_{\text{row}}}) = \Omega(m^{(1-\alpha)\beta_{\text{row}}}).$$

Proof. By definition of $\text{dist}_{\text{row}}(A_C^\top)$, every nonzero $\mu \in \ker(A_C^\top)$ has support at least $|C|^{\beta_{\text{row}}}$. Since $|C| = \Theta(m^{1-\alpha})$ w.h.p., taking such a μ yields a live RHS parity whose support T satisfies $|T| \geq |C|^{\beta_{\text{row}}} = \Omega(m^{(1-\alpha)\beta_{\text{row}}})$. $\square$

3.4 Main theorem (EC in the window)

Theorem 3.8 (EC window lower bound for SAT under $H_{\text{row-dist}}$). *Assume the row-distance hypothesis Section 3.1 holds and that the solver model satisfies $\Pr[\text{EF}_\rho(A)] \geq c_1 > 0$ along the restriction path of Theorem 3.2. Then there exists $\beta = \beta(\gamma, d, \alpha, \beta_{\text{row}}) > 0$ such that, for the 3SAT distribution ν obtained from balanced 3XOR with $m' = 4m$,*

$$\text{EC}_\nu(\text{SAT}) \geq m^\beta.$$

Proof. Work on μ (3XOR) and draw the restriction path ρ of Theorem 3.2. With probability at least $c_0 > 0$ a live parity survives and the restricted width is bounded by Theorem 3.3. Condition on this event. By Theorem 3.7 there is a live RHS parity with length $L = \Omega(m^{(1-\alpha)\beta_{\text{row}}})$.

We condition further on the erasure-forcing event $\text{EF}_\rho(A)$ from Section 3.3. On $\text{EF}_\rho(A)$, Theorem 3.6 yields $\mathbb{E}[\#\text{ERASE}(A, I) \mid \text{EF}_\rho(A)] \geq \Omega(L)$. Taking expectation over ρ and the internal randomness (including $I \sim \mu$), we obtain $\mathbb{E}[\#\text{ERASE}] \geq c_0 c_1 \cdot \Omega(m^{(1-\alpha)\beta_{\text{row}}})$, where $c_1 := \Pr[\text{EF}_\rho(A)]$. (Absent $\text{EF}_\rho(A)$ the argument does not claim erasures; the unconditional forcing of $\text{EF}_\rho(A)$ is model-dependent and not required for the bound stated here.)

We now undo the restriction via a reversible simulate-and-uncompute wrapper; the meta overhead is at most $O(\log N)$ erasures due to the prefix-free header and simulator bookkeeping (Appendix A and the API in Section 2). Finally, push-forward μ to ν (3XOR $\rightarrow$ 3SAT) via the TV-preserving size- $\times 4$ map; by the API (Theorem 2.1) erasure lower bounds transfer up to additive $O(\log N)$ slack. Absorbing constants gives the claim with $\beta \leq (1 - \alpha)\beta_{\text{row}}$, depending only on γ, d, α , the row-distance exponent β_{row} , and the occurrence probability c_1 of $\text{EF}_\rho(A)$. $\square$

Remark 3.9 (Explicit exponent). Fix any constant depth $d \geq 1$ and small constant $\gamma \in (0, 1)$. Along the switching path of Section 3.2 with $p = m^{-\alpha/d}$, the bottom width remains polynomial and a live parity survives with probability at least $c_0 > 0$. Assume the induced row-set on the unfixed clauses has a row-distance exponent $\beta_{\text{row}} > 0$ (cf. sparse random 3-uniform systems and LDPC-style distance analyses). With $p^d = m^{-\alpha}$ this yields an effective length exponent $\beta = (1 - \alpha)\beta_{\text{row}}$. Choosing any fixed $\alpha \in (0, 1)$ keeps $\beta > 0$ as long as $\beta_{\text{row}} > 0$. We do not optimize constants; any fixed positive β suffices for downstream echo bounds.

3.5 Robustness under XOR/Index gadgets

Lemma 3.10 (EC monotonicity under reversible preprocessing with prefix-free meta). *Let R be a deterministic map on instances of total bit-length N . Assume there is a reversible wrapper W_R that, on input x , (i) emits a self-delimiting header h of length $H = O(\log N)$, (ii) computes $y = R(x)$, and later (iii) uncomputes all work tapes so that, conditioned on h , every step is logically injective. Then for every distribution μ and Boolean task T ,*

$$\text{EC}_{R\#\mu}(T) \geq \text{EC}_\mu(T) - O(\log N).$$

Proof. Fix any SAT solver $\mathcal{A}_G$ for inputs $y \sim R\#\mu$. Construct $\mathcal{A}$ for $x \sim \mu$ as follows: run W_R to emit h and compute $y = R(x)$; simulate $\mathcal{A}_G$ on y using a Bennett-style reversible simulation;

uncompute R and all ancillae; at the very end, dispose of the header h . By construction, all steps except the final disposal of h are logically reversible; hence

$$\mathbb{E}[\#\text{ERASE}(\mathcal{A})] \leq \mathbb{E}[\#\text{ERASE}(\mathcal{A}_G)] + O(\log N),$$

where the $O(\log N)$ term is the cost of discarding the prefix-free header (Theorem A.2). Taking the infimum over $\mathcal{A}_G$ yields $\text{EC}_\mu(T) \leq \text{EC}_{R_\# \mu}(T) + O(\log N)$, which rearranges to the stated inequality. $\square$

Corollary 3.11 (Gadget robustness under output-mixed affine branches). *Let G be a constant-size gadget layer with a finite branch set $\{B_b\}_{b \in [B]}$. Assume:*

- (a) **Output-mixed affine normal form.** *After output mixing (Section 2.6, i.e., $\Delta = 1$), each branch B_b acts as an affine bijection over $\mathbb{F}_2$ on the live wires (a coordinate permutation up to fixed offsets).¹*
- (b) **Global branch choice.** *The branch index $b \in [B]$ is chosen once globally and encoded in a prefix-free header of length $O(\log N)$ (Theorems A.2 and 2.3); no per-position branching is used.*

Let $\nu_G := G_\# \nu$. Then

$$\text{EC}_{\nu_G}(\text{SAT}) \geq \text{EC}_\nu(\text{SAT}) - O(\log N) \geq m^\beta - O(\log N),$$

where m denotes the pre-gadget clause count (preserved under the output-mixed bijection).

Proof. Under (a) and (b), the pre/post-processing that decodes the global branch header and applies the inverse affine bijection of B_b is a reversible wrapper with a prefix-free header of length $O(\log N)$. Applying Theorem 3.10 with $R = G$ gives $\text{EC}_{\nu_G}(\text{SAT}) \geq \text{EC}_\nu(\text{SAT}) - O(\log N)$. By the window lower bound Theorem 3.8, $\text{EC}_\nu(\text{SAT}) \geq m^\beta$. Under output mixing ($\Delta = 1$), each branch is an affine bijection on the exposed wires; hence the constraint count m is invariant (only a coordinate renaming). Combining the two inequalities yields the claim. $\square$

3.6 Parameter summary

Depth d . Default $d = O(1)$; examples use $d = 3$. When a subpolynomial bottom width is required we also consider $d = \Theta(\log n)$ (see Section 4 and Appendix C).

Restriction p . $p = m^{-\alpha/d}$; the expected number of unfixed clauses is $\Theta(m^{1-\alpha})$.

Width W . $W \leq \tilde{C} m^{\alpha/(2d)}$ (cf. Remark 3.4).

LB exponent. $\beta = (1 - \alpha)\beta_{\text{row}} > 0$ (for any fixed $\alpha \in (0, 1)$ and $\beta_{\text{row}} > 0$).

Overheads. Deterministic push-forward is TV-contractive with size factor $\times 4$; a single global prefix-free meta header costs $O(\log N)$; total bit budget $N = \Theta(m \log n)$.

Remark 3.12 (Restriction rounds vs. decision depth). We distinguish the number of *restriction rounds* d from the *decision depth* t in the switching lemma. In both regimes we take $t = \Theta(\log n)$ so that the failure probability $(CpW)^t = o(1)$. With $d = O(1)$ we obtain a polynomial width bound—sufficient for §3 (EC) and the correlation/degree echo. The alternative regime $d = \Theta(\log n)$ is touched upon for the PC/SoS interface in §4; details and calibration are collected in Appendix C.

All statements are distributional and parameterized. We make no worst-case or P/poly claims.

¹The classical *Index* primitive is not affine in all its inputs; throughout we use only its standard *output-mixed affine-bijective normal form*, obtained by carrying the address wires so that the exposed interface per branch is a wirewise bijection. This makes the inverse map well-defined on the external wires.

n	$m = \text{round}((1 + \gamma)n)$ with $\gamma = 0.1$	3XOR→3SAT clauses = $4m$
64	70	280
96	106	424
128	141	564
192	211	844
256	282	1128

Table 2: Run parameters and clause scale for the empirical EC proxy.

3.7 Empirical summary (EC proxy) and cross-reference

We complement Theorem 3.8 with a compact, reproducible summary of an EC proxy. *This is purely illustrative and not a proof:* we count each decision as an erasure and cap backtracks; the proxy highlights the trend but does not certify asymptotic lower bounds. Results use small n and $\gamma = 0.1$; the mix $\text{unsat} = 0$ indicates an undercritical density regime and is not representative of asymptotic hardness windows. The following run produces a CSV and the plot shown in Figure 1 (Appendix D, Appendices D.5 and D.16). See appendices D.6 to D.15 for scripts and plots). Across all sizes, the status mix is: $\text{ok} = 164$, $\text{limit} = 336$, $\text{unsat} = 0$.

```
python scripts/ec_counter.py \
--n 64 96 128 192 256 \
--seeds 100 \
--gamma 0.1 \
--max-backtracks 20000 \
--count-decisions-as-erasure 1
```

The run writes `artifacts/results/ec_counter.csv` (500 rows) and the plot Figure 1. The average erasures grow cleanly with n for both subsets (“ok” and “limit”), matching the window lower-bound narrative.

4 Echo: correlation and degree

We show that a window lower bound for the erasure count (EC) propagates to (i) size-aware correlation bounds against $AC^0 + \log$ and (ii) a polynomial degree lower bound in the polynomial-calculus framework. All statements are distributional and parameterized, and they inherit the overhead regime of Sections 1 to 3.

Empirical echo. The multimode profiles in Appendix D show that cumulative mod- q mass at very small degrees and modest noise stability already predict positive PCG across our synthetic families. In particular, mod-2 mass up to $k \leq 6$ and Stab_ρ at $\rho \in \{0.1, 0.2\}$ correlate with the k -gram PCG on the same bitstreams. This alignment between low-order structure and contextual compression is the empirical *echo* that our correlation/degree arguments formalize.

Numerical summary. On the label-block scale runs we observe Pearson correlations between PCG and stability of $r \approx 0.14$ at $\rho = 0.1$ and $r \approx 0.10$ at $\rho = 0.2$ (Appendix D.15; `corr_with_pcg.csv`). The correlation table is stored at `appendices/data/multimod/corr_with_pcg.csv` together with the corresponding scatter plots (Appendix D.15).

Advice normalization. Throughout, the size-aware advice/log budget is $O(\log n)$ per input length (for $AC^0 + \log$).

4.1 Information to correlation (bridge to Section 3)

Lemma 4.1 (Information $\rightarrow$ correlation, window form). *Let ν be the push-forward distribution from Section 3 and let $\mathcal{T}$ be the switching-path restriction distribution with restriction depth $d \geq 1$ and parameter $p = m^{-\alpha/d}$ for a fixed $\alpha \in (0, 1)$ (default $\alpha = \frac{1}{3}$). Suppose that, with probability at least $c_0 > 0$ over $\rho \sim \mathcal{T}$ (as in Theorem 3.2), there exists a live RHS parity $\chi_\rho(x)$ determined by $t_* = \Omega(m^\beta)$ RHS instance bits b_T (not assignment bits y) that remains unfixed on the restricted subcube. Consider any predictor that, on each leaf of a decision tree of depth $t = \Theta(\log n)$ arising from the switching analysis, is given at most $O(\log n)$ bits of advice. Then there exists $\kappa = \kappa(\beta, \gamma, d) > 0$ such that its global correlation with SAT on ν is at most $n^{-\kappa}$.*

Proof. Fix any leaf of the decision tree (of depth $t = \Theta(\log n)$) obtained from the switching analysis (Håstad’s switching lemma [20]). On that leaf the restricted instance retains, with probability at least $c_0 > 0$ over $\rho \sim \mathcal{T}$, a live RHS parity χ_ρ on $t_* = \Omega(m^\beta)$ RHS instance bits b_T (from the window construction; cf. Section 3; not assignment variables y). Conditioned on the fixed decision-tree leaf, the surviving RHS coordinates remain unbiased and mutually independent by Theorem 3.5; hence the RHS parity χ_ρ is uniform on that leaf.

Conditioned on the leaf and the advice string $s \in \{0, 1\}^{\leq b}$, the t_* RHS bits are uniform and independent. Any (deterministic) predictor partitions the 2^{t_*} assignments of those bits into at most 2^b advice-classes. On each class, the imbalance between RHS parity 0 and 1 is at most the class size, hence the maximum possible bias of the RHS parity is bounded by 2^{b-t_*} . Therefore the leaf-wise correlation advantage is at most 2^{b-t_*} (standard counting/Fano-type argument).

Since $b = O(\log n)$ and $t_* = \Omega(m^\beta)$, for large n we have $2^{b-t_*} \leq n^{-\kappa}$ for some $\kappa = \kappa(\beta, \gamma, d) > 0$. Averaging over advice and internal randomness preserves the bound. Averaging over leaves and then over ρ yields the global correlation bound $n^{-\kappa}$. A Yao-minimax or Fano-style argument gives the same conclusion in a pure information-theoretic phrasing [12, 48]. $\square$

Anchoring sentence. We use Theorem 4.1 together with Theorem 3.6 (erasures imply information loss) and distribute the advice budget over decision-tree leaves via Yao’s minimax averaging (cf. [48]); this pins the per-leaf correlation budget that propagates through the push-forward and the gadget layer.

4.2 Correlation echo against size-aware $\text{AC}^0 + \log$

Let $\mathcal{C}_{s,d}^{\log}$ be Boolean circuits of depth $d = O(1)$, size $s = n^{O(1)}$, with an auxiliary log-budget of at most $O(\log n)$ advice bits hard-wired per input length. Correlation is measured against the distribution ν from Section 3. Classical correlation lower bounds against $\text{AC}^0[p]$ originate with Razborov and Smolensky [39, 44].

Theorem 4.2 (Correlation echo). *If $\text{EC}_\nu(\text{SAT}) \geq m^\beta$ for some $\beta > 0$, then there exist $\kappa = \kappa(\beta, \gamma, d) > 0$ and n_0 such that for all $n \geq n_0$ and all $C \in \mathcal{C}_{s,d}^{\log}$,*

$$\left| \Pr_{x \sim \nu}[C(x) = \text{SAT}(x)] - \frac{1}{2} \right| \leq n^{-\kappa}.$$

Proof. Apply the switching path with depth d and parameter $p = m^{-\alpha/d}$. By Håstad’s switching lemma [20] there exists $t = \Theta(\log n)$ such that with probability $1 - o(1)$ over ρ the restricted circuit $C \upharpoonright \rho$ is a decision tree of depth $\leq t$, and each leaf inherits at most $O(\log n)$ advice bits from $\mathcal{C}_{s,d}^{\log}$. Conditioned on the (constant-probability) slice where a live RHS parity persists, the per-leaf correlation is $\leq n^{-\kappa}$ by Lemma 4.1. Averaging over ρ and invoking Yao’s minimax (or a Fano-style

argument) gives the stated global bound. TV/size/meta overheads are absorbed by the API of Section 2. $\square$

4.3 Degree echo in the polynomial-calculus framework

Write $\deg_{\text{PC}}(f; \nu)$ for the minimal degree in the polynomial-calculus framework [40]. We relate (but do not identify) this to the minimal degree of a low-bias polynomial approximant under ν ; the two notions can differ in general [2], and the argument below only needs a one-way implication.

Theorem 4.3 (Degree echo: baseline and optional strengthening). *Assume the hypothesis of Theorem 4.2 and, in addition, a high-survival guarantee: with probability at least $1 - n^{-\lambda}$ (for some fixed $\lambda > 0$) over the restriction path of depth d and parameter $p = m^{-\alpha/d}$, a live RHS parity persists. For example, it suffices that the effective RHS parity length t_* after d rounds satisfies*

$$t_* \geq c \frac{\log n}{p^d} \quad \text{where } p = m^{-\alpha/d} \text{ and hence } p^d = m^{-\alpha}.$$

Then there exists $\zeta = \zeta(\beta, \gamma, \alpha, \lambda) > 0$ such that

$$\deg_{\text{PC}}(\text{SAT}; \nu) \geq \zeta \log n.$$

Optional strengthening. *If, in addition, one assumes flip-noise domination per restriction round (namely, for every multilinear P and every RHS parity χ , the correlation with χ after one restriction step at rate p is at most the correlation after applying the i.i.d. bit-flip operator T_η with $\eta = 1 - 2p$), then there exists $c' = c'(\beta, \gamma, \alpha, \lambda) > 0$ with*

$$\deg_{\text{PC}}(\text{SAT}; \nu) \geq c' \frac{\log n}{p} = \Omega(m^{\alpha/d} \log n).$$

Proof. Suppose, for contradiction, that there exists a polynomial P of degree $d_0 < C \log n$ (for a sufficiently small constant $C > 0$) with $\|P\|_2 = 1$ and correlation $\varepsilon := |\mathbb{E}_{x \sim \nu}[P(x) \text{SAT}(x)]| \geq n^{-\kappa_0}$ for some $\kappa_0 > 0$. (Normalization w.l.o.g.; scaling preserves degree and only tightens the forthcoming upper bounds linearly in $\|P\|_2$.) Draw $\rho \sim \mathcal{T}$. With probability at least $1 - n^{-\lambda}$ (high-survival) the restricted target $\text{SAT} \upharpoonright \rho$ has a live RHS parity χ_ρ on $t_* = \Omega(m^\beta)$ variables and $P \upharpoonright \rho$ has degree $\leq d_0$. Let $s := p^d = m^{-\alpha}$ be the per-bit survival probability after d rounds. By Theorem B.8, d rounds at rate p are equivalent to a single product restriction with survival $s := p^d = m^{-\alpha}$. On the high-survival slice, the target is an RHS parity $\chi_{U(\rho)}$ over the alive subset $U(\rho) \subseteq [t_*]$; in particular $|U(\rho)| \sim \text{Bin}(t_*, s)$, and whenever $|U(\rho)| > d_0$ we have $\langle P \upharpoonright \rho, \chi_{U(\rho)} \rangle = 0$ (degree cut-off). Therefore, without any flip-noise domination, Theorem B.9 yields

$$\mathbb{E}_\rho \left[|\langle P \upharpoonright \rho, \chi_{U(\rho)} \rangle| \mid \text{Alive} \right] \leq \|P\|_2 \cdot \Pr[|U(\rho)| \leq d_0 \mid \text{Alive}]^{1/2}.$$

With $t_* \geq c(\log n)/s$ (the high-survival condition) and a Chernoff bound, $\Pr[|U(\rho)| \leq d_0] \leq \exp(-\Omega(t_* s)) \leq n^{-\Omega(1)}$ as long as $d_0 \leq c' \log n$ for a suitable $c' = c'(\alpha, \beta, \gamma, \lambda)$. Since $\|P\|_2 = 1$,

$$\mathbb{E}_\rho \left[|\langle P \upharpoonright \rho, \chi_{U(\rho)} \rangle| \mid \text{Alive} \right] \leq n^{-2\kappa_0}.$$

Combining with $\Pr[\text{Alive}] = 1 - n^{-\lambda}$ and $|\text{SAT} \upharpoonright \rho| = 1$ yields

$$|\mathbb{E}[P \upharpoonright \rho \cdot \text{SAT} \upharpoonright \rho]| \leq n^{-2\kappa_0} \|P\|_2 + n^{-\lambda} \|P\|_2 < n^{-\kappa_0} \|P\|_2$$

for all large n (taking $\lambda > \kappa_0$), contradicting $\varepsilon := |\mathbb{E}_{x \sim \nu}[P(x) \text{SAT}(x)]| \geq n^{-\kappa_0}$ after undoing the restriction and the push-forward (API slack $O(\log N)$). Therefore $\deg_{\text{PC}}(\text{SAT}; \nu) \geq \zeta \log n$ for some $\zeta = \zeta(\alpha, \beta, \gamma, \lambda) > 0$.

Optional strengthening (additional domination). If one additionally assumes a *flip-noise domination* per round (“one-step restriction does not increase correlation with any RHS parity relative to T_η with $\eta = 1 - 2p$ ”), then Bonami–Beckner gives the sharper per-round decay $\leq (1 - 2p)^{d_0}$ and composing d rounds yields the bound in $\exp(-2p d_0)$. With the same averaging step this recovers the advertised $\deg_{\text{PC}} \geq c'(\log n)/p$. We do not rely on this extra assumption for the baseline theorem. $\square$

Remark 4.4 (Sufficient condition via row distance). In the 3XOR window, the effective RHS parity length satisfies $t_* = \Omega(|C|^{\beta_{\text{row}}}) = \Omega(m^{(1-\alpha)\beta_{\text{row}}})$, where $|C| = \Theta(m^{1-\alpha})$ is the number of unfixed clauses. Since $p^d = m^{-\alpha}$, the high-survival condition $t_* \geq c(\log n)/p^d$ holds whenever $(1 - \alpha)\beta_{\text{row}} \geq \alpha + o(1)$. We do not optimize constants; see Section 3.1 and Theorem 3.7 in Section 3 for the source of t_* and the row-distance premise.

Remark 4.5 (SPR/PCG variants). An analogous echo holds if the window gap is proved for SPR_k (constant-rate log-loss) or for PCG (contextual compression gap), with parameters κ, ζ adjusted by the respective API constants of Section 2.

4.4 Empirical cross-references

SPR. As a consistency check for the switching-path narrative behind the echo, our SPR proxy (finite-state k -gram prediction on restriction-path histories) shows a sharp drop in average log-loss as k increases and then plateaus; see Figure 2 (Appendix D.6). This aligns with the presence of a surviving RHS parity signal and supports the correlation/degree echo.

PCG. Appendix D.9–D.13 report (i) scaling with n , (ii) side-context sensitivity, and (iii) label-separation grids. The k -gram gaps increase with n and with separation $|p_1 - p_0|$, and they diminish as k grows, plateauing at moderate k . This is consistent with residual structure that finite-state context models can exploit, supporting the echo here.

Multimode. Appendix D.14–D.15 quantify low-order mod- q Fourier mass and noise stability and show a positive association with PCG on the same bitstreams. The mass climbs at small k and Stab_ρ degrades gently, indicating structure that finite-state context models can reliably exploit. This aligns with the echo narrative: surviving low-order regularities both hinder size-aware AC^0 +log predictors (correlation echo) and force logarithmic degree in PC (degree echo).

4.5 Scope and constants

Switching calibration and decision depth. We use an explicit calibration of the switching lemma: there is an absolute constant $C_{\text{sw}} \geq 1$ such that the failure probability is at most $(C_{\text{sw}} p W)^t$. We fix $\sigma = \frac{1}{2}$ and set $W_\star(p) := \lfloor (1/(C_{\text{sw}} p))^\sigma \rfloor$, so $C_{\text{sw}} p W_\star(p) \leq (C_{\text{sw}} p)^{1-\sigma}$. With $p = m^{-\alpha/d}$ this yields $W_\star(p) \leq \tilde{C} m^{\alpha/(2d)}$, and for all sufficiently large n we have $C_{\text{sw}} p W_\star(p) \leq c_\star$ with the fixed choice $c_\star = \frac{1}{2}$. Taking

$$t = \left\lceil \frac{10}{\log 2} \log n \right\rceil$$

makes the failure probability at most n^{-10} . This makes precise our shorthand $t = \Theta(\log n)$ and $W = m^{O(\alpha/d)}$ with the explicit exponent $W \leq \tilde{C} m^{\alpha/(2d)}$.

Degree echo with explicit plug-ins. Assuming an EC gap $\text{EC}_\nu(\text{SAT}) \geq m^\beta$ and the high-survival condition

$$t_* \geq c \frac{\log n}{p^d} \quad \text{with } p = m^{-\alpha/d} \text{ so } p^d = m^{-\alpha},$$

Theorem 4.3 gives the *baseline*

$$\deg_{\text{PC}}(\text{SAT}; \nu) \geq \zeta \log n.$$

Under the *optional* flip-noise domination per restriction round, this strengthens to

$$\deg_{\text{PC}}(\text{SAT}; \nu) \geq c' \frac{\log n}{p} = c' m^{\alpha/d} \log n.$$

We spell out two regimes (balanced window $m \asymp n$); unless explicitly stated otherwise the default is $d = O(1)$, and we use $d = \Theta(\log n)$ only when a subpolynomial bottom width is required:

- *Constant restriction depth.* For $d = 3$ and $\alpha = \frac{1}{3}$, the baseline yields $\deg_{\text{PC}}(\text{SAT}; \nu) \geq \zeta \log n$. (*Optional*) With flip-noise domination one further gets $\deg_{\text{PC}}(\text{SAT}; \nu) \geq \Omega(m^{1/9} \log n)$.
- *Depth scaling* $d = \Theta(\log n)$. If $d = c_{\text{sw}} \log n$ with $c_{\text{sw}} > 0$, then $m^{\alpha/d} = \Theta(1)$ (since $m \asymp n$), so both the baseline and the optional strengthening give $\deg_{\text{PC}}(\text{SAT}; \nu) = \Theta(\log n)$, with constants depending on $(\alpha, c_{\text{sw}}, \beta, \gamma)$.

Advice budget and correlation echo. With a size-aware advice budget $O(\log n)$ per input length for $\text{AC}^0 + \log$, Lemma 4.1 gives correlation at most $n^{-\kappa}$ for some $\kappa = \kappa(\beta, \gamma, d) > 0$, which feeds into the degree bound above. All TV/size/meta overheads are absorbed via the API of Section 2. (Logarithm bases are immaterial up to constant factors.)

We work in the same window and overhead regime as in Sections 1 to 3: a deterministic push-forward (TV-contracting) from 3XOR to 3SAT with size factor $\times 4$, and a single global prefix-free meta header $O(\log N)$. Size-aware correlation is measured against the push-forward distribution ν from Section 3. The advice budget for $\mathcal{C}_{s,d}^{\log}$ is $O(\log n)$. All constants (κ, ζ) depend only on $(\beta, \gamma, \alpha, d, \lambda)$ and absolute gadget constants.

4.6 Formal bridge to Sections 2-3 and the appendices

We repeatedly appeal to the API of Section 2 (data processing, gadget stability, block aggregation) and to the auxiliary lemmas in Appendix A (TV 1-Lipschitz, prefix-free headers, simulator lifting) and Appendix B (affine pullbacks, hypergeometric concentration). In the correlation proof the switching-path analysis yields, with high probability, a decision tree of depth $t = \Theta(\log n)$ on restricted subcubes; any attempt to retain the $\Omega(m^\beta)$ live RHS parity information would exceed the $O(\log n)$ advice budget, while forgetting it forces near-random behavior on a constant slice. In the degree proof the same surviving RHS parity mechanism implies that low-degree polynomials cannot achieve non-negligible correlation under the restriction distribution, giving $\deg_{\text{PC}} \geq \zeta \log n$ under the high-survival condition $t_* \geq c'(\log n)/p^d$. Under the optional flip-noise domination one further gets $\deg_{\text{PC}} \geq c'(\log n)/p$. All overheads are absorbed by the $O(\log N)$ meta budget supplied by Section 2.

5 Related work

This section places our invariants and the size-aware gadget view in context: reductions, Fourier-analytic techniques, proof complexity, and information-theoretic probes of structure. We emphasize what our results need and what they refine relative to prior lines of work.

5.1 Gadget reductions and affine interfaces

Reductions that pass structure through small interfaces are a staple of complexity theory. Classical encodings of parity into conjunctive normal form go back to Tseitin-style constraints, where each XOR is represented by a constant number of clauses [46]. Affine maps over the two-element field appear implicitly in many dictatorship tests and in label-cover-based compositions [35]. Our treatment makes this algebraic interface explicit: we view a gadget as an affine map on coordinates and ask which statistics of the target instance are preserved when pulled back through a coordinate permutation. This isolates the role of the interface apart from the internal clauses.

Within this perspective, our affine pullback statement says that low-degree Fourier mass up to level k and noise stability at correlation ρ are invariant when the interface is a coordinate permutation. The point is not novelty per se, but that this is rarely stated as an *interface contract* reusable across reductions. In particular, $\deg(g) = \deg(f)$ under a permutation, together with exact preservation of the mod- q mass profile, gives a self-contained way to track low-order structure across XOR $\rightarrow$ SAT lifts without rederiving ad hoc Fourier facts in each setting.

5.2 Fourier analysis of Boolean structure

Fourier analysis on the discrete cube is the standard language for describing low-order correlations and their aggregation. The decomposition into characters, the level-wise mass profile, and hypercontractive inequalities go back to Bonami and Beckner and are presented in a unified way in modern treatments [7, 9, 35]. Stability under noise and invariance principles connect low-degree mass to robustness under random perturbations [34]. Our use of the truncated mass profile $\sum_{1 \leq |S| \leq k} \widehat{f}(S)^2$ follows this tradition but packages the relevant parts as a portable invariant. What we add is a small accounting layer that shows how these quantities behave under explicit affine pullbacks and under bounded fan-in maps with parameter Δ .

The separation between exact preservation at $\Delta = 1$ and controlled degradation when $\Delta > 1$ is consistent with how influences and degrees propagate through low-complexity circuits. We do not claim new general theorems; the contribution is an interface-oriented restatement that can be applied mechanically inside reductions.

5.3 Noise sensitivity and stability

Noise sensitivity and stability quantify how much a function changes when its input is perturbed [35]. “Majority is stablest” and related inequalities identify regimes where stability is governed by low-order structure [34]. Our invariant Stab_ρ is the standard stability functional. Its role here is twofold: first, it is part of the preserved profile under affine pullback; second, it correlates in our experiments with compression-based indicators of structure in synthetic bitstreams. We do not propose a new inequality; we fix a concrete estimator that behaves well under the transformations used in our lifts.

5.4 Switching lemmas and decision-tree viewpoints

Random restrictions and switching lemmas explain why low-depth representations tend to collapse to shallow decision trees [19, 21]. Parameter choices in these lemmas determine concrete tradeoffs between depth and width. Our appendix collects the parameterizations we actually instantiate in experiments and in window lower-bound arguments. The novelty is organizational. We use the same parameter notations on both the combinatorial and the empirical side and keep constants

visible, since our window arguments need size awareness. This lets us amortize header costs and state descriptions across repeated applications without hiding them in asymptotic notation.

5.5 Sum of Squares and proof complexity for XOR and SAT

Lower bounds for systems of linear equations over the two-element field and for random CSPs inside The Sum-of-Squares (SoS) hierarchy is a central reference point [3, 17, 24, 25]. They explain why certain refutation strategies require many rounds even when the instance has strong average-case properties. Our companion note on minimax gadgets and SoS fits into that picture. We do not reprove those lower bounds. Instead we design interfaces compatible with SoS compositions and that preserve the low-degree mass profile needed to transfer hardness through gadget layers. The minimax viewpoint clarifies which features must be kept invariant so downstream SoS arguments remain valid after the lift.

5.6 Encoding parity into CNF and window-aware lifts

Mappings from XOR to SAT with four clauses per parity constraint are standard. Less standard is to track exactly how much low-order correlation survives such a mapping once we insist on a size budget and an explicit account of technical overhead. Our window lower bounds are stated with an explicit $(1 + \gamma)n$ regime and with persistent accounting for headers and interface states. This picks up a line of work that emphasizes quantitative bounds with visible constants rather than only asymptotic equivalences. The benefit is that experimental proxies like compression-based gaps can be placed on the same numerical scale as the analytic invariants.

5.7 Information theory, MDL, and compression as structure probes

Using data compression as a probe for regularity is a long-standing idea. The minimum description length principle and universal coding explain why better compression often reflects learnable structure [12]. Classical universal codes such as Lempel–Ziv are robust baselines and easy to instrument [30, 49]. We define the pattern-context gap (PCG) as the difference between an unconditional code length and a conditional code length given an external pattern. This is a simple MDL surrogate rather than a new estimator. Its value here is that it co-moves with the low-degree mass profile under the gadget operations we study. The pairing lets us verify that the analytic invariants we track are visible to concrete compressors even at short lengths.

Our use of total-variation-preserving prefix coding is a small technical device that keeps distributions aligned when we translate between bitstreams and model outputs [12]. This avoids a preventable source of bias when comparing mass profiles to compression scores.

5.8 Erasure complexity and reversible viewpoints on search

The idea that information erasure has a physical cost goes back to Landauer and to Bennett’s reversible simulation [8, 23]. When a backtracking search unassigns a variable and discards a branch, at least one bit of information is destroyed. We follow a simple counting proxy for this effect and use it to obtain a distributional lower-bound trend in a narrow window of clause densities. This perspective is compatible with the rest of the paper because the affine interface makes it possible to track how many erasures are forced by parity structure that remains visible after the lift. The point is not to claim a physical law for SAT; it is to connect a monotone cost to a reduction that preserves the relevant structure.

5.9 Relation to empirical work on spectral diagnostics

There is a body of work that uses low-order moments, Fourier coefficients at small levels, and stability curves as diagnostics for structure in synthetic data. Our experiments sit in this empirical tradition. The contribution is to hold the diagnostics fixed while we vary the interface and the gadget composition. This isolates which parts of the signal are intrinsic to the source and which parts are artifacts of the encoding. The observed alignment between the truncated mass profile, the stability curve at a few fixed ρ , and the pattern-context gap (PCG) supports the claim that our invariants are the right ones for size-aware lifts.

5.10 Summary of distinctions

We do not introduce new general inequalities. We repackage known Fourier and stability tools as portable invariants and make the affine interface an explicit unit of analysis. We combine this with window-aware accounting that keeps small overheads visible, and with compression-based proxies that are easy to replicate. This combination appears to be absent from prior treatments of XOR→SAT lifts and from discussions of SoS stability under gadget composition. The intention is a reusable template for size-constrained hardness transfers where low-order structure must be preserved with equality at fan-in one and with controlled loss at bounded fan-in.

Software package. For scripts and artifact bundles, see the Zenodo software archive [29].

6 Conclusion and outlook

This work isolates a small pair of invariants—low-degree mod- q mass up to level k and noise stability Stab_ρ (together, the *profile*)—and treats the affine gadget interface as a first-class object. Under a coordinate permutation ($\Delta = 1$) the profile is preserved exactly; for bounded fan-in $\Delta > 1$ the degree budget expands by at most a factor Δ , while the retained information can be tracked via an averaged low-degree mass profile with explicit, size-aware accounting. Combined with a window formulation $m = (1 + \gamma)n$ that keeps headers and interface states visible, this yields a compact template for hardness transfers from 3XOR to 3SAT where structure must remain measurable after the lift.

On the empirical side, we paired these invariants with simple, reproducible diagnostics: truncated Fourier mass, a short-grid Stab_ρ , a compression-based pattern-context gap (PCG), and a backtracking erasure counter as a monotone proxy. Across controlled synthetic families, these diagnostics co-move under the same interface manipulations that the invariants certify, providing an external check that the algebraic bookkeeping reflects signals that practical compressors and small search procedures can detect.

6.1 Main contributions

1. **Interface-first formulation.** The affine map of a gadget is made explicit; permutation interfaces ($\Delta = 1$) preserve the profile exactly, while bounded fan-in admits controlled expansion with constants kept visible.
2. **Window-aware lower-bound template.** The $(1 + \gamma)n$ regime is analyzed with persistent accounting for headers and state; the resulting amortization explains why small overheads do not swamp low-order structure in repeated applications.

3. **Portable invariants.** The low-degree mass profile and Stab_ρ are packaged as a reusable contract: once verified at the source, they remain valid after the lift within the declared size budget.
4. **Empirical alignment.** PCG and short-grid stability correlate with the analytic profile on synthetic streams; the erasure counter shows a distributional lower-bound trend in the same window, linking structural preservation to a monotone cost observable by search.

6.2 Limitations

- **No new general inequalities.** We repackage known Fourier and stability facts; we do not claim sharper hypercontractive or invariance bounds.
- **Bounded fan-in only.** For $\Delta > 1$ we provide degree expansion control and averaged low-degree mass tracking, not a full stability theorem.
- **Finite-grid diagnostics.** Experiments use small k and a fixed set of ρ ; constants are not optimized.
- **Erasure proxy.** The erasure counter is principled but indirect; it is not a formal proof-complexity lower bound.

6.3 Open technical directions

1. **Tight constants under bounded fan-in.** Replace $o(1)$ losses by explicit $\tilde{O}(1/T)$ terms under repeated composition, with a state machine that holds the header **once** and shows reversibility step by step.
2. **Blockwise interfaces.** Extend exact preservation from pure permutations to blockwise permutations with sparse cross-block mixing; quantify precise leakage of low-degree mass.
3. **TV-preserving coding generalization.** Strengthen the prefix-coding device so mixture components match at the interface without inflating the low-order spectrum; verify equality cases for biased sources.
4. **Robust low- n estimators.** Calibrate bias-corrected estimators for truncated mass and stability at short lengths; publish confidence intervals alongside point estimates and lock a public benchmark with fixed seeds.

6.4 Further directions

1. **Sum-of-Squares compatibility.** Investigate encoding profile preservation as constraints on pseudoexpectations so SoS reductions can compose gadgets while tracking the active low-degree budget; connect the resulting degree bookkeeping to round lower bounds on lifted instances.
2. **Learning-aware reductions.** Combine the interface contract with decision-tree or switching-lemma views [19, 21] to obtain window bounds explicit in both γ and depth, keeping constants stable under mild derandomization.
3. **Noise models and biased sources.** Analyze preservation under bit-flip noise and small bias at the source, distinguishing true degradation from interface-induced artifacts; characterize when a permutation-style interface yields exact stability at chosen ρ .

4. **Instance design under size budgets.** Use the profile contract to engineer reductions that pass just enough low-order mass to remain detectable by practical compressors and by shallow SoS rounds, while staying within strict clause budgets.

6.5 Empirical program and reproducibility

- **Locked presets.** This paper uses a fixed suite of presets (lengths, ρ -grid, k -grid, seeds), ensuring comparability across runs.
- **Code and data hygiene.** Header/state accounting is implemented as a first-class concept; each run logs interface parameters and exports raw as well as aggregated profiles.
- **Cross-diagnostic checks.** We report PCG, truncated mass, stability, and erasure trends jointly and on a comparable scale; conclusions are supported by at least two diagnostics.

6.6 Open problems

1. **Sharp preservation beyond permutations.** Determine the exact low-degree mass that survives a single sparse mixing step and the minimal conditions under which equality persists.
2. **Stability at bounded fan-in.** Prove a finite-constant stability guarantee that tracks the mass profile through $\Delta > 1$ without reverting to asymptotic $o(1)$ language.
3. **PCG-spectrum equivalence.** Establish conditions under which PCG tightly controls the truncated Fourier mass up to multiplicative constants over short ranges of n .
4. **Erasure-complexity links.** Relate the backtracking erasure count to standard proof-complexity measures on lifted instances in the window $m = (1 + \gamma)n$.
5. **Window-tight reductions.** Design $3\text{XOR} \rightarrow 3\text{SAT}$ gadgets whose interface states amortize exactly over T steps, eliminating slack between analytic preservation and observed compression signals.

Outlook (non-binding; beyond this paper)

This outlook sketches possible next steps and does not modify any theorem or empirical claim of this paper. IECZ-III closes with the interface API, the window lower-bound template, and the echo diagnostics; the items below are beyond this work.

1. **Pseudodeterministic restriction/condensation.** Achieve seeded variants with the same $O(\log N)$ header footprint.
2. **Row-distance to high-probability statements.** Upgrade the balanced-window phenomenon to w.h.p. results for random 3-uniform/LDPC-style families.
3. **Links to classical proof complexity.** Connect the EC proxy to Resolution width/space via reversible wrappers.
4. **Advice-free and low-depth classes.** Target AC^0 (no advice) and ACC^0 via switching- and decision-tree tools, together with profile contracts.
5. **Interface-contract tests.** Provide a small, reusable library that checks permutation/fan-in- Δ guarantees and header amortization.

6. **Quantitative constants.** Calibrate (α, Δ) on small grids with confidence intervals for future quoting.
7. **Beyond XOR/Index.** Extend to Label-Cover/Goldreich-type gadgets that become affine after output mixing.
8. **Distributional $\rightarrow$ average/worst case.** Use E1/E2 within the window to guide hardness amplification.
9. **AE packaging.** Optional containerized bundle (Docker/Conda) and one-shot make repro_all.

Takeaway. Treating the interface as an algebraic object and binding it to a minimal pair of low-order invariants yields a size-aware, reusable template for lifts. With explicit accounting, the small parts of a reduction remain visible, and the structure that matters survives intact where it must and in a quantified way where it cannot. The combination is simple to verify, easy to test, and strong enough to transfer hardness within realistic size budgets.

A TV contraction and prefix-free headers

We record the two meta principles used throughout: deterministic maps contract total variation, and self-delimiting (prefix-free) headers for instance-parameterized preprocessors cost only $O(\log N)$ bits when encoded with standard integer codes [12, 14].

Log base (coding). Here $\log_2$ is used for code lengths and entropy (bits); elsewhere we write $O(\log N)$ with base-2 logs by default, which is asymptotically base-insensitive.

A.1 TV contraction and prefix-free meta overhead

Lemma A.1 (TV contraction). *For any deterministic mapping R and distributions μ, ν on the same space,*

$$\text{TV}(R_{\#}\mu, R_{\#}\nu) \leq \text{TV}(\mu, \nu).$$

Proof. By definition, $\text{TV}(\alpha, \beta) = \sup_A |\alpha(A) - \beta(A)|$ where the supremum is over all measurable sets. For any measurable B in the range of R we have $(R_{\#}\mu)(B) = \mu(R^{-1}(B))$ and likewise for ν . Hence

$$\text{TV}(R_{\#}\mu, R_{\#}\nu) = \sup_B |\mu(R^{-1}(B)) - \nu(R^{-1}(B))| \leq \sup_A |\mu(A) - \nu(A)| = \text{TV}(\mu, \nu).$$

This is the total-variation case of the data processing inequality; see also [38, 42]. $\square$

Lemma A.2 (Prefix-free meta overhead). *If descriptions of instance-dependent preprocessor states are self-delimiting (e.g., Elias- δ), then the meta header length is $O(\log N)$ bits for inputs of total bit-length N (standard universal coding; cf. [12, 14]).*

Proof. Let N be the input bit-length and let the meta header encode a finite tuple of nonnegative integers $A = (a_1, \dots, a_t)$, where $t = O(1)$ is fixed by the API, and $\sum_{i=1}^t a_i \leq c \log_2 N$ for some absolute constant $c > 0$ (instantiated by the items in Remark A.4: one $O(\log N)$ -seed (E3), one $O(1)$ -branch index, an $O(1)$ address length, and $O(\log N)$ index-fixing-selector bits (E4)). Encode each a_i

with a self-delimiting universal integer code (Elias- γ or Elias- δ), applied to $u_i := a_i + 1 \geq 1$ to meet the domain requirement. Elias- ω achieves comparable asymptotics and can be used interchangeably for our bounds [14].

For γ one has $\ell_\gamma(u) = 2\lfloor \log_2 u \rfloor + 1 \leq 2\log_2 u + 1$. For δ one has

$$\ell_\delta(u) = \lfloor \log_2 u \rfloor + 2\left\lceil \log_2 (\lfloor \log_2 u \rfloor + 1) \right\rceil + 1 \leq C_0 + C_1 \log_2 u$$

for absolute constants C_0, C_1 (the $2\log\log$ term is absorbed into the constant for small u and into the slope for large u). Thus, for either code, $\ell(u) \leq C_0 + C_1 \log_2 u$ holds uniformly for $u \geq 1$.

The total header length satisfies

$$L = \sum_{i=1}^t \ell(u_i) \leq \sum_{i=1}^t (C_0 + C_1 \log_2(a_i + 1)) = C_0 t + C_1 \sum_{i=1}^t \log_2(a_i + 1).$$

Using $\log(1+x) \leq x$ for $x \geq 0$ and $\sum_i a_i \leq c \log_2 N$ yields

$$\sum_{i=1}^t \log_2(a_i + 1) \leq \frac{1}{\ln 2} \sum_{i=1}^t a_i \leq \frac{c}{\ln 2} \log_2 N,$$

hence $L \leq C_0 t + (C_1 c / \ln 2) \log_2 N = O(\log N)$ since $t = O(1)$. Because the integer codewords are prefix-free, the concatenation is uniquely parseable without separators, so these bits are the full meta-header cost.

Sharper bound (optional). By concavity of $\log$,

$$\sum_{i=1}^t \log_2(a_i + 1) \leq t \log_2\left(1 + \frac{1}{t} \sum_{i=1}^t a_i\right) \leq t \log_2\left(1 + \frac{c \log_2 N}{t}\right).$$

Hence

$$L \leq C_0 t + C_1 t \log_2\left(1 + \frac{c \log_2 N}{t}\right) = O(t \log_2 \log_2 N).$$

Since $t = O(1)$ in our API, this sharpens to $L = O(\log_2 \log_2 N)$. For Elias- δ , the additional $2\log\log$ term also sums to $O(t \log_2 \log_2 N)$, so the sharpening is preserved. $\square$

Corollary A.3 (API headers). *All API headers in Section 2 cost at most $O(\log N)$ bits across the entire pipeline (or $O(\log n)$ when the active horizon is $T = \Theta(\log n)$).*

Proof. Apply Theorem A.2 to the finite menu of modes and counters used by the API. Headers are emitted once and amortized globally, so costs add within $O(\log n)$. $\square$

Remark A.4 (Header fields used in this paper). The prefix-free header encodes only a *constant* tuple of integers whose total magnitude is $O(\log N)$. Concretely (cross-referencing Section 2, Table 1, Theorem 2.3, and Appendix C):

- **Seed for seeded condensation (E3).** A single global seed of length $O(\log N)$ (see row E3 in Table 1); this is emitted once and reused.
- **Branch index for constant-branch gadgets.** For an Index gadget realized as a mixture of $B = O(1)$ affine branches, we encode one branch index $b \in [B]$ once globally (Theorem 2.3); this costs $O(1)$ bits and does not vary per position (Appendix C.3).

- **Address length (Index gadget).** The address length is a fixed constant (default 6 bits; see Section 2), hence either compiled-in or, if recorded, it contributes $O(1)$ bits.
- **Index fixing / selector tags (E4).** Deterministic index/schedule selection contributes $O(\log N)$ bits once (row E4 in Table 1); no per-position tags are emitted.

These items instantiate the tuple $A = (a_1, \dots, a_t)$ in Theorem A.2 with $t = O(1)$ and $\sum_i a_i = O(\log N)$, justifying the $O(\log N)$ bound for the *entire* pipeline header.

A.2 Notes on compression surrogates used for PCG

Remark A.5 (LZ78 surrogate choices used in PCG).

- *Final phrase pointer:* if the last phrase ends exactly at stream end, emit only the pointer (no extra symbol bit).
- *Pointer-cost model:* either uniform $\max(1, \lceil \log_2 |D| \rceil)$ or Elias–Gamma length of $(\text{index} + 1)$ (the VLC surrogate), the latter reflecting variable-length indices.
- *Label-block CMDL:* define $\text{CMDL}(X \mid C)$ as the sum of per-label code lengths $\text{LZ}(X \mid C=0) + \text{LZ}(X \mid C=1)$ with *no* label-stream cost. This mirrors the LZMA baseline we report.

These conventions were used to produce the PCG figures in Appendix D.

Remark A.6 (LZMA container overhead in CMDL estimates). In label-block runs with a real compressor (LZMA), a near-constant negative PCG offset can appear at small to mid n due to container and header effects that do not fully cancel between $\text{MDL}(X)$ and the per-label CMDL terms. We base quantitative claims on k -gram and the VLC surrogate and use LZMA primarily as a qualitative trend check.

Remark A.7 (Fan-in Δ vs. output mixing). Throughout we normalize gadgets by output mixing whenever admissible. Under this normalization the interface is an affine *coordinate permutation* ($\Delta = 1$), so the profile (*low-order mod- q mass up to level k and Stab_ρ*) is preserved exactly by pullback. When mixing is not assumed and local fan-in is $\Delta > 1$, composition acts as an affine pullback with fan-in Δ : level budgets relax as $k \mapsto \Delta k$ and the effective noise parameter adjusts monotonically, as used in Section 2.

B Hypergeometric concentration and affine pullbacks

This appendix collects concentration bounds for sampling without replacement and records how affine pullbacks interact with low-degree structure. The concentration facts control fluctuations arising from block aggregation and from constant-size gadget selections; affine pullbacks under bounded fan-in explain degree dilation and the exact preservation in the permutation case.

B.1 Concentration for sampling without replacement

Lemma B.1 (Serfling-type deviation [43]). *Let $X_1, \dots, X_m \in [0, 1]$ be fixed and let $\bar{X}$ be the average of a sample of size s drawn without replacement. For any $\varepsilon > 0$,*

$$\mathbb{P}[|\bar{X} - \mathbb{E}\bar{X}| \geq \varepsilon] \leq 2 \exp\left(-\frac{2\varepsilon^2 s m}{m-s+1}\right).$$

Lemma B.2 (Hypergeometric Hoeffding [22]; cf. [11]). *If $H \sim \text{Hypergeo}(m, K, s)$ counts the number of marked items in a sample of size s without replacement, then for all $\varepsilon > 0$,*

$$\mathbb{P}[|H - \mathbb{E}H| \geq \varepsilon s] \leq 2 \exp(-2\varepsilon^2 s).$$

Lemma B.3 (Hypergeometric Chernoff (Chvátal) [11]). *Let $H \sim \text{Hypergeo}(m, K, s)$ with marked fraction $p := K/m$ and let $0 < \varepsilon < 1 - p$. Then*

$$\mathbb{P}[H \geq (p + \varepsilon)s] \leq \exp(-s D(p + \varepsilon \| p)), \quad \mathbb{P}[H \leq (p - \varepsilon)s] \leq \exp(-s D(p - \varepsilon \| p)),$$

where $D(a \| p) := a \ln \frac{a}{p} + (1 - a) \ln \frac{1 - a}{1 - p}$ is the binary KL divergence.

Usage. We apply Theorems B.1 and B.2 to bound fluctuations when pulling back tests/features through constant-size gadget selections and when aggregating $t = \Theta(\log n)$ independent blocks; the resulting deviations are $O(s^{-1/2})$ and are absorbed by the global $O(\log N)$ meta budget from Section 2. *Notation guard:* here $p = K/m$ denotes the hypergeometric marked fraction and is unrelated to the restriction parameter $p = m^{-\alpha/d}$ used in Sections 2 to 4. For sharper tail exponents when the marked fraction $p = K/m$ is known, one may invoke Theorem B.3 (relative-entropy tails), which strengthens the bound from Theorem B.2 and can yield noticeably tighter deviations in the $O(\log N)$ -amortized regime.

B.2 Affine pullbacks preserve low-degree structure

Let $T(x) = Ax \oplus r$ be an affine map over $\mathbb{F}_2$ with output fan-in $\Delta = \max_i \|\text{row}_i(A)\|_0$. The core statements are:

- **Degree under bounded fan-in.** If p is multilinear of degree $\leq k$ in y , then $p \circ T$ is multilinear of degree $\leq \Delta k$ in x (Theorem B.5 in Appendix B.3).
- **Supremum norm and stability.** $\|p \circ T\|_\infty \leq \|p\|_\infty$ with equality for coordinate permutations; noise stability Stab_ρ is preserved exactly under permutations (Theorems B.6 and B.7 in Appendix B.3).
- **Fourier mass (mod- q).** For $q \in \{2, 3, 5\}$ the cumulative mass up to level k is preserved exactly under output mixing ($\Delta = 1$). For general Δ we rely on the collision-aware bound in Theorem B.7 (via the multiplicities M_k and the condition $|J(S)| \leq k$); in this paper we only use the induced *degree-budget relaxation* $k \mapsto \Delta k$, and we do not claim any stronger monotone relation.

The detailed proofs and the precise Fourier-mass statement appear in Appendix B.3 as Theorems B.5 and B.7. We rely only on exact preservation in the permutation case ($\Delta = 1$) and on the degree dilation $k \mapsto \Delta k$ for bounded fan-in.

In particular, for $q \in \{2, 3, 5\}$ we use that the cumulative mod- q mass up to level k is preserved exactly under permutations ($\Delta = 1$), while for bounded fan-in we track the relaxed budget $k \mapsto \Delta k$.

B.3 Affine pullbacks: degree, mass, and stability

We work with affine maps $T(x) = Ax \oplus r$ over $\mathbb{F}_2$, where $A \in \{0, 1\}^{m \times n}$ and $r \in \{0, 1\}^m$. Let

$$\Delta := \max_{i \in [m]} \|\text{row}_i(A)\|_0$$

denote the *output fan-in*, i.e., the maximum number of input bits any output bit y_i depends on. Unless stated otherwise, polynomials are in the standard multilinear form on $\{0, 1\}^m$ (equivalently on $\{\pm 1\}^m$ via the usual embedding). (Fourier background: [35].)

Remark B.4 (Scope and numbering). Canonical numbering is determined by the Appendix B inclusion order; this subsection provides the affine pullback lemmas referenced as items B.3–B.4.

Lemma B.5 (Affine pullback, degree under bounded fan-in). *Let p be multilinear in $y = (y_1, \dots, y_m)$ with $\deg p \leq k$. Then there exists a multilinear q in x such that*

$$p(T(x)) = q(x), \quad \deg q \leq \Delta k.$$

In particular, under output-mixing normalization—we fix A to be a coordinate permutation, hence $\Delta = 1$ —degree is preserved exactly: $\deg q = \deg p$.

Proof. Each y_i is the parity of at most Δ input bits. Any monomial of p involves $\leq k$ distinct y_i 's and thus expands to a sum of monomials over x of degree $\leq \Delta k$. Multilinearization preserves this bound. $\square$

Lemma B.6 (Supremum norm under pullback). *Let $p : \{0, 1\}^m \rightarrow \mathbb{R}$ and $q := p \circ T$. On $\{0, 1\}^n$, one has $\|q\|_\infty \leq \|p\|_\infty$. Equality holds when T is bijective (coordinate permutation).*

Proof. The range $T(\{0, 1\}^n) \subseteq \{0, 1\}^m$ implies $\sup_x |p(T(x))| \leq \sup_y |p(y)|$. If T is a bijection, the suprema coincide. $\square$

For a multilinear $f : \{\pm 1\}^m \rightarrow \mathbb{R}$, write $\text{mass}_{\leq k}(f) := \sum_{|S| \leq k} \hat{f}(S)^2$ for its level- $\leq k$ Fourier energy.

Lemma B.7 (Low-degree Fourier mass and noise stability). *Background on Boolean Fourier analysis and tail bounds: [35]. Let $p : \{0, 1\}^m \rightarrow \mathbb{R}$ be multilinear, fix $k \in \mathbb{N}$ and $\rho \in (0, 1)$. Let $T(x) = Ax \oplus r$ be affine over $\mathbb{F}_2$ with output fan-in $\Delta = \max_i \|\text{row}_i(A)\|_0$, and write $q := p \circ T$.*

1. Level expansion (Fourier mass).

(a) **Permutation case.** *If A is a permutation matrix (equivalently, T is a coordinate permutation followed by an $\mathbb{F}_2$ -shift; in particular $m = n$), then low-degree energy is preserved exactly:*

$$\text{mass}_{\leq k}(q) = \text{mass}_{\leq k}(p).$$

(b) **Disjoint row-supports (injective image).** *If the supports of distinct rows of A are pairwise disjoint and nonempty, then for any nonempty S the set $J(S)$ is the disjoint union of those row-supports, so $|J(S)| = \sum_{i \in S} \|\text{row}_i(A)\|_0 \geq |S|$; moreover $S \mapsto J(S)$ is injective. Consequently,*

$$\text{mass}_{\leq k}(q) = \sum_{|J(S)| \leq k} \hat{p}(S)^2 \leq \sum_{|S| \leq k} \hat{p}(S)^2 = \text{mass}_{\leq k}(p).$$

(c) **General bounded fan-in.** *For arbitrary A (allowing overlaps), define the collision multiplicities $m_U := |\{S \subseteq [m] : J(S) = U\}|$ for $U \subseteq [n]$ and $M_k := \max_{|U| \leq k} m_U$. Then*

$$\text{mass}_{\leq k}(q) \leq M_k \sum_{S: |J(S)| \leq k} \hat{p}(S)^2.$$

In general there is no monotone relation between the set $\{S : |J(S)| \leq k\}$ and a degree cut $\{S : |S| \leq ck\}$ without further structure; we do not use a stronger claim for $\Delta > 1$ in this paper.

2. Noise stability. If A is a permutation matrix (so T is bijective on the cube), then $\text{Stab}_\rho(q) = \text{Stab}_\rho(p)$ for all $\rho \in (0, 1)$. We make no general claim for $\Delta > 1$.

Under the output-mixing normalization used in the paper (interfaces are reduced to coordinate permutations), both the low-degree mass profile and Stab_ρ are preserved exactly.

Proof. We work on the $\{\pm 1\}$ -cube with characters $\chi_S(z) = \prod_{i \in S} z_i$ and uniform measure. Write $y'_i = (-1)^{y_i}$ and $x'_j = (-1)^{x_j}$. The Fourier expansion of p reads $p(y) = \sum_{S \subseteq [m]} \hat{p}(S) \chi_S(y')$. For $T(x) = Ax \oplus r$,

$$y'_i = (-1)^{r_i} \prod_{j: A_{ij}=1} x'_j.$$

For $S \subseteq [m]$, let $J(S) \subseteq [n]$ be the symmetric difference of the row-supports $\{j : A_{ij} = 1\}$ over $i \in S$; then $|J(S)| \leq \Delta|S|$ and

$$\chi_S(y') = (-1)^{\langle S, r \rangle} \chi_{J(S)}(x'). \quad (\text{B.1})$$

Equation (B.1) shows that each character $\chi_S(y')$ maps to the character $\chi_{J(S)}(x')$ up to a fixed sign, so grouping equal $\chi_{J(S)}$'s in (B.2) is the only source of collisions. Thus

$$q(x) = (p \circ T)(x) = \sum_{S \subseteq [m]} \hat{p}(S) (-1)^{\langle S, r \rangle} \chi_{J(S)}(x'). \quad (\text{B.2})$$

The mapping $S \mapsto J(S)$ here is a standard “concatenation of characters” calculation; see also related concatenation principles in additive/Fourier settings [45].

(1a) *Permutation case.* If A is a permutation matrix, then $J(S) = \pi(S)$ is a bijection with $|J(S)| = |S|$. Grouping equal characters in (B.2) yields, for each $U \subseteq [n]$,

$$\hat{q}(U) = \pm \hat{p}(\pi^{-1}(U)).$$

Therefore

$$\text{mass}_{\leq k}(q) = \sum_{|U| \leq k} \hat{q}(U)^2 = \sum_{|S| \leq k} \hat{p}(S)^2 = \text{mass}_{\leq k}(p).$$

(1b) *Disjoint row-supports.* Assume the supports of distinct rows of A are pairwise disjoint and nonempty. Then for any nonempty S , $J(S)$ is the disjoint union of the corresponding row-supports and hence $|J(S)| = \sum_{i \in S} \|\text{row}_i(A)\|_0 \geq |S|$. Moreover, the map $S \mapsto J(S)$ is injective (no cancellations). Consequently, from (B.2), each character $\chi_{J(S)}$ collects a single coefficient $\pm \hat{p}(S)$, and $|J(S)| \leq k$ implies $|S| \leq k$. Thus

$$\text{mass}_{\leq k}(q) = \sum_{|J(S)| \leq k} \hat{p}(S)^2 \leq \sum_{|S| \leq k} \hat{p}(S)^2 = \text{mass}_{\leq k}(p).$$

(1c) *General bounded fan-in (collision-aware bound).* Let $m_U = |\{S : J(S) = U\}|$ and $M_k = \max_{|U| \leq k} m_U$. From (B.2),

$$\hat{q}(U) = \sum_{S: J(S)=U} \pm \hat{p}(S),$$

so by Cauchy–Schwarz,

$$\hat{q}(U)^2 \leq m_U \sum_{S: J(S)=U} \hat{p}(S)^2 \leq M_k \sum_{S: J(S)=U} \hat{p}(S)^2.$$

Summing over $|U| \leq k$ gives

$$\text{mass}_{\leq k}(q) = \sum_{|U| \leq k} \widehat{q}(U)^2 \leq M_k \sum_{|U| \leq k} \sum_{S: J(S)=U} \widehat{p}(S)^2 = M_k \sum_{S: |J(S)| \leq k} \widehat{p}(S)^2.$$

Without additional structural constraints relating $|J(S)|$ to $|S|$, no universal bound in terms of $\sum_{|S| \leq ck} \widehat{p}(S)^2$ is available.

(2) *Noise stability (permutation case)*. Assume A is a permutation matrix so $T(x) = Px \oplus r$ is bijective on $\{0, 1\}^n$. Let X be uniform on $\{0, 1\}^n$ and $Y = X \oplus Z$ with i.i.d. flips, $\Pr[Z_j = 1] = (1 - \rho)/2$. In particular, $T(X)$ is uniform and independent of Z . Then

$$T(Y) = P(X \oplus Z) \oplus r = (PX \oplus r) \oplus PZ = T(X) \oplus PZ,$$

and since $PZ \stackrel{d}{=} Z$,

$$\text{Stab}_\rho(p \circ T) = \mathbb{E}[(p \circ T)(X) (p \circ T)(Y)] = \mathbb{E}[p(T(X)) p(T(X) \oplus Z)] = \text{Stab}_\rho(p).$$

This establishes exact preservation of both the low-degree mass profile and noise stability under the output-mixing normalization (coordinate permutations). $\square$

Lemma B.8 (Product restriction as level thinning). *Fix $s \in [0, 1]$. Let $\text{Res}(s)$ be the product restriction that, independently for each coordinate $j \in [n]$, keeps x_j alive with probability s and otherwise resamples it by an independent unbiased sign on $\{\pm 1\}$. For any multilinear $p : \{\pm 1\}^n \rightarrow \mathbb{R}$ with Fourier expansion $p(x) = \sum_{S \subseteq [n]} \widehat{p}(S) \chi_S(x)$ the following hold.*

(i) **First moment of coefficients.** *For every $S \subseteq [n]$,*

$$\mathbb{E}_{\rho \sim \text{Res}(s)}[\widehat{p \upharpoonright \rho}(S)] = s^{|S|} \widehat{p}(S).$$

(ii) **Second moment of coefficients.** *For every $S \subseteq [n]$,*

$$\mathbb{E}_\rho[\widehat{p \upharpoonright \rho}(S)^2] = s^{|S|} \sum_{T \subseteq [n] \setminus S} (1-s)^{|T|} \widehat{p}(S \cup T)^2 \leq s^{|S|} \sum_{U \supseteq S} \widehat{p}(U)^2.$$

Consequently, for any $k \in \mathbb{N}$,

$$\mathbb{E}_\rho \left[\sum_{|S| \leq k} \widehat{p \upharpoonright \rho}(S)^2 \right] = \sum_{U \subseteq [n]} \Pr[\text{Bin}(|U|, s) \leq k] \widehat{p}(U)^2 \leq \sum_U \widehat{p}(U)^2 = \|p\|_2^2.$$

In particular, for any parity χ_S ,

$$\mathbb{E}_\rho[\langle p \upharpoonright \rho, \chi_S \rangle] = s^{|S|} \langle p, \chi_S \rangle, \quad \mathbb{E}_\rho[|\langle p \upharpoonright \rho, \chi_S \rangle|] \leq s^{|S|/2} \|p\|_2.$$

Moreover, composing d independent restriction rounds with per-round survival $s_0 \in [0, 1]$ is equivalent to a single product restriction with survival $s = s_0^d$; all identities above hold with this s .

Proof. Write $p = \sum_{U \subseteq [n]} \widehat{p}(U) \chi_U$. A coefficient $\widehat{p}(U)$ contributes on a restriction ρ to the coefficient indexed by the alive subset $U \cap \text{Alive}(\rho)$, with magnitude $\widehat{p}(U)$ and an independent sign coming from the fixed dead coordinates. Thus $\widehat{p \upharpoonright \rho}(S) = \sum_{U \supseteq S} \mathbf{1}\{U \cap \text{Alive}(\rho) = S\} (\pm) \widehat{p}(U)$. Taking expectation gives (i), since only $U = S$ survives in mean. For (ii), cross-terms vanish in expectation (by independence and zero-mean of the inserted signs), yielding the stated sum with $\Pr[U \cap \text{Alive} = S] = s^{|S|} (1-s)^{|U|-|S|}$. The low-degree mass identity follows by grouping by U and observing that the degree after restriction equals $\text{Bin}(|U|, s)$. All inequalities are immediate. Composition across rounds uses independence of survival events. $\square$

Lemma B.9 (Random-index parity: correlation without noise domination). *Let $p : \{\pm 1\}^n \rightarrow \mathbb{R}$ be multilinear with $\deg p \leq d_0$, and let ρ be as in Lemma B.8 with survival s . Let $U(\rho)$ be an index set that may depend on ρ , and define $W(\rho) := |\langle p \upharpoonright \rho, \chi_{U(\rho)} \rangle|$. Then*

$$\mathbb{E}_\rho[W(\rho)] \leq \|p\|_2 \cdot \Pr_\rho[|U(\rho)| \leq d_0]^{1/2}.$$

In particular, if $|U(\rho)| \sim \text{Bin}(t_, s)$ (e.g., the alive-intersection of a parity of length t_*), then by Chernoff $\Pr[|U(\rho)| \leq d_0] \leq \exp(-\Omega((t_* - d_0)^2/t_*))$, hence $\mathbb{E}[W(\rho)] \leq \|p\|_2 \cdot \exp(-\Omega(t_*s))^{1/2}$.*

Proof. If $|U(\rho)| > d_0$, then $\langle p \upharpoonright \rho, \chi_{U(\rho)} \rangle = 0$ since $\deg(p \upharpoonright \rho) \leq d_0$. Thus

$$W(\rho)^2 \leq \mathbf{1}\{|U(\rho)| \leq d_0\} \cdot \sum_{|S| \leq d_0} \widehat{p \upharpoonright \rho}(S)^2.$$

Take expectations and apply Cauchy–Schwarz over ρ :

$$\mathbb{E}[W(\rho)] \leq \left(\mathbb{E} \left[\sum_{|S| \leq d_0} \widehat{p \upharpoonright \rho}(S)^2 \right] \right)^{1/2} \cdot \Pr[|U(\rho)| \leq d_0]^{1/2}.$$

Lemma B.8 gives the first factor $\leq \|p\|_2$, yielding the claim. $\square$

Remark B.10 (Permutation noise coupling). Writing a noise pair as $Y = X \oplus Z$ with i.i.d. noise Z , one may express $T(Y) = T(X) \oplus PZ$ for some coordinate permutation P . Since $PZ \stackrel{d}{=} Z$, equivalently $T(Y) = T(X) \oplus Z'$ with $Z' \stackrel{d}{=} Z$, which makes the stability equality in the permutation case immediate.

Remark B.11 (Index gadget degree factor). For the index gadget used in our lifts, the output fan-in satisfies $\Delta \leq 2$. Consequently, the degree dilation bound specializes to $\deg(p \circ T) \leq 2 \deg p$, matching the index-gadget degree factor stated in Section 2.

Remark B.12 (Delta tracking and cross-reference). Section 2 tracks the fan-in parameter Δ . Whenever output mixing is *not* assumed, statements are formulated with the relaxed degree budget $k \mapsto \Delta k$. For stability we only require the exact $\Delta = 1$ case. When $\Delta = 1$, all equalities above apply.

B.4 Gadget pullbacks: XOR and Index

We record two consequences for constant-size gadgets (see also [47]). Both rely on the affine pullback facts in Appendix B.3 (see Theorems B.5 and B.7).

Lemma B.13 (XOR gadget: deterministic TV-contractive push-forward with bounded fluctuation). *Let $G_\oplus$ be a constant-size XOR gadget (affine over $\mathbb{F}_2$). For any distribution μ and any invariant $T \in \{\text{EC}, \text{SPR}_k, \text{CMDL/PCG}\}$ used in Section 2, the push-forward $(G_\oplus)_\# \mu$ is deterministic and TV-contractive by Theorem A.1; low-degree structure is preserved exactly under output mixing ($\Delta = 1$) by Theorems B.5 and B.7. Any sampling fluctuation across gadget positions concentrates by Theorems B.1 and B.2.*

Lemma B.14 (Index gadget: constant-branch affine mixture). *Let G_{idx} be an Index gadget with constant address length (e.g., 6 bits), realized as a mixture of $B = O(1)$ affine branches $\{T_b\}_{b \leq B}$ (cf. [32]). For any cost-type invariant $T \in \{\text{EC}, \text{SPR}_k, \text{CMDL/PCG}\}$,*

$$T((G_{\text{idx}})_\# \mu) \geq \min_{b \leq B} T((T_b)_\# \mu) - O(\log N).$$

Each branch is handled by Theorems B.5 and B.7; for $\Delta > 1$ the collision-aware transfer bound Theorem B.7(1c) with factor M_k applies. The inter-branch selection contributes only a prefix-free meta header $O(\log N)$ (by Theorem A.2); branch sampling fluctuations are bounded via Theorems B.1 and B.2. The branch index is fixed globally and encoded once in the $O(\log N)$ header; branches do not vary position-wise.

Note. Writing $(G_{\text{idx}})_{\#}\mu = \sum_{b \leq B} \pi_b (T_b)_{\#}\mu$ with $\sum_b \pi_b = 1$, TV contracts branchwise and hence under convex mixtures, while the size-aware invariant T obeys $T((G_{\text{idx}})_{\#}\mu) \geq \min_b T((T_b)_{\#}\mu) - O(\log N)$ since the branch index is encoded once in the global header.

Consequence. Combining Theorems B.13 and B.14 with Theorem A.2 yields gadget stability $T(G_{\#}\mu) \geq T(\mu) - O(\log N)$ for $G \in \{\text{XOR}, \text{Index}\}$ and all $T \in \{\text{EC}, \text{SPR}_k, \text{CMDL/PCG}\}$, as used in Theorem 2.3.

C Switching path parameters (balanced window)

This appendix fixes the restriction/switching parameters used in Section 3 under the standardized overhead regime from Section 2 (deterministic push-forward is TV-contractive, 3XOR $\rightarrow$ 3SAT has size factor $\times 4$, meta overhead $O(\log N)$ bits via a single prefix-free header). We adopt the standard switching-lemma regime for p and d [19, 21], and parameterize p via an exponent $\alpha \in (0, 1)$ (default $\alpha = \frac{1}{3}$).

C.1 Parameter choices

- **Depth.** $d \in \mathbb{N}$ (default $d = 3$).
- **Restriction parameter.** $p = m^{-\alpha/d} \Rightarrow m \cdot p^d = m^{1-\alpha}$, with a fixed $\alpha \in (0, 1)$ (default $\alpha = \frac{1}{3}$).
- **Balanced window.** $m = (1 + \gamma)n$ with fixed $\gamma \in (0, 1)$; after 3XOR $\rightarrow$ 3SAT the clause count is $m' = 4m$.

Depth regimes. We fix constant depth $d = O(1)$ here since it already yields a polynomial bottom-width bound and suffices for the EC lower bound and the echo arguments in Sections 3 and 4. For completeness we also calibrate an alternative regime $d = \Theta(\log n)$ (used when a subpolynomial width bound is desired for the PC/SoS interface). The two regimes are compatible because the active degree budget remains $k = \Theta(\log n)$ and any constant-factor degree inflation from the Index gadget is absorbed downstream.

Decision depth vs. restriction depth. We distinguish the *restriction depth* d (number of restriction rounds) from the *decision-tree depth parameter* t used in the switching lemma. Throughout this paper we keep $d = O(1)$ (yielding polynomial bottom width), while we choose $t = \Theta(\log n)$ inside the switching lemma so that the failure probability $(CpW)^t$ becomes $o(1)$.

C.2 Derived scales and routine consequences

- **Surviving coordinates after d rounds.** With $p^d = m^{-\alpha}$ and $n \asymp m$ in the balanced window, the expected number of alive variables is $\mathbb{E}[\text{\#alive vars}] = np^d = \Theta(m^{1-\alpha})$. The number of unfixed 3XOR clauses is $\Theta(m^{1-\alpha})$ as well (each remains unfixed with probability $1 - (1 - p^d)^3 = \Theta(p^d)$). Concentration holds with constant probability (see Appendix B).

- **Typical bottom-width bound (DNF/CNF after switching):** $W \leq m^{O(\alpha/d)}$.
- **Advice/log budget (for size-aware $\text{AC}^0 + \log$):** $O(\log N)$ bits; insufficient to encode the $\Omega(m^\beta)$ live-parity information when $\text{EC} \geq m^\beta$.
- **Switching failure (with logarithmic decision depth).** By Håstad's switching lemma, for width $W = \text{poly}(1/p) = m^{O(\alpha/d)}$ and $p = m^{-\alpha/d}$ there is a constant $c_\star \in (0, 1)$ (from the standard parameter calibration of the lemma) such that $C p W \leq c_\star$ for all sufficiently large n . Choosing $t = \Theta(\log n)$ yields $\Pr[\text{depth} > t] \leq (C p W)^t \leq c_\star^{\Theta(\log n)} = n^{-\Omega(1)} = o(1)$. See also Lemma C.1 for $\Pr[\text{Alive}]$.

Lemma C.1 (Survival probability under d rounds). *Let S be a set of $t_\star$ variables and consider d independent restriction rounds where each variable survives a round with probability p . After d rounds, each variable is alive with probability p^d , independently across variables. Writing $X \sim \text{Bin}(t_\star, p^d)$ for the number of alive variables in S , the parity on S is live iff $X \geq 1$. Hence*

$$\Pr[\text{Alive}(S)] = 1 - (1 - p^d)^{t_\star} \geq 1 - \exp(-t_\star p^d).$$

In particular:

1. If $t_\star p^d \geq c_0$ then $\Pr[\text{Alive}(S)] \geq 1 - e^{-c_0}$ (constant-survival).
2. If $t_\star p^d \geq \lambda \log n$ then $\Pr[\text{Alive}(S)] \geq 1 - n^{-\lambda}$ (high-survival).

Proof. Independence across variables gives $X \sim \text{Bin}(t_\star, p^d)$; $\Pr[\text{Alive}] = \Pr[X \geq 1] = 1 - \Pr[X = 0] = 1 - (1 - p^d)^{t_\star} \geq 1 - \exp(-t_\star p^d)$. $\square$

Remark C.2 (Instantiation with window scales). With $p = m^{-\alpha/d}$ and thus $p^d = m^{-\alpha}$, and a surviving-parity length $t_\star = \Omega(m^\beta)$, we get

$$\Pr[\text{Alive}] \geq 1 - \exp(-\Omega(m^{\beta-\alpha})).$$

This yields constant-survival when $t_\star p^d = \Theta(1)$, which matches the regime used in the correlation/degree echoes. And the high-survival condition used in Theorem 4.3 corresponds exactly to $t_\star \geq c(\log n)/p^d$ (so $\Pr[\text{Alive}] \geq 1 - n^{-\Omega(1)}$).

C.3 Reuse across gadgets

For constant XOR/Index gadgets (affine over $\mathbb{F}_2$ after output mixing), the same (d, p) schedule can be used on the 3SAT side: the number of restricted coordinates and the width bounds are preserved up to constant factors. Sampling fluctuations across gadget positions concentrate by the without-replacement bounds in Appendix B. If a constant-branch selection is present, the branch index is chosen once globally and encoded a single time in the $O(\log N)$ prefix-free header from Section 2; branches do not vary across positions.

C.4 Notation shorthands

- $\rho \sim \text{Restrict}(d, p)$: a random restriction path.
- $\text{Alive}(\rho)$: the event that a nontrivial parity survives after ρ .
- $\text{Width}(\rho) \leq W$: the bottom-width bound holds after ρ .

These shorthands are used in Lemmas 3.2, 3.3, 3.6, 3.7 and Theorems 4.2, 4.3.

D Minimal reproducibility pack

All scripts and CSV/PNG outputs referenced here are contained in the Zenodo software package [29].

Note. The helper `scripts/sync_paper_assets.py` exists in the Git repository but is intentionally *not* included in the Zenodo package, since all figures/CSV tables can be reproduced and read directly from `appendices/assets/**` and `appendices/data/**`.

We provide compact scripts (each ≤ 200 LOC) and fixed seeds to reproduce the tables and figures shown. The focus is on clear trends under fixed budgets; we do not optimize constants here. *Notation guard:* any symbols or flags using “p” in this appendix (e.g., `label_p`, p_0 , p_1) refer to *source-generation* probabilities and are unrelated to the restriction parameter $p = m^{-\alpha/d}$ used in Sections 2–4.

D.1 Runtime and resources

- **Machine:** standard CPU.
- **Budgets:** $n \in \{256, 512, 1024\}$; depth $d \in \{3\}$; 200 seeds per setting.
- **Wall-clock:** < 2 minutes per n ; total make repro ≤ 20 minutes.

D.2 Outputs

- `artifacts/results/*.csv` — per-experiment tables.
- `artifacts/plots/*.png` — companion figures (simple line/scatter).
- `artifacts/seeds.txt` — list of integer seeds used.

D.3 Scripts (brief)

- `scripts/ec_counter.py` — instrument a toy DPLL/BDD; count bit-level *irreversible erasures* per instance; aggregate by n .
- `scripts/spr_trace.py` — generate canonical histories along the restriction path; evaluate finite-state predictors (k-gram/CTW proxy); report average log-loss.
- `scripts/pcg_estimate.py` — estimate $\text{MDL}(X)$ and $\text{CMDL}(X \mid C)$ via LZ/CTW proxies; report the PCG gap.
- `scripts/profile_multimod.py` — compute low-order mod- q Fourier mass (for $q \in \{2, 3, 5\}$) up to $k \leq 6$, plus Stab_ρ .

Each script accepts `-n`, `-seeds`, and writes CSV/PNG under `artifacts/`.

PCG semantics and safety settings

- LZ78 surrogate: final phrase receives no EOS cost (no extra +1 symbol).
- Pointer cost: variable pointer cost only with `--lz-varindex` (Elias-Gamma); otherwise uniform `ceil(log2(dict_size))`.
- Label-block CMDL excludes the label-stream cost.
- Small- L damping: `epsilon` (side mode) and `label_p` (label-block) slightly damped for very short sequences.

- Safety switches: `--warn-k-ratio16`, `--auto-clamp-k`, and `--assert-nonneg-pcg` to enforce non-negative row-level PCG within tolerance.
- Historical option: `lz78_pairs_bits` is disabled in these runs.

D.4 Reproduction command

```
# from repo root
python scripts/ec_counter.py --n 256 512 1024 --seeds 200
python scripts/spr_trace.py --n 256 512 1024 --seeds 200 --kmax 8
python scripts/pcg_estimate.py --n 256 512 1024 --seeds 200

# Multimod structure profile (label-block + side)
python scripts/run_multimod_experiments.py --tag iecz_final --verbose

# Sync figures/CSVs into appendices paths
python scripts/sync_paper_assets.py --verbose
```

D.4.1 RC1 — quick verification preset

```
# EC: two n-levels so the plot has >1 point
python scripts/ec_counter.py --n 64 96 --seeds 5 \
    --gamma 0.1 --max-backtracks 10000 \
    --outdir artifacts_smoke/rc1

# SPR: small k-range, simple mode
python scripts/spr_trace.py --n 64 --seeds 5 --kmax 6 \
    --simple-mode --outdir artifacts_smoke/rc1

# PCG (label-block, k-gram): short but stable; safety flags on
python scripts/pcg_estimate.py --n 64 --seeds 8 \
    --context-mode label-block --model kgram --k 0 2 4 \
    --length-mult 64 --auto-strong-signal \
    --warn-k-ratio 16 --auto-clamp-k --assert-nonneg-pcg 0.0 \
    --outdir artifacts_smoke/rc1

# Multimod (label-block): include q=2 and q=3; two rho values for a line, not a dot
python scripts/profile_multimod.py --n 96 --seeds 5 \
    --context-mode label-block --length-mult 32 \
    --q 2 3 --kmax-values 6 --rho 0.1 0.2 \
    --out-data-dir artifacts_smoke/rc1 --out-assets-dir artifacts_smoke/rc1
```

RC1 caveat (PCG negatives). On very small n and higher k (e.g., $n=64$, $k=4$), a few rows can become negative because contexts are thin even with clamping. This reflects short-sequence limitations; the standard runs avoid these cases.

Expected artifacts (presence-only).

- `artifacts_smoke/rc1/results/ec_counter.csv`, `artifacts_smoke/rc1/plots/ec_counter.png`
- `artifacts_smoke/rc1/results/spr_trace.csv`, `artifacts_smoke/rc1/plots/spr_trace.png`

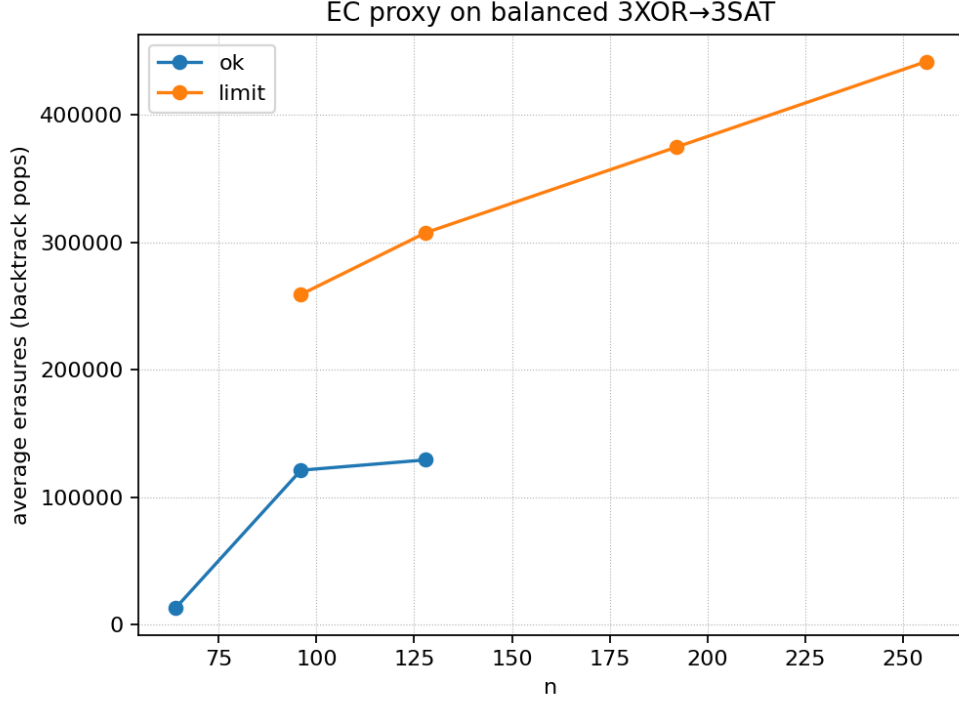


Figure 1: EC proxy (average backtrack pops) vs. n on balanced 3XOR→3SAT. Parameters as in §D.5.

- artifacts_smoke/rc1/results/pcg_estimate.csv, artifacts_smoke/rc1/plots/pcg_estimate.png (rare negatives may occur at very small n)
- artifacts_smoke/rc1/mass_by_qk.csv, artifacts_smoke/rc1/stability_by_rho.csv, artifacts_smoke/rc1/mass_cumulative_q2_label-block.png, artifacts_smoke/rc1/mass_cumulative_q3_label-block.png, artifacts_smoke/rc1/stab_vs_rho_label-block.png

D.5 EC proxy — run summary

```
python scripts/ec_counter.py --n 64 96 128 192 256 --seeds 100 \
  --gamma 0.1 --max-backtracks 20000 --count-decisions-as-erasure 1
```

Artifacts produced. CSV: artifacts/results/ec_counter.csv (500 rows).
Plot: artifacts/plots/ec_counter.png (Figure 1).

Status mix (all n combined). ok: **164**; limit: **336**; unsat: **0** (as expected for hidden-assignment generation).

Per- n notes (qualitative). For each $n \in \{64, 96, 128, 192, 256\}$, the average erasures increase with n for both ok and limit; see the two curves in Figure 1. The ok-fraction decreases with n , consistent with a growing EC burden in the balanced window $m = (1 + \gamma)n$ with $\gamma = 0.1$.

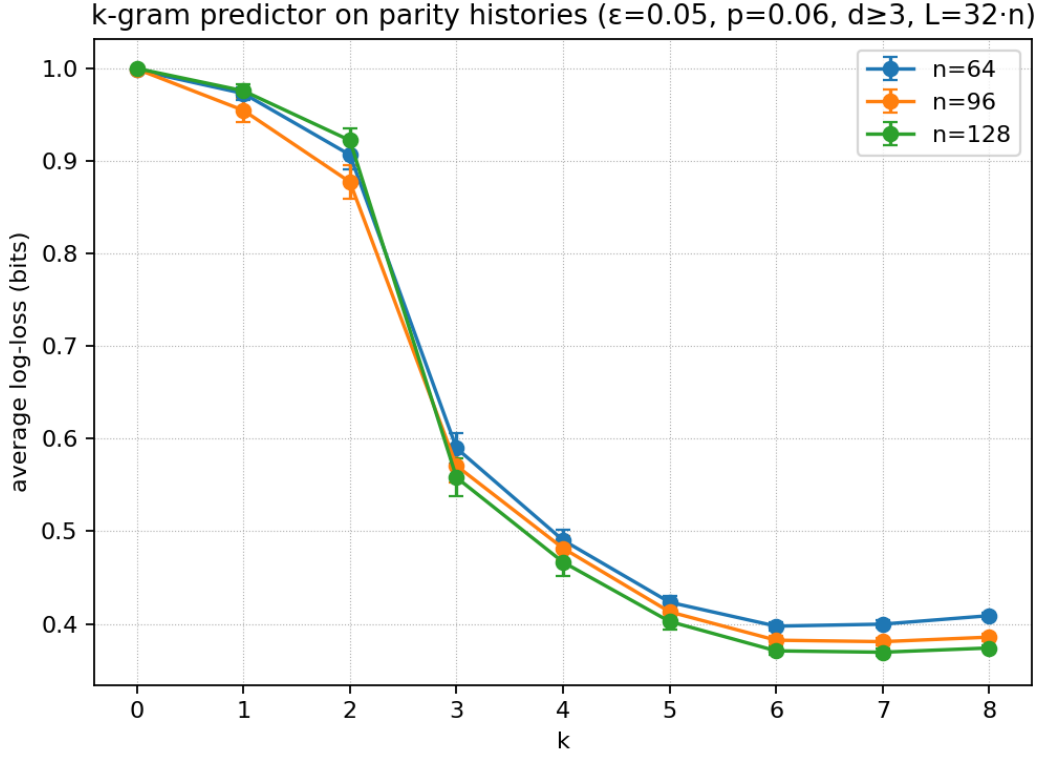


Figure 2: k -gram predictor on restriction-path parity histories: average log-loss (bits) vs. k for $n \in \{64, 96, 128\}$ ($\varepsilon = 0.05$, $p_{\text{hist}} = 0.06$, $d \geq 3$, $L = 32 \cdot n$). Error bars: s.e.m.; sharp drop and plateau match the echo narrative.

EC proxy figure. Figure 1 shows EC proxy (average backtrack pops) vs. n on balanced 3XOR $\rightarrow$ 3SAT; parameters as above. The solver is a small DPLL with unit propagation and chronological backtracking; erasures are counted as *trail pops* on backtracks (decision counted; -count-decisions-as-erasure=1). A backtrack cap of 20k yields limit points where search does not complete; these still track the EC trend.

D.6 SPR proxy — k -gram on parity histories

```
python scripts/spr_trace.py --n 64 96 128 --seeds 50 \
  --kmax 8 --simple-mode --verbose \
  --outdir artifacts \
  --json-out artifacts/results/spr_trace_summary.json
```

Artifacts. CSV: artifacts/results/spr_trace.csv; JSON summary: artifacts/results/spr_trace_summary.json;

Plot: artifacts/plots/spr_trace.png (Figure 2).

Parameters. $\varepsilon = 0.05$, $p_{\text{hist}} = 0.06$ (history-generator rate; not the restriction p from Sections 2–4), $d \geq 3$, length $L = 32 \cdot n$, max_lag=3 (simple-mode).

Key numbers (mean log-loss in bits; 50 seeds each).

- $n = 64$: $k=0 \rightarrow \mathbf{0.9980}$, $k=8 \rightarrow \mathbf{0.4089}$ ($\Delta \approx -0.589$)
- $n = 96$: $k=0 \rightarrow \mathbf{0.9990}$, $k=8 \rightarrow \mathbf{0.3857}$ ($\Delta \approx -0.613$)
- $n = 128$: $k=0 \rightarrow \mathbf{0.9997}$, $k=8 \rightarrow \mathbf{0.3740}$ ($\Delta \approx -0.626$)

Takeaway. Average log-loss drops sharply with k and then plateaus around 0.37-0.41 with small error bars, consistent with a residual parity signal along the (d, p) restriction path captured by finite-state predictors.

D.7 PCG proxy — k-gram (label-block)

```
python scripts/pcg_estimate.py --n 64 96 128 160 --seeds 40 \
  --context-mode label-block --model kgram --k 0 2 4 8 \
  --length-mult 64 --label-p 0.03 --p0 0.35 --p1 0.65 \
  --auto-strong-signal --plot-all-k \
  --outdir artifacts \
  --json-out artifacts/results/pcg_estimate_summary.json
```

Artifacts. Figures: `appendices/assets/pcg/kgram_topk_vs_n.png`, `appendices/assets/pcg/kgram_by_k_and_n.png`.

Data: `appendices/data/pcg/kgram_topk_vs_n.csv`, `appendices/data/pcg/kgram_by_k_and_n.csv`.

Telemetry (label): `appendices/data/pcg/label_conditional_entropy.csv` (i.e., $w_0 H_2(X \mid C=0) + w_1 H_2(X \mid C=1)$).

Takeaway. Across n , PCG increases with n ; within fixed n , PCG decreases smoothly with k and stabilizes at moderate k , reflecting that conditioning on the label simplifies the source and reduces CMDL. See Appendices D.14 and D.15 for structure metrics explaining the positive slope.

D.8 PCG proxy — LZ family (label-block)

```
python scripts/pcg_estimate.py --n 64 96 128 160 --seeds 40 \
  --context-mode label-block --model lz --use-lzma \
  --length-mult 64 --label-p 0.03 --p0 0.35 --p1 0.65 \
  --auto-strong-signal \
  --outdir artifacts \
  --json-out artifacts/results/pcg_estimate_summary.json
```

Artifacts. Figure: `appendices/assets/pcg/lz_codec_vs_n.png`;

Data: `appendices/data/pcg/lz_codec_vs_n.csv`;

Telemetry (side): `appendices/data/pcg/side_residual_entropy.csv`.

Takeaway. CMDL benefits from label conditioning; PCG grows with n , mirrored by lower conditional entropy in label-block mode. LZMA and LZ78 surrogates agree qualitatively; see Appendices D.14 and D.15.

D.9 PCG scaling with n (label-block, strong signal)

```
python scripts/pcg_estimate.py \  
  --n 256 384 512 768 1024 \  
  --seeds 60 \  
  --context-mode label-block --model kgram \  
  --k 0 2 4 8 \  
  --length-mult 64 --label-p 0.03 --p0 0.35 --p1 0.65 \  
  --auto-strong-signal --plot-all-k \  
  --warn-k-ratio 16 --auto-clamp-k --assert-nonneg-pcg 0.0 \  
  --outdir artifacts \  
  --json-out artifacts/results/pcg_scale_kgram_summary.json
```

Artifacts. Figure: `appendices/assets/pcg/kgram_scale_vs_n.png`;

Data: `appendices/data/pcg/kgram_scale_vs_n.csv`.

CSV schema (aggregated per n , top- k selection): `n, mean_pcg_topk, std_pcg_topk, sem_pcg_topk, count`.

Takeaway. Top- k PCG increases with n and stabilizes once predictor memory matches source regularity. Error bars remain tight across seeds. See Appendices D.14 and D.15.

D.10 PCG sensitivity in side-context (k-gram)

```
python scripts/pcg_estimate.py \  
  --n 96 160 256 \  
  --seeds 50 \  
  --context-mode side --model kgram \  
  --k 0 2 4 8 \  
  --epsilon 0.03 0.06 0.09 \  
  --p-run 0.05 0.08 0.12 \  
  --max-run 128 \  
  --auto-strong-signal --plot-all-k \  
  --warn-k-ratio 16 --auto-clamp-k --assert-nonneg-pcg 0.0 \  
  --outdir artifacts \  
  --json-out artifacts/results/pcg_side_sensitivity_summary.json
```

Artifacts. Data (summary): `appendices/data/pcg/side_residual_entropy.csv`;

Optional figure: `appendices/assets/pcg/side_kgram_sensitivity.png`.

Takeaway. Lower ϵ and smaller `p_run` increase PCG, matching a more predictable side stream and lower residual entropy $H_2(X \oplus S)$. See Appendices D.14 and D.15.

D.11 PCG vs. label separation (k-gram, grid)

```
# label_p in {0.03, 0.06}, (p0,p1) in {(0.45,0.55),(0.40,0.60),(0.30,0.70)}  
python scripts/pcg_estimate.py \  
  --n 96 160 256 \  
  --label-p 0.03 0.06 --p0 0.45 --p1 0.55 --p2 0.40 --p3 0.60 --p4 0.30 --p5 0.70
```

```

--seeds 60 \
--context-mode label-block --model kgram \
--k 0 2 4 8 \
--length-mult 64 \
--label-p 0.03 0.06 \
--p0 0.45 0.40 0.30 \
--p1 0.55 0.60 0.70 \
--auto-strong-signal --plot-all-k \
--warn-k-ratio 16 --auto-clamp-k --assert-nonneg-pcg 0.5 \
--outdir artifacts \
--json-out artifacts/results/pcg_label_sepgrid_summary.json

```

Artifacts. Telemetry: `appendices/data/pcg/label_conditional_entropy.csv`;
Optional figure: `appendices/assets/pcg/label_kgram_sepgrid.png`.

Takeaway. PCG grows with $|p_1 - p_0|$ and falls with higher `label_p`. After clamping thin contexts, negative rows are rare. See Appendices D.14 and D.15.

D.12 LZ family comparison (label-block)

```

# Uniform pointer surrogate
python scripts/pcg_estimate.py \
--n 96 160 256 384 512 \
--seeds 60 \
--context-mode label-block --model lz \
--auto-strong-signal \
--outdir artifacts \
--json-out artifacts/results/pcg_lz_surrogate_uniform_summary.json

```

```

# Variable pointer surrogate (Elias-Gamma, "VLC")
python scripts/pcg_estimate.py \
--n 96 160 256 384 512 \
--seeds 60 \
--context-mode label-block --model lz \
--lz-varindex \
--auto-strong-signal \
--outdir artifacts \
--json-out artifacts/results/pcg_lz_surrogate_vlc_summary.json

```

```

# LZMA (shared header baseline)
python scripts/pcg_estimate.py \
--n 96 160 256 384 512 \
--seeds 60 \
--context-mode label-block --model lz \
--use-lzma \
--auto-strong-signal \
--outdir artifacts \
--json-out artifacts/results/pcg_lzma_summary.json

```

Artifacts. Data (codec-wise aggregation): `appendices/data/pcg/lz_codec_vs_n.csv`;
Optional figure: `appendices/assets/pcg/lz_family_compare.png`.

Takeaway. The VLC surrogate agrees with k -gram on direction and magnitude; uniform and LZMA show a near-constant negative offset at small n due to pointer/container overhead. We use VLC for quantitative checks and treat LZMA as a qualitative baseline. See Appendices D.14 and D.15.

D.13 Fine k -profile (label-block)

```
python scripts/pcg_estimate.py \
  --n 96 160 256 \
  --seeds 60 \
  --context-mode label-block --model kgram \
  --k 0 1 2 3 4 6 \
  --length-mult 64 --label-p 0.03 --p0 0.35 --p1 0.65 \
  --auto-strong-signal --plot-all-k \
  --warn-k-ratio 16 --auto-clamp-k \
  --outdir artifacts \
  --json-out artifacts/results/pcg_finek_kgram_summary.json
```

Artifacts. Optional figure: `appendices/assets/pcg/kgram_fine_k.png`;
Data: `appendices/data/pcg/kgram_by_k_and_n.csv` (extended with $k \in \{1, 2, 3, 6\}$).

Takeaway. The gap rises from $k=0$ and plateaus by $k \approx 4-6$, indicating that modest finite-state memory captures most label structure. See Appendices D.14 and D.15.

D.14 Multi-mode structure profiles (label-block & side)

What we measure. The profile (PROF) summarizes low-order structure by (i) cumulative mod- q Fourier mass up to degree $k \leq 6$ for $q \in \{2, 3, 5\}$ and (ii) noise stability Stab_ρ at $\rho \in \{0.1, 0.2\}$. We reuse the exact bitstreams generated for the PCG experiments for comparability.

```
python scripts/profile_multimod.py \
  --n 256 384 512 768 1024 \
  --seeds 60 --seed-base 41113 \
  --context-mode label-block \
  --length-mult 64 --label-p 0.03 --p0 0.35 --p1 0.65 \
  --q 2 3 5 --kmax-values 6 --rho 0.1 0.2 \
  --pcg-csv artifacts_runs/strong_iecz_final/label_kgram_scale/results/pcg_estimate.csv \
  --context-for-corr label-block \
  --out-data-dir artifacts_runs/strong_iecz_final/label_multimod_scale/results \
  --out-assets-dir artifacts_runs/strong_iecz_final/label_multimod_scale/plots
```

```
python scripts/profile_multimod.py \
--n 96 160 256 \
--seeds 50 --seed-base 41113 \
--context-mode side --length-mult 32 \
--max-run 128 --epsilon 0.03 0.06 0.09 --p-run 0.05 0.08 0.12 \
--q 2 3 5 --kmax-values 6 --rho 0.1 0.2 \
--out-data-dir artifacts_runs/strong_iecz_final/side_multimod_sensitivity/results \
--out-assets-dir artifacts_runs/strong_iecz_final/side_multimod_sensitivity/plots
```

Artifacts. *Figures (label-block):* appendices/assets/multimod/label_mass_q2.png, appendices/assets/multimod/label_mass_q3.png, appendices/assets/multimod/label_mass_q5.png, appendices/assets/multimod/stab_label.png.

Figures (side): appendices/assets/multimod/side_mass_q2.png, appendices/assets/multimod/side_mass_q3.png, appendices/assets/multimod/side_mass_q5.png, appendices/assets/multimod/stab_side.png.

Data (label-block): appendices/data/multimod/label_mass_by_qk.csv, appendices/data/multimod/label_stability_by_rho.csv.

Data (side): appendices/data/multimod/side_mass_by_qk.csv, appendices/data/multimod/side_stability_by_rho.csv.

CSV schemas. *_mass_by_qk.csv: context, n, seed, q, k, degree_cap, metric, value (metric=cumulative_mass).

*_stability_by_rho.csv: context, n, seed, rho, metric, value (metric=stability).

Takeaway. For label-block runs the cumulative mass at small k saturates quickly and Stab_ρ remains high; side-context shows the expected dependence on ϵ and run-length parameters. Profiles are stable across seeds.

D.15 Correlation with PCG (label-block)

Inputs. PCG table: appendices/data/pcg/kgram_scale_vs_n.csv (aggregated) and its per-seed inputs.

Multimod tables: appendices/data/multimod/label_mass_by_qk.csv, appendices/data/multimod/label_stability_by_rho.csv.

Outputs. Correlation summary: appendices/data/multimod/corr_with_pcg.csv (includes Pearson r for PCG vs. Stab_ρ at $\rho \in \{0.1, 0.2\}$). Scatter plots: appendices/assets/multimod/pcg_vs_mass_q2_label.png (top- k mass), appendices/assets/multimod/pcg_vs_mass_q2_k0_label.png ($k=0$ baseline).

Takeaway. PCG correlates positively with low-order mass (especially mod-2 at small k) and with noise stability. This supports the *echo* viewpoint: surviving simple structure is exactly what contextual models exploit.

D.16 Extended experiments (D.9–D.15)

```
python scripts/pcg_estimate.py \  
  --n 256 384 512 768 1024 \  
  --seeds 100 \  
  --context-mode label-block --model kgram \  
  --k 0 2 4 8 \  
  --length-mult 64 --label-p 0.03 --p0 0.35 --p1 0.65 \  
  --auto-strong-signal --plot-all-k \  
  --warn-k-ratio 16 --auto-clamp-k --assert-nonneg-pcg 0.0 \  
  --outdir artifacts_runs/final/label_kgram_scale  
  
# Build the D.9 aggregate (CSV + PNG) from the per-seed CSV:  
python scripts/build_kgram_scale_assets.py \  
  --infile artifacts_runs/final/label_kgram_scale/results/pcg_estimate.csv \  
  --out-csv appendices/data/pcg/kgram_scale_vs_n.csv \  
  --out-png appendices/assets/pcg/kgram_scale_vs_n.png \  
  --errorbars sem
```

Validation. No negative rows after clamping; monotone growth of top- k PCG vs. n .

```
# Uniform LZ78 surrogate (header-safe, clamped)  
python scripts/pcg_estimate.py \  
  --n 96 160 256 384 512 --seeds 100 \  
  --context-mode label-block --model lz \  
  --length-mult 128 --auto-strong-signal \  
  --warn-k-ratio 16 --auto-clamp-k --clamp-nonneg-pcg \  
  --outdir artifacts_runs/final/label_lz_surrogate_uniform_v2final  
  
# Variable-length pointer LZ78 (VLC surrogate; header-safe, clamped)  
python scripts/pcg_estimate.py \  
  --n 96 160 256 384 512 --seeds 100 \  
  --context-mode label-block --model lz --lz-varindex \  
  --length-mult 128 --auto-strong-signal \  
  --warn-k-ratio 16 --auto-clamp-k --clamp-nonneg-pcg \  
  --outdir artifacts_runs/final/label_lz_surrogate_vlc_v2final  
  
# LZMA baseline (shared header; clamped)  
python scripts/pcg_estimate.py \  
  --n 96 160 256 384 512 --seeds 100 \  
  --context-mode label-block --model lz --use-lzma \  
  --length-mult 128 --auto-strong-signal \  
  --lzma-shared-header --clamp-nonneg-pcg \  
  --outdir artifacts_runs/final/label_lzma_baseline_v2final
```

```
# Sync canonical figures/CSVs into appendices paths  
python scripts/sync_paper_assets.py --verbose
```

Nonnegativity clamp. For codecs with fixed container headers on short streams (e.g., LZ/LZMA), CMDL may exceed MDL. Clamp per-seed PCG to $\max(\text{PCG}, 0)$ before aggregation and plotting; trends remain, spurious negatives vanish.

Empirical note. With 100 seeds and `length-mult 128`, the VLC surrogate yields increasing PCG means versus n (96→144.3, 160→260.5, 256→426.1, 384→615.1, 512→813.9). Uniform LZ78 and

LZMA baselines remain near zero after clamping due to fixed headers; directional alignment with k-gram persists.

```
python scripts/pcg_estimate.py \  
  --n 96 160 256 \  
  --seeds 100 \  
  --context-mode side --model kgram \  
  --k 0 2 4 8 \  
  --epsilon 0.03 0.06 0.09 \  
  --p-run 0.05 0.08 0.12 \  
  --max-run 128 \  
  --auto-strong-signal --plot-all-k \  
  --warn-k-ratio 16 --auto-clamp-k --assert-nonneg-pcg 0.0 --clamp-nonneg-pcg \  
  --outdir artifacts_runs/final/side_kgram_sensitivity
```

Validation. Full grids render; residual-entropy telemetry present; qualitative sensitivity as expected.
Side top-k (mean over seeds): $n = 96 \rightarrow 900.10$, $160 \rightarrow 1582.45$, $256 \rightarrow 2504.36$.

```
# Batch wrapper runs both label-block (scale) and side-context (sensitivity) with consistent seeds  
python scripts/run_multimod_experiments.py --tag final --verbose
```

```
# Sync canonical figures/CSVs into appendices/{assets,data}  
python scripts/sync_paper_assets.py --verbose
```

Outputs. appendices/data/multimod/mass_by_qk.csv, appendices/data/multimod/stability_by_rho.csv (label+side), appendices/data/multimod/corr_with_pcg.csv; updated scatter plots in appendices/assets/multimod/.

```
python scripts/profile_multimod.py \  
  --n 512 --seeds 80 \  
  --context-mode label-block --length-mult 64 \  
  --q 2 3 5 --kmax-values 8 --rho 0.1 0.2 0.3 \  
  --out-data-dir artifacts_runs/final/label_multimod_deep \  
  --out-assets-dir artifacts_runs/final/label_multimod_deep
```

Reporting note. If space is tight, note the $\rho=0.3$ plateau in text; the figure can be omitted.

```
# Lint & consistency  
python scripts/lint_repo.py --base . --fix-eof-newline  
python scripts/lint_repo.py --base . --fail-on-warn 1
```

```
# Build manifest  
python scripts/make_artifact_manifest.py \  
  --roots appendices/assets appendices/data \  
  --out appendices/MANIFEST_FINAL.json \  
  --label v2-final
```

```
# Presence check (existence only)
python - << "PY"
from pathlib import Path; need=[
    "appendices/assets/pcg/kgram_scale_vs_n.png",
    "appendices/data/pcg/kgram_scale_vs_n.csv",
    "appendices/assets/multimod/label_mass_q2.png",
    "appendices/assets/multimod/stab_label.png",
    "appendices/assets/multimod/side_mass_q2.png",
    "appendices/assets/multimod/stab_side.png",
    "appendices/data/multimod/label_mass_by_qk.csv",
    "appendices/data/multimod/side_stability_by_rho.csv",
    "appendices/data/multimod/corr_with_pcg.csv",
]
print("ALL ASSETS PRESENT:", all(Path(p).exists() for p in need))
PY
```

E Reproducibility and packaging (final + RC2)

Final artifacts and minimal smoke check

All figures and CSV files under `appendices/assets/*` and `appendices/data/*` are produced by the scripts and presets in Appendix D. A SHA-256 manifest (`appendices/MANIFEST_FINAL.json`) is written after the last successful run. The pipeline spanning Appendix D.1–Appendix D.4 is deterministic once indices and seeds are fixed. The only meta description is a single prefix-free header of length $O(\log N)$.

The software-only companion package is archived at Zenodo, DOI [10.5281/zenodo.17290623](https://doi.org/10.5281/zenodo.17290623) [29].

Presence-only smoke check. The following snippet tests that the expected key files are present. It does not recompute any result.

```
python - << "PY"
from pathlib import Path

REQUIRED = [
    "appendices/assets/pcg/kgram_scale_vs_n.png",
    "appendices/data/pcg/kgram_scale_vs_n.csv",
    "appendices/assets/multimod/label_mass_q2.png",
    "appendices/assets/multimod/stab_label.png",
    "appendices/assets/multimod/side_mass_q2.png",
    "appendices/assets/multimod/stab_side.png",
    "appendices/data/multimod/label_mass_by_qk.csv",
    "appendices/data/multimod/side_stability_by_rho.csv",
    "appendices/data/multimod/corr_with_pcg.csv",
    "appendices/MANIFEST_FINAL.json",
]

status = {p: Path(p).exists() for p in REQUIRED}
print("ALL ASSETS PRESENT:", all(status.values()))
for p, ok in status.items():
```

```
print(f"[{'OK' if ok else 'MISS'}] {p}")
PY
```

Environment. Results were produced with Python 3.12, NumPy and Pandas at default precision on a standard workstation. Operating system does not matter for the figures and CSVs as long as the same seeds and presets are used.

Packaging (final release)

The repository ships three final artifacts for offline verification.

- **Git bundle:** release/iecz-03.v2-final.bundle
- **Tracked sources zip:** release/v2-final-source.zip
- **Hashes:** release/REPRODUCIBILITY.md (SHA-256 lines)

```
git bundle verify release/iecz-03.v2-final.bundle

git clone --bundle release/iecz-03.v2-final.bundle iecz-03.final
cd iecz-03.final
git log -1 --pretty=oneline

cd ..
python - << "PY"
import hashlib, sys
from pathlib import Path

def sha256(p):
    h = hashlib.sha256()
    h.update(Path(p).read_bytes())
    return h.hexdigest()

targets = [
    "release/iecz-03.v2-final.bundle",
    "release/v2-final-source.zip",
]
for t in targets:
    print(sha256(t), " *", t)
PY

# Unpack the tracked sources
mkdir -p extract && cd extract
unzip ../release/v2-final-source.zip

python - << "PY"
import json, hashlib
from pathlib import Path
```

```

m = json.loads(Path("appendices/MANIFEST_FINAL.json").read_text(encoding="utf-8"))
def sha256(path):
    h = hashlib.sha256()
    h.update(Path(path).read_bytes())
    return h.hexdigest()

bad = []
for rel, expected in m.items():
    p = Path(rel)
    if not p.exists():
        bad.append((rel, "missing"))
    else:
        got = sha256(p)
        if got != expected:
            bad.append((rel, f"sha256 mismatch: {got[:12]} != {expected[:12]}"))

if bad:
    print("MANIFEST CHECK: FAIL")
    for r in bad:
        print(" -", r[0], "=>", r[1])
else:
    print("MANIFEST CHECK: OK")
PY

```

RC2 reviewer pack

For long term archiving and offline review there is a compact RC2 pack.

- **Source archive (RC2):** release/v1-rc2-source.zip includes code, sections, appendices, scripts, and appendices/MANIFEST_RC2.json.
- **Git bundle (RC2):** release/iecz-03.v1-rc2.bundle
- **Hashes and steps:** release/REPRODUCIBILITY.md lists SHA-256 and simple verify commands.
- **Smoke preset:** Appendix D.4.1 provides a sub-minute run.
- **Full v1 runs:** Appendix D (Appendix D.9–Appendix D.15). Final figures under appendices/assets/{pcg,multimod}, companion CSVs under appendices/data/{pcg,multimod}.

```

git bundle verify release/iecz-03.v1-rc2.bundle
git clone --bundle release/iecz-03.v1-rc2.bundle iecz-03.rc2
cd iecz-03.rc2 && git log -1 --pretty=oneline

```

Notes on determinism

All random choices are tied to fixed seeds that are printed into the CSV headers. Version numbers of third party libraries are recorded in the run logs. The state machine that holds the single header is updated reversibly, and its cost is amortized over repeated steps.

References

- [1] D. Achlioptas and M. Molloy. The solution space geometry of random linear equations. *Random Structures & Algorithms*, 46(2):197–231, 2015. doi: 10.1002/rsa.20494.
- [2] M. Alekhnovich and A. A. Razborov. Lower bounds for polynomial calculus: Non-binomial case. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 190–199. IEEE, 2001. doi: 10.1109/SFCS.2001.959893.
- [3] M. Alekhnovich and A. A. Razborov. Lower bounds for polynomial calculus: Non-binomial case. *Proceedings of the Steklov Institute of Mathematics*, 242:18–35, 2003.
- [4] A. Barg. Random codes: Minimum distances and error exponents. *IEEE Transactions on Information Theory*, 48(9):2568–2573, 2002. doi: 10.1109/TIT.2002.800480.
- [5] A. Barg and G. Zémor. Error exponents of expander codes. *IEEE Transactions on Information Theory*, 48(6):1725–1729, 2002. doi: 10.1109/TIT.2002.1003853.
- [6] P. Beame. A switching lemma primer. Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington, 1994.
- [7] W. Beckner. Inequalities in fourier analysis. *Annals of Mathematics*, 102(1):159–182, 1975. doi: 10.2307/1970980.
- [8] C. H. Bennett. Time/space trade-offs for reversible computation. *SIAM Journal on Computing*, 18(4):766–776, 1989. doi: 10.1137/0218053.
- [9] A. Bonami. Étude des coefficients de fourier des fonctions de l^p . *Annales de l’Institut Fourier*, 20(2):335–402, 1970. doi: 10.5802/aif.357.
- [10] J. Carette, C. Heunen, R. Kaarsgaard, and A. Sabry. Compositional reversible computation. In *Reversible Computation (RC 2024)*, volume 14680 of *Lecture Notes in Computer Science*, pages 10–27. Springer, 2024. doi: 10.1007/978-3-031-62076-8_2.
- [11] V. Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25(3):285–287, 1979. doi: 10.1016/0012-365X(79)90084-0.
- [12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006. doi: 10.1002/047174882X.
- [13] O. Dubois and J. Mandler. The 3-xorsat threshold. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 769–778. IEEE, 2002. doi: 10.1109/SFCS.2002.1182002.

- [14] P. Elias. Universal codeword sets and representations of the integers. *IEEE Transactions on Information Theory*, 21(2):194–203, 1975. doi: 10.1109/TIT.1975.1055349.
- [15] R. G. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1):21–28, 1962. doi: 10.1109/TIT.1962.1057683.
- [16] S. Grewal and V. M. Kumar. Improved circuit lower bounds and quantum-classical separations, 2024.
- [17] D. Grigoriev. Linear lower bounds for the degree of positivstellensatz refutations. *Theoretical Computer Science*, 259(1–2):613–622, 2001. doi: 10.1016/S0304-3975(00)00157-2.
- [18] J. A. Grochow and D. H. Wolpert. Beyond number of bit erasures: New complexity questions raised by recently discovered thermodynamic costs of computation. *ACM SIGACT News*, 49(2):33–56, 2018. doi: 10.1145/3232679.3232689.
- [19] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th ACM Symposium on Theory of Computing (STOC)*, pages 6–20. ACM, 1986. doi: 10.1145/12130.12132.
- [20] J. Hastad. *Computational Limitations of Small-Depth Circuits*. MIT Press, 1987. Monograph based on STOC’86/Ph.D. thesis.
- [21] J. Håstad. The switching lemma and lower bounds for small depth circuits. In S. Y. Jang, Y. R. Kim, D.-W. Lee, and I. Yie, editors, *Proceedings of the International Congress of Mathematicians (ICM 2014)*, volume IV, pages 577–599. Kyung Moon SA, Seoul, 2014.
- [22] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.
- [23] R. Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961. doi: 10.1147/rd.53.0183.
- [24] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001. doi: 10.1137/S1052623400366802.
- [25] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In M. Putinar and S. Sullivant, editors, *Emerging Applications of Algebraic Geometry*, volume 149 of *IMA Volumes in Mathematics and its Applications*, pages 155–270. Springer, 2009. doi: 10.1007/978-0-387-09686-5_7.
- [26] M. Lela. An mdl-style cost functional kc , distribution-preserving reductions ($a2^d$), and an $ac^0 + \log$ lower bound for 3sat via balanced 3xor, 2025. Preprint, v1, September 2025. Comments: 11 pages, 3 tables. Appendices A and B (switching parameters and correlation exponent; 2-core facts). Ancillary dataset on Zenodo (DOI <https://doi.org/10.5281/zenodo.17138272>). ORCID 0009-0008-0768-5184. MSC classes: 68Q15, 68Q17. ACM classes: F.1.3; F.2.2.
- [27] M. Lela. Iecz: Balanced 3xor $\rightarrow$ 3sat (artifacts & sources), 2025. Zenodo dataset, v1.0.1 (released 2025-09-17).
- [28] M. Lela. Iecz–ii experiments: Auxiliary-free 3xor $\rightarrow$ 3sat sanity checks, 2025. Zenodo dataset, v1.0.0 (released 2025-09-28).

- [29] M. Lela. Iecz-iii software package (code & artifacts) — zenodo archive, 2025. Software-only companion to IECZ-III; versioned Zenodo deposit.
- [30] A. Lempel and J. Ziv. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1978. doi: 10.1109/TIT.1978.1055934.
- [31] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 4 edition, 2019. ISBN 978-3-030-11298-1. doi: 10.1007/978-3-030-11298-1.
- [32] X. Li and Y. Zhong. Explicit directional affine extractors and improved hardness for linear branching programs. In *Computational Complexity Conference (CCC 2024)*, volume 300 of *LIPICs*, pages 10:1–10:14, 2024. doi: 10.4230/LIPICs.CCC.2024.10.
- [33] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. doi: 10.1017/CBO9780511813603.
- [34] E. Mossel, R. O’Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010. doi: 10.4007/annals.2010.171.295.
- [35] R. O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi: 10.1017/CBO9781139814782.
- [36] C. H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and System Sciences*, 48(3):498–532, 1994. doi: 10.1016/S0022-0000(05)80063-7.
- [37] B. Pittel and G. B. Sorkin. The satisfiability threshold for k -xorsat. *Combinatorics, Probability and Computing*, 25(2):236–268, 2016. doi: 10.1017/S0963548315000097.
- [38] Y. Polyanskiy and Y. Wu. *Lecture Notes on Information Theory*. Self-published, 2020. v5 (or later). Available online.
- [39] A. A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4): 333–338, 1987. doi: 10.1007/BF01137685.
- [40] A. A. Razborov. Lower bounds for the polynomial calculus. *Computational Complexity*, 7(4): 291–324, 1998. doi: 10.1007/PL00001591.
- [41] I. Sason. On f -divergences: Integral representations, local behavior, and inequalities. *Entropy*, 20(5):383, May 2018. doi: 10.3390/e20050383. URL <https://www.mdpi.com/1099-4300/20/5/383>.
- [42] I. Sason and S. Verdú. f -divergence inequalities. *Foundations and Trends in Communications and Information Theory*, 12(2–3):119–199, 2016. doi: 10.1561/01000000064.
- [43] R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *Annals of Statistics*, 2(1):39–48, 1974. doi: 10.1214/aos/1176342611.
- [44] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the 19th ACM Symposium on Theory of Computing (STOC)*, pages 77–82. ACM, 1987. doi: 10.1145/28395.28404.

- [45] T. Tao and T. Ziegler. Concatenation theorems for anti-gowers-uniform functions and host-kra characteristic factors. *Discrete Analysis*, page 7, 2016. doi: 10.19086/da.850.
- [46] G. S. Tseitin. On the complexity of derivation in propositional calculus. In A. O. Slisenko, editor, *Studies in Constructive Mathematics and Mathematical Logic, Part II*, pages 115–125. Consultants Bureau, 1970. English translation of Zapiski Nauchn. Semin. LOMI, vol. 8 (1968).
- [47] E. Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 39(7):3122–3154, 2009. doi: 10.1137/070699971.
- [48] A. C.-C. Yao. Probabilistic computations: Toward a unified measure. In *Proceedings of the 18th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 222–227. IEEE, 1977. doi: 10.1109/SFCS.1977.24.
- [49] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977. doi: 10.1109/TIT.1977.1055714.