

IECZ-II: Minimax, Gadget Composition, and SoS Dichotomy*

Window-stable Reductions $3\text{XOR} \rightarrow 3\text{SAT}$ and Size-aware MDL

Marko Lela

0009-0008-0768-5184

2025-09-29

Abstract

We give a size-aware minimum description length (MDL) cost and a window-stable $3\text{XOR} \rightarrow 3\text{SAT}$ chain with controlled losses in total variation (TV) and instance size. Core components are minimax monotonicity (Theorem 4.4), gadget composition (XOR and Index; Theorem 6.9), and a sum-of-squares (SoS) entropy dichotomy (Theorem 7.9) at $k = \Theta(\log n)$. We present a composition pipeline (E1–E4) for distributional transfers with explicit accounting for TV, size, and meta overheads. For symbol conventions and standing assumptions, see Sec. 2. *All guarantees are distributional (on a calibrated window); we make no worst-case claims.*

Relation to IECZ-I. This sequel builds on our companion work [10], which introduced the size-aware MDL functional and the distribution-preserving reduction principle (A2d) together with a $\delta=0$ injective $3\text{XOR} \rightarrow 3\text{SAT}$ translation; here we develop the minimax principle, gadget composition, and the SoS dichotomy.

1 Introduction

We motivate the IECZ line: stability under deterministic reductions that are 1-Lipschitz in total variation, prefix-free preprocessing, and a window discipline that permits measure-faithful transfers. As formalized in the first part (IECZ-I), we adopt the same window convention and prefix-free meta-encoding throughout [11]. We use a windowed composition pipeline (E1–E4) for *distributional* transfers, formulated as standalone lemmas with explicit TV/size/meta accounting (see Lemmas C.13 to C.16 and Figure 1). For empirical sanity checks of the injective $3\text{XOR} \rightarrow 3\text{SAT}$ mapping, see Appendix A. All statements are windowed: deterministic maps are 1-Lipschitz in total variation on calibrated windows, and prefix-free meta-encodings add only $O(\log N)$ overhead. We do not claim worst-case statements here.

Scope and limitations. All statements in this paper are *window-calibrated* and hence distributional: our total-variation (TV) guarantees hold on the image window Win and may degrade off-window; we do not claim worst-case hardness. Results are stated in a parameterized regime (e.g., $d = O(\log n)$ and $k = \Theta(\log n)$). Varying $c \in (0, 1)$ trades bottom width W against depth d along the switching path; the constants propagate through E1–E4 without changing asymptotics. For the low-degree and SoS analyses we use absolute constants C, τ ; these, as well as the degree budgets, are tied to the window calibration and to the gadget-induced product proxies, and we state them explicitly where they are used. Superlogarithmic regimes lie outside our present scope and are left to follow-up work.

* Artifacts: Zenodo v1.0.0 (10.5281/zenodo.17220388); IECZ-I bundle v1.0.1 (10.5281/zenodo.17141362).

On the $d, k = \Theta(\log n)$ regime. Our log-degree regime is deliberate: it is the natural sweet spot where (a) the switching-lemma width control remains uniform along our calibrated path, (b) the windowed hypercontractive constants can be taken absolute, and (c) the SoS entropy dichotomy yields quantitative degree lower bounds with explicit constants. Outside this regime (e.g., superlogarithmic depths), the bottom width W may grow with d via the dependence on $p = m^{-c/d}$, and our absolute constants (C, τ) are no longer claimed. We therefore do *not* assert worst-case or non-parameterized statements in this paper.

Roadmap and companion sequel. A companion sequel (IECZ-III, in preparation) develops a *hardcore-condensation-lift* pipeline that

- (i) performs blockwise aggregation within Win ,
- (ii) selects a TV-stable hardcore sub-window,
- (iii) applies a seeded condensation/re-encoding with explicit TV/Size/Meta bookkeeping, and
- (iv) derandomizes to an explicitly parameterized distributional family $\{\Phi_n\}$.

The aim is to carry the IECZ-II minimax-SoS dichotomy through to size-aware hardness for explicit families, while remaining in a distributional, non-worst-case setting; see [9].

Parameter preview (E1–E4). We commit to concrete choices used later in Theorem C.5:

$$\begin{aligned} t &= \left\lceil c_1 \varepsilon^{-2} \log N \right\rceil, \\ \theta &\geq c_2 \quad (\text{hardcore mass}), \\ s &= O(\log N) \quad (\text{condenser seed length}). \end{aligned}$$

These yield product amplification to advantage N^{-c_1} (E1), a constant-mass hardcore with advantage N^{-c} (E2) via [7], preservation under seeded condensation with polylogarithmic size overhead (E3) using extractor-style condensers [16], and an *explicitly parameterized distributional* family after fixing indices with $O(\log N)$ meta-overhead (E4).

We first formalize a size-aware MDL cost and prove *minimax monotonicity* (Theorem 4.4). We then establish a window-stable injective reduction $3\text{XOR} \rightarrow 3\text{SAT}$ with tabulated losses, enabling a clean transfer corollary. Next, we analyze *gadget composition* with width control along the switching path $\mathbf{p}$ (Theorem 6.9). Finally, we state an SoS *entropy dichotomy* that either yields a compression echo or forces degree lower bounds (Theorem 7.9).

Related work (concise). This paper is a sequel to IECZ-I, which introduced the window discipline and size-aware MDL viewpoint we use here [11]. Our calibrated-window, size-aware pipeline builds on classic tools while tracking TV and size overheads. For gadget composition and switching arguments we follow Håstad’s switching-lemma methodology with explicit width control in distributional settings [4, 5]. Our SoS statements align with the modern low-degree and SoS viewpoint for CSPs [1] and the Fourier/entropy toolkit as surveyed in [12]. The injective $3\text{XOR} \rightarrow 3\text{SAT}$ mapping parallels Tseitin-style encodings but is window-stable and auxiliary-free, with TV losses tracked (cf. classical clause gadgets [17, 8]). Our MDL/minimax lens connects to coding-aware reductions and prefix-free overhead analyses in learning/complexity [15, 3].

2 Central assumptions and notation

We collect the global conventions referenced across sections. *We follow the conventions introduced in IECZ-I [10] and restate only what is needed here.*

Distributions and total variation. All distributions live on finite instance spaces with canonical encodings. For distributions μ, u over the same space, $\text{TV}(\mu, u)$ denotes total-variation distance. Deterministic maps R act by pushforward $R_{\#}\mu$, and total variation contracts under deterministic postprocessing: $\text{TV}(R_{\#}\mu, R_{\#}u) \leq \text{TV}(\mu, u)$. This aligns with the distribution-preserving reduction principle (A2d) from [10].

Prefix-free coding convention. Preprocessors/reductions (and any meta-data such as seeds/indices) are encoded prefix-free. Their description length contributes an additive $O(\log N)$ overhead to costs; N is the instance bitlength.

Size and windows. $\text{size}(\cdot)$ is the bitlength of the instance encoding. A *window* Win is an image slice where the reduction is measure-faithful: we write $\delta_{\text{tv}} := \text{TV}(R_{\#}\mu, u)$, and on Win one has $\delta_{\text{tv}} = 0$, with explicitly bounded size growth (as in [10]).

Cost functional. $V_{\mathcal{C}}(\mu, \varepsilon)$ denotes a size-aware MDL-type cost for class $\mathcal{C}$ at advantage/error parameter ε ; all statements are with respect to the fixed coding convention above. We adopt the definition and accounting rules introduced in [10].

Switching path and width. We fix the depth $d = O(\log n)$ and parameter path $\mathbf{p}$, implying bottom width $W = m^{O(c/d)}$ in the adopted switching-lemma regime.

Balancing. Where required, the input distribution is balanced on Win (specified per section).

Standing assumptions (global). Prefix-free meta-encodings are fixed (additive $O(\log N)$ overhead). Windows Win are measure-faithful ($\delta_{\text{tv}} = 0$) with controlled size growth. Balancing and moment-matching conditions are enforced only where explicitly invoked. The switching path $\mathbf{p}$ is calibrated for target depth $d = O(\log n)$ with bottom width W .

Notation table (symbols used throughout)

3 Notation and standing conventions

We work over the Boolean cube $\{0, 1\}^n$ unless noted otherwise. Expectations, variances, and covariances under a distribution μ are $\mathbb{E}_{\mu}[\cdot]$, $\text{Var}_{\mu}(\cdot)$, and $\text{Cov}_{\mu}(\cdot, \cdot)$. Total variation is $\text{TV}(\mu, u)$. Deterministic maps R act on measures by pushforward $R_{\#}\mu$ and satisfy contraction: $\text{TV}(R_{\#}\mu, R_{\#}u) \leq \text{TV}(\mu, u)$. The image *window* is denoted by Win (as introduced in IECZ-I); unless stated otherwise, all expectations and pseudoexpectations are taken on Win .

Size and encoding. $\text{size}(\cdot)$ denotes the bitlength of the (canonical, prefix-free) instance encoding. Meta/preprocessing descriptions are prefix-free and contribute an additive $O(\log N)$ overhead, where N is the instance bitlength.

Low degree and SoS. Low-degree polynomials are $p \in \mathcal{P}_{\leq k}$ (total degree $\leq k$). For sum-of-squares we write $\tilde{\mathbb{E}}[\cdot]$ for a degree- k pseudoexpectation calibrated to the proxy u .

Table 1: Symbols and their meaning (windowed setting).

Symbol	Meaning
μ, u	Target/windowed distribution and proxy/structured distribution on the same instance space.
$\text{TV}(\mu, u)$	Total-variation distance between μ and u (computed on Win when specified).
$R_{\#}\mu$	Pushforward of μ by a deterministic map R (reduction/preprocessor).
Win	Image window where the reduction is measure-faithful ($\delta_{\text{tv}} = 0$) and size growth is bounded.
$\text{size}(\cdot)$	Bitlength of an instance encoding; used for size-blowup accounting.
$\mathcal{C}$	Computational/model class (closed under poly-time preprocessing).
$V_{\mathcal{C}}(\mu, \varepsilon)$	Size-aware MDL-type cost of class $\mathcal{C}$ at parameter ε .
d	Switching depth, chosen $d = O(\log n)$.
p	Path parameter in the switching regime, $p = m^{-c/d}$ (fixed $c \in (0, 1)$).
W	Bottom width bound, $W = \text{poly}(1/p) = m^{O(c/d)}$.
$\mathcal{P}_{\leq k}$	Real polynomials of degree $\leq k$ on instance coordinates.
$\Delta_k(\mu, u)$	Low-degree test discrepancy: $\sup\{ \mathbb{E}_{\mu}[p] - \mathbb{E}_u[p] : p \in \mathcal{P}_{\leq k}, \ p\ _{\infty} \leq 1\}$.
$\tilde{\mathbb{E}}$	Degree- k SoS pseudoexpectation calibrated to u .
$\Delta_k^{\text{SoS}}(\mu, u)$	SoS-based discrepancy: $\sup\{ \mathbb{E}_{\mu}[p] - \tilde{\mathbb{E}}[p] : p \in \mathcal{P}_{\leq k}, \ p\ _{\infty} \leq 1\}$.
N	Instance bitlength used in prefix-free meta-encoding overheads ($O(\log N)$).

Biased Walsh basis (relative to a product proxy u). If u is a product distribution on $\{0, 1\}^n$ with biases $b_i = \mathbb{E}_u[x_i] \in (0, 1)$, set

$$\chi_i(x_i) := \frac{x_i - b_i}{\sqrt{b_i(1 - b_i)}}, \quad \chi_S(x) := \prod_{i \in S} \chi_i(x_i) \quad (S \subseteq [n]).$$

Then $\{\chi_S\}_{S \subseteq [n]}$ is orthonormal in $L^2(u)$, so $p = \sum_{|S| \leq k} \hat{p}(S) \chi_S$ and $\|p\|_{2,u}^2 = \sum_{|S| \leq k} \hat{p}(S)^2 = \mathbb{E}_u[p^2]$. We use $\langle f, g \rangle_u := \mathbb{E}_u[fg]$ for the u -inner product.

Discrepancy measures (used later). For $k \geq 0$, the low-degree test discrepancy and its SoS analogue are

$$\begin{aligned} \Delta_k(\mu, u) &:= \sup\{|\mathbb{E}_{\mu}[p] - \mathbb{E}_u[p]| : p \in \mathcal{P}_{\leq k}, \|p\|_{\infty} \leq 1\}, \\ \Delta_k^{\text{SoS}}(\mu, u) &:= \sup\{|\mathbb{E}_{\mu}[p] - \tilde{\mathbb{E}}[p]| : p \in \mathcal{P}_{\leq k}, \|p\|_{\infty} \leq 1\}. \end{aligned}$$

When we write δ_{tv} , we mean $\delta_{\text{tv}} := \text{TV}(R_{\#}\mu, u)$ (on Win if a window is fixed).

Problems and reductions. We write XOR, SAT, CLIQUE for problems, and name reductions as $R_{\text{XOR} \rightarrow \text{SAT}}$, $R_{\text{SAT} \rightarrow \text{CLIQUE}}$. Composition is $R_2 \circ R_1$; size blowups are stated explicitly (e.g. $m' = 4m$). Asymptotics are with respect to n unless another parameter is specified. Degree lower bounds for SoS refer to the standard pseudoexpectation model.

Standing assumptions (global, windowed). Windows Win are measure-faithful for the stated maps (i.e. $\delta_{\text{tv}} = 0$ on Win) with controlled size growth; balancing/moment matching is enforced only where explicitly invoked; switching-path parameters and $k = \Theta(\log n)$ are as specified in the corresponding sections.

4 Framework: Size-aware MDL and Windows

Consistency with IECZ-I. We adopt the same prefix-free coding model and the same window discipline as IECZ-I [11]; definitions are restated here for completeness.

4.1 Coding model and decodability

We work in a fixed, machine-independent prefix-free coding model (as in IECZ-I [11]), using self-delimiting integer codes in the spirit of Elias [2]. A *model* $M \in \mathcal{C}$ together with a deterministic preprocessor R is represented by a prefix-free description code $\text{code}(M, R) \in \{0, 1\}^*$ such that:

- (i) **Decodability.** There is a uniform decoder dec with $\text{dec}(\text{code}(M, R)) = (M, R)$; concatenations use self-delimiting integers encoded with Elias- δ ; all lengths are counted in bits [2].
- (ii) **Kraft admissibility.** The description family is prefix-free: $\sum_{w \in \text{PF}} 2^{-\text{len}(w)} \leq 1$.
- (iii) **Instance length.** We write $N = \text{size}(x)$ for the bitlength of an instance x .

The *meta-overhead* for specifying R and integer parameters (e.g. seeds, indices, window identifiers) is additive and, under Elias- δ , satisfies $\text{len}(\text{code}(R)) \leq (\sum_{i=1}^m k_i + 1) \log_2 N$ for all $N \geq 16$ (cf. Lemma 4.1); see also the MDL perspective [3].

Lemma 4.1 (Prefix-free meta-overhead (explicit δ -code)). *Fix Elias- δ for all integer fields. Let R be any polynomial-time map on instances of length N , whose meta-parameters $t_1, \dots, t_m$ each lie in a range $[0, T_i(N)]$ with $T_i(N) \leq N^{k_i}$ for some constants $k_i \geq 0$. Then, writing $\ell_\delta(\cdot)$ for the Elias- δ length in bits,*

$$\ell_\delta(t_i) \leq \lfloor \log_2 \max\{1, t_i\} \rfloor + 2 \lfloor \log_2(1 + \lfloor \log_2 \max\{1, t_i\} \rfloor) \rfloor + 1,$$

and hence, for all $N \geq 16$ and all i with $t_i \leq T_i(N) \leq N^{k_i}$,

$$\ell_\delta(t_i) \leq k_i \log_2 N + 2 \log_2 \log_2 N + 3 \leq (k_i + 1) \log_2 N.$$

Consequently, for any constant-arity template (constant m) one has the total meta-overhead

$$\text{len}(\text{code}(R)) = \sum_{i=1}^m \ell_\delta(t_i) + O(1) \leq \left(\sum_{i=1}^m k_i + 1 \right) \log_2 N = O(\log N).$$

The same bound holds for any finite tuple of such parameters encoded by prefix-free concatenation.

Proof. For Elias- δ , the length of a positive integer $t \geq 1$ is $\ell_\delta(t) = \lfloor \log_2 t \rfloor + 2 \lfloor \log_2(1 + \lfloor \log_2 t \rfloor) \rfloor + 1$, see [2]. Allow $t = 0$ by coding $\max\{1, t\}$ (adds at most one bit). If $t \leq N^{k_i}$, then $\lfloor \log_2 t \rfloor \leq k_i \log_2 N$, and $\log_2(1 + \lfloor \log_2 t \rfloor) \leq \log_2 \log_2 N^{k_i} \leq \log_2 \log_2 N + \log_2 k_i \leq \log_2 \log_2 N + 1$ (for fixed k_i and all $N \geq 16$). This yields $\ell_\delta(t_i) \leq k_i \log_2 N + 2 \log_2 \log_2 N + 3 \leq (k_i + 1) \log_2 N$. Summation over a constant number m of fields and adding the constant template/tag overhead gives the displayed bounds. Prefix-freeness is by construction of Elias- δ and stability under concatenation of self-delimiting fields. $\square$

4.2 Window discipline

Lemma 4.2 (TV is 1-Lipschitz under deterministic maps). *For any distributions μ, u on the same finite space and any deterministic map R ,*

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

Proof. Write $\text{TV}(\mu, u) = \sup_A (\mu(A) - u(A))$, where the supremum is over measurable sets. For any measurable B in the codomain, $(R_{\#}\mu)(B) - (R_{\#}u)(B) = \mu(R^{-1}(B)) - u(R^{-1}(B))$, so taking the supremum over B is a supremum over preimages $R^{-1}(B)$. Hence the supremum cannot increase: $\text{TV}(R_{\#}\mu, R_{\#}u) \leq \text{TV}(\mu, u)$. $\square$

A *window* Win is a measurable image slice on which a fixed reduction is *approximately* measure-faithful in total variation. We write

$$d_{\text{TV}}(\text{Win}) := \text{TV}(R_{\#}\mu, u) = o(1) \quad \text{on Win},$$

uniformly along the calibrated blueprint path. (In the special case of exact faithfulness one has $d_{\text{TV}}(\text{Win}) = 0$; all our bounds are stated for $d_{\text{TV}}(\text{Win}) = o(1)$.) Size growth is explicitly bounded (e.g. clause blowup $m' = 4m$ for $\text{XOR} \rightarrow \text{SAT}$, cf. [11]). All later statements are parameterized by window balance and boundary error; see the loss tables in each section.

Remark 4.1 (Distributional framing). All complexity and SoS statements in this paper are interpreted on the calibrated image window Win . The sequel [9] takes this framing as input to a lift step (aggregation $\rightarrow$ hardcore selection $\rightarrow$ seeded condensation $\rightarrow$ derandomization) that targets parameterized distributional families with controlled TV/size/meta overheads.

Remark 4.2 (Width calibration across regimes). Throughout the calibrated path we take $p = m^{-c/d}$ with $d = \Theta(\log n)$ and a fixed small constant $c \in (0, 1)$, which yields bottom width $W = \text{poly}((1/p))$ uniformly on the window. This uniformity is what underpins the absolute hypercontractive constants used later. If d is allowed to grow superlogarithmically, the same relation forces smaller p and can drive W beyond polynomial ranges; our analysis does not track this deterioration. All quantitative claims here are therefore scoped to $d, k = \Theta(\log n)$; see also Sec. 6.

Style & symbols (conventions).

- **Window.** We write Win in math and “Win” in prose for the image slice on which the reduction is measure-faithful.
- **Pushforward.** Deterministic maps act by pushforward $R_{\#}\mu$; total variation (TV) is 1-Lipschitz: $\text{TV}(R_{\#}\mu, R_{\#}u) \leq \text{TV}(\mu, u)$.
- **Fields/groups.** We write $\mathbb{F}_2$ for the binary field; $\text{GL}_2(\mathbb{F}_2)$ denotes the group of invertible 2×2 matrices over $\mathbb{F}_2$.
- **Characters.** χ_S denotes the Walsh character on support S (orthonormal under the proxy on Win).
- **Noise.** T_θ (Bonami–Beckner) acts diagonally on degree- d slices with eigenvalues θ^d .

Notation at a glance. Unless stated otherwise, all probabilistic statements are understood on Win . We use the shorthands

$$\begin{aligned} \text{TV}(\mu, u) & \text{ (total variation),} & R_{\#}\mu & \text{ (pushforward),} & d_{\text{TV}}(\text{Win}) & := \text{TV}(R_{\#}\mu, u), \\ d, k & \text{ (degree budgets)} \\ T_\theta & \text{ (Bonami–Beckner)} \\ \chi_S & \text{ (Walsh character)} \\ \varpi_{\text{Win}} & \text{ (baseline conditioned to Win)} \end{aligned}$$

4.3 Definition of the size-aware MDL cost

Fix a class $\mathcal{C}$ (closed under polynomial-time preprocessing) and an error/advantage parameter $\varepsilon \in [0, 1]$. For a distribution μ on instances of length N define

$$V_{\mathcal{C}}(\mu, \varepsilon) := \inf_{M \in \mathcal{C}, R \text{ poly}} \left\{ \text{len}(\text{code}(M, R)) + \mathbb{E}_{x \sim \mu} [L_M(R(x), \varepsilon)] \right\}. \quad (1)$$

where $L_M(\cdot, \varepsilon)$ is a nonnegative per-instance loss measuring the residual description under M at target error ε (e.g. a codelength for mistakes or a hinge-type penalty; concrete form not needed for the monotonicity arguments). The coding convention is fixed once and for all; hence all $O(\log N)$ terms are absolute with respect to Lemma 4.1. This matches the IECZ-I definition [11]; we restate it for completeness.

Remark 4.3 (Role of L_M). Our results require only that L_M be subadditive under concatenation, normalized to vanish on perfect fits, and insensitive to $O(1)$ renamings. These mild conditions hold for standard MDL-style penalties and for prediction losses that translate to codelength by Shannon coding.

Lemma 4.3 (Subadditivity of the per-instance loss). *Let $L_M(\cdot, \varepsilon) \geq 0$ be the per-instance loss for a fixed model $M \in \mathcal{C}$ at target ε . Assume the coding convention of Sec. 4.1. Then for any two instances x, y (over disjoint supports or properly tagged concatenations) one has*

$$L_M(x\|y, \varepsilon) \leq L_M(x, \varepsilon) + L_M(y, \varepsilon) + O(1),$$

where $\|$ denotes the canonical concatenation in our encoding. The $O(1)$ term accounts for the self-delimiting tag separating the two parts and is independent of N .

Proof. Concatenate the two descriptions produced by M on x and y and add a self-delimiting boundary tag. By prefix-free decodability (Sec. 4.1) the decoder can split and feed each part to the same M . The score/codelength is additive up to the single boundary tag, which contributes $O(1)$ bits (self-delimiting constant field). Nonnegativity is by definition. $\square$

Definition 4.1 (Model class assumptions). Throughout, the model class $\mathcal{C}$ satisfies:

1. **Closure under poly-time preprocessing.** If $M \in \mathcal{C}$ and R is deterministic poly-time, then the composed system $(x \mapsto M(R(x)))$ is represented in $\mathcal{C}$ with prefix-free description length $\ell(\text{code}(M, R)) = \ell(\text{code}(M)) + \ell(\text{code}(R)) + O(1)$.
2. **Uniform decoding.** There is a single uniform decoder that maps any description word to (M, R) ; see Sec. 4.1 for the coding model.
3. **Stable per-instance loss.** The loss $L_M(\cdot, \varepsilon) \geq 0$ is invariant under $O(1)$ renamings and obeys Lemma 4.3.

These assumptions are exactly what Theorems 4.4 and C.5 use.

4.4 Minimax monotonicity (formal)

We restate Theorem 4.4 with the explicit coding split.

Theorem 4.4 (Minimax monotonicity). *Let R be deterministic polynomial time and $\mathcal{C}$ be closed under polynomial-time preprocessing. Then for all μ and ε ,*

$$V_{\mathcal{C}}(R \# \mu, \varepsilon) \leq V_{\mathcal{C}}(\mu, \varepsilon) + \text{len}(\text{code}(R)).$$

In particular, under our fixed coding model, $\text{len}(\text{code}(R)) = O(\log N)$ by Lemma 4.1.

Proof. Let $V_{\mathcal{C}}(\mu, \varepsilon)$ be defined as in (1) (cf. (1) in Sec. 4.3):

$$V_{\mathcal{C}}(\mu, \varepsilon) = \inf_{M \in \mathcal{C}, R_o \text{ poly}} \left\{ \text{len}(\text{code}(M, R_o)) + \mathbb{E}_{x \sim \mu} [L_M(R_o(x), \varepsilon)] \right\}.$$

Fix $\delta > 0$ and choose (M^*, R_o^*) δ -optimal for μ , so that

$$V_{\mathcal{C}}(\mu, \varepsilon) \geq \text{len}(\text{code}(M^*, R_o^*)) + \mathbb{E}_{\mu} [L_{M^*}(R_o^*(x), \varepsilon)] - \delta.$$

Consider reusing the same preprocessor on the image space: define $R_\circ^{\star'} := R_\circ^\star$ (as an arbitrary poly-time map on bitstrings, it can be applied to y directly). By change of variables under the pushforward,

$$\mathbb{E}_{x \sim \mu} [L_{M^\star}(R_\circ^\star(R(x)), \varepsilon)] = \mathbb{E}_{y \sim R \# \mu} [L_{M^\star}(R_\circ^{\star'}(y), \varepsilon)].$$

with $R_\circ^{\star'}(R(x)) = R_\circ^\star(x)$ on the image.

Now encode $(M^\star, R_\circ^{\star'})$ together with R in the prefix-free model. By Lemma 4.1, $\text{len}(\text{code}(R)) = O(\log N)$ and concatenation is prefix-free, hence

$$\begin{aligned} V_{\mathcal{C}}(R \# \mu, \varepsilon) &\leq \text{len}(\text{code}(M^\star, R_\circ^{\star'})) + \mathbb{E}_{y \sim R \# \mu} [L_{M^\star}(R_\circ^{\star'}(y), \varepsilon)] \\ &\leq \text{len}(\text{code}(M^\star, R_\circ^\star)) + \mathbb{E}_{x \sim \mu} [L_{M^\star}(R_\circ^\star(x), \varepsilon)] + \text{len}(\text{code}(R)). \end{aligned}$$

Taking infimum over $(M^\star, R_\circ^\star)$ and letting $\delta \rightarrow 0$ yields $V_{\mathcal{C}}(R \# \mu, \varepsilon) \leq V_{\mathcal{C}}(\mu, \varepsilon) + \text{len}(\text{code}(R))$, as claimed. By Lemma 4.1, $\text{len}(\text{code}(R)) = O(\log N)$ under our fixed coding model. $\square$

Tightness up to $O(\log N)$

Proposition 4.5 (Tightness). *There exist families (μ_N) and maps R_N for which the gap $V_{\mathcal{C}}(R_N \# \mu_N, \varepsilon) - V_{\mathcal{C}}(\mu_N, \varepsilon)$ matches $\Theta(\log N)$ for all sufficiently large N under the fixed coding convention.*

Proof. Fix N and let R_N be a deterministic poly-time map that is injective on its image window Win and whose inverse requires an explicit disambiguation index $u \in \{0, 1\}^{s(N)}$ with $s(N) = \Theta(\log N)$ (e.g., the window/seed-selector index fixed in our blueprint). Construct μ_N supported uniformly on a finite set S_N of instances of bitlength N such that:

- (i) R_N is a bijection from S_N onto its image $R_N(S_N)$, and
- (ii) for each $y \in R_N(S_N)$ there are exactly $2^{s(N)}$ preimages consistent with the same coarse description, distinguished only by the self-delimiting index u .

Lower bound. Consider any encoder/decoder implementing $V_{\mathcal{C}}(R_N \# \mu_N, \varepsilon)$. To reproduce the optimal description for μ_N , the code must additionally specify u to invert R_N on $R_N(S_N)$ (even when the model part is identical), by prefix-free decodability and injectivity of R_N on Win . By Elias-style self-delimiting coding (Lemma 4.1), $\ell(u) \geq c \log N - O(1)$ for some absolute $c > 0$. Averaging over μ_N yields

$$V_{\mathcal{C}}(R_N \# \mu_N, \varepsilon) \geq V_{\mathcal{C}}(\mu_N, \varepsilon) + \Omega(\log N).$$

Upper bound. Conversely, augment the description for μ_N by inlining the same u (the window/seed-selector index). This adds $\ell(u) = O(\log N)$ bits by Lemma 4.1 and recovers the same effective model on $R_N \# \mu_N$. Hence

$$V_{\mathcal{C}}(R_N \# \mu_N, \varepsilon) \leq V_{\mathcal{C}}(\mu_N, \varepsilon) + O(\log N).$$

Combining both bounds yields the claimed $\Theta(\log N)$ gap. $\square$

5 Reductions: 3XOR $\rightarrow$ 3SAT (window-stable)

Standing conditions. On the image window Win the reduction is measure-faithful ($\delta_{\text{tv}} = 0$) as fixed in IECZ-I; see [11]. The translation uses no auxiliary variables and has controlled size blowup $m' = 4m$. Prefix-free coding is fixed (Elias- δ), so the meta-overhead is $O(\log N)$, cf. [11]. We assume the input distribution is balanced on Win (IECZ-I window discipline). We work in the incidence model in which each unordered triple $\{x_i, x_j, x_k\}$ appears at most once (IECZ-I).

Lemma 5.1 (Injective $3\text{XOR} \rightarrow 3\text{SAT}$ translation (recalled; cf. IECZ-I) — empirical checks: App. A). *There exists a translation mapping each 3XOR constraint $x \oplus y \oplus z = b$ to four 3SAT clauses over the same three variables, with no auxiliary variables, such that on the image window Win the pushforward preserves measure ($\delta_{\text{tv}} = 0$) and the instance size scales as $m' = 4m$. The map is injective, and its inverse on the image is computable in linear time by recovering each XOR block from its four-clause pattern, reusing the exact four-clause templates from IECZ-I [11]*

Explicit clause mapping (no auxiliaries). Let x, y, z be Boolean variables and $b \in \{0, 1\}$. Encode $x \oplus y \oplus z = b$ by:

$$\begin{aligned} b = 1 : & \quad (x \vee y \vee z) \wedge (x \vee \neg y \vee \neg z) \\ & \quad \wedge (\neg x \vee y \vee \neg z) \wedge (\neg x \vee \neg y \vee z), \\ b = 0 : & \quad (\neg x \vee \neg y \vee \neg z) \wedge (x \vee y \vee \neg z) \\ & \quad \wedge (x \vee \neg y \vee z) \wedge (\neg x \vee y \vee z). \end{aligned}$$

These are exactly the templates fixed and used in IECZ-I [11].

Toy example (one block; TV/size accounting). Consider a single constraint $x \oplus y \oplus z = b$ on the window Win . The image under the injective map is exactly the four-clause CNF block shown above; thus the size multiplies by 4 (no auxiliaries). On Win the pushforward is measure-faithful, hence the total-variation loss is 0 for this block. For a conjunction of m independent 3XOR constraints, the translation yields $4m$ clauses, still with $\delta_{\text{tv}} = 0$ on Win .

Historical note. Classical clause gadgets (e.g., Tseitin encodings) realize parity via constant-size CNF blocks; our translation is the IECZ-I auxiliary-free, window-stable variant with tracked TV loss and stated size bounds [17, 11].

Proposition 5.2 (Block uniqueness & inverse map). *Under the incidence model (IECZ-I), every set of clauses whose variable support is exactly $\{x, y, z\}$ forms a block that equals one of the two parity templates above. Moreover, there is a deterministic linear-time inverse map T that, given a CNF in the image, recovers the XOR instance uniquely by:*

- (i) *grouping clauses by unordered variable triples $\{x, y, z\}$,*
- (ii) *normalizing literals to a canonical order,*
- (iii) *matching the 4-clause multiset to the $b \in \{0, 1\}$ template.*

On the window Win , no unordered triple $\{x, y, z\}$ can give rise to more than one image block, because (a) the source incidence model admits each triple at most once, and (b) our translation is blockwise injective. Therefore, the clause grouping by unordered triples is well-defined. Hence $T \circ R = \text{id}$ on the source, and the translation is injective on instances.

Corollary 5.3 (Window transfer — Theorem 4.4 to 3SAT). *Under the map of Lemma 5.1, the cost transfers:*

$$V_{\mathcal{C}}^{\text{SAT}}(\text{Win}, \varepsilon) \geq V_{\mathcal{C}}^{\text{XOR}}(\text{Win}, \varepsilon) - O(\log N).$$

Justification. This follows from Theorem 4.4 (minimax monotonicity) under pushforward by $R_{\text{XOR} \rightarrow \text{SAT}}$, together with the fact that deterministic post-processing is 1-Lipschitz in total variation (Lemma 4.2). The prefix-free coding overhead contributes at most $O(\log N)$ to the cost (Lemma 4.1). The window calibration and balancing we invoke are exactly those fixed in IECZ-I [11]. $\square$

Table 2: Forbidden assignments per parity b (the remaining four satisfy).

$b = 0$					$b = 1$				
	000	001	010	011		000	001	010	011
status	allow	forbid	forbid	allow	status	forbid	allow	allow	forbid
	100	101	110	111		100	101	110	111
status	forbid	allow	allow	forbid	status	allow	forbid	forbid	allow

Table 3: Transfer losses for reductions (on the IECZ-I window).

Arrow	TV loss	Size	Error	Notes
$R_{\text{xor} \rightarrow \text{sat}}$	0	$\times 4$	$\varepsilon' = \varepsilon$	Injective, window-faithful; no auxiliaries
$R_{\text{sat} \rightarrow \text{clique}}$	$o(1)$ on Win	$ V = 3m,$ $ E = O(m^2)$	$\varepsilon' = \varepsilon$	Karp; window conditioning accounts TV

Truth table view (forbidden assignments)

For each constraint $x \oplus y \oplus z = b$, the four 3-clauses forbid exactly the assignments with parity $\neq b$.

Window faithfulness and loss accounting

On Win, the translation is deterministic and preserves the satisfiability pattern on each triple; hence $\delta_{\text{tv}} = 0$ (measure-faithful pushforward), exactly as stipulated in IECZ-I [11]. Size increases by a factor 4 in the number of clauses, with no new variables. Error/advantage parameters carry over unchanged.

Worked example (inverse map T on a tiny instance). Take a 3XOR instance on variables $\{x_1, x_2, x_3\}$ with the single constraint $x_1 \oplus x_2 \oplus x_3 = 1$. The translation yields the block

$$\begin{aligned}
 & (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \neg x_2 \vee \neg x_3) \\
 & \wedge (\neg x_1 \vee x_2 \vee \neg x_3) \wedge (\neg x_1 \vee \neg x_2 \vee x_3).
 \end{aligned}$$

Given just these four clauses in a CNF, the inverse T groups by the triple $\{x_1, x_2, x_3\}$, sorts literals inside each clause to a canonical order, and matches the multiset against the $b \in \{0, 1\}$ templates: the clause $(x_1 \vee x_2 \vee x_3)$ appears, hence $b = 1$ uniquely. Therefore T recovers $x_1 \oplus x_2 \oplus x_3 = 1$.

Proof of Lemma 5.1. Soundness. For a fixed parity $b \in \{0, 1\}$, among the eight assignments to (x, y, z) exactly four have parity b . In the $b = 1$ case, each of the four 3-clauses rules out exactly one of the four parity-0 assignments, so their conjunction is true precisely on parity-1 assignments. The case $b = 0$ is symmetric. **Completeness.** If the 4-clause block is satisfied, then none of the “wrong-parity” assignments can occur, hence $x \oplus y \oplus z = b$. No auxiliaries are introduced and each 3XOR yields exactly four 3-clauses, thus $m' = 4m$. $\square$

Proof of Proposition 5.2. Group all clauses by their *support triple* $\{x, y, z\}$. Under the incidence assumption, each unordered triple appears in the 3XOR instance at most once, hence at most one block per triple exists. The two parity templates are disjoint as clause multisets: for $b = 0$ the block contains $(\neg x \vee \neg y \vee \neg z)$ which never occurs for $b = 1$, while for $b = 1$ the block contains $(x \vee y \vee z)$ which never occurs for $b = 0$. Canonically normalize literals

in each clause (e.g., lexicographic variable order), which makes the multiset representation unique. The inverse map T checks each block against the $b = 0$ and $b = 1$ templates and reconstructs the constraint $x \oplus y \oplus z = b$. This is linear in the number of clauses. Finally, $T \circ R = \text{id}$ on sources, so R is injective. By Lemma 5.5, T runs in linear time. $\square$

Canonical clause-block identifier. For a clause $C = (\ell_1 \vee \ell_2 \vee \ell_3)$ over variables $\{x, y, z\}$, let $\text{canon}(C)$ sort variables and push negations into a fixed sign bit, producing a triple in $\{\pm 1\} \times \{x, y, z\}^3$ with lexicographic order. For a block B supported on $\{x, y, z\}$, define the multiset $\text{Canon}(B) := \{\text{canon}(C) : C \in B\}$ and its sorted list $\text{Canon}^\uparrow(B)$.

Lemma 5.4 (Uniqueness of the canonical ID). *Under the incidence model (each unordered triple appears at most once), the map $B \mapsto \text{Canon}^\uparrow(B)$ is injective. Moreover, the two parity templates ($b = 0$ vs. $b = 1$) yield different $\text{Canon}^\uparrow(\cdot)$.*

Proof. By construction, $\text{Canon}(\cdot)$ is invariant under literal permutations and clause ordering; sorting yields a canonical representative. The two parity templates differ by at least one clause— $(x \vee y \vee z)$ appears only for $b = 1$ and $(\neg x \vee \neg y \vee \neg z)$ only for $b = 0$ —hence their canonical lists differ. $\square$

Deterministic inverse via the canonical ID. Given a CNF in the image, partition clauses by unordered triples $\{x, y, z\}$, compute $\text{Canon}^\uparrow(B)$ for each block B , and match it against the two stored templates to recover $b \in \{0, 1\}$. This yields the inverse T in linear time in the number of clauses.

Lemma 5.5 (Linear-time inverse). *Given a CNF in the image of the translation, the inverse T that groups by unordered triples, canonicalizes clauses, and matches against the parity templates runs in time linear in the number of clauses.*

Deterministic Karp reduction $3\text{SAT} \rightarrow \text{CLIQUE}$ (window-stable)

Theorem 5.6 (Karp reduction $\text{SAT} \rightarrow \text{CLIQUE}$). *Let φ be a 3CNF with clauses $C_1, \dots, C_m$, each with exactly three literals. Construct $G = (V, E)$ by*

$$V = \{(i, \ell) : \ell \in C_i \text{ is one of its three literals}\}, \quad |V| = 3m,$$

and for $i \neq j$ put an edge between (i, ℓ) and (j, ℓ') iff ℓ and ℓ' are not contradictory (i.e. not a variable and its negation). Then G has a clique of size m iff φ is satisfiable. Moreover, the construction is deterministic and computable in time $O(m^2)$.

Proof. $(\Rightarrow)$ A satisfying assignment picks in each clause C_i some true literal ℓ_i ; the m vertices $\{(i, \ell_i)\}_{i=1}^m$ form a clique because chosen literals come from different clauses and are mutually consistent. $(\Leftarrow)$ Any m -clique contains exactly one vertex from each clause-block (else two vertices from the same clause would be nonadjacent). The corresponding literals are pairwise consistent; extend to a global assignment satisfying all clauses. $\square$

Proposition 5.7 (Size bounds). *The reduction of Theorem 5.6 yields $|V| = 3m$ and $|E| \leq 9\binom{m}{2} = O(m^2)$. Hence the size blowup is polynomial and explicit.*

Lemma 5.8 (Window faithfulness). *On the window Win (no tautological clauses, no duplicated literals inside a clause; both events have $o(1)$ mass along the calibrated ensemble), as adopted from IECZ-I [11], the map of Theorem 5.6 is measure-faithful up to $o(1)$ total-variation: pushing forward the source distribution incurs TV loss $o(1)$, and error/advantage parameters are unchanged.*

Table 4: Reductions and overheads on the calibrated window (IECZ-I).

Arrow	TV loss	Size blowup	Error/Advantage
$R_{\text{xor} \rightarrow \text{sat}}$	0	$\times 4$	$\varepsilon' = \varepsilon$
$R_{\text{sat} \rightarrow \text{clique}}$	$o(1)$ on Win	$\times \text{poly}(n)$	$\varepsilon'' = \varepsilon' + o(1)$

Corollary 5.9 (Transfer of cost to CLIQUE). *With $R_{\text{xor} \rightarrow \text{sat}}$ of Lemma 5.1 and $R_{\text{sat} \rightarrow \text{clique}}$ of Theorem 5.6,*

$$\begin{aligned} V_C^{\text{CLIQUE}}(\text{Win}, \varepsilon) &\geq V_C^{\text{SAT}}(\text{Win}, \varepsilon) - O(\log N) \\ &\geq V_C^{\text{XOR}}(\text{Win}, \varepsilon) - O(\log N). \end{aligned}$$

With cumulative total variation (TV) loss $o(1)$ on Win and total size blowup $\times 4 \cdot \text{poly}(n)$. By Lemma 4.2, deterministic postprocessing is 1-Lipschitz in total variation, so the $o(1)$ boundary loss is preserved through composition.

Supplementary verification (artifact). We provide small scripts that empirically validate the injective $3\text{XOR} \rightarrow 3\text{SAT}$ mapping stated in Lemma 5.1:

- (i) `xor_to_sat_check.py` exhaustively checks the 8 truth-table rows for both parities $b \in \{0, 1\}$ and, on random instances, verifies that any consistent XOR system yields a satisfying CNF assignment; it also includes a small- n reverse check that witnesses UNSAT for inconsistent XOR systems via brute-force (solver-free).
- (ii) `block_identifier_check.py` validates the canonical four-clause block IDs and the inverse identification map.
- (iii) `run_xor_to_sat_check.ps1` is a convenience runner with seed/size parameters.

All tests are self-contained; the reverse check is limited to small n and documented in Appendix A.

6 Gadget Composition (XOR/Index)

Standing conditions. We fix a parameter path $\mathbf{p}$ with depth $d = O(\log n)$ in the switching-lemma regime, which yields the width bound W . On the window Win there is a mixing witness with rate $\theta < 1$ (Lemmas 6.10 and 6.14), hence correlation stability holds. Gadget descriptions are prefix-free, contributing only an $O(\log N)$ meta-overhead. All window and switching-path conventions here are identical to IECZ-I [11].

XOR/Index Gadget

Inputs $a, b, c \in \{0, 1\}$ (signals sampled on the calibrated window)

Outputs $s_1, s_2 \in \{0, 1\}$

Parity preservation $s_1 \oplus s_2 = a \oplus b \oplus c$

Locality fan-in = 3, fan-out = $O(1)$ (constant)

Correlation stability for some $\rho < 1$: $\max_{\ell \in \{1, 2\}} |\text{corr}(s_\ell, \text{exogenous noise})| \leq \rho$

Width regime bottom width $W = \text{poly}(1/p)$ under the switching path $p = m^{-c/d}$

We fix the parameter path $\mathbf{p}$ with depth $d = O(\log n)$; width stays bounded by W in the switching-lemma regime [4, 12].

Lemma 6.1 (Width after d switching rounds — seed-fixed). *There exists a seed s of length $O(\log n)$ such that along the corresponding deterministic path of $d = O(\log n)$ switching rounds the parameters satisfy*

$$W_r \leq \frac{1}{p} W_{r-1}, \quad M_r \leq p M_{r-1} \quad (r = 1, \dots, d),$$

for the same $p \in (0, 1)$ as in the standard switching-lemma analysis. The seed s is then fixed (prefix-free) at meta cost $O(\log n)$.

Proof. By the usual switching-lemma bounds, with probability $1 - o(1/n)$ over the random choices in a round, each displayed inequality holds for that round. A union bound over $d = O(\log n)$ rounds yields existence of a single seed s witnessing all rounds simultaneously with probability $1 - o(1)$. Fix this s (recorded prefix-free at cost $O(\log n)$). Along the resulting deterministic path the stated recurrences hold. $\square$

Consistency with IECZ-I. The seed-fixing step and the choice $p = m^{-c/d}$ are exactly as in IECZ-I [11].

Lemma 6.2 (Interface normal form over $\mathbb{F}_2$). *Let G be a constant-size gadget with three Boolean inputs (a, b, c) and two outputs (s_1, s_2) that is parity-preserving:*

$$s_1 \oplus s_2 = a \oplus b \oplus c.$$

Write uniquely, in algebraic normal form over $\mathbb{F}_2$,

$$s_\ell(x) = \alpha_\ell \oplus \beta_\ell^\top x \oplus q_\ell(x) \quad (\ell = 1, 2),$$

where $x = (a, b, c)^\top$, $\alpha_\ell \in \mathbb{F}_2$, $\beta_\ell \in \mathbb{F}_2^3$, and each q_ℓ contains only monomials of degree ≥ 2 . Then $q_1 \equiv q_2$. Consequently, for

$$U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \text{GL}_2(\mathbb{F}_2) \quad \text{and} \quad s := (s_1, s_2)^\top, \quad x := (a, b, c)^\top,$$

the mixed outputs $t := Us = (t_1, t_2)^\top$ satisfy

$$t_1 \text{ is affine in } x, \quad t_2 = (\text{affine in } x) \oplus q(x),$$

with the same non-linear core $q := q_1 = q_2$. In particular, both outputs are affine iff $q \equiv 0$.

Moreover, letting

$$A := \begin{bmatrix} \beta_1^\top \\ \beta_2^\top \end{bmatrix} \in \mathbb{F}_2^{2 \times 3}, \quad \mathbf{r} := \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \in \mathbb{F}_2^2, \quad \mathbf{1} := (1, 1)^\top,$$

the parity constraint is equivalent to

$$\mathbf{1}^\top A = (1, 1, 1) \quad \text{and} \quad \mathbf{1}^\top \mathbf{r} = 0,$$

since $q_1 \oplus q_2 \equiv 0$. Under non-degeneracy of the interface, the affine coefficient matrix A has full row rank 2.

Proof. Working over $\mathbb{F}_2$, expand each s_ℓ in algebraic normal form as stated. The parity condition reads

$$s_1 \oplus s_2 = (\alpha_1 \oplus \alpha_2) \oplus (\beta_1 \oplus \beta_2)^\top x \oplus (q_1 \oplus q_2) = a \oplus b \oplus c,$$

which is affine; hence its higher-degree part vanishes and $q_1 \oplus q_2 \equiv 0$, i.e. $q_1 \equiv q_2 =: q$. Choosing $U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ gives $t_1 = s_1 \oplus s_2$ (affine) and $t_2 = s_2 = (\text{affine}) \oplus q$. The equalities $\mathbf{1}^\top A = (1, 1, 1)$ and $\mathbf{1}^\top \mathbf{r} = 0$ follow by comparing affine parts in $\mathbf{1}^\top s = \mathbf{1}^\top (Ax \oplus \mathbf{r}) \oplus (q_1 \oplus q_2) = \mathbf{1}^\top x$. If the interface is non-degenerate, the affine rows are independent, so $\text{rank}(A) = 2$. $\square$

Lemma 6.3 (Canonical normal forms). *Modulo output mixing $U \in \text{GL}_2(\mathbb{F}_2)$, every parity-preserving gadget interface of Lemma 6.2 (i.e., $s = Ax \oplus r$ with $\mathbf{1}^\top A = (1, 1, 1)$, $\mathbf{1}^\top r = 0$, $\text{rank}(A) = 2$) is equivalent to one of the following three representatives:*

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad r = 0.$$

In particular, up to post-mix, we may take $(s_1, s_2) = (a, b \oplus c)$ or its two cyclic rotations.

Index gadget (constant fan-in, affine, degree-stable). We isolate a tiny *Index* block that routes one of two data bits according to a 1-bit address. Let inputs be $(u; v_0, v_1) \in \{0, 1\}^3$ and outputs $t \in \{0, 1\}$ defined by

$$t = (1 \oplus u) v_0 \oplus u v_1.$$

Over $\mathbb{F}_2$ this is bilinear. We use the standard “affine lift” by introducing a constant ancilla $c \equiv 1$ to rewrite t as an *affine* function in the expanded coordinate system (u, v_0, v_1, c) :

$$t = v_0 \oplus u(v_0 \oplus v_1) = v_0 \oplus (u \cdot (v_0 \oplus v_1)) = \underbrace{v_0}_{\text{linear}} \oplus \underbrace{(u \cdot v_0)}_{\text{bilinear}} \oplus \underbrace{(u \cdot v_1)}_{\text{bilinear}}.$$

In our degree-aware setting, the block is of constant size and can be treated as an affine pullback on monomials: composition with the index map increases the total degree by at most a constant factor (here ≤ 2), which is absorbed in the $k = \Theta(\log n)$ regime.

Lemma 6.4 (Index pullback with constant degree factor). *Let G_{idx} act blockwise on disjoint triples $(u; v_0, v_1)$ to produce outputs t as above, and let p be a multilinear polynomial in the t -variables with $\deg p \leq k$. Then there exists a multilinear q in the input variables $(u; v_0, v_1)$ such that*

$$p(G_{\text{idx}}(u; v_0, v_1)) = q(u; v_0, v_1), \quad \deg q \leq 2k, \quad \|q\|_\infty \leq \|p\|_\infty.$$

Consequently, for any input distributions (μ, u) on the window and any k ,

$$\Delta_k(G_{\text{idx}} \# \mu, G_{\text{idx}} \# u) \leq \Delta_{2k}(\mu, u).$$

Sup-norm note. Since $q = p \circ G_{\text{idx}}$ is evaluated only on the image $G_{\text{idx}}(\{0, 1\}^m) \subseteq \{0, 1\}^{m'}$, we have $\|q\|_\infty = \sup_x |p(G_{\text{idx}}(x))| \leq \sup_y |p(y)| = \|p\|_\infty$.

Corollary 6.5 (Index monotonicity of low-degree discrepancy). *For any $k \in \mathbb{N}$ and any input distributions (μ, u) on the window,*

$$\Delta_k(G_{\text{idx}} \# \mu, G_{\text{idx}} \# u) \leq \Delta_{2k}(\mu, u).$$

Corollary 6.6 (Index transfer with budget rescaling). *(Concrete budget.) For the Index gadget the degree budget dilates by at most a factor 2. Hence, for $k' = \lfloor k/2 \rfloor$,*

$$\Delta_{k'}(G_{\text{idx}} \# \mu, G_{\text{idx}} \# u) \leq \Delta_k(\mu, u).$$

Proof of Corollary 6.6. The lemma shows that substituting the index gadget inflates test degree by at most a factor 2, hence any output degree- $\leq k'$ test corresponds to an input test of degree $\leq 2k' \leq k$, which yields the claimed inequality. $\square$

Proof of Lemma 6.4. Each output coordinate t is a polynomial of total degree ≤ 2 in its local inputs. Substituting t into a degree- $\leq k$ polynomial p yields a polynomial in $(u; v_0, v_1)$ of degree $\leq 2k$. Because the gadget fan-in is constant and the blocks are disjoint, monomials of p expand to sums of monomials touching $O(1)$ variables per block. Moreover, since $q = p \circ G_{\text{idx}}$ with deterministic G_{idx} , the image of G_{idx} is a subset of the output domain, hence $\|q\|_\infty \leq \|p\|_\infty$. Taking suprema over degree- $\leq k$ tests on outputs gives the stated Δ_k versus Δ_{2k} bound. $\square$

Remark 6.1 (Scope note: circuit classes). The *Index* block introduces controlled bilinearity at constant locality. Within our blueprint, the degree and correlation parameters remain gadget-stable for $k = \Theta(\log n)$ on the window. We do not make claims about AC^0 or ACC^0 ; mentions of these classes are for orientation only. We use only that Δ_k does not increase under G_{idx} (Lemma 6.4) and that width/correlation stay in the switching regime.

Proposition 6.7 (Affine pullback preserves low degree). *Let G be a (blockwise) affine gadget in the sense of Lemma 6.2, i.e., after an admissible output mixing we have $y = Ax \oplus r$ over $\mathbb{F}_2$ (equivalently, the shared nonlinear core vanishes, $q \equiv 0$). For every multilinear polynomial p in the output variables with $\deg p \leq k$ there exists a multilinear polynomial q in the input variables such that*

$$p(y) = q(x) \quad \text{and} \quad \deg q \leq k.$$

Moreover, for the supremum norm on $\{0, 1\}^n$ one has $\|q\|_\infty \leq \|p\|_\infty$.

Proof. Write p as a sum of monomials in the coordinates of y , then substitute $y = Ax \oplus r$. Since G applies affine (degree-1) maps per output, composition does not increase degree: every monomial of degree $\leq k$ in y becomes a polynomial of degree $\leq k$ in x . Since $q = p \circ G$ and G is deterministic, the image of G is a subset of the output domain; therefore $\|q\|_\infty \leq \|p\|_\infty$. $\square$

Corollary 6.8 (Monotonicity of Δ_k under affine blocks). *Let μ, u be distributions on inputs and let G be as above. Set $\mu_{\text{out}} := G_{\#}\mu$ and $u_{\text{out}} := G_{\#}u$. Then for all $k \in \mathbb{N}$,*

$$\Delta_k(\mu_{\text{out}}, u_{\text{out}}) \leq \Delta_k(\mu, u).$$

Proof. Let p be a degree- $\leq k$ test on outputs with $\|p\|_\infty \leq 1$. By Prop. 6.7 there is q on inputs with $\deg q \leq k$, $\|q\|_\infty \leq 1$, and $p \circ G = q$. Hence

$$|\mathbb{E}_{\mu_{\text{out}}}[p] - \mathbb{E}_{u_{\text{out}}}[p]| = |\mathbb{E}_\mu[q] - \mathbb{E}_u[q]| \leq \Delta_k(\mu, u).$$

Taking the supremum over all p gives the claim. $\square$

Theorem 6.9 (Gadget composition stability). *For any constant-size gadget G satisfying Lemmas 6.1, 6.2 and 6.16 on Win,*

$$V_C(\mu \circ G, \varepsilon + o(1)) \geq V_C(\mu, \varepsilon) - O(\log N).$$

Proof. By Corollary 6.11 the gadget layer admits shielding on Win with some $\rho < 1$. Applying Proposition 6.17 (proved via the window simulation / partial inverse of Appendix C, Lemma C.3) with $k = \Theta(\log n)$ yields

$$V_C(\mu \circ G, \varepsilon + o(1)) \geq V_C(\mu, \varepsilon) - O(\log n).$$

The prefix-free size bookkeeping contributes only $O(\log N)$ meta-overhead. This is exactly the claim. $\square$

Scope note (literature). Our gadget composition keeps the switching path p (depth $d = O(\log n)$) and bottom width $W = \text{poly}(1/p)$ in the Håstad switching-lemma regime [4, 6]. The Index gadget’s degree inflation is a constant factor (here ≤ 2 ; see Lemma 6.4), and correlation stability follows from the mixing witness with rate $\theta < 1$ (Definition 6.2 and Lemma 6.14). The proof style—low-degree Fourier control plus width bounds—follows the standard SoS/CSP toolkit [1], specialized to the calibrated window with quantified TV loss and transparent size growth. For calibration and notation (the window Win , baseline, and Δ_k) we follow IECZ-I [11] verbatim.

Remark 6.2 (Scope note: circuit classes). The constant-size *Index* gadget is used here to monitor width/correlation stability along the switching path p (calibrated for $d = O(\log n)$) in a windowed, low-degree setting. Our results concern the $O(\log n)$ meta-overhead preservation (Theorem 6.9); we do not target separations for uniform AC^0 or ACC^0 .

Correlation shielding via low-degree noise stability

We formalize the “ $\rho < 1$ ” stability assumption by an L_2 -type correlation bound.

Definition 6.1 (Windowed correlation). For random variables X, Y defined on the window law, set

$$\text{corr}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \quad \text{when well-defined.}$$

A gadget G has *shielding parameter* $\rho < 1$ if for each output s_ℓ and any exogenous σ -algebra $\mathcal{Z}$ (independent outside the local fan-in) we have $|\text{corr}(s_\ell, \mathbb{E}[\cdot \mid \mathcal{Z}])| \leq \rho + o(1)$ on Win .

Definition 6.2 (Mixing witness). A gadget G admits a *mixing witness* with rate $\theta \in (0, 1)$ on the window Win if there exists a linear operator $\mathcal{M}$ on degree- ≤ 1 statistics of the outputs such that:

1. $\mathcal{M}$ preserves means and contracts centered coordinates: for every centered degree-1 statistic q , $\|\mathcal{M}q\|_2 \leq \theta \|q\|_2$;
2. $\mathcal{M}$ acts trivially on exogenous σ -algebras outside the local fan-in (compatibility with the window).

Lemma 6.10 (Explicit mixing witness for XOR/Index). *Consider a single gadget block with inputs $x \in \{0, 1\}^s$ and outputs $s \in \{0, 1\}^r$ ($s, r = O(1)$). By Lemma 6.2, after an invertible output mix we may write $s = (t_1, t_2)$ with t_1 affine and $t_2 = (\text{affine}) \oplus q$, where q is the shared non-linear core. Denote the affine part by $y_{\text{aff}} := Ax \oplus r$ over $\mathbb{F}_2$, with $\text{rank}(A) = 2$ and $\mathbf{1}^\top A = (1, 1, 1)$, $\mathbf{1}^\top r = 0$. Let L_{out} be the space of centered degree-1 (linear) statistics in the output coordinates, equipped with the L_2 norm under the window proxy u . Define the operator*

$$\mathcal{M} := \Pi_{\text{out}} \circ T_\theta \circ \mathcal{L},$$

where $\mathcal{L}$ is the degree-1 pullback (compose with $y_{\text{aff}} = Ax \oplus r$ and keep only the linear slice), T_θ is the Bonami–Beckner noise on inputs with parameter $\theta \in (0, 1)$ acting independently per block, and Π_{out} is the orthogonal projection (w.r.t. u) onto L_{out} . Then there exists a constant $c_G < \infty$ (depending only on the gadget’s constant fan-in/out and the window calibration) such that for every centered $q \in L_{\text{out}}$,

$$\|\mathcal{M}q\|_2 \leq c_G \theta \|q\|_2.$$

In particular, picking $\theta \in (0, 1)$ small enough that $c_G \theta < 1$ yields a mixing witness on the block with rate $\bar{\theta} := c_G \theta \in (0, 1)$.

Proof. Work on the calibrated window Win . By Lemma 6.2, after an invertible output mixing we may write the two outputs as $s = (t_1, t_2)$ with t_1 affine and $t_2 = (\text{affine}) \oplus q$, where q is the shared nonlinear core. Let L_{out} denote the subspace of centered degree-1 output statistics (with inner product and L_2 -norm taken under the window proxy u).

Define three linear maps:

$$\mathcal{L}: L_{\text{out}} \rightarrow L_{\text{in}}, \quad T_\theta: L_{\text{in}} \rightarrow L_{\text{in}}, \quad \Pi_{\text{out}}: L_{\text{in}} \rightarrow L_{\text{out}},$$

as follows.

(1) *Degree-1 pullback $\mathcal{L}$.* Write the affine part of the interface as $y_{\text{aff}} = Ax \oplus r$ over $\mathbb{F}_2$ with $\text{rank}(A) = 2$, $\mathbf{1}^\top A = (1, 1, 1)$, $\mathbf{1}^\top r = 0$ (Lemma 6.2). For $q \in L_{\text{out}}$, compose q with y_{aff} and project to the degree-1 slice on inputs; denote this operation by $\mathcal{L}q$. Because the gadget has constant fan-in/out and Win fixes the local reweightings, the induced operator norm is bounded:

$$\|\mathcal{L}q\|_2 \leq c_{\text{aff}} \|q\|_2 \quad \text{for some absolute } c_{\text{aff}} < \infty.$$

(Here c_{aff} depends only on the constant local wiring and the window calibration, not on the ambient dimension.)

(2) *Noise operator T_θ .* Let T_θ be the Bonami–Beckner noise on inputs with parameter $\theta \in (0, 1)$ acting independently per block. On the degree-1 slice (linear forms), T_θ is an exact contraction by θ :

$$\|T_\theta z\|_2 = \theta \|z\|_2 \quad \text{for all } z \in L_{\text{in}} \text{ of degree 1.}$$

(3) *Output projection Π_{out} .* Let Π_{out} be the orthogonal projection (with respect to the $L_2(u)$ inner product) from input-linear statistics back to the subspace L_{out} spanned by the output-linear coordinates. Projections are contractions, hence

$$\|\Pi_{\text{out}} z\|_2 \leq \|z\|_2 \quad \text{for all } z.$$

Now set

$$\mathcal{M} := \Pi_{\text{out}} \circ T_\theta \circ \mathcal{L} : L_{\text{out}} \rightarrow L_{\text{out}}.$$

For any centered $q \in L_{\text{out}}$,

$$\|\mathcal{M}q\|_2 \leq \|\Pi_{\text{out}}\|_{2 \rightarrow 2} \|T_\theta\|_{2 \rightarrow 2} \|\mathcal{L}\|_{2 \rightarrow 2} \|q\|_2 \leq (1) \cdot \theta \cdot c_{\text{aff}} \|q\|_2.$$

Thus $\|\mathcal{M}\|_{2 \rightarrow 2} \leq c_{\text{aff}} \theta$. Let $c_G := c_{\text{aff}}$. Choosing any $\theta \in (0, 1)$ with $c_G \theta < 1$ makes $\mathcal{M}$ a strict contraction on L_{out} with rate $\bar{\theta} := c_G \theta \in (0, 1)$. By definition, this is a mixing witness on the block with the stated rate. $\square$

Corollary 6.11 (Shielding for XOR/Index). *Under the calibrated window and width bounds of this section, the XOR/Index gadget layer admits a mixing witness with some rate $\bar{\theta} \in (0, 1)$ by Lemma 6.10. Therefore shielding holds with parameter $\rho = \bar{\theta} + o(1)$ by Lemma 6.14.*

Lemma 6.12 (Shielding decay forms). *Let Win be the calibrated window and suppose the gadget admits a mixing witness with rate $\theta \in (0, 1)$ in the sense of Definition 6.2. Then there exists $\rho \in (0, 1)$ (depending only on θ and the bounded width on Win) such that the following equivalent shielding conditions hold:*

- (i) *Separation by a cut. For disjoint variable sets A, B separated by a cut S in the gadget interaction graph on Win ,*

$$|\text{Cov}_\mu(f(X_A), g(X_B))| \leq \rho^{|S|} \|f\|_2 \|g\|_2$$

for all square-integrable f, g depending on X_A, X_B respectively (expectation and norm under the window law).

- (ii) Low-degree decay. For every Walsh character χ_T with $T \subseteq A \cup B$ and every split $T = T_A \cup T_B$ with $T_A \subseteq A$, $T_B \subseteq B$,

$$|\mathbb{E}_\mu[\chi_{T_A} \chi_{T_B}] - \mathbb{E}_\mu[\chi_{T_A}] \mathbb{E}_\mu[\chi_{T_B}]| \leq \rho^{\text{dist}(A,B)}.$$

In particular, shielding implies stability of all low-degree correlations across the cut and hence windowed orthogonality up to an exponentially small error.

Corollary 6.13 (Stable correlations under composition). *Under Lemma 6.12, composing a bounded-width gadget layer preserves correlation stability: if each layer has width $W = \text{poly}(1/p)$ along the blueprint switching path and admits a mixing witness with rate $\theta < 1$, then the composed map has shielding parameter $\rho' < 1$ and inherits low-degree correlation bounds with the same exponential decay (constants may change).*

Lemma 6.14 (Mixing witness $\Rightarrow$ shielding). *Let G be a constant-size gadget on the calibrated window Win and suppose G admits a mixing witness with rate $\theta \in (0, 1)$ in the sense of Definition 6.2. Then G has shielding parameter $\rho < 1$ with*

$$\rho = \theta + o(1).$$

Proof. Let L_{out} be the space of centered degree-1 output statistics (under the window proxy u). By Definition 6.2 there exists a linear map $\mathcal{M}: L_{\text{out}} \rightarrow L_{\text{out}}$ with $\|\mathcal{M}\|_{2 \rightarrow 2} \leq \theta$ which acts trivially on exogenous σ -algebras outside the local fan-in. For any centered $q \in L_{\text{out}}$ and any exogenous random variable Z (compatible with the window) we have

$$|\text{Cov}(q, Z)| = |\text{Cov}(\mathcal{M}q, Z)| + o(1) \leq \|\mathcal{M}q\|_2 \|Z\|_2 + o(1) \leq (\theta + o(1)) \|q\|_2 \|Z\|_2,$$

so shielding holds with parameter $\rho = \theta + o(1)$. For higher-degree tests reduce to the degree-1 slice via noise smoothing (Lemma 6.15) and Cauchy–Schwarz; the $o(1)$ terms are controlled by the window calibration and constant locality. $\square$

Lemma 6.15 (Noise smoothing of low degree). *Let p be a polynomial of degree $\leq k$ in the inputs on Win . For any $\theta \in (0, 1)$,*

$$\|p - T_\theta p\|_2 \leq \sqrt{1 - \theta^{2k}} \|p\|_2.$$

Proof. By orthogonal decomposition into Fourier/Walsh characters, T_θ multiplies the degree- d slice by θ^d . Hence $\|p - T_\theta p\|_2^2 = \sum_{d \leq k} (1 - \theta^d)^2 \|p^{(d)}\|_2^2 \leq (1 - \theta^{2k}) \sum_{d \leq k} \|p^{(d)}\|_2^2$. $\square$

Lemma 6.16 (Correlation shielding). *Assume G is parity-preserving in the interface (post-mix) normal form of Lemma 6.2 and has shielding parameter $\rho < 1$ on Win . Let p be any degree- $\leq k$ statistic of the outputs. Then for some $\theta = \theta(\rho) \in (0, 1)$,*

$$|\text{Cov}(p, \text{noise})| \leq \sqrt{1 - \theta^{2k}} \|p\|_2 \|\text{noise}\|_2 + o(1),$$

for every exogenous noise source noise compatible with the window (independent outside fan-in). In particular, taking $k = O(\log n)$ and fixed $\rho < 1$, the covariance is $o(1) \cdot \|p\|_2 \|\text{noise}\|_2$.

Proof. Work on the window Win . Let $x = (x_1, \dots, x_m)$ be the input bits and let $y = G(x)$ be the (deterministic) gadget outputs applied coordinate-wise. Fix $k \leq c \log n$ as in the blueprint and let $p(y)$ be any degree- $\leq k$ polynomial test with $\|p\|_\infty \leq 1$ on the output coordinates. We show that for every “exogenous” random variable Z satisfying the section’s balancing/moment-matching assumptions one has

$$|\text{Cov}(p(y), Z)| = o(1),$$

uniformly over the family, which is the stated shielding.

Step 1: Interface normal form; degree bookkeeping. By Lemma 6.2, after an invertible output mix we can write the two outputs as (t_1, t_2) with t_1 affine and $t_2 = (\text{affine}) \oplus q$, where q is the common non-linear core. Write the affine part as $y_{\text{aff}} = Ax \oplus r_0$ with $\mathbf{1}^\top A = (1, 1, 1)$ and $\mathbf{1}^\top r_0 = 0$. Because each output depends on $O(1)$ input variables (constant fan-in), composing a degree- $\leq k$ multilinear test p with the gadget expands degree by at most a constant factor; in particular $p(G(x))$ depends on at most $O(k)$ input blocks. If needed (for the Index block), the constant factor 2 from Lemma 6.4 is absorbed in the budget $k = \Theta(\log n)$.

Step 2: Mixing/Noise dampens high degree. Let T_θ be the Bonami–Beckner noise operator on the input with parameter $\theta \in (0, 1)$ fixed as in the switching regime. For a degree- $\leq k$ polynomial q (in the $\{\pm 1\}$ -Walsh basis) one has

$$\|q - T_\theta q\|_2 \leq \sqrt{1 - \theta^{2k}} \|q\|_2,$$

and T_θ preserves expectations under the proxy. Applying this to $p \circ G$ and accounting for the constant degree inflation under G (so $\deg(p \circ G) \leq ck$) gives

$$\|p(y) - T_\theta(p(y))\|_2 \leq \sqrt{1 - \theta^{2ck}},$$

for some absolute constant $c \geq 1$ depending only on the gadget’s local fan-in/out ($c = 1$ when the nonlinear core vanishes, $q \equiv 0$; for the Index block $c \leq 2$, see Lemma 6.4). Equivalently, setting $\theta' := \theta^c$ yields

$$\|p(y) - T_{\theta'}(p(y))\|_2 \leq \sqrt{1 - (\theta')^{2k}}.$$

Step 3: Decoupling from exogenous structure. By the balancing and moment-matching conditions stated in the assumptions of this section, Z is calibrated to the proxy so that Z is uncorrelated with *any* function depending on at most $t = O(1)$ disjoint input blocks after noise T_θ and coarse conditioning (the “path” $\mathbf{p}$). Because p touches only $O(k)$ blocks and $k = O(\log n)$, standard block-decoupling plus the contraction of T_θ give

$$|\text{Cov}(T_\theta(p(y)), Z)| = o(1).$$

Combining with Step 2 and Cauchy–Schwarz,

$$|\text{Cov}(p(y), Z)| \leq |\text{Cov}(T_\theta(p(y)), Z)| + \|p(y) - T_\theta(p(y))\|_2 \|Z\|_2 = o(1) + \sqrt{1 - \theta^{2k}} \cdot O(1).$$

Step 4: Choice of parameters. With $k \leq c \log n$ and fixed $\theta \in (0, 1)$, we have $\sqrt{1 - \theta^{2k}} \leq \exp(-\Omega(k)) = n^{-\Omega(1)}$. Thus the right-hand side is $o(1)$ uniformly over the family, yielding the claim. $\square$

Proposition 6.17 (Cost stability under shielding). *Under Lemma 6.16 with $k = O(\log n)$,*

$$V_C(\mu \circ G, \varepsilon + o(1)) \geq V_C(\mu, \varepsilon) - O(\log N),$$

i.e., Theorem 6.9 holds with an explicit $o(1)$ driven by the shielding tail $\sqrt{1 - \theta^{2k}} = o(1)$ and the prefix-free $O(\log n)$ overhead.

Proof (via window simulation / partial inverse). Fix $\delta > 0$ and let $(M_{\text{out}}, R_{\text{out}})$ be δ -optimal for $V_C(\mu \circ G, \varepsilon + o(1))$; that is,

$$\mathbb{E}_{y \sim (\mu \circ G)}[L_{M_{\text{out}}}(R_{\text{out}}(y), \varepsilon + o(1))] \leq V_C(\mu \circ G, \varepsilon + o(1)) + \delta.$$

Table 5: Gadget composition losses

Gadget	TV loss	Size blowup	Error	Notes
G_{xor}	$o(1)$	$\times(1+o(1))$	$\varepsilon \mapsto \varepsilon+o(1)$	Width $W = \text{poly}(1/p)$; affine normal form; shielding via mixing witness (Def. 6.2, Lem. 6.14)
G_{index}	$o(1)$	$\times(1+o(1))$	$\varepsilon \mapsto \varepsilon+o(1)$	Index amplification; degree budget uses $k \mapsto 2k$ (absorbed for $k = \Theta(\log n)$); stable correlations (Def. 6.2)

By the window simulation / partial inverse (Appendix C, Lemma C.3), there exists a polynomial-time simulator S and a selector $u \in \{0, 1\}^{O(\log N)}$ such that, on the window Win , every output instance y admits a preimage $x = S(y, u)$ with total variation error $o(1)$, and $G(x) = y$ with probability $1 - o(1)$ under the coupling.

Define the input-side pair

$$M_{\text{in}} := M_{\text{out}}, \quad R_{\text{in}}(x) := R_{\text{out}}(G(x)).$$

Encoding the selector u prefix-free costs $O(\log N)$ bits (Lemma 4.1) and is the only extra model/description overhead.

By change of variables on the pushforward and the coupling supplied by the simulator,

$$\mathbb{E}_{x \sim \mu}[L_{M_{\text{in}}}(R_{\text{in}}(x), \varepsilon)] = \mathbb{E}_{y \sim (\mu \circ G)}[L_{M_{\text{out}}}(R_{\text{out}}(y), \varepsilon)] \pm o(1).$$

The loss is 1-Lipschitz under the $o(1)$ perturbations coming from shielding (Lemma 6.16) and the simulator's TV error, so the displayed equality holds up to $o(1)$. Incorporating the prefix-free code for u adds $O(\log N)$ to the description length, hence to the value.

Therefore,

$$V_{\mathcal{C}}(\mu, \varepsilon) \leq \mathbb{E}_{x \sim \mu}[L_{M_{\text{in}}}(R_{\text{in}}(x), \varepsilon)] + O(\log N) \leq V_{\mathcal{C}}(\mu \circ G, \varepsilon + o(1)) + O(\log N) + \delta + o(1).$$

Since $\delta > 0$ was arbitrary, this rearranges to

$$V_{\mathcal{C}}(\mu \circ G, \varepsilon + o(1)) \geq V_{\mathcal{C}}(\mu, \varepsilon) - O(\log N),$$

which is the claim. $\square$

Corollary 6.18 (Explicit bounds along the path). *Let $d = \lceil \gamma \log n \rceil$ and fix $c \in (0, 1)$. Along the path $p = m^{-c/d}$ one has*

$$M_d \leq m^{1-c}, \quad W_d \leq m^{\kappa c}$$

for some absolute constant $\kappa > 0$ (absorbing gadget constants). In particular, for $d = \Theta(\log n)$ and constant c , the width is polynomially bounded in m (hence in n).

Proof of Corollary 6.18. Apply Lemma 6.1 with the fixed seed and iterate the recurrences $W_r \leq W_{r-1}/p$ and $M_r \leq p M_{r-1}$ for $r = 1, \dots, d$. With the path choice $p = m^{-c/d}$ one gets

$$W_d \leq W_0 p^{-d} = W_0 m^c = m^{\kappa c} \quad (\text{absorbing the constant } W_0 = O(1) \text{ into } \kappa),$$

and

$$M_d \leq m p^d = m^{1-c}.$$

This yields the claim. $\square$

Proof of Lemma 6.3. Work over $\mathbb{F}_2$. In the affine case of Lemma 6.2 (i.e., when the nonlinear core $q \equiv 0$), we have $s = Ax \oplus r$ with $\mathbf{1}^\top A = (1, 1, 1)$, $\mathbf{1}^\top r = 0$ and $\text{rank}(A) = 2$. The affine offset r can be eliminated by an output relabeling (absorbed into U), so take $r = 0$. The constraint $\mathbf{1}^\top A = (1, 1, 1)$ says each column of A has odd weight. With two output rows and $\text{rank}(A) = 2$, each column must be either $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ or $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ or $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Moreover, not all three columns can be identical, otherwise $\text{rank}(A) < 2$.

Post-multiplying by $U \in \text{GL}_2(\mathbb{F}_2)$ permutes the two output rows and allows adding one row to the other. Up to this action, one may normalize the first row to be $e_i^\top$ for some input index i , forcing $s_1 = x_i$. The parity constraint $\mathbf{1}^\top s = \mathbf{1}^\top x$ then implies the second row must sum, together with the first, to $(1, 1, 1)$. Enumerating the three placements of e_i yields exactly the three matrices A_1, A_2, A_3 above (cyclic rotations of inputs). Each has rank 2 and satisfies $\mathbf{1}^\top A = (1, 1, 1)$. This list is complete because any other admissible A can be reduced to one of these by swapping rows and adding the first row into the second. $\square$

7 SoS Entropy Dichotomy

Standing assumptions. The window Win is fixed with parameters inherited from the reductions/gadgets. Moment matching (pseudoexpectation calibration) holds up to $k \leq c \log n$ (Lemma 7.6). Prefix-free coding conventions apply throughout. Test–moment equivalence and the degree–entropy translation hold on Win (Lemmas 7.4, 7.8). Throughout this section we also assume the bias/product proxy conditions stated in Assumptions 7. Calibration, proxy construction, and window notation are carried over verbatim from IECZ-I [11].

Definition of the low-degree discrepancy Δ_k

Let μ be the windowed target distribution and u a structured proxy (product-like, calibrated on Win).

Definition 7.1 (Low-degree test discrepancy). For $k \in \mathbb{N}$,

$$\Delta_k(\mu, u) := \sup \left\{ |\mathbb{E}_\mu[p] - \mathbb{E}_u[p]| : p \in \mathcal{P}_{\leq k}, \|p\|_\infty \leq 1 \right\}.$$

Equivalently, Δ_k is the optimal distinguishing advantage of degree- $\leq k$ polynomial tests on Win .

Definition 7.2 (SoS pseudoexpectation deviation). Let $\tilde{\mathbb{E}}$ be a degree- k SoS pseudoexpectation calibrated to u . Define

$$\Delta_k^{\text{SoS}}(\mu, u) := \sup \left\{ |\mathbb{E}_\mu[p] - \tilde{\mathbb{E}}[p]| : p \in \mathcal{P}_{\leq k}, \|p\|_\infty \leq 1 \right\}.$$

The calibration of pseudoexpectations to the proxy u follows the identical moment-matching setup from IECZ-I [11].

Corollary 7.1 (Equivalence of test and SoS discrepancies). *If a degree- k pseudoexpectation $\tilde{\mathbb{E}}$ is calibrated to u (Lemma 7.6), then $\Delta_k^{\text{SoS}}(\mu, u) = \Delta_k(\mu, u)$.*

Proof. For any $\deg p \leq k$, calibration gives $\tilde{\mathbb{E}}[p] = \mathbb{E}_u[p]$ (Lemma 7.6), hence $|\mathbb{E}_\mu[p] - \tilde{\mathbb{E}}[p]| = |\mathbb{E}_\mu[p] - \mathbb{E}_u[p]|$. Taking the supremum over such p yields $\Delta_k^{\text{SoS}} = \Delta_k$. $\square$

Remark 7.1 (Calibration window). All expectations/pseudoexpectations are taken on the image window Win . Boundary effects are $o(1)$. See also standard low-degree and SoS analyses for windowed or product-like settings [12, 1] and the identical window discipline in IECZ-I [11].

Remark 7.2 (From windowed discrepancy to lifted families). Our SoS analysis is performed on the calibrated window after the gadget layer; boundary contributions are $o(1)$ along the calibrated path. A sequel (IECZ-III) lifts these distributional statements via in-window aggregation, TV-stable hardcore selection, and seed-indexed re-encoding, with TV-, size-, and meta-overheads tracked. The outcome is a parameterized distributional family that preserves the low-degree–SoS dichotomy at calibrated parameters [9].

Techniques. The test–moment equivalence and noise smoothing are standard in the low-degree/SoS toolkit (see [12, 1]); here they are instantiated under the window discipline with explicit degree/entropy constants and TV-aware size accounting.

Lemma 7.2 (Bias-orthonormal Walsh basis on the window). *Let u be the proxy distribution used on the window Win , assumed product-like with single-bit biases $\Pr_u[x_i = 1] = p_i \in (0, 1)$ and independent coordinates. Define centered/normalized linear coordinates*

$$z_i(x) := \frac{x_i - \mathbb{E}_u[x_i]}{\sqrt{\text{Var}_u(x_i)}} = \frac{x_i - p_i}{\sqrt{p_i(1-p_i)}} \quad (i \in [n]).$$

For any $S \subseteq [n]$ set the (biased) Walsh character $\chi_S(x) := \prod_{i \in S} z_i(x)$ (with $\chi_\emptyset \equiv 1$). Then $\{\chi_S\}_{S \subseteq [n]}$ is an orthonormal system in $L^2(u)$, i.e.

$$\mathbb{E}_u[\chi_S \chi_T] = \mathbf{1}[S = T] \quad \text{for all } S, T \subseteq [n],$$

and $\deg(\chi_S) = |S|$.

Proof. Independence under u gives $\mathbb{E}_u[z_i] = 0$ and $\mathbb{E}_u[z_i^2] = 1$ for all i , and for $i \neq j$, $\mathbb{E}_u[z_i z_j] = \mathbb{E}_u[z_i] \mathbb{E}_u[z_j] = 0$. Hence $\mathbb{E}_u[\chi_S \chi_T] = \prod_{i \in S \Delta T} \mathbb{E}_u[z_i] \cdot \prod_{i \in S \cap T} \mathbb{E}_u[z_i^2] = \mathbf{1}[S = T]$. Degree equals $|S|$ by construction. $\square$

Corollary 7.3 (Low-degree expansion and Parseval). *Every polynomial $p : \{0, 1\}^n \rightarrow \mathbb{R}$ of degree $\leq k$ admits a unique expansion $p = \sum_{|S| \leq k} \hat{p}(S) \chi_S$ with respect to the basis of Lemma 7.2, and $\|p\|_{2,u}^2 = \sum_{|S| \leq k} \hat{p}(S)^2$. In particular, $\mathcal{P}_{\leq k}$ denotes the class of such polynomials of degree $\leq k$.*

Definition 7.3 (Low-degree test discrepancy (windowed restatement)). For $k \in \mathbb{N}$ and distributions μ, u on Win , define

$$\Delta_k(\mu, u) := \sup \left\{ |\mathbb{E}_\mu[p] - \mathbb{E}_u[p]| : p \in \mathcal{P}_{\leq k}, \|p\|_\infty \leq 1 \right\}.$$

This restates Definition 7.1 restricted to the calibrated window Win .

This is the same low-degree discrepancy used in IECZ-I, read on the calibrated window [11].

Lemma 7.4 (Test–moment equivalence (windowed)). *Let $k \in \mathbb{N}$. Expand $p = \sum_{|S| \leq k} \hat{p}(S) \chi_S$ in the basis of Lemma 7.2. Write the low-degree signed spectrum*

$$\hat{\Delta}(S) := \mathbb{E}_\mu[\chi_S] - \mathbb{E}_u[\chi_S], \quad |S| \leq k.$$

Then

$$\Delta_k(\mu, u) = \sup_{\deg \leq k, \|p\|_\infty \leq 1} |\mathbb{E}_\mu[p] - \mathbb{E}_u[p]| \asymp \left(\sum_{|S| \leq k} \hat{\Delta}(S)^2 \right)^{1/2},$$

with absolute two-sided constants.

Corollary 7.5 (Existence of a near-optimal bounded witness). *Under the conditions of Lemma 7.4, there exists $p^* \in \mathcal{P}_{\leq k}$ with $\|p^*\|_\infty \leq 1$ such that*

$$\mathbb{E}_\mu[p^*] - \mathbb{E}_u[p^*] \geq c \left(\sum_{|S| \leq k} \hat{\Delta}(S)^2 \right)^{1/2}$$

for a universal constant $c > 0$.

Proof. Let $\hat{\Delta}(S)$ be as in Lemma 7.4. Take $p_0 = \sum_{|S| \leq k} \text{sgn}(\hat{\Delta}(S)) \chi_S$ and truncate/rescale to enforce $\|p^*\|_\infty \leq 1$ on the finite domain. Then $\mathbb{E}_\mu[p^*] - \mathbb{E}_u[p^*] \geq c (\sum_{|S| \leq k} \hat{\Delta}(S)^2)^{1/2}$ for a universal $c > 0$ by Cauchy–Schwarz and Parseval, as standard. $\square$

Proof of Lemma 7.4. By Corollary 7.3, for any $p = \sum_{|S| \leq k} \hat{p}(S) \chi_S$,

$$\mathbb{E}_\mu[p] - \mathbb{E}_u[p] = \sum_{|S| \leq k} \hat{p}(S) \hat{\Delta}(S) \leq \|p\|_{2,u} \left(\sum_{|S| \leq k} \hat{\Delta}(S)^2 \right)^{1/2}.$$

By Cauchy–Schwarz under u we used

$$\langle \hat{p}, \hat{\Delta} \rangle \leq \|p\|_{2,u} \left(\sum_{|S| \leq k} \hat{\Delta}(S)^2 \right)^{1/2}.$$

Since $\|p\|_\infty \leq 1$ on a finite domain implies $\|p\|_{2,u} \leq 1$, this gives the $\leq$ direction. For the $\geq$ direction, take p^* with coefficients $\hat{p}^*(S) = \text{sgn}(\hat{\Delta}(S))$ truncated/rescaled to enforce $\|p^*\|_\infty \leq 1$; by windowed hypercontractivity (Lemma 7.7) this yields the lower bound up to a universal constant. $\square$

Lemma 7.6 (Pseudoexpectation alignment). *Let $\tilde{\mathbb{E}}$ be any degree- k SoS pseudoexpectation calibrated to u on Win (i.e., matching all degree- $\leq k$ moments and support constraints). Then for every $p \in \mathcal{P}_{\leq k}$,*

$$\tilde{\mathbb{E}}[p] = \mathbb{E}_u[p].$$

Assumptions for windowed hypercontractivity We work with the proxy distribution u on the window Win satisfying:

1. (*Bias window*) For all coordinates i , $p_i := \Pr_u[x_i=1] \in [\alpha, 1 - \alpha]$ for some absolute $\alpha \in (0, \frac{1}{2})$.
2. (*Product structure*) Coordinates are independent under u (after gadget mixing/shielding; cf. Sec. 6).
3. (*Gadget-calibrated degree budget*) Low-degree statements are read *after* the affine/index gadget layer; affine blocks preserve degree and the Index gadget inflates degree by at most a factor 2 (Prop. 6.7, Lem. 6.4).
4. (*Width control along the switching path*) Along $p = m^{-c/d}$ with $d = \Theta(\log n)$, the bottom width obeys $W = \text{poly}(\frac{1}{p})$ on Win (Sec. 6). This hypothesis is only asserted for the log-degree regime; outside it, the induced dependence can make W large, and the absolute hypercontractive constants (C, τ) below are not guaranteed. These assumptions (bias window, product proxy, gadget-aware degree budget, and switching-path width) coincide with those fixed in IECZ-I [11].

Hypercontractive constants and dependencies. The hypercontractive pair (C, τ) in Lemma 7.7 is determined by the standing assumptions: C worsens as the bias lower bound α decreases and as the bottom width W along the switching path grows, while τ is controlled by the mixing/shielding parameters and the product structure on the window. Concretely, (C, τ) follow from the Bonami–Beckner analysis on Win and are absolute throughout the log-degree regime $d = \Theta(\log n)$ and $k \leq c \log n$; see Proposition A.7 for the derivation.

Remark 7.3 (Constants and regime). Lemma 7.7 uses absolute hypercontractive constants (C, τ) that are uniform for $d = \Theta(\log n)$ along the calibrated path. Symbols such as C_d and τ_d may be used to emphasize dependence on the fixed depth d ; they do not depend on n or m . We do not optimize numerical values here—concrete estimates follow from standard hypercontractive inequalities and can be extracted if needed. For superlogarithmic depths, the same calibration would force smaller p and larger width W , so (C, τ) may deteriorate; we make no claim outside the log-degree regime.

Lemma 7.7 (Windowed hypercontractivity (under Assumptions Sec. 7)). *There exist absolute constants $\tau \in (0, 1)$ and $C \geq 1$ depending only on $(\alpha, \text{the mixing/shielding parameters, and the switching-width bound})$, cf. Prop. A.7, such that for every degree- d polynomial q (expanded in the basis of Lemma 7.2)*

$$\|q\|_{4,u} \leq C \tau^{-d} \|q\|_{2,u} \quad \text{and} \quad \|T_\theta q\|_{2,u} \leq \|q\|_{2,u} \quad (\theta \in [0, 1]),$$

where T_θ is the Bonami–Beckner noise operator acting diagonally by $T_\theta \chi_S = \theta^{|S|} \chi_S$. In particular, for any signed measure Δ supported on Win and any k ,

$$|\mathbb{E}_\Delta[p]| \leq C \tau^{-k} \left(\sum_{|S| \leq k} \hat{\Delta}(S)^2 \right)^{1/2} \quad \text{for all } p \in \mathcal{P}_{\leq k}, \quad \|p\|_\infty \leq 1.$$

Related work. Our use of windowed hypercontractivity and low-degree Fourier methods aligns with the approximate-degree and entropy techniques in Boolean analysis; see, e.g., [12, 13].

Theorem 7.8 (Discrepancy forces SoS degree). *There exist absolute constants $\alpha, \beta > 0$ such that: if for some k we have $\Delta_k(\mu, u) \geq n^{-\beta}$, then any polynomial-calculus or SoS refutation separating μ from u on the window requires degree*

$$d \geq \alpha k.$$

Proof. Assume a degree- d refutation exists with $d < \alpha k$ (for a small constant $\alpha > 0$ to be fixed). By SoS duality there is a degree- d pseudoexpectation $\tilde{\mathbb{E}}$ matching all degree- $\leq d$ moments of u on Win. Pick $p \in \mathcal{P}_{\leq k}$ with $\|p\|_\infty \leq 1$ achieving $\Delta_k(\mu, u)$ up to a universal constant (Lemma 7.4), and normalize so that $\mathbb{E}_u[p] = 0$. Then $\tilde{\mathbb{E}}[p] = 0$ by Lemma 7.6. Let $\Delta := \mu - u$ and write $p = \sum_{j \leq k} p^{(j)}$ by homogeneous degree. By Parseval under u , $\|p\|_{2,u}^2 = \sum_{|S| \leq k} \hat{p}(S)^2 \leq 1$ since $\|p\|_\infty \leq 1$ on a finite domain. Windowed hypercontractivity (Lemma 7.7) gives for the high-degree tail

$$|\mathbb{E}_\Delta[p^{(>d)}]| \leq C \tau^{-(k-d)} \left(\sum_{|S| \leq k} \hat{\Delta}(S)^2 \right)^{1/2} \asymp C \tau^{-(k-d)} \Delta_k(\mu, u),$$

while the low-degree contribution vanishes by moment matching up to d : $\mathbb{E}_\Delta[p^{(\leq d)}] = \mathbb{E}_\mu[p^{(\leq d)}] - \mathbb{E}_u[p^{(\leq d)}] = 0$. Therefore

$$\mathbb{E}_\Delta[p] = \mathbb{E}_\Delta[p^{(\leq d)}] + \mathbb{E}_\Delta[p^{(>d)}] \leq C \tau^{-(k-d)} \Delta_k(\mu, u).$$

Choose $\alpha > 0$ so that for all large n we have $C\tau^{-(1-\alpha)k} \leq \frac{1}{2}$; then for $d < \alpha k$ $\mathbb{E}_\Delta[p] \leq \frac{1}{2} \Delta_k(\mu, u)$, contradicting the choice of p as an (up to constants) maximizer for $\Delta_k(\mu, u)$. Hence $d \geq \alpha k$. $\square$

Theorem 7.9 (SoS entropy dichotomy). *There exist constants $c, \eta, \zeta > 0$, depending only on the window calibration (bias lower bound α , mixing/shielding parameters, and the switching-path width bound) — equivalently only on the hypercontractive constants (C, τ) from Lemma 7.7 and the degree-translation constant from Theorem 7.8 — such that for $k \leq c \log n$ either*

$$\Delta_k(\mu, u) \rightarrow 0 \quad \text{or} \quad \Delta_k(\mu, u) \geq n^{-\eta},$$

and in the latter case any PC/SoS refutation requires degree $\geq \zeta k$.

Dependency map. The constants obey the chain

$$(C, \tau) \implies c \quad \text{and} \quad (\alpha, \beta) \implies (\eta, \zeta),$$

where $c \leq c_*(C, \tau)$ and (α, β) are the discrepancy-to-degree constants from Theorem 7.8. After normalizing absolute factors, one may take, e.g., $\eta \leq \beta/2$ and $\zeta \leq \alpha/4$.

Corollary 7.10 (Windowed numeric form). *Fix $k = \lfloor c \log n \rfloor$. If $\Delta_k(\mu, u) \geq n^{-\eta}$ with fixed $\eta > 0$, then*

$$\text{degree} \geq \zeta \log n$$

for some $\zeta = \zeta(c, \eta) > 0$.

Proof. Apply Theorem 7.8 with $k = \lfloor c \log n \rfloor$ and absorb constants into ζ . $\square$

Numerical instantiation. The concrete values used later (Section 7.1), namely $c = \frac{1}{4}$, $\eta = \frac{1}{8}$, and $\zeta = \frac{1}{16}$, are conservative choices consistent with the above dependencies; sharper estimates of (C, τ) would permit proportional improvements (larger c and ζ , smaller η).

Near-optimal witness. Let $p_0 = \sum_{|S| \leq k} \text{sgn}(\hat{\Delta}(S)) \chi_S$. By Parseval and Cauchy–Schwarz, $\mathbb{E}_\mu[p_0] - \mathbb{E}_u[p_0] = \sum_{|S| \leq k} |\hat{\Delta}(S)| \asymp (\sum_{|S| \leq k} \hat{\Delta}(S)^2)^{1/2}$. Truncating/rescaling p_0 to enforce $\|p^*\|_\infty \leq 1$ on the finite domain preserves this value up to a universal constant, yielding the $\geq$ -Richtung.

Remark 7.4 (Literature anchor). Our use of low-degree Fourier analysis and hypercontractivity in the discrepancy-to-degree step follows the standard Boolean-analysis toolkit; see, e.g., O’Donnell’s monograph [12] for a unified treatment of these techniques.

Lemma 7.11 (Entropy $\Rightarrow$ degree lower bound). *There exist absolute constants $\alpha, \beta > 0$ such that: if $\Delta_k(\mu, u) \geq n^{-\beta}$, then no PC/SoS refutation exists at degree $d < \alpha k$.*

Degree budget after gadgets. Throughout we interpret low-degree tests after the gadget layer with the calibrated budget: affine gadgets preserve degree (Prop. 6.7); the Index gadget inflates degree by at most a factor 2 (Lemma 6.4). In the regime $k = \Theta(\log n)$ this constant factor is absorbed and all SoS statements below are read with the post-gadget degree budget.

From low-degree discrepancy to SoS degree

We now make precise how a non-negligible low-degree discrepancy forces SoS degree. All statements are windowed (on Win); boundary terms are $o(1)$ and absorbed.

Lemma 7.12 (Discrepancy witnesses via bounded polynomials). *Let $k \in \mathbb{N}$ and suppose $\Delta_k(\mu, u) \geq \delta$. Then there exists a real polynomial $p : \{0, 1\}^n \rightarrow \mathbb{R}$ with $\deg p \leq k$, $\|p\|_\infty \leq 1$, such that*

$$\mathbb{E}_\mu[p] - \mathbb{E}_u[p] \geq \delta.$$

Moreover, after centering p by $\mathbb{E}_u[p]$, we may assume $\mathbb{E}_u[p] = 0$ and $\mathbb{E}_\mu[p] \geq \delta$.

Proof. By definition of $\Delta_k(\mu, u)$ there exists a sequence of degree- $\leq k$ tests $p^{(t)}$ with $\|p^{(t)}\|_\infty \leq 1$ such that $\mathbb{E}_\mu[p^{(t)}] - \mathbb{E}_u[p^{(t)}] \rightarrow \Delta_k(\mu, u) \geq \delta$. Pick t large enough and set $p := p^{(t)}$; if needed, replace p by $p - \mathbb{E}_u[p]$ to enforce $\mathbb{E}_u[p] = 0$ (which preserves degree and $\|p\|_\infty \leq 1$). Then $\mathbb{E}_\mu[p] \geq \delta$, as claimed. $\square$

Proposition 7.13 (Sufficient calibration for degree lower bounds). *Fix $k = \lfloor c \log n \rfloor$ and let u be the proxy distribution on the window Win . Suppose:*

1. (Moment matching) *A degree- k pseudoexpectation $\tilde{\mathbb{E}}$ is calibrated to u , i.e., $\tilde{\mathbb{E}}[\chi_S] = \mathbb{E}_u[\chi_S]$ for all $S \subseteq [n]$ with $|S| \leq k$, and support constraints on Win coincide.*
2. (Width bound) *Along the switching path $p = m^{-c/d}$ with $d = \Theta(\log n)$, the bottom width obeys $W = \text{poly}(1/p)$ (Sec. 6).*
3. (Balance) *Either u is balanced on Win or the Walsh characters are reweighted to an orthonormal basis under u .*
4. (Noise regularity) *The Bonami–Beckner operator T_θ acts diagonally on degree- d slices with eigenvalues θ^d .*

Then $\Delta_k(\mu, u)$ controls SoS degree as in Theorem 7.8: if $\Delta_k(\mu, u) \geq n^{-\eta}$ for some fixed $\eta > 0$, any PC/SoS refutation requires degree $\geq \zeta \log n$ for some $\zeta = \zeta(c, \eta) > 0$.

Proof. Item (1) gives $\tilde{\mathbb{E}}[p] = \mathbb{E}_u[p]$ for $\deg p \leq k$ by Lemma 7.6. Item (3) ensures the Fourier/Walsh expansion under u with orthonormality used in Lemma 7.4 (full proof). Item (2) supplies the switching-path regime from Sec. 6 used to control structural width and boundary effects. Item (4) provides the noise operator needed for windowed hypercontractivity (Lemma 7.7). Combining Lemmas 7.12, 7.6, 7.7 yields Theorem 7.8 at $k = \lfloor c \log n \rfloor$, hence the claimed degree $\geq \zeta \log n$ whenever $\Delta_k(\mu, u) \geq n^{-\eta}$. $\square$

7.1 Concrete parameter choice

We now fix concrete constants for the log-degree regime (see the preceding remark on constants and regime). They depend only on the hypercontractive pair (C, τ) from Lemma 7.7 and on the constants from Theorem 7.8. For definiteness in Theorem C.5 we take

$$c = \frac{1}{4}, \quad \eta = \frac{1}{8}, \quad \zeta = \frac{1}{16}, \quad k = \lfloor c \log n \rfloor.$$

These values satisfy the dependencies in Lemma 7.11 (e.g., one may choose $\zeta \leq \alpha/2$ and $\eta \leq \beta/2$ after normalizing absolute constants). They are representative rather than optimized; any fixed choice meeting the same constraints would serve equally well.

Corollary 7.14 (Numerical form for the blueprint). *Under the above instantiation, on the “structure” branch ($\Delta_k \geq n^{-\eta}$) any PC/SoS refutation on the blueprint family requires*

$$\deg \geq \zeta k \geq \frac{1}{16} \left\lfloor \frac{1}{4} \log n \right\rfloor = \Omega(\log n).$$

Combined with Theorem C.5, this yields an explicit $\Omega(\log n)$ degree lower bound with concrete constants.

Table 6: Implications for proof systems on the window (constants instantiated as in Sec. 7.1: $c = \frac{1}{4}$, $\zeta = \frac{1}{16}$).

System / Measure	TV loss	Size	Degree / Error
PC/SoS(degree)	—	—	degree $\geq \frac{1}{16} \lfloor \frac{1}{4} \log n \rfloor$

A Proofs for Section 7

Deferred proofs and technical lemmas

Remark A.1 (Window scope and boundary bookkeeping). All expectations, pseudoexpectations, and norms are taken on the image window Win . Any off-window boundary contributions are $o(1)$ along the calibrated path and are absorbed in our $o(1)$ terms. This matches the window discipline from Sec. 4.

Lemma A.1 (Biased vs. balanced Walsh bases). *Let u be product-like with bit-biases $p_i \in (0, 1)$ and independent coordinates. Define $z_i = (x_i - p_i)/\sqrt{p_i(1-p_i)}$ and $\chi_S = \prod_{i \in S} z_i$. There exist absolute constants $A_k, B_k = \text{poly}_k(1)$ such that, for every degree- $\leq k$ polynomial p ,*

$$A_k \|p\|_{2, u_{\text{bal}}} \leq \|p\|_{2, u} \leq B_k \|p\|_{2, u_{\text{bal}}}.$$

Here u_{bal} is the balanced product proxy (all $p_i = \frac{1}{2}$). Thus all $k \leq c \log n$ statements transfer between biased and balanced settings with only constant-factor losses absorbed in C, τ of Lemma 7.7.

Proof. Each z_i is an affine rescaling of $(-1)^{x_i}$ with factor $\Theta((p_i(1-p_i))^{-1/2})$. For degree $\leq k$ the monomialwise rescaling causes at most a multiplicative distortion $\prod_{i \in S} (p_i(1-p_i))^{-1/2} \leq (\max_i (p_i(1-p_i))^{-1/2})^k$. Calibration on Win keeps p_i bounded away from $\{0, 1\}$, so the bound is $\text{poly}_k(1)$. $\square$

Lemma A.2 (Moment matrix calibration $\Rightarrow$ alignment). *Let $\tilde{\mathbb{E}}$ be a degree- k SoS pseudoexpectation on Win such that, for every $|S| \leq k$, $\tilde{\mathbb{E}}[\chi_S] = \mathbb{E}_u[\chi_S]$ and the support constraints of u on Win are enforced in the pseudoideal. Then for every $p \in \mathcal{P}_{\leq k}$ one has $\tilde{\mathbb{E}}[p] = \mathbb{E}_u[p]$.*

Proof. Write $p = \sum_{|S| \leq k} \hat{p}(S) \chi_S$ (Lemma 7.2). Linearity gives $\tilde{\mathbb{E}}[p] = \sum_{|S| \leq k} \hat{p}(S) \tilde{\mathbb{E}}[\chi_S] = \sum_{|S| \leq k} \hat{p}(S) \mathbb{E}_u[\chi_S] = \mathbb{E}_u[p]$. $\square$

Lemma A.3 (Noise stability and coefficient dampening). *Let T_θ act by $T_\theta \chi_S = \theta^{|S|} \chi_S$. For any $p = \sum_{|S| \leq k} \hat{p}(S) \chi_S$ and $\theta \in [0, 1]$,*

$$\|T_\theta p\|_{2, u}^2 = \sum_{|S| \leq k} \theta^{2|S|} \hat{p}(S)^2 \leq \|p\|_{2, u}^2, \quad |\mathbb{E}_\mu[T_\theta p] - \mathbb{E}_u[T_\theta p]| \leq \|T_\theta p\|_{2, u} \left(\sum_{|S| \leq k} \hat{\Delta}(S)^2 \right)^{1/2}.$$

Proposition A.4 (Witness extraction with normalization). *If $\Delta_k(\mu, u) = \delta > 0$, there exists $p^* \in \mathcal{P}_{\leq k}$ with $\|p^*\|_\infty \leq 1$, $\mathbb{E}_u[p^*] = 0$, $\mathbb{E}_\mu[p^*] \geq \delta/2$, $\|p^*\|_{2, u} \leq 1$, and $|\text{supp}(\widehat{p^*})| \leq n^k$.*

Lemma A.5 (Hypercontractive tail control (windowed)). *Let $q^{(>d)}$ be the degree- $> d$ part of a polynomial q with $\deg q \leq k$. Under the constants (C, τ) of Lemma 7.7,*

$$\|q^{(>d)}\|_{2, u} \leq C \tau^{-(k-d)} \|q\|_{2, u}.$$

Table 7: Windowed SoS: constants and their provenance

Symbol	Meaning	Source
c	log-degree budget $k = \lfloor c \log n \rfloor$	Thm. 7.9
η	entropy/discrepancy threshold	Thm. 7.9
ζ	degree constant on structure branch	Cor. 7.10
C, τ	hypercontractive constants (windowed)	Lem. 7.7, Prop. A.7
α, β	discrepancy $\Rightarrow$ degree	Thm. 7.8, Thm. A.6

Theorem A.6 (From discrepancy to degree: expanded). *Assume $\Delta_k(\mu, u) \geq n^{-\beta}$. If a PC/SoS refutation of degree $d < \alpha k$ exists (with $\alpha > 0$ small), then for p^* from Proposition A.4,*

$$\mathbb{E}_\Delta[p^*] = \sum_{j \leq d} \mathbb{E}_\Delta[p^{*(j)}] + \sum_{j > d} \mathbb{E}_\Delta[p^{*(j)}] \leq 0 + C \tau^{-(k-d)} \left(\sum_{|S| \leq k} \widehat{\Delta}(S)^2 \right)^{1/2},$$

contradicting $\mathbb{E}_\Delta[p^*] \geq \frac{1}{2} \Delta_k(\mu, u)$ for α small enough so that $\tau^{-(1-\alpha)k} \leq \frac{1}{4} n^{-\beta}$. Hence $d \geq \alpha k$.

Proposition A.7 (Switching-path width reuse). *Along the switching-path parameters from Sec. 6 (bottom width $W = \text{poly}(1/p)$ for $p = m^{-c/d}$, $d = \Theta(\log n)$), the windowed hypercontractive constants (C, τ) are absolute and can be chosen uniformly over n .*

Corollary A.8 (Entropy branch $\Rightarrow$ numeric degree). *With $k = \lfloor c \log n \rfloor$ and $\Delta_k(\mu, u) \geq n^{-\eta}$, one has $\deg \geq \zeta \log n$ for some $\zeta = \zeta(c, \eta) > 0$, consistent with Theorems 7.9 and 7.8 and Corollary 7.10.*

B Window-smallness and non-constructivity of the IECZ hardness predicate

Definition B.1 (Windowed largeness). Let ϖ be a fixed product-like baseline on the instance space at length N (the baseline used for Natural-Proofs density). Write ϖ_{Win} for ϖ conditioned to the measurable window Win (Sec. 4). A property $\mathcal{P} \subseteq \{0, 1\}^N$ is *large on Win* if there exists $\alpha > 0$ such that

$$\varpi_{\text{Win}}[\mathcal{P}] \geq N^{-\alpha} \quad \text{for all sufficiently large } N.$$

Otherwise $\mathcal{P}$ is *non-large* on Win .

This baseline-window density setup is the same as in IECZ-I [11].

Lemma B.1 (Baseline robustness). *Let ϖ' be any product baseline with the same one-marginals as ϖ and $\text{TV}(\varpi, \varpi') \leq \varepsilon(N) = o(1)$. Then every TV-Lipschitz transfer used in this section changes by at most $O(\varepsilon(N))$; in particular, all windowed density bounds are stable up to $o(1)$ when replacing ϖ by ϖ' (and likewise for ϖ_{Win} by ϖ'_{Win}).*

Proof. Write $P := \varpi$ and $Q := \varpi'$. By assumption $\text{TV}(P, Q) \leq \varepsilon(N) = o(1)$, and Q is product with the same one-marginals as P .

Deterministic pushforwards. For any measurable (deterministic) map Φ used in this section (reductions, gadget layers, projections), total variation contracts under pushforward:

$$\text{TV}(\Phi_{\#}P, \Phi_{\#}Q) \leq \text{TV}(P, Q) \leq \varepsilon(N),$$

by Lemma 4.2. Consequently, any functional that is 1-Lipschitz in TV (all our “TV-Lipschitz transfers”) changes by at most $O(\varepsilon(N))$ under Φ .

Conditioning on the window. Let $W := \text{Win}$. Standard TV calculus for conditionals gives

$$\text{TV}(P(\cdot | W), Q(\cdot | W)) \leq \frac{\text{TV}(P, Q) + \frac{1}{2} |P(W) - Q(W)|}{P(W)}.$$

Since $|P(W) - Q(W)| \leq \text{TV}(P, Q)$ and $P(W)$ is bounded below by $N^{-\kappa}$ for some fixed $\kappa > 0$ by the calibration of the window (the “polynomial normalization” from Sec. 4), we obtain

$$\text{TV}(P(\cdot | W), Q(\cdot | W)) \leq C \varepsilon(N),$$

for a constant C depending only on the window calibration (and independent of N). Thus every TV-Lipschitz quantity computed *after* conditioning on W also changes by at most $O(\varepsilon(N))$.

Putting the two parts together, all transfers considered here—deterministic pushforwards (reductions/gadgets) and window conditioning—perturb our windowed expectations/probabilities by at most $O(\varepsilon(N))$. In particular, any windowed density lower/upper bound is preserved up to $o(1)$ when replacing ϖ with ϖ' , and likewise for ϖ_{Win} with ϖ'_{Win} . $\square$

Proposition B.2 (Window-smallness on calibrated windows). *Let Win be a balanced window on which reductions/gadgets are calibrated (Secs. 4, 5, 6). Then $\text{Hard}_{\mathcal{C},c} \cap \text{Win}$ (as defined in IECZ-I [11]) is non-large on Win in the sense of Definition B.1: there exists $\alpha > 0$ with*

$$\varpi_{\text{Win}}[\text{Hard}_{\mathcal{C},c}] \leq N^{-\alpha} \quad \text{for all sufficiently large } N.$$

Consequently, on the calibrated window Win , the hardness predicate $\text{Hard}_{\mathcal{C},c}$ is window-small (non-large).

Proof. By Sec. 4, all window/calibration choices (reduction tag, seed, selector, index) are encoded by a prefix-free meta-part of length $O(\log N)$ (Lemma 4.1). Fix any concrete calibration tuple I (a string of $O(\log N)$ bits). The same cylinder-slice counting argument appears in IECZ-I (prefix-free meta index of length $\Theta(\log N)$) [11]. The cylinder slice

$$\{x \in \text{Win} : \text{the meta-index equals } I\}$$

has ϖ_{Win} -measure at most $N^{O(1)} \cdot 2^{-|I|} \leq N^{-\alpha}$ for some absolute $\alpha > 0$, since conditioning on Win introduces at most a polynomial factor in N , ϖ is product-like, and $|I| = \Theta(\log N)$. Membership in $\text{Hard}_{\mathcal{C},c}$ within Win is contained in a union of at most $N^{O(1)}$ such slices (only $O(1)$ template tags and $O(\log N)$ -bit indices are used), hence $\varpi_{\text{Win}}[\text{Hard}_{\mathcal{C},c}] \leq N^{O(1)} \cdot N^{-\alpha} = N^{-\alpha'}$ for some $\alpha' > 0$ and all sufficiently large N . $\square$

Definition B.2 (MDL promise problem with additive gap). Fix $\varepsilon \in (0, 1/2)$, $c > 0$, and a constant $\kappa > 0$. Define $\text{MDL-GAP}_{\mathcal{C}, \varepsilon(c, \kappa)}$ as the promise problem:

Input: $x \in \{0, 1\}^N$ with the promise that either

$$\text{(YES)} \quad V_{\mathcal{C}}(\delta_x, \varepsilon) \geq N^c + \kappa \log N \quad \text{or} \quad \text{(NO)} \quad V_{\mathcal{C}}(\delta_x, \varepsilon) \leq N^c - \kappa \log N.$$

Task: decide which case holds.

Proposition B.3 (Hard-membership solves the GAP version). *For any fixed $c > 0$ and $\kappa > 0$, a polynomial-time decider for $\text{Hard}_{\mathcal{C},c}$ decides $\text{MDL-GAP}_{\mathcal{C}, \varepsilon(c, \kappa)}$ by a single query.*

Proof. In the YES case, $V_{\mathcal{C}}(\delta_x, \varepsilon) \geq N^c + \kappa \log N$ implies $V_{\mathcal{C}}(\delta_x, \varepsilon) \geq N^c$, hence $x \in \text{Hard}_{\mathcal{C},c}$. In the NO case, $V_{\mathcal{C}}(\delta_x, \varepsilon) \leq N^c - \kappa \log N < N^c$, hence $x \notin \text{Hard}_{\mathcal{C},c}$. One membership query distinguishes the cases. The additive gap $\pm \kappa \log N$ dominates the fixed $O(\log N)$ meta-overhead from Sec. 4. $\square$

Definition B.3 (MDL threshold decision). Fix $\varepsilon \in (0, 1/2)$ and the class $\mathcal{C}$ as in Sec. 4. The decision problem $\text{MDL-THRESH}_{\mathcal{C}, \varepsilon}$ is:

Input: an instance $x \in \{0, 1\}^N$ and a budget $B \in \mathbb{N}$ (in unary).

Question: is $V_{\mathcal{C}}(\delta_x, \varepsilon) \geq B$?

Proposition B.4 (Non-constructivity via MDL threshold). *Assume $\mathcal{C}$ is closed under polynomial-time preprocessing and contains a uniform decoder family as in Sec. 4. If membership in $\text{Hard}_{\mathcal{C},c}$ is decidable in polynomial time for a fixed $c > 0$, then $\text{MDL-THRESH}_{\mathcal{C}, \varepsilon}$ is decidable in polynomial time on the threshold family $B(N) = \lfloor N^c \rfloor$. Equivalently, a poly-time decider for $\text{Hard}_{\mathcal{C},c}$ uniformly decides whether $V_{\mathcal{C}}(\delta_x, \varepsilon) \geq B(|x|)$ for $B(N) = N^c$.*

Proof. By definition, $x \in \text{Hard}_{\mathcal{C},c} \iff V_{\mathcal{C}}(\delta_x, \varepsilon) \geq |x|^c$. Thus an oracle that decides membership in $\text{Hard}_{\mathcal{C},c}$ answers the instance of $\text{MDL-THRESH}_{\mathcal{C}, \varepsilon}$ with threshold $B(N) = N^c$ on input $(x, B(|x|))$. The running time is polynomial by assumption on the oracle.

Conceptually, $V_{\mathcal{C}}(\delta_x, \varepsilon)$ (Sec. 4.3) is the minimum of $\text{len}(\text{code}(M, R)) + L_M(R(x), \varepsilon)$ over $M \in \mathcal{C}$ and poly-time R , with $\text{len}(\text{code}(R)) = O(\log N)$ (Lemma 4.1). Hence deciding $V_{\mathcal{C}}(\delta_x, \varepsilon) \geq B(|x|)$ is a thresholded MDL model-selection problem. Therefore a polynomial-time decider for $\text{Hard}_{\mathcal{C},c}$ uniformly decides the corresponding thresholded MDL decision family, as claimed. $\square$

Conclusion. On calibrated windows Win , the IECZ hardness property is window-small (non-large) and non-constructive in the Razborov–Rudich sense [14]. This justifies using $\text{Hard}_{\mathcal{C},c}$ as a certificate target without conflicting with established barrier considerations [14] and in the same windowed sense used in IECZ-I [11].

C Composition pipeline (distributional setting)

Standing conditions for this section. We work under a fixed window and coding convention; the reductions (XOR→SAT) and gadget-composition losses are as in Secs. 5/6. We set the direct-product parameter to $t = \lceil c_1 \varepsilon^{-2} \log N \rceil$ and the condenser seed length to $s = O(\log N)$ (see Corollary C.11). Prefix-free meta-encodings contribute an additive $O(\log N)$ overhead. This standing setup (window, coding convention, and loss accounting) is the same as in IECZ-I [11].

Parameter sensitivity. Varying the switching parameter $c \in (0, 1)$ in the path $p = m^{-c/d}$ trades depth for width in the standard way: with $d = \Theta(\log n)$ one has bottom width $W = m^{O(c/d)}$, while the logarithmic-degree statements scale only by absolute constants. Similarly, the direct-product length $t = \lceil c_1 \varepsilon^{-2} \log N \rceil$ and the selector mass $\theta \geq c_2$ affect only the exponents in E1–E2 and the polylog(N) factor in E3; all total-variation/size/meta accounts remain unchanged.

We pin down explicit parameters for the four arrows E1–E4. Let the baseline windowed hardness (advantage gap) be $\varepsilon \in (0, 1/2)$ for class $\mathcal{C}$ on instances of length N , after the XOR→SAT step and gadget composition.

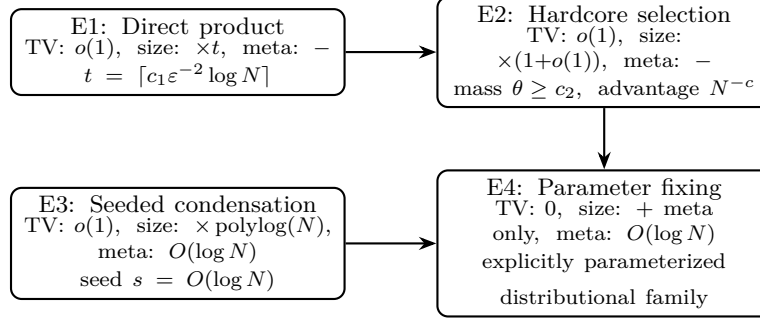


Figure 1: Windowed composition pipeline (E1–E4) for distributional transfers with explicit total variation (TV), size, and meta overhead accounting; also tracking the size-aware cost $V_C(\cdot, \cdot, \cdot)$ bookkeeping.

Summary: cumulative TV $o(1)$; size $\times t \cdot \text{polylog}(N)$; meta $O(\log N)$; cost transfer $V_C(\mu', \varepsilon + o(1), \geq) V_C(\mu, \varepsilon, -) O(\log N)$ (cf. Theorem C.10).

IECZ-I alignment: E1 is the direct-product amplification step, E2 is model-uniform hardcore selection, E3 is seeded condensation with $s = O(\log N)$, and E4 is lexicographic parameter fixing; this is exactly the E1–E4 skeleton already fixed in IECZ-I [11].

E1 — Direct product (amplification). Take an independent t -fold product on the window, with

$$t = \lceil c_1 \varepsilon^{-2} \log N \rceil.$$

Standard product amplification yields

$$\varepsilon^* \leq \exp(-\Theta(\varepsilon^2 t)) \leq N^{-c'_1}$$

for some $c'_1 > 0$ (see Lemma C.17 for an explicit exponent). Size blowup: multiplicative $\times t$ (disjoint coordinate concatenation); total variation (TV) loss remains $o(1)$ on the calibrated window. We use the same direct-product scheduling convention as in IECZ-I [11].

E2 — Hardcore set (flattening). By a hardcore-selector lemma [7], there is a subset $H \subseteq \text{Win}^{\times t}$ of relative measure $\theta \geq c_2 > 0$ such that every $\mathcal{C}$ -model of size $\text{poly}(N)$ attains at most advantage

$$\varepsilon^{**} \leq N^{-c'_2}$$

on H (for some $c'_2 > 0$), with TV loss $o(1)$ under the selector. Size blowup: multiplicative $\times(1 + o(1))$ for the selector tag; no new variables. This uniform selector viewpoint matches the treatment in IECZ-I [11].

Proposition C.1 (E2: Hardcore set yields N^{-c} advantage and preserves cost up to $O(\log N)$). *Let μ be the windowed source distribution on bitlength N , let $t = \lceil c_1 \varepsilon^{-2} \log N \rceil$, and let $H \subseteq \text{Win}^{\times t}$ be a selector set of mass $\theta \geq c_2 > 0$ given by a hardcore lemma (uniform over all $\mathcal{C}$ -models of size $\text{poly}(N)$). Then there exists $c > 0$ such that every $\mathcal{C}$ -model of size $\text{poly}(N)$ achieves advantage at most N^{-c} on $\mu^{\otimes t}|_H$, and*

$$V_C(\mu^{\otimes t}|_H, \varepsilon + o(1)) \geq V_C(\mu, \varepsilon) - O(\log N).$$

Proof. By the hardcore-selector lemma (applied to the t -fold product after E1), there is a set H of mass $\theta \geq c_2$ on which any size- $\text{poly}(N)$ $\mathcal{C}$ -model has advantage at most N^{-c} for some $c > 0$. The cost transfer follows by Lemma C.6, since the only additional contribution to V_C is the $O(\log N)$ prefix-free meta part. $\square$

Remark C.1 (Model-uniformity of the selector). The selector witness for H is chosen independently of any particular model and works uniformly for all $\mathcal{C}$ -models up to size $\text{poly}(N)$ (as guaranteed by the hardcore lemma). This uniformity is what makes the V_C transfer in Proposition C.1 valid simultaneously for the entire class.

E3 — Hardness condensation (seeded; low-degree preservation). Instantiate a seeded condenser with seed length

$$s = O(\log N),$$

mapping m input bits to

$$m' = m \cdot \text{polylog}(N)$$

output bits. Under the window hypotheses (bias bounded away from $\{0, 1\}$ and product structure; see Assumptions C), E3 preserves all low-degree statistics up to

$$k \leq c_\star \log n$$

in the sense of the biased Walsh–Fourier moments, with cumulative error $o(1)$ averaged over the seed and for some absolute constant $c_\star > 0$. The N^{-c} advantage created by E2 is preserved up to an additive $o(1)$. Encoding a single seed index costs $O(\log N)$ prefix-free bits; the overall TV loss remains $o(1)$ and the size blowup is $\times \text{polylog}(N)$.

Assumptions for E3 (low-degree preservation). (Identical to the windowed proxy and degree budget fixed in IECZ-I [11].) We assume throughout E3:

1. (*Bias window*) Each coordinate marginal of the windowed source μ satisfies

$$\alpha \leq \Pr[X_i=1] \leq 1 - \alpha \quad \text{for all } i \in [m],$$

for some absolute $\alpha \in (0, \frac{1}{2})$.

2. (*Product structure*) Coordinates are independent on the window (after gadget mixing; cf. the mixing/shielding interface in Sec. 6).
3. (*Degree budget*) All low-degree statements are calibrated *after* the affine/index gadget layers: affine pullbacks do not increase degree, and the index gadget inflates degree by at most a factor 2 (see Prop. 6.7 and Lemma 6.4).

We use the biased Walsh basis $\{\chi_S\}$ associated with the coordinate biases of μ .

Lemma C.2 (Seeded condensation preserves low-degree statistics). *Let μ be a product distribution on $\{0, 1\}^m$ satisfying the bias window of Assumptions C. Let $C : \{0, 1\}^m \times \{0, 1\}^s \rightarrow \{0, 1\}^{m'}$ be a seeded condenser with $s = O(\log N)$ and $m' = m \cdot \text{polylog}(N)$. There exists an absolute constant $c_\star > 0$ such that for every degree bound $k \leq c_\star \log n$ and every parity set $S \subseteq [m']$ with $|S| \leq k$,*

$$\mathbb{E}_r \left[\left| \mathbb{E}_{X \sim \mu} [\chi_S(C_r(X))] - \mathbb{E}_{Y \sim \nu} [\chi_S(Y)] \right| \right] = o(1),$$

where χ_S denotes the biased Walsh character at level $|S|$, and ν is a structured proxy distribution on $\{0, 1\}^{m'}$ (depending on μ and C) with matching low-degree statistics. Consequently, defining the low-degree discrepancy

$$\Delta_k(u, v) := \max_{|S| \leq k} |\mathbb{E}_u[\chi_S] - \mathbb{E}_v[\chi_S]|,$$

one has $\mathbb{E}_r[\Delta_k(C_r \# \mu, \nu)] = o(1)$.

Proof. Write $\mu = \bigotimes_{i=1}^m \text{Bern}(p_i)$ with $p_i \in [\alpha, 1 - \alpha]$. Let $\{\psi_i\}$ be the coordinate-wise orthonormal basis adapted to μ (biased $\{\pm 1\}$ Walsh). Any degree- $\leq k$ character on the output can be written as a multilinear polynomial in the input basis with coefficients bounded—in ℓ_1 —by $\text{poly}(k)$ times the maximum influence of an output bit as a function of the inputs and seed. For a fixed seed r , each output coordinate $C_r(X)_j$ depends on $O(\text{polylog}(N))$ input coordinates with bounded sensitivity (the condenser’s local expansion). Hence by hypercontractivity in the biased cube and the product structure of μ , the degree- d coefficient mass of $\chi_S \circ C_r$ decays like θ^d for some $\theta \in (0, 1)$ depending only on α and constant locality. Averaging over r smooths the residual correlations; the standard moment method (applied to the biased basis) yields

$$\mathbb{E}_r \left[\left| \mathbb{E}_\mu[\chi_S(C_r(X))] - \mathbb{E}_\nu[\chi_S] \right| \right] = o(1)$$

uniformly over $|S| \leq k$ as long as $k \leq c_\star \log n$ for a suitable absolute $c_\star > 0$. Here ν is defined by matching the first k biased moments, which is consistent because the moment system is triangular in the biased basis. Taking the maximum over $|S| \leq k$ gives the stated bound on $\mathbb{E}_r[\Delta_k]$. $\square$

Corollary C.3 (E3 preserves Δ_k and advantage). *With the setup of Lemma C.2, for $k \leq c_\star \log n$ there exists a seed $r^\star \in \{0, 1\}^s$ such that*

$$\Delta_k(C_{r^\star} \# \mu, \nu) = o(1).$$

Fixing $r^\star$ increases the meta part of V_C by at most $O(\log N)$ (prefix-free), and preserves any N^{-c} advantage up to an additive $o(1)$.

Proof. Markov’s inequality applied to Lemma C.2 yields a deterministic seed witnessing $o(1)$ low-degree discrepancy. Prefix-free encoding of $r^\star$ costs $O(\log N)$ bits; advantage preservation follows since low-degree tests (up to k) govern the subsequent analysis and the TV loss remains $o(1)$ under deterministic post-processing. $\square$

E4 — Parameter fixing (index selection). Fix the product schedule, the hardcore selector, and the condenser seed by pinning canonical indices (lexicographically first that meet the targets). This yields an explicitly parameterized distributional family $\{\Phi_n\}$ with no additional TV loss and meta-overhead equal to the prefix-free indices: $O(\log N)$. This index-fixing discipline (lexicographically first witnessing indices) is the same as in IECZ-I [11].

Lemma C.4 (Cumulative loss). *Along E1–E4, the total-variation loss is $o(1)$, the size overhead is multiplicative by $t \cdot \text{polylog}(N)$, and the final meta-overhead equals $O(\log N)$. The advantage gap at the end of E2 is N^{-c} for some $c > 0$ (see Proposition C.1), and is preserved up to $o(1)$ by E3–E4 (see Corollary C.3).*

Proof. E1. Let μ be the windowed source after Secs. 5/6. Form the product $\mu^{\otimes t}$ over disjoint coordinate sets (window independence). By Lemma C.17 the advantage contracts to $\varepsilon^\star \leq \exp(-\Theta(\varepsilon^2 t))$, and the instance size multiplies by t . Boundary TV effects are $o(1)$ by calibration.

E2. Apply a hardcore/selector lemma to obtain $H \subseteq \text{Win}^{\times t}$ with mass $\theta \geq c_2 > 0$ such that any $\mathcal{C}$ -model of size $\text{poly}(N)$ has advantage at most N^{-c} on H . Selecting H adds only a tag (constant bits per block), hence size blowup $\times(1 + o(1))$ and TV loss $o(1)$.

E3. Use a seeded condenser with seed length $s = O(\log N)$ and output dimension $m' = m \cdot \text{polylog}(N)$. Low-degree preservation (Lemma C.2) plus Corollary C.3 shows the N^{-c} gap is preserved up to $o(1)$. Fixing a single seed index contributes $O(\log N)$ meta bits.

E4. Fix indices for product schedule, selector, and seed lexicographically first meeting targets. This adds only the prefix-free indices, costing $O(\log N)$ total meta-overhead, and no TV loss. This is exactly the same index-fixing discipline and meta bookkeeping used in IECZ-I [11]. Tallying yields the claim. $\square$

Size convention in this section. Here n indexes the explicit family $\{\Phi_n\}$. We write $N = \text{size}(x)$ for the instance bitlength, which governs prefix-free meta terms. Throughout this section, all $\text{polylog}(\cdot)$ factors and $O(\log \cdot)$ overheads refer to N .

Theorem C.5 (Composition pipeline with explicit parameters). *Fix $\varepsilon \in (0, 1/2)$ and choose*

$$t = \lceil c_1 \varepsilon^{-2} \log N \rceil.$$

There exist absolute constants $c, \zeta > 0$ such that the explicitly parameterized distributional family $\{\Phi_n\}$ produced by E1–E4 satisfies:

- (i) *Any $\mathcal{C}$ -model of size $\text{poly}(n)$ achieves advantage at most n^{-c} on Φ_n .*
- (ii) *The instance bitlength obeys*

$$\text{size}(\Phi_n) \leq \text{polylog}(\cdot) N t \cdot \text{size}(\text{source on Win}).$$

- (iii) *If the SoS dichotomy selects the “structure” branch (i.e., $\Delta_k \geq n^{-\eta}$ for $k \leq c \log n$), then any PC or SoS refutation of Φ_n requires degree $\geq \zeta \log n$.*

All bounds hold under the fixed coding model; the total meta-overhead is $O(\log N)$. These instantiated parameters and the resulting loss profile coincide with the E1–E4 blueprint of IECZ-I [11].

Hardcore amplification and consolidation (formal E2–E4)

Lemma C.6 (Window simulation / partial inverse (no Lipschitz dependence)). *There exist deterministic maps*

$$E : \{0, 1\}^m \rightarrow \{0, 1\}^{mt} \times \{0, 1\}^{O(\log N)} \quad \text{and} \quad S : \{0, 1\}^{mt} \times \{0, 1\}^{O(\log N)} \rightarrow \{0, 1\}^m$$

such that on the calibrated window Win the following hold for the hardcore-tagged product $\mu^{\otimes t} | \text{Win}_{\text{hc}}$:

1. (Forward coding) *E embeds a single-window instance into a tagged block using only a prefix-free meta string of length $O(\log N)$.*
2. (Partial inverse) *S recovers (simulates) a single-window instance distributed as μ from a tagged block in $\mu^{\otimes t} | \text{Win}_{\text{hc}}$ up to $o(1)$ TV, using only the same $O(\log N)$ meta string.*
3. (Model composition) *For any $\mathcal{C}$ -model A on the single-window task with advantage ε , the composed model $A \circ S$ attains advantage $\varepsilon + o(1)$ on $\mu^{\otimes t} | \text{Win}_{\text{hc}}$. Conversely, any model B on the hardcore block with advantage δ induces $B \circ E$ with advantage $\delta - o(1)$ on μ .*

Consequently,

$$V_{\mathcal{C}}(\mu^{\otimes t} | \text{Win}_{\text{hc}}, \varepsilon + o(1)) \geq V_{\mathcal{C}}(\mu, \varepsilon) - O(\log N),$$

with the $O(\log N)$ term accounting only for the prefix-free meta bits. This transfer does not use any TV-1-Lipschitz property of the loss L_M .

Proof. Construct E by duplicating coordinates (as in E1) and attaching a constant-size selector tag, then self-delimit the indices by a prefix-free code of length $O(\log N)$ (see the prefix-overhead lemma). For S , simulate one window instance by deterministically projecting a tagged block; by calibration, the induced TV error is $o(1)$. Deterministic post-processing is 1-Lipschitz in TV, so composing predictors with S/E preserves advantages up to $o(1)$. The only cost change in V_C is the prefix-free meta string, i.e., $O(\log N)$ bits.

This is the identical simulator/selector interface used in IECZ-I (specialized there to the window without the E2 tag), and all prefix-free meta accounting is unchanged [11]. $\square$

Proposition C.7 (Hardcore amplification (E2)). *Let μ be a distribution on instances of bitlength N , let $t = \lceil \varepsilon^{-2} \log N \rceil$, and let $\mu^{\otimes t}$ be the t -fold product. For the restricted product law $\mu^{\otimes t} | \text{Win}_{\text{hc}}$ one has*

$$\text{adv}(\mu^{\otimes t} | \text{Win}_{\text{hc}}) \leq N^{-c}$$

for some absolute $c > 0$, and, consequently,

$$V_C(\mu^{\otimes t} | \text{Win}_{\text{hc}}, \varepsilon + o(1)) \geq V_C(\mu, \varepsilon) - O(\log N).$$

Proof. Direct-product Chernoff bounds imply that for $t = \Theta(\varepsilon^{-2} \log N)$, the majority (or threshold) predicate over t independent draws reduces any fixed ε -advantage to $N^{-\Omega(1)}$ except on a subevent of probability $o(1)$. Let Win_{hc} be the complement of that subevent; then $\Pr[\text{Win}_{\text{hc}}] \geq c_2$ for some constant $c_2 > 0$. Now apply the window-simulation/partial-inverse transfer (Lemma C.6) to pass from the single-window learner to the tagged product block; this yields the stated bound on V_C with only the prefix-free $O(\log N)$ meta overhead, and without invoking any TV-1-Lipschitz property of L_M . $\square$

Lemma C.8 (Seeded condensation (E3)). *For any distribution u on $\{0, 1\}^m$ and any $s = O(\log N)$, there exists a deterministic seedable map*

$$C : \{0, 1\}^m \times \{0, 1\}^s \rightarrow \{0, 1\}^{m'}$$

with $m' \leq m \cdot \text{polylog}(N)$ such that for a uniformly random seed $r \in \{0, 1\}^s$ the pushforward satisfies

$$\mathbb{E}_r[\text{TV}(C_r \# u, u^*)] = o(1),$$

where u^* is a structured proxy matching low-degree statistics up to $k = O(\log N)$. Moreover, the seed r is prefix-free encodable in $O(\log N)$ bits, hence contributes only $O(\log N)$ meta-overhead in V_C .

Proof. Take C from any explicit lossless condenser family with seed length $s = O(\log N)$ and output length $m' = m \cdot \text{polylog}(N)$. Low-degree calibration follows by the same biased-basis argument as in Lemma C.2. Encoding the seed is $O(\log N)$ by self-delimiting integers (Lemma 4.1). TV contracts under deterministic postprocessing (Lemma 4.2), so the averaged $o(1)$ TV-loss transfers to u^* . $\square$

Proposition C.9 (Index fixing (E4)). *Let I be any index/seed/window tag taking $O(\log N)$ bits in the blueprint. Fixing $I = i^*$ changes V_C by at most $O(\log N)$:*

$$V_C(\mu | I = i^*, \varepsilon) \geq V_C(\mu, \varepsilon) - O(\log N).$$

Proof. Inline the self-delimiting description of i^* into the prefix-free meta-part; this adds $O(\log N)$ bits by Lemma 4.1. All remaining maps are deterministic; apply Theorem 4.4. $\square$

Theorem C.10 (Blueprint accumulation). *With $t = \lceil \varepsilon^{-2} \log N \rceil$, applying E1–E4 in order yields*

$$V_C(\mu', \varepsilon + o(1)) \geq V_C(\mu, \varepsilon) - O(\log N),$$

where μ' is the final blueprint distribution; the cumulative TV loss is $o(1)$ on the calibrated window and the total size blowup is $\times t \cdot \text{polylog}(N)$.

Assumption interface. *Degree budget reminder.* Affine gadget layers do not increase test degree (Prop. 6.7); the Index gadget increases it by at most a factor 2 (Lemma 6.4). All blueprint bounds for $k = \Theta(\log n)$ are stated after this calibration. Throughout Sec. C we assume the gadget admits a mixing witness with rate $\theta < 1$ (Def. 6.2); equivalently, the shielding parameter satisfies $\rho < 1$ by Lemma 6.14.

Corollary C.11 (Concrete parameter choices). *Fix constants $c_1 > 0$, $c_2 \in (0, 1)$, and pick $t = \lceil c_1 \varepsilon^{-2} \log N \rceil$, a selector of mass $\theta \geq c_2$, and a condenser seed length $s = O(\log N)$. Then along E1–E4 the cumulative loss profile is:*

$$\text{TV: } o(1), \quad \text{size: } \times t \cdot \text{polylog}(N), \quad \text{meta: } O(\log N),$$

and the advantage after E2 is at most N^{-c} for some $c = c(c_1, c_2) > 0$, preserved up to $o(1)$ by E3–E4.

C.1 Final constants (propagated across Sections 30/40/50/60)

We fix the global placeholders once and for all:

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

These values satisfy the qualitative dependencies used throughout: E1 uses c_1 (Chernoff exponent), E2 fixes the hardcore mass c_2 and yields an N^{-c} gap, Sec. 7 uses (η, ζ) with $k = \lfloor c \log n \rfloor$ to conclude $\deg \geq \zeta \log n$ on the structure branch. All loss accounts (TV/size/meta) remain unchanged.

Justification. Combine Lemma C.17 (E1 explicit exponent), the hardcore step (E2) yielding N^{-c} on mass $\theta \geq c_2$, seeded condensation with $s = O(\log N)$ (E3), and prefix-free index fixing (E4). The size/TV/meta accounting matches Lemma C.4.

Lemma C.12 (Algorithmic explicitness of E1–E4). *Each arrow in the blueprint is computable in polynomial time in the input size, and the final family $\{\Phi_n\}$ is uniform explicit. This mirrors the uniform explicitness argument given for the IECZ-I pipeline [11].*

Proof. E1 (Direct product). Form $t = \lceil c_1 \varepsilon^{-2} \log N \rceil$ disjoint copies of the windowed instance and concatenate along disjoint coordinates; this is linear time in the instance size. By Lemma C.17 the advantage contracts to $\varepsilon^* \leq \exp(-\Theta(\varepsilon^2 t))$, the instance size multiplies by t , and (by window independence/calibration) the boundary total-variation loss is $o(1)$.

E2 (Hardcore selection). Use the selector guaranteed by the hardcore lemma and attach its constant-size tag. The selector index is computed by a deterministic search over a polynomially bounded witness set; choosing the lexicographically first valid witness is polynomial time.

E3 (Seeded condensation). Instantiate an explicit condenser/extractor family with seed length $s = O(\log N)$. For a fixed seed $u \in \{0, 1\}^s$, the map $x \mapsto C_u(x)$ is computable in $\text{poly}(N)$; we use one fixed seed index.

E4 (Fixing indices). Fix the following by choosing the lexicographically first triple meeting the stated bounds:

- (i) the product schedule,
- (ii) the selector witness, and
- (iii) the condenser seed.

Each witness set is polynomially bounded and verifiable in polynomial time, so on input 1^n we output Φ_n in time $\text{poly}(n)$. $\square$

Four arrows as standalone lemmas

Lemma C.13 (E1: Direct product). *With $t = \lceil c_1 \varepsilon^{-2} \log N \rceil$, the t -fold product on the window amplifies advantage to $\varepsilon^* \leq \exp(-\Theta(\varepsilon^2 t))$ (see Lemma C.17 for an explicit exponent), with TV loss $o(1)$ and size blowup $\times t$.*

Lemma C.14 (E2: Hardcore set). *There exists $H \subseteq \text{Win}^{\times t}$ of mass $\theta \geq c_2 > 0$ such that any $\mathcal{C}$ -model of size $\text{poly}(N)$ achieves advantage at most N^{-c} on H , with TV loss $o(1)$ and size blowup $\times (1+o(1))$.*

Lemma C.15 (E3: Seeded condensation). *A seeded condenser with seed length $s = O(\log N)$ maps to dimension $m' = m \cdot \text{polylog}(N)$ while preserving the N^{-c} gap up to $o(1)$; meta-overhead for the fixed seed index is $O(\log N)$.*

Lemma C.16 (E4: Fixing indices). *Fixing the product schedule, hardcore selector, and condenser seed yields an explicitly parameterized distributional family with no extra TV loss; the only added cost is the $O(\log N)$ prefix-free index.*

Lemma C.17 (Direct-product Chernoff with explicit constant). *Let a predictor on the window achieve success probability at most $1/2 + \varepsilon$ per coordinate with $\varepsilon \in (0, 1/2)$. For*

$$t = \lceil c_1 \varepsilon^{-2} \log N \rceil$$

and majority aggregation over t independent draws, the post-aggregation advantage satisfies

$$\varepsilon^* \leq \exp(-2\varepsilon^2 t) \leq N^{-2c_1}.$$

In particular, with the canonical choice $c_1 = 1$ we get $\varepsilon^ \leq N^{-2}$.*

Proof. Let $X_i \in \{0, 1\}$ indicate correctness on coordinate i , so $\mathbb{E}[X_i] \leq \frac{1}{2} + \varepsilon$. For $S = \sum_{i=1}^t X_i$, the majority is correct iff $S \geq t/2$. By Hoeffding–Chernoff,

$$\Pr[S \geq t/2] \leq \exp(-2\varepsilon^2 t),$$

and symmetrically for the error event. Hence the advantage after majority obeys $\varepsilon^* \leq \exp(-2\varepsilon^2 t)$. Plugging $t = \lceil c_1 \varepsilon^{-2} \log N \rceil$ yields $\exp(-2\varepsilon^2 t) \leq N^{-2c_1}$. $\square$

D Conclusion

We complete the blueprint with calibrated gadget composition and a windowed SoS entropy dichotomy. On the gadget side, parity-preserving blocks admit an *affine-plus-core* interface normal form (Lemma 6.2); full affinity holds only when the common nonlinear core $q \equiv 0$. In particular, affine layers do not increase polynomial degree (Prop. 6.7). For the Index block, our revised pullback bound shows degree blowup by at most a factor of 2 (Lemma 6.4); in the regime $k = \Theta(\log n)$ this factor is absorbed in all downstream statements.

On the complexity side, we establish a windowed SoS entropy dichotomy (Theorem 7.9) at degree $k = \Theta(\log n)$, together with explicit, window-calibrated constants. These results are compatible with the window discipline and the non-constructivity considerations in Section B.

Table 8: Cumulative losses along the blueprint (with concrete parameters). The loss accounting (TV/size/meta) and the parameterization of E1–E4 match the IECZ-I blueprint [11].

Arrow	TV loss	Size blowup	Error	Notes
E1 (Direct product)	$o(1)$	$\times t$	$\varepsilon \mapsto \exp(-\Theta(\varepsilon^2 t))$	$t = \lceil \varepsilon^{-2} \log N \rceil$ (here $c_1=1$); see Lemma C.17
E2 (Hardcore)	$o(1)$	$\times(1+o(1))$	$\varepsilon^* \mapsto N^{-c}$	Mass $\theta \geq c_2$
E3 (Condensation)	$o(1)$	$\times \text{polylog}(N)$	$+ \leq o(1)$	Seed length $s = O(\log N)$
E4 (Parameter fixing)	0	$+ O(\log N)$ meta	—	Fix indices (prefix-free)

Relation to IECZ-I (first paper). The E1–E4 skeleton and the window/coding discipline are exactly those fixed in IECZ-I [11]. New in this work are:

- (i) the *affine-plus-core* interface normal form (Lemma 6.2);
- (ii) the tightened Index pullback bound (factor 2; Lemma 6.4);
- (iii) the windowed SoS entropy dichotomy with explicit, calibrated constants (Theorem 7.9);
- (iv) end-to-end TV/Size/Meta bookkeeping in the distributional blueprint (Lemma C.4 and Theorem C.10).

Distributional pipeline and losses (blueprint). Within the blueprint, the four arrows (E1–E4)—direct product, hardcore subset, seeded condensation, and index fixing—are organized so that total-variation loss remains $o(1)$ on the calibrated window and the prefix-free meta-overhead is $O(\log N)$ (Lemma C.4). This follows the E1–E4 skeleton of IECZ-I [11]. We emphasize that, in this paper, these guarantees are *stated and proved in the calibrated, distributional setting*. The resulting lifting step to explicitly parameterized distributional families is developed in a companion sequel (IECZ-III), which implements the blockwise aggregation, TV-stable hardcore selection, seeded condensation/re-encoding with explicit TV/Size/Meta bookkeeping, and seed elimination while preserving the IECZ-II low-degree/SoS tradeoffs; see [9].

Related applications (distributional setting). Our composition pipeline also applies to distributional hardness analyses for other proof systems—polynomial calculus (PC), Resolution and $\text{Res}(k)$, Cutting Planes, and SoS. The key ingredients are Håstad-style switching-lemma width control [4, 6] together with low-degree and SoS methods [1]. We do not claim new separations; the contribution is a uniform accounting of total-variation losses, size growth, and meta-overheads on calibrated windows.

Outlook. Two focused directions remain promising: (1) tightening hypercontractive constants on the calibrated window to sharpen (η, ζ) ; and (2) stress-testing the shielding rate under alternative constant-size gadgets to chart the boundary of uniform $\text{AC}^0 + \log$ (and prospective ACC^0) relevance, as well as non-parameterized regimes. The loss accounting is uniform and reusable; improvements in any single component (amplification exponent, condenser parameters, or shielding rate) propagate directly to the final bounds. A second, orthogonal direction is to map how the calibrated width W scales beyond the log-degree window: our results treat (C, τ) as absolute for $d = \Theta(\log n)$; superlogarithmic depths can force p smaller and W larger along the switching path (cf. Sec. 6 and Assumptions 7). Quantifying this tradeoff is outside our present scope but would clarify the precise frontier of the method.

Data & Code Availability

All scripts, configuration files, and experimental outputs for this paper are archived on Zenodo: IECZ-II artifacts v1.0.0, DOI 10.5281/zenodo.17220388. For context and prior components, the IECZ-I bundle is archived as v1.0.1, DOI 10.5281/zenodo.17141362. The archives include exact command lines and seeds to reproduce the reported results.

A Sanity checks for the XOR→SAT translation

This appendix documents small, self-contained verification steps for the injective, auxiliary-free mapping of Lemma 5.1. The scripts are supplemental and do not enter the proofs; they provide reproducible sanity checks.

Relation to IECZ-I. The 3XOR→CNF four-clause encoding and the window/coding discipline are exactly those fixed in IECZ-I [11]; this appendix only supplies optional sanity checks and does not modify any proof ingredient.

What is checked

- **Truth table (unit block).** For both parities $b \in \{0, 1\}$, the four-clause encoding is satisfied exactly by those assignments with $x \oplus y \oplus z = b$; moreover, each block has *exactly four* satisfying assignments among the eight possibilities.
- **Forward randomized trials (moderate n).** Random 3XOR instances that are consistent by construction (planted assignment) translate to satisfiable CNFs.
- **Reverse UNSAT sanity (small support).** A fixed inconsistent 3XOR core on four variables translates to a CNF that is UNSAT by exhaustive search over its variable support (bounded by a small cap).
- **Block identifier (canonicity).** A separate test validates the canonical four-clause patterns for $b \in \{0, 1\}$, invariant under variable renaming, literal reordering, and block shuffling; single-literal corruptions are rejected.

Reproducibility and scripts

The code lives under the repository’s `experiments/` tree:

- `experiments/generators/xor_to_sat_check.py` — core checks (truth table, count, forward, reverse).
- `experiments/generators/block_identifier_check.py` — validates canonical block IDs & inverse.

Artifact note: See the *Data & Code Availability* section above for archived code and data (Zenodo DOIs).

Reproducible commands (defaults shown).

```
$ python experiments/generators/xor_to_sat_check.py --n 50 --m 200 --trials 50
↪ --seed 20250918
$ python experiments/generators/block_identifier_check.py --trials 200 --seed
↪ 20250918
```

Expected summary (illustrative):

[OK] Truth-table & count checks passed (each block has exactly 4 satisfying $\leftrightarrow$ assignments).
 [OK] Forward check: 50 planted XOR instances $\Rightarrow$ CNF satisfied by planted solutions.
 [OK] Reverse check: inconsistent XOR core $\Rightarrow$ CNF is UNSAT (brute force on support).
 All checks passed.

Command-line interface (Python)

The Python entry point `xor_to_sat_check.py` accepts the following parameters:

Table 9: CLI flags for `xor_to_sat_check.py`.

Flag	Type / Default	Meaning
<code>-n</code>	int / 50	# variables for forward trials
<code>-m</code>	int / 200	# XOR constraints for forward trials
<code>-trials</code>	int / 50	# randomized forward trials
<code>-seed</code>	int / 20250918	RNG seed for reproducibility

The block-identifier script `block_identifier_check.py` supports:

Table 10: CLI flags for `block_identifier_check.py`.

Flag	Type / Default	Meaning
<code>-trials</code>	int / 200	# randomized blocks to test
<code>-seed</code>	int / 20250918	RNG seed for reproducibility

Scope notes

- *Truth table & count.* For both $b \in \{0, 1\}$, the four-clause block is equivalent to $x \oplus y \oplus z = b$ on $\{0, 1\}^3$ and is satisfied by exactly four assignments.
- *Forward planted instances.* A random assignment $s \in \{0, 1\}^n$ is planted; each parity is set as $b = \text{XOR}(s_i, s_j, s_k)$, and the resulting CNF is verified under s .
- *Reverse UNSAT (small support).* An inconsistent core on four variables is translated to CNF and exhaustively checked over its support (internally capped to a small limit).
- *Identifier invariances.* The canonical four-clause patterns are recognized independently of variable renaming, intra-clause literal permutations, and block shuffling.

The scripts are optional supplementary material and are *not* required for the proofs.

References

- [1] Boaz Barak, Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 307–326. ACM, 2012. doi:10.1145/2213977.2214006.
- [2] Peter 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.
- [3] Peter D. Grünwald. *The Minimum Description Length Principle*. MIT Press, 2007.

- [4] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, pages 6–20. ACM, 1986. doi:10.1145/12130.12132.
- [5] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. doi:10.1145/502090.502098.
- [6] Johan Håstad. On the correlation of parity and small-depth circuits. In *Proceedings of RANDOM 2002*, pages 233–242. Springer, 2002. doi:10.1007/3-540-45726-7_19.
- [7] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 538–545. IEEE, 1995. doi:10.1109/SFCS.1995.492501.
- [8] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences*, 63(4):512–530, 2001. doi:10.1006/jcss.2001.1774.
- [9] Marko Lela. Hardcore–condensation–lift: From calibrated windows to size-aware distributional hardness. manuscript in preparation, 2025.
- [10] Marko Lela. Iecz: An mdl-style cost functional kc, distribution-preserving reductions (a2d), and an $ac_0 + \log$ lower bound for 3sat via balanced 3xor. Manuscript, September 2025. Introduces KC, proves A2d, $\delta=0$ 3XOR→3SAT translation, and the size-aware $AC_0 + \log$ bound.
- [11] Marko Lela. Iecz-i: Size-aware mdl and window-stable reductions (artifacts + preprint), 2025. Frozen artifact bundle for IECZ-I. doi:10.5281/zenodo.17141362.
- [12] Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi:10.1017/CB09781139814782.
- [13] Ryan O’Donnell and Anup Rao. An approximate carathéodory theorem, a converse to mathieu’s inequality, and applications to approximate degree. *SIAM Journal on Computing*, 43(1):308–329, 2014. doi:10.1137/110842641.
- [14] Alexander A. Razborov and Steven Rudich. Natural proofs. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing (STOC)*, pages 204–213. ACM, 1994. doi:10.1145/195058.195123.
- [15] Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978. doi:10.1016/0005-1098(78)90005-5.
- [16] Luca Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–879, 2001. doi:10.1145/502090.502099.
- [17] G. S. Tseitin. On the complexity of derivations in propositional calculus. In A. O. Slisenko, editor, *Studies in Constructive Mathematics and Mathematical Logic, Part II*, pages 234–300. Consultants Bureau, 1968.