Electronic Colloquium on Computational Complexity, Report No. 201 (2026)

Obfuscation and the Limits of Witness Isolation: Affine Tests and Membership-Query Bounds

Sebastian Ben Daniel
Ben-Gurion University of the Negev

September 2026

Abstract

Assume indistinguishability obfuscation (iO) and one-way functions, both secure against nonuniform polynomial-size adversaries. We show that no randomized polynomial-size pruning procedure isolates a witness with probability at least $a/\log L$, for any constant $a > 0$, where L is the length of the circuit description. Dell, Kabanets, van Melkebeek, and Watanabe (DKMW) proved without cryptographic assumptions that success $2/3 + 1/\operatorname{poly}(L)$ implies $\mathrm{NP} \subseteq \mathrm{P/poly}$ [1]. Under iO alone we get the same collapse from success $a/\log L$, and the guarantee only has to hold on nonempty affine-subspace inputs. The isolator is given no affine basis, it may use the circuit description in any way, and the obfuscator may have negligible correctness error.

The reduction hides a known affine subspace inside a larger solution space whose dimension does not depend on the scale being tested. Obfuscation then lets us compare the isolator's output, computationally, with an independent reference output. Along the way we prove an unconditional preprocessing criterion. It characterizes presentation-invariant isolation and, unless $\mathrm{NP} \subseteq \mathrm{P/poly}$, gives common and efficiently testable witnesses that a constant-success isolator depends on the presentation. For isolators that see the target only through adaptive membership queries, we determine the optimal tradeoff between queries and success up to absolute constants. Finally, a matching restriction-law construction shows why tests on the planted region stop at the logarithmic scale.

1 Introduction

Valiant–Vazirani isolation prunes a satisfiable circuit on n variables so that, with probability $\Omega(1/n)$, exactly one witness survives [3]. We show that indistinguishability obfuscation limits how far any polynomial-size pruning procedure can improve this guarantee: success $a/\log L$, for a constant $a > 0$ and description length L, would already imply $\mathrm{NP} \subseteq \mathrm{P/poly}$, even if it is required only on circuits whose solution sets are affine.

The affine promise comes without a basis. If we were handed an origin and direction basis, $v + \operatorname{span}\{u_1, \ldots, u_m\}$, Gaussian elimination would find the lexicographically least point and isolation would be deterministic. Here the space is given only as a Boolean circuit. The hard part is getting at its structure from that description; the linear algebra afterwards is easy.

Theorem 1.1 (Isolation under indistinguishability obfuscation)**.** *Assume indistinguishability obfuscation for polynomial-size Boolean circuits, with negligible statistical correctness error and security against nonuniform polynomial-size adversaries. Let A be any randomized polynomial-size pruning procedure. If, for some constant $a > 0$, it isolates every nonempty affine-subspace input of length L with probability at least $a/\log L$ for all sufficiently large L, then* $\mathrm{NP} \subseteq \mathrm{P/poly}$*. No presentation invariance of A is assumed.*

ISSN 1433-8092

By statistical correctness we mean that, except with negligible probability over the obfuscator's coins, the obfuscated circuit agrees with the original on every assignment. We do not need perfect correctness. Remark 3.13 tracks the error and also handles the weaker pointwise convention. The proof in Section 3.6 uses security only through one efficient test and never needs the obfuscations to be statistically close.

Corollary 1.2 (Cryptographic impossibility of inverse-logarithmic isolation)**.** *Assume iO as in Theorem 1.1 and a one-way function secure against nonuniform polynomial-size adversaries. Then, for every constant $a > 0$, no randomized polynomial-size pruning procedure has isolation guarantee $a/\log L$ on all nonempty affine-subspace inputs of length L for all sufficiently large L.*

The corollary holds because NP $\subseteq$ P/poly lets polynomial-size circuits invert any polynomial-time function, by searching for a preimage one prefix bit at a time (Section 3.6). One-way functions play no role in Theorem 1.1 itself.

The unconditional preprocessing criterion. Theorem 3.5 needs no cryptographic assumption. For each distribution on source inputs it lets us choose an *equivalence-preserving preprocessing* map, and it is enough that the chosen map produces an inverse-polynomial average gap in an explicit affine test. Minimax and a small-support argument then turn these choices into polynomial-size decision circuits. The obfuscation proof runs the same comparison, with an allowance for rare functional errors.

The presentation-invariant application. Valiant–Vazirani intersects the solution set with a random affine hash predicate, so equivalent circuits on the same ordered variables induce the same distribution of surviving sets. We call this property *presentation invariance*; only descriptions of the same length are compared. For presentation-invariant procedures the identity preprocessing already meets the criterion, and no assumption is needed.

Corollary 1.3 (Characterization of presentation-invariant isolation)**.** *The following statements are equivalent.*

(i) NP $\subseteq$ P/poly.

(ii) There are a constant $a > 0$ and a randomized polynomial-size pruning procedure that is presentation-invariant on affine-subspace inputs and isolates every nonempty such input of length L with probability at least $a/\log L$, for all sufficiently large L.

(iii) There is a deterministic polynomial-size presentation-invariant pruning procedure that isolates every satisfiable Boolean circuit.

In (ii), pruning is required on every input; the remaining requirements concern only nonempty affine-subspace inputs. The proof, in Section 3.4, combines the affine test with the self-reducibility of SAT. Under iO, Theorem 1.1 shows that the invariance requirement in (ii) can be dropped. Corollary 3.6 applies the criterion to canonical preprocessing.

Common witnesses of presentation dependence. Suppose NP $\not\subseteq$ P/poly and fix a pruning procedure with constant success. At infinitely many lengths there is then a single polynomial-size list of equivalent presentation pairs that stays distinguishable after *every* preprocessing map in a prescribed polynomially bounded family. The list is fixed before the map is chosen, and at fixed accuracy each test uses only constantly many output evaluations (Theorem 3.17 in Section 3.7). For the identity map this becomes the simpler statement below, where $D_A(H)$ is the random surviving set on input H.

Corollary 1.4 (Necessary dependence on the presentation)**.** *Let A be a randomized polynomial-size pruning isolator with constant success guarantee $p > 0$. If* NP $\not\subseteq$ P/poly*, then, for every constant $\eta \in (0, p)$, infinitely many description lengths admit equivalent circuits H, H' on the same variables, with a common affine-subspace satisfying set, such that*

$$d_{\mathrm{TV}}\big(\mathrm{Law}(D_A(H)), \mathrm{Law}(D_A(H'))\big) > p - \eta. \tag{1.1}$$

The circuits may be chosen from the construction in Section 3.1.

Adaptive membership access. When the target can be reached only through membership queries, the same affine comparison gives an unconditional bound. Suppose q_{mem} queries are allowed each time an output bit is evaluated, and the target dimension is not supplied. The best worst-case isolation probability over all nonempty affine targets in $\mathbb{F}_2^n$ is then

$$\mathrm{Iso}(n, q_{\mathrm{mem}}) = \Theta\left(\frac{1}{1 + \log(2^n/(q_{\mathrm{mem}} + 1))}\right), \qquad 0 \le q_{\mathrm{mem}} \le 2^n - 1.$$

So a subexponential query budget leaves the optimal success at $\Theta(1/n)$, even if computation between queries is free. Theorem 5.5 gives explicit finite bounds and a construction that attains the matching order.

1.1 The affine test and the distributional criterion

The source circuit is promised to have zero or one satisfying assignment. From it the reduction builds a new circuit together with a known affine subspace K. If the source is unsatisfiable, K is the whole solution set of the new circuit. If it is uniquely satisfiable, K sits inside a much larger affine solution set S, whose dimension stays fixed while the dimension m of K varies. A random affine change of coordinates makes K uniform among the m-dimensional affine subspaces of S (Figure 1).

The test accepts when exactly one point of K survives. On an unsatisfiable source it accepts with probability at least p, the isolation guarantee. On a uniquely satisfiable source, presentation invariance makes the output independent of the planted choice once S is fixed, and the argument comes down to one estimate:

$$\sum_{m=0}^{d} \Pr[|R \cap K_m| = 1] \le 2 \qquad \text{for every fixed } R \subseteq S,$$

where K_m is a uniformly random affine m-subspace. Choosing m uniformly therefore keeps acceptance on uniquely satisfiable sources at most $2/(d+1)$. With $d = O(\log L)$ the largest tested region still has polynomial size, and this gives the inverse-logarithmic bound. Note the direction: high acceptance points to an *unsatisfiable* source, even though the constructed circuits are satisfiable in both cases.

The general criterion asks for less than invariance, namely a gap in average acceptance. For each distribution on uniquely satisfiable sources we may pick a short preprocessing map that pushes the average acceptance below the NO-source guarantee. Minimax combines these choices into one mixed strategy, and the small-support theorem of Lipton and Young [13] makes that strategy polynomially describable; we include the concentration and advice-size calculations. There is also a sufficient condition in terms of conditional mutual information: at constant isolation success, it is enough that the output carry $o(\log \log L)$ bits about the planted choice. In the iO application, exact independence of the actual output is replaced by computational indistinguishability from an independent reference output, and the affine flag bound is applied only to that reference output.

1.2 Previous work and the scope of the results

Dell, Kabanets, van Melkebeek, and Watanabe (DKMW) prove, without cryptographic assumptions, that an unrestricted randomized polynomial-size pruning isolator with success $2/3 + 1/\operatorname{poly}(L)$ implies NP $\subseteq$ P/poly [1, 2]. Theorem 1.1 lowers the required success to $a/\log L$, at the price of assuming iO. The isolator is still arbitrary in how it uses the circuit description, and its success guarantee has to hold only on nonempty affine-subspace inputs. Adding nonuniformly secure one-way functions turns this into the impossibility result of Corollary 1.2.

DKMW also prove an unconditional $O(1/n)$ bound in a target-oblivious model, in which a distribution over pruning sets is fixed before the target is chosen [1, Definition 4.1 and Theorem 4.4] (Definition 2 in the ECCC version [2, Section 4]). A target-oblivious procedure uses the same distribution for every presentation and is therefore presentation-invariant. The converse fails, since presentation invariance still allows dependence on the solution set itself, for example lexicographically least selection. The DKMW bound does not cover distributions chosen from the syntax of the input. Section 5 extends the target-oblivious bound to adaptive membership access, where subexponential query budgets still allow only $O(1/n)$ success on affine targets. That result is about query access and says nothing about procedures that read a circuit description.

For presentation-invariant isolation a quantitative gap remains, between $\Omega(1/n)$ and $a/\log L$, where n is the witness length and L the description length. Our restriction theorem shows that the known-region test cannot close it just by switching to another statistic. At success $\Theta(1/\log k)$ there are valid abstract output laws that agree on every membership bit in the known k-point region, so even joint and adaptive observations inside that region see no difference. The theorem concerns those observations only, and leaves other uses of the circuit description untouched.

The preprocessing theorem takes the distributional gap as a hypothesis; minimax does not derive it from isolation alone. Its hard lists are existential and nonuniform, for resource bounds fixed in advance. Throughout, presentation invariance refers to the unconditioned distribution of the output set, failures included, and the finite-order results of Appendix B are sufficient conditions rather than optimal thresholds. The unrestricted DKMW question remains open.

Other aspects of isolation include randomness efficiency [5], single-valued selection [6], parity pruning [7, 8], and derandomization of list-style isolation [9]. Witness-query lower bounds appear in [10]; logspace consequences are studied in [11]. We use the random-weight isolation lemma [4] as another example of invariance, and we discuss almost-uniform isolation [12] in Appendix B. The harmonic-sum method of [14] appears only in Appendix A.

For the obfuscation result we use the standard notion of iO [15], in the nonuniform sampler-and-distinguisher formulation of [16, Section 3.2]. Subspace-hiding and compute-and-compare obfuscation [16, 17] motivate the restricted preprocessing question at the end of Section 3.6, but neither directly supplies the transformation we would need on circuit-presented affine spaces.

Organization. Section 2 defines the model and recurring notation. Section 3 proves the distributional criterion, its canonical, invariant, and obfuscation applications, and the common-pair theorem. Section 4 proves the order limitation of the planted-region test. Section 5 gives the unconditional adaptive-query bound and its matching tradeoff; Section 6 concludes. A reader interested only in the cryptographic result can read the construction and flag bound in Sections 3.1–3.2 and then skip to Section 3.6; that proof does not use the minimax or information-theoretic refinements. The leading-constant calculations, finite-order and single-origin consequences, local amplification bound, and retention polytope are in Appendices A–D, none of which is needed for Theorem 1.1.

2 Preliminaries and notation

2.1 Circuits, pruning, and the promise problem

For a Boolean circuit C on n input variables, write

$$\mathrm{Sol}(C) = \{x \in \{0,1\}^n : C(x) = 1\}.$$

We reserve SAT for the decision problem of circuit satisfiability. The notation $C \equiv C'$ means that the circuits have the same ordered variables and compute the same Boolean function. The binary description of C is denoted by $\langle C \rangle$. Padding a description does not change its variables or its function. We use a fixed bounded-fan-in circuit encoding in which a length-ℓ description specifies at most ℓ input variables. Gates and descriptions are different size measures. Gate count enters iO security, while L always denotes description length; when the security definition requires it, the constructions pad both.

A randomized *polynomial-size circuit transformation* is a possibly nonuniform family of Boolean circuits mapping $(\langle H \rangle, \rho)$ to an output circuit description. The transformation size, output length, and length of the uniform random string ρ are polynomially bounded in the input-description length. We write $A(H;\rho)$ for the circuit represented by this output. The transformation acts on descriptions; P/poly retains its usual meaning as a class of decision problems.

A *pruning procedure* A outputs a circuit on the same variables as H and satisfies $\mathrm{Sol}(A(H;\rho)) \subseteq \mathrm{Sol}(H)$ for every random string. We distinguish the output circuit from its satisfying set by writing

$$D_A(H;\rho) = \mathrm{Sol}(A(H;\rho)), \qquad D_A(H) = D_A(H;\rho) \quad \text{for uniform random } \rho.$$

The procedure has isolation guarantee $p(L) \in (0,1]$ on a class of length-L inputs if

$$\Pr_{\rho}[|D_A(H;\rho)| = 1] \geq p(L)$$

for every circuit in the class with a nonempty satisfying set. The function p is a lower bound; the actual success probability may depend on the input. When the guarantee holds for every satisfiable circuit, we call A a pruning isolator. An *affine-subspace input* is a circuit whose satisfying set is an affine subspace over $\mathbb{F}_2$, without a supplied basis. A promise-restricted isolator remains defined and pruning on every circuit; only the success requirement is restricted.

The promise problem USAT has a YES instance when $|\mathrm{Sol}(C)| = 1$ and a NO instance when $\mathrm{Sol}(C) = \varnothing$. We use the following length-specific sets:

$$\mathcal{C}^{(1)}_\ell = \{C : |\langle C \rangle| = \ell,\ |\mathrm{Sol}(C)| = 1\}, \qquad \mathcal{C}^{(0)}_\ell = \{C : |\langle C \rangle| = \ell,\ \mathrm{Sol}(C) = \varnothing\}.$$

A circuit family deciding USAT must be correct on these two sets; its answers on circuits with several satisfying assignments are unrestricted. Throughout the paper, YES and NO refer to this *source* promise. The circuits built by our reduction are satisfiable in both cases, so telling them apart is not a satisfiability test.

2.2 Probability and invariance

For a finite set Ω, let $\Delta(\Omega)$ be its set of probability distributions, and let $\mathrm{Unif}(\Omega)$ be the uniform distribution. $\mathrm{Law}(X)$ denotes the distribution of a random variable X. $\mathbf{1}_{\mathsf{Ev}}$ is the indicator of an event Ev. We write log for the base-two logarithm and ln for the natural logarithm. For probability distributions μ, ν on the same finite set,

$$d_{\mathrm{TV}}(\mu,\nu) = \frac{1}{2}\sum_x |\mu(x) - \nu(x)| = \max_{\mathsf{Ev}} |\mu(\mathsf{Ev}) - \nu(\mathsf{Ev})|, \qquad D_{\mathrm{KL}}(\mu\|\nu) = \sum_x \mu(x) \log \frac{\mu(x)}{\nu(x)}.$$

The usual conventions apply when a probability is zero. Binary entropy is $h_2(a) = -a\log a - (1-a)\log(1-a)$, and binary relative entropy is

$$\mathrm{kl}(a\|b) = a\log(a/b) + (1-a)\log((1-a)/(1-b)).$$

Thus $\mathrm{kl}(a\|b) = D_{\mathrm{KL}}(\mathrm{Bern}(a)\|\,\mathrm{Bern}(b))$, where $\mathrm{Bern}(a)$ has probability a of returning one. We use $\mathrm{Ent}(X \mid Y)$ and $I(X;Z \mid Y)$ for conditional entropy and conditional mutual information, in bits. The operator Ent distinguishes entropy from a circuit denoted by H. In particular,

$$I(X;Z \mid Y) = \mathbb{E}_Y\Big[D_{\mathrm{KL}}\big(\mathrm{Law}(X,Z \mid Y)\|\,\mathrm{Law}(X \mid Y)\otimes\mathrm{Law}(Z \mid Y)\big)\Big].$$

The notation $X \perp\!\!\!\perp Z \mid Y$ means conditional independence.

Definition 2.1 (Presentation invariance)**.** A pruning procedure is *presentation-invariant* if

$$H \equiv H', \quad |\langle H\rangle| = |\langle H'\rangle| \quad\Longrightarrow\quad \mathrm{Law}(D_A(H)) = \mathrm{Law}(D_A(H')). \tag{2.1}$$

It is *k-wise retention-invariant* if, under the same hypotheses, the joint distribution of

$$\big(\mathbf{1}_{\{x_1\in D_A(H)\}},\dots,\mathbf{1}_{\{x_j\in D_A(H)\}}\big)$$

is unchanged for every $j \le k$ and every choice of distinct common satisfying assignments $x_1,\dots,x_j$. These conditions may be required only on a stated input class. The case $k = 1$ is called retention invariance.

Presentation invariance implies every finite-order condition. Equality of the one-point retention probabilities is strictly weaker: choosing uniformly from $\{\{x\},\{y\}\}$ or from $\{\varnothing,\{x,y\}\}$ retains each point with probability $1/2$, but gives different joint distributions.

Affine hashing is not the only presentation-invariant rule. Random-weight pruning is another, and so is testing a fixed polynomial-size candidate set before falling back to Valiant–Vazirani. Appendix B gives both examples with their success guarantees. The second depends on the solution set and so is not target-oblivious.

2.3 Nonuniform transfer and parameter conventions

Lemma 2.2 (Transfer from unique satisfiability)**.** *If* USAT *has polynomial-size circuits, then* NP $\subseteq$ P/poly*. More generally, for each fixed $a_{\exp} \ge 1$, circuits of size $2^{O((\log n)^{a_{\exp}})}$ for* USAT *imply the same size bound for every language in* NP*. A bounded-error randomized circuit family suffices in either premise.*

Proof. The polynomial-size statement is DKMW's Lemma 3.2 [1, 2]; we recall the proof to track the larger size bound. For length-n SAT inputs, repeat Valiant–Vazirani polynomially many times. A union bound over at most 2^n input descriptions gives one fixed list of random strings such that every satisfiable input has a uniquely satisfiable pruning in the list. Every pruning of an unsatisfiable input remains unsatisfiable.

Apply the promised decider to each pruning and take the OR of its answers. A satisfiable input has a promised YES pruning, which forces acceptance regardless of the answers on unpromised prunings. For an unsatisfiable input, all prunings are promised NO instances and all answers are zero.

Each resulting description has length $n^{O(1)}$. Polynomially many circuits of size $2^{O((\log(n^{O(1)}))^{a_{\exp}})}$ still have size $2^{O((\log n)^{a_{\exp}})}$. Polynomial-time reductions give the assertion for any fixed language in NP. For randomized deciders, first amplify the error below 2^{-n-1} on each length-n promised input and fix the random strings by a union bound. This costs only a polynomial factor. □

Symbol	Meaning	Scope
ℓ; L_0, L_1	Source, raw, and preprocessed description lengths	Sections 2–4
N; $Q = 2^N$	Extension degree and large solution-set size	Affine construction
$\mathcal{C}_\ell^{(1)}, \mathcal{C}_\ell^{(0)}$	Source circuits with one or zero solutions	Sections 2–4
A; ρ	Isolator and its random string	Throughout
$\mathrm{Sol}(H)$; $D_A(H)$	Solution set; random surviving set	Throughout
$T \in \mathcal{T}_\ell$; $b(\ell)$	Preprocessing map, family, and description bound	Section 3
m, M; d	Fixed/random planted dimension; largest allowed dimension	Sections 3–4
$k = 2^d$; $j = 2^m$	Maximum and selected number of inspected points	Sections 3–4
$K \subseteq S$	Known planted subspace inside the large solution set	Sections 3–4
$R \subseteq S$	Fixed retained subset in an extremal bound	Sections 3–4, App. A
χ_N; g	Irreducible polynomial; random affine bijection	Section 3.1
Test_K; $a_\ell(C, T)$	Exactly-one test; its acceptance probability	Section 3
p_ℓ; γ_ℓ	Isolation guarantee; decision gap	Section 3
$p_{\mathrm{ind}}(d)$	Independent-reference acceptance bound $2/(d+1)$	Section 3
μ; λ; t	Source law; preprocessing mixture; stored list size	Section 3
$\mathcal{I}_\ell(\mu, T)$; $I_{\mathrm{bd}}(\ell)$	Conditional information and its common upper bound	Section 3
$\delta_{A,d}$; η	Output-law diameter; accuracy parameter	Section 3
$\pi \in \Pi_d$	Distribution of planted dimensions	Section 4
$c_{\pi,Q}$; $u_{\pi,Q}$	Exact maximum acceptance; chosen upper bound	Section 4
$\sigma_{\pi,Q}$	Acceptance for a singleton retained set	Section 4
$p^\star_{\pi,Q}$; $p^{\mathrm{opt}}_{d,Q}$	Certified threshold; best such threshold	Section 4
ϑ; τ	Bernoulli inclusion probability; density cutoff	Section 4, App. A
$\mathbb{E}[\cdot]$; Acc	Expectation; acceptance event	Throughout; Section 3

Table 1: Recurring notation. Additional quantities are defined locally; obfuscation notation is introduced in Section 3.6; appendix notation is summarized locally.

We use three description lengths. The source length is ℓ. The affine construction sets $N = 2\ell$, uses $3N$ witness variables, and has common padded description length $L_0(\ell)$. After preprocessing, the common length is $L_1(\ell)$. Both length functions are polynomial in ℓ and are fixed before the number of scales is chosen. For the identity preprocessing, $L_1 = L_0$. An unsubscripted L denotes a generic circuit-description length.

A positive function is inverse-polynomially bounded below if it is at least $\ell^{-a_{\exp}}$ for some constant $a_{\exp} > 0$ at all sufficiently large ℓ. Expressions such as $1/\operatorname{poly}(\ell)$ refer to a specified bound of this kind, not to a single universal polynomial. Length-dependent numerical thresholds may be included as dyadic advice to precision smaller than the gap used by the test.

The proofs first analyze exact uniform sampling from finite groups. The algorithms use bounded random strings and incur only exponentially small statistical error. A uniform $r \times r$ binary matrix is invertible with probability $\prod_{j=1}^{r}(1 - 2^{-j}) > 1/4$. Test invertibility by Gaussian elimination and repeat until an invertible matrix is obtained. Conditional on success, its distribution is uniform. After a sufficiently large constant times ℓ attempts, use a fixed invertible matrix as a fallback; this changes the distribution by at most $2^{-\Omega(\ell)}$. We use the same truncation convention to sample a scale whose range is not a power of two. These errors are smaller than every decision gap used below.

3 A distributional criterion for isolation

We first build the affine instances and their known-region test. The distributional preprocessing criterion then yields the canonical and invariant applications. A test-level replacement lemma extends

the argument to indistinguishability obfuscation. Finally, the dual minimax argument gives common distinguishing pairs.

3.1 The balanced affine construction

Fix a source circuit $C \in \mathcal{C}_\ell^{(1)} \cup \mathcal{C}_\ell^{(0)}$. We first turn its possible unique witness into a linear space, then add known witnesses so that the two source cases have different dimensions. Extend the witness variables of C to $N = 2\ell$ bits, forcing each added bit to zero, and denote the resulting circuit by $\widehat{C}$. This preserves the number of solutions and increases the description length only polynomially.

Choose a monic irreducible polynomial $\chi_N \in \mathbb{F}_2[X]$ of degree N and identify N-bit strings with $\mathbb{F} = \mathbb{F}_2[X]/(\chi_N)$. The $O(N)$ coefficients of χ_N may be included in nonuniform advice. Remark B.1 gives an explicit alternative. Addition and multiplication in this field have polynomial-size Boolean circuits. On a nonzero input, inversion can be computed by exponentiation to $2^N - 2$ with $O(N)$ field multiplications. Its value at zero may be fixed arbitrarily.

For $\alpha, z \in \mathbb{F}$, define

$$F_C(\alpha, z) = \begin{cases} \mathbf{1}_{\{z=0\}}, & \alpha = 0, \\ \widehat{C}(\alpha^{-1} z), & \alpha \neq 0. \end{cases} \tag{3.1}$$

If $C \in \mathcal{C}_\ell^{(0)}$, the only satisfying assignment of F_C is $(0, 0)$. If $C \in \mathcal{C}_\ell^{(1)}$ and $w \in \mathbb{F}$ is its padded witness, then

$$\mathrm{Sol}(F_C) = \mathcal{L}_w := \{(\alpha, \alpha w) : \alpha \in \mathbb{F}\}.$$

For each value of α, exactly one value of z is accepted. Thus the set has $Q = 2^N$ points, although the circuit describing it has polynomial size. Multiplication by the fixed field element w is linear over $\mathbb{F}_2$, so this is an N-dimensional linear subspace, including when $w = 0$.

For each $0 \le m \le N$, let $U_m \le \mathbb{F}_2^N$ and $U_{N-m}^{\mathrm{fld}} \le \mathbb{F}$ be fixed coordinate subspaces of dimensions m and $N - m$, respectively. On $3N$ variables, set

$$H_{C,m}(x_0, \alpha, z) = \mathbf{1}_{\{x_0 \in U_m\}} \mathbf{1}_{\{\alpha \in U_{N-m}^{\mathrm{fld}}\}} F_C(\alpha, z). \tag{3.2}$$

Its satisfying set is

$$C \in \mathcal{C}_\ell^{(0)} : \quad K_m^0 = U_m \times \{0\} \times \{0\}, \qquad \dim K_m^0 = m, \tag{3.3}$$

$$C \in \mathcal{C}_\ell^{(1)} : \quad W_{m,w} = U_m \times \{(\alpha, \alpha w) : \alpha \in U_{N-m}^{\mathrm{fld}}\}, \qquad \dim W_{m,w} = N. \tag{3.4}$$

All dimensions are over $\mathbb{F}_2$. The reduction knows K_m^0, which is contained in $W_{m,w}$. Raising m adds known coordinates and removes the same number of free coordinates from the field line, so the NO-source dimension is m while the YES-source dimension stays at N. That balance is the point of the construction. Figure 1 shows the two cases.

Choose a uniform affine bijection $g \in \mathrm{AGL}(3N, 2)$, where $\mathrm{AGL}(3N, 2)$ is the group of maps $g(x) = \mathrm{Lin}(g)x + g(0)$ with $\mathrm{Lin}(g)$ an invertible matrix over $\mathbb{F}_2$. Define

$$J_{C,m,g} = H_{C,m} \circ g^{-1}, \qquad K = g(K_m^0). \tag{3.5}$$

Applying g to the known origin and basis vectors gives an affine basis for K, so its points can be listed in time $2^m \,\mathrm{poly}(\ell)$ without knowing the source witness w or a basis for $W_{m,w}$.

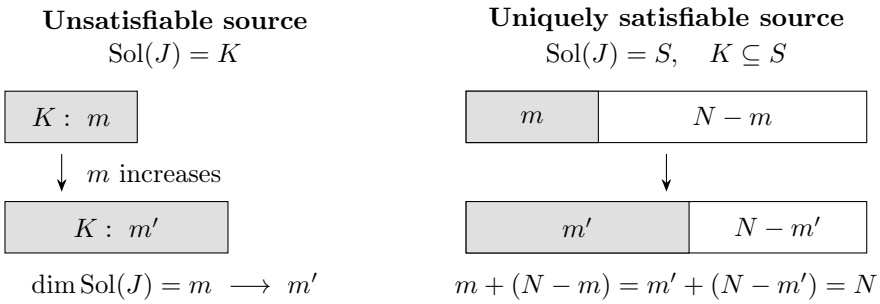

Figure 1: The dimension-balanced construction, schematically, for $m < m'$. Shading marks the known part. A larger planted dimension increases the NO-source solution dimension but leaves the YES-source dimension equal to N. The affine map g randomizes the placement while preserving dimensions. The blocks represent coordinate factors before applying g, not the relative numbers of points or a common nesting of separate trials.

Padding before choosing the scales. Fix one strictly increasing polynomial length function $L_0(\ell)$ large enough for every circuit $J_{C,m,g}$, uniformly over all length-ℓ sources, all $m \le N$, and all g. Pad the descriptions to this length:

$$|\langle J_{C,m,g}\rangle| = L_0(\ell). \tag{3.6}$$

The bound covers every $m \le N$ and is fixed before the maximum tested dimension d is chosen. Listing the points of K is part of the test and does not enlarge the isolator's input. This matters because the success guarantee depends on description length: a more explicit construction, with one copy of the source per known witness, would make the length grow with the number of points tested.

Lemma 3.1 (Distribution of the planted subspace)**.** *Fix $C \in \mathcal{C}_\ell^{(1)}$. For any fixed m, the set $S = g(W_{m,w})$ is uniform among affine N-subspaces of $\mathbb{F}_2^{3N}$, and this distribution does not depend on m. Conditional on S and m, the set $K = g(K_m^0)$ is uniform among affine m-subspaces of S.*

Proof. The affine group acts transitively on pairs (K, S) satisfying $K \subseteq S$, $\dim K = m$, and $\dim S = N$. To map one pair to another, choose an origin in K and extend a basis of its direction space first to one for S, and then to an ambient basis. The map between the chosen origins and bases is the required affine bijection.

A uniform group element therefore sends the original pair to a uniform pair of these dimensions: each pair has the same number of preimages. The marginal on S is uniform, and conditional on S, so is the choice of K. Because the marginal on S is the same for every m, it is also independent of any initial random choice of m. □

3.2 Preprocessing maps and the test

A preprocessing map changes the description of a circuit without changing its Boolean function. We may select a preprocessing map for a distribution of source inputs, but we then apply that same transformation to every input drawn from the distribution. The choice may depend on the distribution and the length, but not on which source is drawn.

Definition 3.2 (Admissible preprocessing maps)**.** For each source length ℓ, let $\mathcal{T}_\ell$ be a nonempty finite family of deterministic transformations. Each transformation has a binary description of at most $b(\ell)$ bits, where $b \ge 1$ is a fixed polynomial. A common interpreter evaluates all descriptions

within a fixed polynomial time bound. For every $T \in \mathcal{T}_\ell$ and every presentation $J_{C,m,g}$ arising from a promised source input,

$$T(J_{C,m,g}) \equiv J_{C,m,g}.$$

All transformed circuits have the same ordered variables as the original and a common padded length $L_1(\ell) \geq \ell$, polynomial in ℓ. The description, time, and length bounds are fixed for the whole range $m \leq N$, before the maximum tested dimension is chosen.

Admissibility restricts the family for the purposes of the analysis; no algorithm has to decide whether a given description is admissible. For example, one may take all function-preserving transformations within fixed size and output-length bounds. Those bounds are chosen before any source distribution or common hard list. The singleton family containing only the identity has $L_1 = L_0$.

Set $p_\ell = p(L_1(\ell))$ and fix $2 \leq d \leq N$ with $d = O(\log \ell)$. A trial first chooses a planted dimension $M \sim \mathrm{Unif}(\{0, \ldots, d\})$ and an independent affine map $g \sim \mathrm{Unif}(\mathrm{AGL}(3N, 2))$. It forms $J = J_{C,M,g}$, applies the chosen preprocessing T, and runs the isolator with fresh randomness ρ. The test on the resulting circuit B is

$$\mathrm{Test}_K(B) = \mathbf{1}_{\{|\,\mathrm{Sol}(B) \cap K| = 1\}}, \qquad K = g(K^0_M). \tag{3.7}$$

The test accepts when exactly one of the known points survives. All evaluations within a trial use the same sampled output circuit; the isolator is not rerun for each point. The trial's acceptance probability is

$$a_\ell(C, T) = \Pr_{M,g,\rho}\big[\mathrm{Test}_K(A(T(J_{C,M,g}); \rho)) = 1\big]. \tag{3.8}$$

The test evaluates at most 2^d known points. If $C \in \mathcal{C}^{(0)}_\ell$, the constructed satisfying set is exactly K. Preprocessing preserves this set, and every successful isolation passes the test. Thus, for all admissible T,

$$a_\ell(C, T) \geq p_\ell \qquad (C \in \mathcal{C}^{(0)}_\ell). \tag{3.9}$$

The guaranteed high acceptance is on NO sources. To decide the promise we will push acceptance on YES sources lower, so passing this auxiliary test should not be confused with the source being a YES instance of USAT.

Lemma 3.3 (Affine flag bound)**.** *Let S be a finite affine space over $\mathbb{F}_2$, and let $0 \leq d \leq \dim S$. For each m, let K_m be a uniform affine m-subspace of S. Then, for every fixed $R \subseteq S$,*

$$\sum_{m=0}^{d} \Pr[|R \cap K_m| = 1] \leq 2. \tag{3.10}$$

In particular, a uniform choice of dimension gives acceptance probability at most $2/(d+1)$.

Proof. The sum depends only on the marginal distribution at each dimension, so we may analyze all dimensions on one random flag. Choose K_d uniformly, then repeatedly choose a uniform affine hyperplane inside the preceding space. The resulting sequence $K_d \supset K_{d-1} \supset \cdots \supset K_0$ has the required uniform marginals by affine symmetry.

Along this descending flag, the intersection with R can only shrink. If it reaches a singleton, each subsequent hyperplane retains that point with conditional probability $1/2$. After the point is lost, every later intersection is empty. Conditional on ever reaching a singleton, the expected number of singleton levels is therefore at most $1 + 1/2 + 1/4 + \cdots = 2$. The sum in (3.10) is exactly the unconditional expected number of such levels, which proves the bound. □

When R is a singleton and $d = \dim S$ the sum equals $2 - 2^{-d}$, so the constant 2 cannot be improved in general. Appendix A.2 looks at the regime $d \ll \dim S$.

3.3 The distributional criterion

We now drop invariance. The first step is to bound how strongly the output must depend on the planted choice for the test to accept often. Fix a source distribution $\mu \in \Delta(\mathcal{C}_\ell^{(1)})$ and a preprocessing T, run the trial with $C \sim \mu$, and put

$$D = \mathrm{Sol}(A(T(J_{C,M,g}); \rho)), \qquad S = g(W_{M,w}).$$

Define

$$\mathcal{I}_\ell(\mu, T) = I((M, K); D \mid C, S). \tag{3.11}$$

It measures how much the surviving set reveals about the planted pair (M, K) once the source circuit and the constructed solution space are fixed. The quantity appears only in the analysis; the test never reconstructs D as a set or finds a basis for S.

Lemma 3.4 (An information bound for the acceptance probability)**.** *Let $\alpha = \mathbb{E}_{C\sim\mu}[a_\ell(C, T)]$ and $p_{\mathrm{ind}}(d) = 2/(d + 1)$. Then*

$$\mathcal{I}_\ell(\mu, T) \geq \alpha \log \frac{d+1}{2} - h_2(\alpha) \geq \alpha \log \frac{d+1}{2} - 1. \tag{3.12}$$

If $\alpha \geq p_{\mathrm{ind}}(d)$, then also $\mathcal{I}_\ell(\mu, T) \geq \mathrm{kl}(\alpha \| p_{\mathrm{ind}}(d))$.

Proof. Compare the actual joint distribution P of (C, S, M, K, D) with a reference distribution $P^{\otimes}$. The reference keeps the same (C, S) marginal and the same conditional marginals of (M, K) and D, but samples these last two objects independently conditional on C, S. By the definition of conditional mutual information,

$$D_{\mathrm{KL}}(P \| P^{\otimes}) = \mathcal{I}_\ell(\mu, T).$$

The acceptance event is $\mathsf{Acc} = \{|D \cap K| = 1\}$, with $P(\mathsf{Acc}) = \alpha$. Under $P^{\otimes}$, condition on C, S, D. The dimension is still uniform, and K is a uniform subspace of that dimension, now independent of D. Hence the affine flag bound gives $\beta := P^{\otimes}(\mathsf{Acc}) \leq p_{\mathrm{ind}}(d)$.

It remains to compare the probabilities of this event in the two distributions. The log-sum inequality, applied to Acc and its complement, yields

$$\mathcal{I}_\ell(\mu, T) \geq \mathrm{kl}(\alpha \| \beta) = -h_2(\alpha) + \alpha \log(1/\beta) + (1 - \alpha) \log(1/(1 - \beta)).$$

Discard the last nonnegative term and use $\beta \leq p_{\mathrm{ind}}(d)$ to obtain (3.12). When $\alpha \geq p_{\mathrm{ind}}(d)$, binary relative entropy decreases as its second argument increases toward α, which gives the last assertion. Boundary cases follow by continuity. □

Theorem 3.5 (Distributional preprocessing criterion)**.** *Let A be a randomized polynomial-size pruning procedure with success at least p_ℓ on nonempty affine-subspace inputs of length $L_1(\ell)$. Fix an admissible family $\mathcal{T}_\ell$ and $2 \leq d \leq N$ with $d = O(\log \ell)$. Suppose p_ℓ and γ_ℓ are inverse-polynomially bounded below, $0 < \gamma_\ell < p_\ell$, and, at every sufficiently large length,*

$$\text{for every } \mu \in \Delta(\mathcal{C}_\ell^{(1)}) \text{ there is } T \in \mathcal{T}_\ell \text{ with } \mathbb{E}_{C\sim\mu}[a_\ell(C, T)] \leq p_\ell - \gamma_\ell. \tag{3.13}$$

Then NP $\subseteq$ P/poly. *A list of $O((\ell + 1)/\gamma_\ell^2)$ preprocessing maps suffices, and their descriptions occupy $O(b(\ell)(\ell + 1)/\gamma_\ell^2)$ advice bits.*

In particular, (3.13) *follows if there is a common information bound $I_{\mathrm{bd}}(\ell) \geq 0$ such that*

$$\begin{gathered}\text{for every } \mu \in \Delta(\mathcal{C}_\ell^{(1)}) \text{ there is } T \in \mathcal{T}_\ell \text{ with } \mathcal{I}_\ell(\mu, T) \leq I_{\mathrm{bd}}(\ell), \\ \frac{I_{\mathrm{bd}}(\ell) + 1}{\log((d+1)/2)} \leq p_\ell - \gamma_\ell.\end{gathered} \tag{3.14}$$

Proof. A mixed preprocessing strategy that works for every source. If $\mathcal{C}_\ell^{(1)}$ is empty, every promised source at that length is a NO instance. Otherwise form the finite matrix with rows indexed by preprocessing maps, columns by $C \in \mathcal{C}_\ell^{(1)}$, and entry $a_\ell(C,T)$. The preprocessing player minimizes acceptance, while the source player maximizes it. The assumed distributional bound and minimax give

$$\begin{aligned} v_\ell &:= \min_{\lambda\in\Delta(\mathcal{T}_\ell)} \max_{C\in\mathcal{C}_\ell^{(1)}} \mathbb{E}_{T\sim\lambda}[a_\ell(C,T)] \\ &= \max_{\mu\in\Delta(\mathcal{C}_\ell^{(1)})} \min_{T\in\mathcal{T}_\ell} \mathbb{E}_{C\sim\mu}[a_\ell(C,T)] \le p_\ell - \gamma_\ell. \end{aligned} \tag{3.15}$$

A polynomial-size list. Sample t preprocessing maps independently from an optimal mixed strategy λ. For a fixed source, the probability that the empirical average exceeds its mean by $\gamma_\ell/2$ is at most $\exp(-t\gamma_\ell^2/2)$, by Hoeffding's inequality. There are at most 2^ℓ source descriptions. A union bound therefore shows that choosing

$$t > \frac{2(\ell+1)\ln 2}{\gamma_\ell^2}$$

gives a fixed list $T_1,\dots,T_t$ whose mean acceptance is at most $p_\ell - \gamma_\ell/2$ for every $C \in \mathcal{C}_\ell^{(1)}$. The sample size may be chosen to be a power of two from the outset.

The decision procedure. Store this list as advice. Choose a uniform entry and run the test with fresh affine and isolator randomness. Every NO source has acceptance at least p_ℓ by (3.9); every YES source has acceptance at most $p_\ell - \gamma_\ell/2$. The storage cost is the one stated in the theorem, and each trial is polynomial-time because $2^d = \ell^{O(1)}$. Approximate sampling changes the probabilities by only $2^{-\Omega(\ell)}$. Repetition estimates the acceptance probability well enough to distinguish the two intervals. A high value is interpreted as NO. Fixing the randomness at each length gives circuits for the promise problem, and Lemma 2.2 yields NP $\subseteq$ P/poly.

The sufficient information condition. Rearranging Lemma 3.4 gives

$$\mathbb{E}_{C\sim\mu}[a_\ell(C,T)] \le \frac{\mathcal{I}_\ell(\mu,T)+1}{\log((d+1)/2)}.$$

This proves that (3.14) is sufficient. □

The small-support step is the finite-game theorem of Lipton and Young [13]; the proof above supplies the concentration calculation. Selecting a list entry uses $O(\log \ell + \log(1/\gamma_\ell))$ random bits, but ordinary advice must store the whole list. The isolator's advice and field data add only polynomial storage. The information bound is used to bound acceptance before minimax is applied; no linearity of information is assumed.

Bibliographic note. A related quantifier exchange appears in the advice-complexity manuscript [19, Section 4]. The present argument is self-contained and uses the published small-support theorem.

For constant $p > 0$, take $d = \lfloor \log \ell \rfloor$. Suppose each YES-source distribution admits a preprocessing with $\mathcal{I}_\ell(\mu,T) \le I_{\mathrm{bd}}(\ell)$, for one common bound $I_{\mathrm{bd}}(\ell) = o(\log\log \ell)$. The ratio in (3.14) is then $o(1)$, so Theorem 3.5 applies with $\gamma_\ell = p/2$. So the criterion tolerates some leakage of information, even an amount that grows with the length, and exact invariance is not needed.

Corollary 3.6 (Canonical preprocessing within a fixed-source family)**.** *Suppose that, at every sufficiently large source length, one admissible preprocessing T_ℓ satisfies the following condition: for each fixed $C \in \mathcal{C}_\ell^{(1)}$ and all $m, m' \le N$ and affine bijections g, g',*

$$J_{C,m,g} \equiv J_{C,m',g'} \quad\Longrightarrow\quad \langle T_\ell(J_{C,m,g})\rangle = \langle T_\ell(J_{C,m',g'})\rangle. \tag{3.16}$$

Let A be any randomized polynomial-size pruning procedure with isolation probability at least $a/\log L$ on nonempty affine-subspace inputs of length L, for a fixed $a > 0$. Then NP $\subseteq$ P/poly. *No presentation invariance of A is required.*

Proof. Fix C, S. By (3.16), every possible planted choice gives the same transformed description. Even a syntax-dependent isolator therefore has the same output distribution for every such choice, and fresh randomness gives $D \perp\!\!\!\perp (M, K) \mid C, S$.

Set $p_0 = a/\log L_1(\ell)$ and $d = \lceil 8/p_0 \rceil$. Because the common transformed length was fixed before d, we have $d = O(\log \ell)$; for sufficiently large lengths, the test is within the permitted range and runs in polynomial time. The affine flag bound gives $a_\ell(C, T_\ell) \le 2/(d+1) \le p_0/4$ for every $C \in \mathcal{C}_\ell^{(1)}$. The one-element preprocessing family $\{T_\ell\}$ therefore meets Theorem 3.5 with $\gamma_\ell = p_0/2$. □

This hypothesis constrains the preprocessing map, not the isolator. One transformation T_ℓ must work for every source of length ℓ, but its outputs need agree only across equivalent presentations of the same source. It may assign different descriptions to the same set when the sources differ.

Remark 3.7 (Global canonicalization already decides unique satisfiability). Global canonicalization of all raw presentations, including those from NO sources, is stronger and by itself implies NP $\subseteq$ P/poly. For a promised source C, compare the canonical descriptions of $J_{C,0,\mathrm{id}}$ and $J_{C_\emptyset,0,\mathrm{id}}$, where $C_\emptyset$ is a known unsatisfiable source padded to the same length. They are equal exactly when C is unsatisfiable: otherwise their satisfying sets have cardinalities 2^N and one, and an equivalence-preserving map cannot give them the same output. Lemma 2.2 applies. The fixed-source hypothesis in (3.16) does not require agreement across these two sources.

3.4 The presentation-invariant special case

Presentation invariance gives the same independence without preprocessing: the identity preprocessing already has low acceptance on YES sources. We first state the reference-distribution argument precisely. Its finite-order version will also be used for tests with a fixed query budget.

Lemma 3.8 (Replacement by a reference distribution)**.** *Fix $C \in \mathcal{C}_\ell^{(1)}$ and an associated satisfying space S. Choose one padded presentation of S, and let D_0 have the output distribution of A on that presentation, with fresh isolator randomness. Both the chosen presentation and the law of D_0 are understood to depend on the fixed pair (C, S). Under presentation invariance, the actual surviving set D satisfies*

$$D \perp\!\!\!\perp (M, K) \mid C, S, \qquad \mathrm{Law}(D \mid C, S) = \mathrm{Law}(D_0). \tag{3.17}$$

If only k-wise retention invariance is assumed and $2^m \le k$, then, for fixed dimension m, replacement by the reference distribution remains valid for $D \cap K$. In particular,

$$\Pr[|D \cap K| = 1 \mid C, S, M = m] = \mathbb{E}_{D_0}\big[\Pr_K[|D_0 \cap K| = 1 \mid C, S, M = m]\big], \tag{3.18}$$

where D_0 and the uniform affine m-subspace K are independent on the right.

Proof. All the presentations have the same variables, padded length, and satisfying set S. Full invariance therefore fixes the conditional output distribution for every (m, g), which proves (3.17).

For the finite-order assertion, first fix an actual (m, g) and its planted set K. The joint law of the 2^m membership bits on K agrees with the corresponding law for the reference presentation. The event $|D \cap K| = 1$ consequently has the same probability. Average over the uniform K from Lemma 3.1 to obtain (3.18). The reduction need not find the reference presentation; it serves only to analyze these probabilities. □

Proof of Corollary 1.3. The nontrivial implication (ii)⇒(i). Use the family containing the identity map, so $L_1 = L_0$, and put $p_0 = a/\log L_0(\ell)$ and $d = \lceil 8/p_0 \rceil$. For large ℓ, $2 \le d \le N$ and $d = O(\log \ell)$. By the reference-distribution lemma and the affine flag bound, every $C \in \mathcal{C}_\ell^{(1)}$ satisfies

$$a_\ell(C, \mathrm{id}) \le \frac{2}{d+1} \le \frac{p_0}{4}. \tag{3.19}$$

The general criterion holds with $\gamma_\ell = p_0/2$, so Theorem 3.5 gives the collapse. Each test uses at most $2^d \le 2L_0(\ell)^{8/a}$ evaluations. Every queried circuit on a promised source input is a nonempty affine-subspace input, so no success or invariance hypothesis outside that class is used. Although the information in (3.11) is zero in this case, we use the exact independence bound (3.19), not the coarser entropy estimate.

The reverse implications. Assume (i). Use SAT decision circuits to determine whether the input is satisfiable; otherwise output the zero circuit. On a satisfiable input, choose witness bits from left to right. Choose zero whenever the current prefix has a satisfying extension with that next bit zero, and choose one otherwise. These polynomially many SAT queries have polynomial length, so composing their decision circuits gives a polynomial-size transformation. It outputs the point predicate of the lexicographically least witness, which depends only on the satisfying set. This proves (iii). Finally, (iii)⇒(ii) holds with $a = 1$ for sufficiently large lengths. □

Corollary 3.9 (Success probability and circuit size)**.** *Under the affine-input invariance hypothesis of Corollary 1.3(ii), a success guarantee $p(L) \ge a/(\log L)^{a_{\exp}}$, for fixed $a > 0$ and $a_{\exp} \ge 1$, implies*

$$\mathrm{NP} \subseteq \mathrm{SIZE}\big(2^{O((\log n)^{a_{\exp}})}\big). \tag{3.20}$$

Here n is the input length of the language being decided; the implicit constant may depend on that language, A, a, and $a_{\exp}$.

Proof. Take $d = \lceil 8(\log L_0(\ell))^{a_{\exp}}/a \rceil < N$ for sufficiently large ℓ. The same identity test separates the source cases, but listing K now costs $2^{O((\log \ell)^{a_{\exp}})}$. Amplification, fixing the randomness, and the larger-size statement in Lemma 2.2 preserve this bound. We use the test directly: when $a_{\exp} > 1$, its running time need not satisfy the polynomial-time hypothesis of the general preprocessing theorem. □

3.5 Test stability under preprocessing

The comparison with a reference presentation needs control of one test, not statistical closeness of the whole output set. We state this weaker condition first. It also permits randomized preprocessing without changing Definition 3.2: a random preprocessing is a probability distribution over the deterministic maps in that definition.

For $C \in \mathcal{C}_\ell^{(1)}$ and an associated satisfying space S, let $\mathcal{J}(C, S, d)$ be the family of raw padded presentations $J_{C,m,g}$ with $m \le d$ and $g(W_{m,w}) = S$. Choose a reference $J_0(C, S) \in \mathcal{J}(C, S, d)$, depending only on C, S.

Lemma 3.10 (Replacement at the level of the test)**.** *Let A be a randomized polynomial-size pruning procedure with isolation guarantee p_ℓ on nonempty affine-subspace inputs at length $L_1(\ell)$. Fix an admissible preprocessing family $\mathcal{T}_\ell$, a distribution $\lambda_\ell \in \Delta(\mathcal{T}_\ell)$, and $2 \le d \le N$. Suppose that for every $C \in \mathcal{C}_\ell^{(1)}$ and every $J = J_{C,m,g} \in \mathcal{J}(C, S, d)$, with $K = g(K_m^0)$,*

$$\Big| \Pr_{T \sim \lambda_\ell, \rho} [\mathrm{Test}_K(A(T(J); \rho)) = 1] - \Pr_{T \sim \lambda_\ell, \rho} [\mathrm{Test}_K(A(T(J_0(C, S)); \rho)) = 1] \Big| \le \varepsilon_{\mathrm{test}}(\ell). \tag{3.21}$$

The map T and the isolator randomness are sampled independently of the planted choice in both experiments. Then

$$\begin{aligned} &\mathbb{E}_{T\sim\lambda_\ell}[a_\ell(C,T)] \geq p_\ell && (C \in \mathcal{C}_\ell^{(0)}), \\ &\mathbb{E}_{T\sim\lambda_\ell}[a_\ell(C,T)] \leq \frac{2}{d+1} + \varepsilon_{\mathrm{test}}(\ell) && (C \in \mathcal{C}_\ell^{(1)}). \end{aligned} \tag{3.22}$$

If $d = O(\log \ell)$ and $p_\ell - 2/(d+1) - \varepsilon_{\mathrm{test}}(\ell)$ is eventually bounded below by an inverse polynomial, then NP $\subseteq$ P/poly.

Proof. Every deterministic map in the family preserves the function, so (3.9) gives the first inequality. For a YES source, condition on C, S. Run A on a fresh transformation $T(J_0(C,S))$, and call its surviving set D_0. The reference presentation and the distribution λ_ℓ do not depend on (M,K) after C,S are fixed. Hence $D_0 \perp\!\!\!\perp (M,K) \mid C,S$, and $D_0 \subseteq S$. The affine flag bound gives $\Pr[|D_0 \cap K| = 1 \mid C,S] \leq 2/(d+1)$. Averaging (3.21) proves the second inequality.

Choose an inverse-polynomial γ_ℓ smaller than the resulting gap and than p_ℓ. For any source distribution μ, the average over T is at most $p_\ell - \gamma_\ell$. Some deterministic T therefore satisfies (3.13) for that μ. Theorem 3.5 applies. This last step requires only existence of λ_ℓ, not an efficient sampler for it. □

The statistical criterion is a special case. Define the diameter of the output-set laws on equivalent raw presentations by

$$\delta_{A,d}(C,S) = \max_{H,H' \in \mathcal{J}(C,S,d)} d_{\mathrm{TV}}\big(\mathrm{Law}(D_A(H)), \mathrm{Law}(D_A(H'))\big). \tag{3.23}$$

This quantity compares surviving sets rather than output encodings.

Corollary 3.11 (Approximate-invariance criterion)**.** *Let A be a randomized polynomial-size pruning procedure with success at least p_0 at length $L_0(\ell)$. Suppose that $\delta_{A,d}(C,S) \leq \delta_\ell$ for every associated YES source and space. The uniform-scale test satisfies*

$$\begin{aligned} &\Pr[|D \cap K| = 1] \geq p_0 && (C \in \mathcal{C}_\ell^{(0)}), \\ &\Pr[|D \cap K| = 1] \leq 2/(d+1) + \delta_\ell && (C \in \mathcal{C}_\ell^{(1)}). \end{aligned} \tag{3.24}$$

If $d = O(\log \ell)$ and $p_0 - 2/(d+1) - \delta_\ell$ is bounded below by an inverse polynomial, eventually at every length, then NP $\subseteq$ P/poly.

Proof. Use the family containing the identity map and fix a reference presentation for each C, S. A total-variation bound of δ_ℓ controls the acceptance difference for every event, including $\mathrm{Test}_K \circ A$. Thus (3.21) holds with $\varepsilon_{\mathrm{test}}(\ell) = \delta_\ell$. Lemma 3.10 gives the bounds and the collapse. □

For example, fix $a > 0$ and $\eta_{\mathrm{rel}} \in (0,1)$. If $p(L) \geq a/\log L$ and $\delta_{A,d}(C,S) \leq (1-\eta_{\mathrm{rel}})p(L)$ with $d = \lceil 8/(\eta_{\mathrm{rel}} p(L)) \rceil$, the gap is at least $3\eta_{\mathrm{rel}} p(L)/4$. Thus the inverse-logarithmic implication also holds with this amount of variation between equivalent presentations.

3.6 Indistinguishability obfuscation as preprocessing

Exact canonicalization is enough for the test replacement lemma, but the lemma asks for less. Indistinguishability obfuscation (iO) offers a computational substitute: no efficient test distinguishes obfuscations of equivalent circuits of equal size [15]. We use classical security against nonuniform adversaries, which lets us put both the isolator and the test's known subspace into the adversary's polynomial-size advice.

Security convention. The security parameter is λ_{sec}; it is distinct from a mixed strategy λ_ℓ. A function is negligible in λ_{sec} if it is eventually smaller than $\lambda_{\mathrm{sec}}^{-a_{\mathrm{exp}}}$ for every fixed $a_{\mathrm{exp}} > 0$. The obfuscator's coins are ρ_{obf}, independent of the isolator's coins ρ. We give the sampler formulation explicitly; [16, Definition 3.2] states the corresponding formulation even against quantum adversaries, which is stronger than needed here.

Definition 3.12 (Nonuniformly secure indistinguishability obfuscation)**.** An obfuscator iO is a uniform randomized polynomial-time algorithm that, on $(1^{\lambda_{\mathrm{sec}}}, J)$, outputs a circuit on the same ordered variables. Its running time, output length, and number of random bits are polynomial in λ_{sec} and the input-description length. We require:

(i) *Statistical correctness.* There is a negligible function $\varepsilon_{\mathrm{corr}}$ such that, for every circuit J of at most λ_{sec} gates,

$$\Pr_{\rho_{\mathrm{obf}}}\big[\mathsf{iO}(1^{\lambda_{\mathrm{sec}}}, J; \rho_{\mathrm{obf}}) \not\equiv J\big] \le \varepsilon_{\mathrm{corr}}(\lambda_{\mathrm{sec}}).$$

The error bound is uniform over these circuits. The event means that the obfuscated circuit is wrong on at least one input, rather than on an independently sampled input.

(ii) *Indistinguishability.* Let $(\mathsf{Samp}, \mathsf{Dist})$ be any pair of nonuniform randomized polynomial-size circuit families. Suppose $\mathsf{Samp}(1^{\lambda_{\mathrm{sec}}})$ always outputs (J_0, J_1, aux), where J_0, J_1 are equivalent circuits on the same variables, with the same number of gates, at most λ_{sec} gates, and descriptions of polynomial length. The auxiliary string aux has polynomial length. Then the two ensembles

$$(\mathsf{aux}, \mathsf{iO}(1^{\lambda_{\mathrm{sec}}}, J_0)) \quad \text{and} \quad (\mathsf{aux}, \mathsf{iO}(1^{\lambda_{\mathrm{sec}}}, J_1)) \tag{3.25}$$

have negligible acceptance-probability difference for Dist. In both experiments the sampler is run first, and the obfuscator then uses fresh independent coins. The negligible bound may depend on the two adversary families, but not on the sampled pair.

Remark 3.13 (Statistical and pointwise correctness). Functional error $\varepsilon_{\mathrm{corr}}$ changes the NO-source bound only to $p_0(1 - \varepsilon_{\mathrm{corr}})$; the reference flag bound is unchanged when applied to $D_0 \cap S$. The proof below tracks this error. Pointwise correctness also suffices: once the error on every fixed input is uniformly at most $1/4$, the pointwise majority of $t_{\mathrm{amp}} = 8(n_{\mathrm{in}} + \lambda_{\mathrm{sec}}) + 1$ independent obfuscations of an n_{in}-variable circuit has whole-function error at most

$$2^{n_{\mathrm{in}}} e^{-t_{\mathrm{amp}}/8} \le e^{-\lambda_{\mathrm{sec}}},$$

by Hoeffding and a union bound. A polynomial hybrid preserves iO security. This uses correctness on every fixed input, not merely most inputs, and does not produce perfect correctness; compare [18].

Averaging over many equivalent presentations requires the security bound to be uniform. The next lemma shows that this follows from nonuniform security; no further assumption on the obfuscator is made.

Lemma 3.14 (A uniform bound for bounded auxiliary tests)**.** *Fix polynomial bounds on descriptions and auxiliary strings, and a nonuniform polynomial-size randomized distinguisher family* Dist*. Under Definition 3.12, there is one negligible function bounding its acceptance-probability difference for every equivalent, equal-size pair* (J_0, J_1) *of at most* λ_{sec} *gates within those bounds and every allowed fixed auxiliary string* aux *at each security parameter.*

Proof. For each parameter, maximize the absolute acceptance difference over the finite set of allowed triples. If these maxima were not negligible, at infinitely many parameters a maximizing triple would have inverse-polynomial advantage. A nonuniform sampler can store that triple in its advice and output it deterministically. At the remaining parameters it outputs two identical circuits. This gives one admissible sampler family violating iO security with the fixed distinguisher. Hence the maxima are negligible. The argument uses neither an algorithm for finding the triples nor a union bound over them. □

Proof of Theorem 1.1. Fix the lengths before choosing the test. For each source length ℓ, choose a strictly increasing polynomial $\lambda_{\mathrm{sec}} = \lambda_{\mathrm{sec}}(\ell) \geq L_0(\ell)$ large enough for the following padding. The deterministic map $\mathrm{pad}^{\mathrm{in}}_\ell$ makes all raw $J_{C,m,g}$, over the whole range $m \leq N$, have the same gate count, at most λ_{sec}, and the same description length. It preserves their variables and functions. Equal description length alone would not ensure equal gate count.

The polynomial output bound gives a common padded length $L_1(\ell) = \mathrm{poly}(\ell)$ for all obfuscations, uniformly over the source, m, g, and the obfuscator coins. Define the randomized preprocessing

$$\mathsf{Obf}_\ell(J; \rho_{\mathrm{obf}}) = \mathrm{pad}^{\mathrm{out}}_\ell\big(\mathsf{iO}(1^{\lambda_{\mathrm{sec}}(\ell)}, \mathrm{pad}^{\mathrm{in}}_\ell(J); \rho_{\mathrm{obf}})\big). \tag{3.26}$$

All seeds have a common polynomial length $b_{\mathrm{obf}}(\ell)$. With statistical correctness, a seed-fixed map need not be admissible in Definition 3.2. We therefore keep the coins random and analyze the resulting decision procedure directly.

Only now set

$$p_0 = \frac{a}{\log L_1(\ell)}, \qquad d = \left\lceil \frac{8}{p_0} \right\rceil. \tag{3.27}$$

For sufficiently large ℓ, $2 \leq d \leq N$ and $d = O(\log \ell)$. The test uses at most $2^d \leq 2L_1(\ell)^{8/a}$ output evaluations. The hidden constant in the polynomial running time may depend on the fixed success constant a, but not on the source circuit. Let its acceptance probability, with fresh independent coins for the obfuscator and the isolator, be

$$a^{\mathrm{obf}}_\ell(C) = \Pr_{M,g,\rho_{\mathrm{obf}},\rho}[\mathrm{Test}_K(A(\mathsf{Obf}_\ell(J_{C,M,g}; \rho_{\mathrm{obf}}); \rho)) = 1], \qquad K = g(K^0_M). \tag{3.28}$$

Apply security to the actual efficient test. Fix a YES source C, its solution space S, and a reference $J_0(C,S) \in \mathcal{J}(C,S,d)$. For an actual presentation $J = J_{C,m,g}$ of S, the padded input circuits for J and $J_0(C,S)$ are equivalent and have equal gate count. Give the distinguisher an affine origin and direction basis for K as auxiliary input. It pads the challenge obfuscation to $L_1(\ell)$, runs A with fresh coins, and applies Test_K to the output. It tests only K, not all of S.

The basis has $O(N(d+1))$ bits. The distinguisher has polynomial size in λ_{sec} and may hardwire A's length-specific circuit. By Lemma 3.14, one negligible function $\varepsilon_{\mathrm{iO}}(\ell)$ bounds

$$\left| \Pr_{\rho_{\mathrm{obf}},\rho}[\mathrm{Test}_K(A(\mathsf{Obf}_\ell(J; \rho_{\mathrm{obf}}); \rho)) = 1] - \Pr_{\rho_{\mathrm{obf}},\rho}[\mathrm{Test}_K(A(\mathsf{Obf}_\ell(J_0(C,S); \rho_{\mathrm{obf}}); \rho)) = 1] \right| \leq \varepsilon_{\mathrm{iO}}(\ell) \tag{3.29}$$

for all these choices at once. Negligibility transfers from λ_{sec} to ℓ because the parameters are polynomially related and $\lambda_{\mathrm{sec}} \geq \ell$. No factor proportional to the number of presentations is lost. The reference is needed only in the analysis; the reduction does not find it or reconstruct S.

Account for correctness errors. On a NO source, each raw presentation has solution set K. With probability at least $1 - \varepsilon_{\mathrm{corr}}(\lambda_{\mathrm{sec}}(\ell))$ its obfuscation computes the same function. Conditional on any such correct obfuscation, A receives a nonempty affine-subspace input and accepts the test with probability at least p_0. On other obfuscations its test acceptance is simply nonnegative.

For a YES source, condition on C, S and run A on a fresh obfuscation of $J_0(C, S)$. Let D_0 be the resulting surviving set. This random set is independent of (M, K) conditional on C, S, whether the obfuscation is correct or not. On an erroneous obfuscation D_0 need not be a subset of S, but $D_0 \cap S$ is. Since $K \subseteq S$, the flag bound applies to $(D_0 \cap S) \cap K = D_0 \cap K$ and gives acceptance at most $2/(d+1)$. Average (3.29) to obtain

$$\begin{aligned} a_\ell^{\mathrm{obf}}(C) &\geq p_0\big(1 - \varepsilon_{\mathrm{corr}}(\lambda_{\mathrm{sec}}(\ell))\big) \quad (C \in \mathcal{C}_\ell^{(0)}), \\ a_\ell^{\mathrm{obf}}(C) &\leq \frac{2}{d+1} + \varepsilon_{\mathrm{iO}}(\ell) \qquad\quad (C \in \mathcal{C}_\ell^{(1)}). \end{aligned} \tag{3.30}$$

Thus the acceptance gap is at least

$$p_0 - \frac{2}{d+1} - p_0\varepsilon_{\mathrm{corr}}(\lambda_{\mathrm{sec}}(\ell)) - \varepsilon_{\mathrm{iO}}(\ell).$$

For all sufficiently large ℓ, the NO acceptance is at least $3p_0/4$ and the YES acceptance is at most $p_0/2$. The gap is inverse-logarithmic. Independent repetition estimates the acceptance probability, with a high value interpreted as NO. The finite-group sampling error is exponentially small and does not affect this gap. Amplify the decision error and fix the *complete decision procedure's* randomness as in Lemma 2.2. This gives USAT circuits of polynomial size and hence NP $\subseteq$ P/poly. □

The argument is the test-level replacement from Lemma 3.10, with an explicit allowance for functional error and an efficient randomized sampler. It does not invoke that lemma's admissibility hypothesis for erroneous seeds. Under perfect correctness, every seed-fixed map in (3.26) is admissible; averaging over seeds for each source distribution then recovers a literal application of Theorem 3.5. With statistical correctness, fixing the final decision randomness avoids any need to find or recognize individually correct obfuscator seeds.

We never assume that a fixed-seed obfuscator is secure, and we do not infer from iO that surviving-set laws are statistically close or carry little Shannon information. Security enters only through (3.29), and exact independence is used only in the reference experiment. Unlike the global canonicalization of Remark 3.7, iO gives no equality test for obfuscated descriptions and no way to compare the inequivalent singleton and large affine targets.

Proof of Corollary 1.2. An isolator as in the corollary would imply NP $\subseteq$ P/poly by Theorem 1.1. Let f be a polynomial-time function. The predicate on $(1^n, y, u)$ asking whether there exists $x \in \{0,1\}^n$ extending u with $f(x) = y$ belongs to NP; the unary length bound makes the witness length polynomial in the predicate's input. Polynomial-size circuits for this predicate allow successive prefix queries to recover a preimage of every image of f at that input length. The resulting inverse has polynomial size, contradicting one-wayness against nonuniform polynomial-size adversaries. The one-wayness assumption is not used elsewhere in the obfuscation proof. □

A restricted preprocessing problem. The proof uses iO only on the padded affine presentations from one fixed source, with the planted K available as auxiliary information. This raises the question of whether a function-preserving randomized preprocessing for this restricted family can satisfy (3.21) under a weaker assumption, such as learning with errors (LWE), without obfuscating general circuits. The input must be the circuit itself, not a basis of its solution space. The condition could be weakened further to an average test gap as in Theorem 3.5.

Two known obfuscation notions do not give an immediate construction. Subspace-hiding obfuscation takes a subspace basis as input [16, Section 3.4]. Here obtaining such a basis for a YES space would already reveal the unique source witness: after undoing g, a basis yields a point outside

K_m^0 when $m < N$, with field coordinates $\alpha \neq 0$ and $z = \alpha w$, hence $w = \alpha^{-1} z$. The LWE-based compute-and-compare obfuscator of Wichs and Zirdelis requires suitable conditional pseudoentropy of the comparison target, given the computation and the allowed auxiliary information [17]. In the direct encoding $\mathbf{1}_{\{f(x)=y\}}$, an accepting point $x \in K$ determines the target as $y = f(x)$. Thus the known planted region defeats the required hiding condition. Whether a dedicated preprocessing of this kind exists is left open; we do not claim a construction from LWE.

Remark 3.15 (The planted dimension and the planted location). The multiscale comparison concerns the pair (M, K), not merely the coordinate map g. For any distinct field elements w, w', the spaces in (3.4) satisfy

$$W_{m,w} \cap W_{m,w'} = K_m^0.$$

Indeed, a point in both spaces satisfies $\alpha(w - w') = 0$, hence $\alpha = 0$. Every bijection preserving $W_{m,w}$ for all w must therefore preserve K_m^0 as a set. In particular, witness-independent shears $(x_0, \alpha, z) \mapsto (x_0 + \Lambda(\alpha), \alpha, z)$, with linear $\Lambda : \mathbb{F} \to U_m$, fix K_m^0 pointwise. Composing g with such a symmetry changes its description but leaves the planted set K unchanged. A single bijection cannot map every $W_{m,w}$ to $W_{m',w}$ when $m \neq m'$: it would have to map their common intersections, of sizes 2^m and $2^{m'}$, bijectively.

There is also a computational obstruction to explicit resampling. Suppose two equivalent presentations of the *same* uniquely satisfiable source, $J_{C,m,g}$ and $J_{C,m',g'}$, are supplied together with m, g, m', g', and have different planted spaces K, K'. Affine linear algebra finds a point in their symmetric difference. For $x \in K' \setminus K$, the coordinates $g^{-1}(x) = (x_0, \alpha, z)$ satisfy $\alpha \neq 0$ and $z = \alpha w$, so $w = \alpha^{-1} z$. The other difference is handled with g'^{-1}. Thus an explicit change of the planted region reveals the source witness. This statement does not apply to an arbitrary output circuit with no exposed affine decomposition.

Neither component may be ignored. If a reference output can depend freely on m, it can choose a different retained set at every scale, and the single-scale lower bound in Proposition 4.2(a) prevents the flag argument from giving a vanishing bound. Conversely, selecting a uniform point of K gives a singleton that is uniform on S at every fixed m, yet passes Test_K with probability one. Hiding m alone therefore does not suffice. The iO proof compares the actual test with one whose reference output is independent of the *entire* planted pair conditional on (C, S); it does not merely randomize coordinates or claim information-theoretic secrecy of that pair.

3.7 Common distinguishing pairs

We now keep track of the explicit test when applying the dual minimax argument. This gives more than a lower bound on total variation: the same small list of presentation pairs has efficiently computable tests that work against every allowed preprocessing.

Remark 3.16 (A common hard input distribution). Assume NP $\not\subseteq$ P/poly. Fix an admissible family, $2 \leq d \leq N$ with $d = O(\log \ell)$, and an inverse-polynomial $0 < \eta_\ell < p_\ell$. Infinitely many lengths admit a uniform multiset of $O((b(\ell) + 1)/\eta_\ell^2)$ YES inputs whose distribution μ_ℓ^{hard} satisfies $\mathbb{E}_{C \sim \mu_\ell^{\mathrm{hard}}}[a_\ell(C, T)] > p_\ell - \eta_\ell$ for every $T \in \mathcal{T}_\ell$. Indeed, the game value in (3.15) exceeds $p_\ell - \eta_\ell/2$ infinitely often, or Theorem 3.5 would imply collapse. At such a length, sample a maximizing input distribution and use concentration and a union bound over at most $2^{b(\ell)+1}$ descriptions. For constant p, constant $\eta \in (0, p)$, and $d = \lfloor \log \ell \rfloor$, the same ensemble satisfies $I((M, K); D \mid C, S) \geq (p - \eta) \log((d + 1)/2) - 1 = (p - \eta) \log \log L_1(\ell) - O(1)$ for every preprocessing map, by Lemma 3.4. The distribution is existential and nonuniform, and the preprocessing size bound is fixed before it is chosen.

Both logarithmic scales come from the same place: a polynomial query budget $k = 2^d$ allows only $d = O(\log \ell)$. The information lower bound is then proportional to $\log(d+1) = O(\log \log L)$ at constant acceptance, while Theorem 4.3 matches the entire observed restriction at success of order $1/(d+1)$. The information estimate is only what this argument certifies; it is not an upper limit on what an isolator can transmit.

Theorem 3.17 (A common list of distinguishable presentation pairs)**.** *Let A be a randomized polynomial-size pruning procedure with constant success $p > 0$ on nonempty affine-subspace inputs. Assume* NP $\not\subseteq$ P/poly. *Fix an admissible preprocessing family and $\eta \in (0, p)$, and set $d = \lceil 8/\eta \rceil$. At infinitely many source lengths there is a list of $t = O((b(\ell)+1)/\eta^2)$ triples (J_i, J_i', K_i) such that:*

(i) $J_i \equiv J_i'$ have the same raw length and affine satisfying set S_i, and $K_i \subseteq S_i$ is an explicitly specified affine subspace of dimension at most d;

(ii) the list is independent of the preprocessing map, and every $T \in \mathcal{T}_\ell$ satisfies

$$\frac{1}{t}\sum_{i=1}^{t}\left(\Pr_{\rho}[\mathrm{Test}_{K_i}(A(T(J_i);\rho)) = 1] - \Pr_{\rho}[\mathrm{Test}_{K_i}(A(T(J_i');\rho)) = 1]\right) > p - \eta. \tag{3.31}$$

Each test uses at most 2^d evaluations of one output circuit. An affine origin and basis for K_i require at most $3N(d+1)$ bits. For fixed η, the query bound is constant and this side information is linear in the number of witness variables.

Proof. Choose a distribution hard for all preprocessing maps. If $v_\ell \le p - \eta/4$ eventually at every length, Theorem 3.5 would imply the excluded collapse. Thus $v_\ell > p - \eta/4$ infinitely often. Fix such a length and a maximizing distribution μ on $\mathcal{C}_\ell^{(1)}$. Every preprocessing has average test acceptance greater than $p - \eta/4$ under this one distribution.

Compare with an independent presentation of the same space. Draw $C \sim \mu$ and the actual presentation $J = J_{C,M,g}$, with its known K. Conditional on C, S, independently draw $J' = J_{C,M',g'}$ from the same conditional presentation distribution. Both circuits represent S. This sampling is used only to prove that the list exists; no efficient reconstruction of S from C is assumed.

For any fixed T, the output on $T(J')$ is independent of the original (M, K) conditional on C, S. The flag bound makes its acceptance on the original K less than $\eta/4$. The mean signed gap on (J, J', K) is therefore greater than $p - \eta/2$ for every preprocessing map. What makes this work is that the distribution of the triple does not depend on T.

Choose one small list. For a fixed triple and preprocessing, the signed gap lies in $[-1, 1]$. Sample $O((b(\ell)+1)/\eta^2)$ triples. Hoeffding's inequality and a union bound over at most $2^{b(\ell)+1}$ preprocessing descriptions give one fixed list whose average loses at most $\eta/4$ for every preprocessing map. This proves (3.31). An origin and at most d direction vectors specify each K_i; listing its points gives the test and storage bounds. □

Deduction of Corollary 1.4. Use the family containing the identity map. Some pair in the list has test gap greater than $p - \eta$. For a fixed K_i, the test is an event determined by the surviving set, so the two output distributions have total-variation distance greater than $p - \eta$. Strictly increasing raw lengths $L_0(\ell)$ give infinitely many description lengths. □

For each preprocessing map, at least one pair in this same list has the required gap; the successful index may depend on the preprocessing. The list as a whole has polynomial storage cost. The linear side-information bound applies to one selected test at fixed accuracy, not to the complete list.

4 Limitations of the planted-region test

The test in Section 3 inspects the isolator's output only on the known region K. We show that its inverse-logarithmic scale is not an artifact of choosing the singleton indicator: below that scale, valid output laws can agree on the *entire* restriction to K. We first compare scale distributions, then prove the restriction theorem.

Throughout this section, $Q = 2^N$, $k = 2^d < Q$, and all presentations have length $L_0(\ell)$. No preprocessing is applied. Finite-order consequences are collected in Appendix B; the more detailed leading-constant analysis is in Appendix A.

4.1 A common formulation for scale distributions

The identity test need not choose its dimension uniformly. To describe all choices in one notation, fix $d \le N$, $k = 2^d$, and $Q = 2^N$. Write $\Pi_d = \Delta(\{0, \dots, d\})$. For $\pi \in \Pi_d$ and $R \subseteq S$, define

$$f_{\pi,Q}(R) = \sum_{m=0}^{d} \pi_m \Pr_{K_m}[|R \cap K_m| = 1], \tag{4.1}$$

$$c_{\pi,Q} = \max_{R \subseteq S} f_{\pi,Q}(R), \qquad \sigma_{\pi,Q} = \frac{1}{Q} \sum_{m=0}^{d} \pi_m 2^m. \tag{4.2}$$

For each dimension, K_m is uniform among the affine m-subspaces of S. The value $c_{\pi,Q}$ is the largest acceptance probability over all fixed retained subsets. For a singleton, the acceptance probability is $\sigma_{\pi,Q}$, so $c_{\pi,Q} \ge \sigma_{\pi,Q}$. A point mass at m always uses that scale; the uniform distribution chooses one of the permitted scales at random in each trial. Both require at most k evaluations per trial.

Proposition 4.1 (Acceptance bound for a distribution of scales)**.** *Suppose A has k-wise retention invariance on the balanced presentations and success at least p at their common length. Choose $M \sim \pi$ and let D be its surviving set. Let $u_{\pi,Q}$ be an upper bound on the maximum acceptance probability $c_{\pi,Q}$. Thus*

$$u_{\pi,Q} \ge c_{\pi,Q} \ge \sigma_{\pi,Q}.$$

The bound $u_{\pi,Q} \ge c_{\pi,Q}$ controls all retained sets, while $u_{\pi,Q} \ge \sigma_{\pi,Q}$ makes the success–failure bound monotone in the actual isolation probability. We obtain

$$\begin{aligned} &\Pr[|D \cap K| = 1] \ge p && (C \in \mathcal{C}_\ell^{(0)}), \\ &\Pr[|D \cap K| = 1] \le p\sigma_{\pi,Q} + (1-p)u_{\pi,Q} && (C \in \mathcal{C}_\ell^{(1)}). \end{aligned} \tag{4.3}$$

Consequently these tests imply NP $\subseteq$ P/poly *when k is polynomial in ℓ, π can be sampled in polynomial time to negligible error, and*

$$p(1 + u_{\pi,Q} - \sigma_{\pi,Q}) - u_{\pi,Q} \ge 1/\operatorname{poly}(\ell) \tag{4.4}$$

for all sufficiently large lengths. The threshold certified by this particular upper bound is

$$p^\star_{\pi,Q} = \frac{u_{\pi,Q}}{1 + u_{\pi,Q} - \sigma_{\pi,Q}}. \tag{4.5}$$

Proof. Pruning and successful isolation give the NO-source bound. For a YES source, condition on C, S and choose a reference output D_0. At every permitted scale, $2^m \le k$, so Lemma 3.8 replaces the actual restriction law by the reference restriction law. Averaging over π gives $\mathbb{E}[f_{\pi,Q}(D_0)]$.

Write $p_{\mathrm{act}} = \Pr[|D_0| = 1] \geq p$. On the singleton event the test accepts with probability $\sigma_{\pi,Q}$; otherwise its acceptance is at most $u_{\pi,Q}$. The resulting bound is $p_{\mathrm{act}}\sigma_{\pi,Q} + (1 - p_{\mathrm{act}})u_{\pi,Q}$. It is nonincreasing in p_{act} because $u_{\pi,Q} \geq \sigma_{\pi,Q}$, so replacing p_{act} by p gives (4.3). The stated gap permits amplification and Lemma 2.2. □

For uniform scales and $d \geq 1$, the flag bound permits

$$u_{\mathrm{unif},Q} = \frac{2}{d+1}, \qquad \sigma_{\mathrm{unif},Q} = \frac{2^{d+1} - 1}{(d+1)Q}. \tag{4.6}$$

The notation "unif" means $\pi = \mathrm{Unif}(\{0, \ldots, d\})$. Substitution gives

$$p^{\mathrm{mix}}_{k,Q} = \frac{2}{\log k + 3 - (2k-1)/Q}, \qquad k = 2^d \geq 2. \tag{4.7}$$

This estimate also uses the isolation guarantee on the reference output; the simpler flag bound in the main proof does not need it. When $k/Q \to 0$, the leading term is $2/(\log k + 3)$. For example, $k = 32$ gives the sufficient guarantee $1/4 + 1/\operatorname{poly}(L)$. The next subsection shows how far this acceptance-bound argument can go; Appendix B.1 records its finite-order applications.

4.2 A lower bound for every distribution of dimensions

The next proposition bounds acceptance from below for some retained subset, regardless of the chosen scale distribution. The maxima range over all subsets; no circuit-size restriction is imposed.

Proposition 4.2 (Lower bounds for singleton-intersection probabilities)**.** *Let $|S| = Q = 2^N$ and $k = 2^d < Q$. Set $\beta_1 = 1$ and $\beta_j = (1 - 1/j)^{j-1}$ for $j \geq 2$.*

(a) For $j = 2^m$ with $1 \leq m \leq d$, define the exact single-scale maximum $c^*_{m,Q} = \max_{R \subseteq S} \Pr[|R \cap K_m| = 1]$. Then

$$c^*_{m,Q} \geq \beta_j, \qquad \frac{c^*_{m,Q}}{1 + c^*_{m,Q} - j/Q} \geq \frac{\beta_j}{1 + \beta_j - j/Q} \geq \frac{1}{1+e}. \tag{4.8}$$

The maximum restricted to $|R| \neq 1$ is at least $\beta_j - (1 - 1/j)^{Q-1}$. Its corresponding threshold therefore remains at least $1/(1+e) - o(1)$ when $Q/j \to \infty$.

(b) For any $\pi \in \Pi_d$, choose $m_* \in \arg\max_m \pi_m$ and write $j_* = 2^{m_*}$. Then

$$c_{\pi,Q} \geq \pi_{m_*}\beta_{j_*} > \frac{1}{e(d+1)}. \tag{4.9}$$

If singleton retained sets are excluded, define $c^-_{\pi,Q} = \max_{R \subseteq S,\ |R| \neq 1} f_{\pi,Q}(R)$. Then

$$\begin{aligned} c^-_{\pi,Q} &\geq \pi_{m_*}\big(\beta_{j_*} - (1 - 1/j_*)^{Q-1}\big) \\ &\geq \frac{e^{-1} - \exp(-(Q-1)/k)}{d+1}. \end{aligned} \tag{4.10}$$

When $Q/k \to \infty$, the last bound is $(e^{-1} - o(1))/(d+1)$, uniformly over π.

Proof. Include each point of S in a random set R independently with probability ϑ. For every fixed K_m of size j,

$$\mathbb{E}_R\big[\Pr_{K_m}[|R \cap K_m| = 1]\big] = j\vartheta(1-\vartheta)^{j-1}. \tag{4.11}$$

The identity is exact: the membership choices in R are independent even though the points of K_m have affine relations. At $\vartheta = 1/j$, the expectation is β_j, so some fixed subset attains at least that value. Since $j < Q$, the map $c \mapsto c/(1+c-j/Q)$ is increasing. Together with $\beta_j = (1+1/(j-1))^{-(j-1)} > e^{-1}$ for $j \geq 2$, this proves the displayed bounds in (a).

At density $1/j$, the contribution to (4.11) from globally singleton sets is

$$\Pr[|R| = 1]\frac{j}{Q} = (1-1/j)^{Q-1}.$$

Subtracting this contribution leaves an expectation supported on nonsingleton sets. Their maximum is at least that remaining expectation, which gives the failure-only bound. For $j = 1$, simply take $R = S$: the intersection is always a singleton, while R itself is not.

For (b), use density $1/j_*$. All terms in the scale average are nonnegative, so the term at the heaviest scale alone gives $\mathbb{E}_R[f_{\pi,Q}(R)] \geq \pi_{m_*}\beta_{j_*}$. Since $\pi_{m_*} \geq 1/(d+1)$, this proves the all-mixture bound. To exclude singleton outputs, apply the single-scale correction only to this same term. For $j_* > 1$,

$$(1-1/j_*)^{Q-1} \leq \exp(-(Q-1)/j_*) \leq \exp(-(Q-1)/k).$$

For $j_* = 1$ the correction is zero. The bracket in the last line of (4.10) is nonnegative because $k < Q$; hence $\pi_{m_*} \geq 1/(d+1)$ gives that line. □

DKMW use the same Bernoulli identity in their known-density black-box analysis [1, Theorem 4.2]. Here it provides a retained subset on which the chosen affine test accepts often. Appendix B gives the associated finite-order collapse bounds; Appendix A refines the leading constant for mixtures without changing the logarithmic order.

The best threshold certified by this method. Combining the preceding lower bound with uniform-scale sampling gives

$$\frac{1}{e(d+1)} \leq \inf_{\pi \in \Pi_d} c_{\pi,Q} \leq \frac{2}{d+1}. \tag{4.12}$$

Using the exact maxima, rather than chosen upper estimates, define the best success threshold certified by the acceptance-bound argument:

$$p^{\mathrm{opt}}_{d,Q} := \inf_{\pi \in \Pi_d} \frac{c_{\pi,Q}}{1 + c_{\pi,Q} - \sigma_{\pi,Q}}.$$

Because $\sigma_{\pi,Q} \geq 0$, each ratio is at least $c_{\pi,Q}/(1+c_{\pi,Q})$. Using uniform scales for the upper bound gives, for $d \geq 1$,

$$\frac{1}{e(d+1)+1} \leq p^{\mathrm{opt}}_{d,Q} \leq \frac{2}{d+3-(2k-1)/Q}, \qquad k = 2^d < Q. \tag{4.13}$$

The two sides have the same order. Even after optimizing the dimension distribution, this acceptance-bound argument certifies a success threshold of order $1/(d+1) = 1/(1+\log k)$.

A sharper separation of success and failure outputs does not change the order. Consider a distribution that returns a singleton with probability p and a nonsingleton set attaining $c^-_{\pi,Q}$ otherwise. Its isolation probability is exactly p, and its test acceptance is exactly $p\sigma_{\pi,Q} + (1-p)c^-_{\pi,Q}$.

Therefore an upper bound based only on these constraints cannot place YES acceptance below the NO guarantee p unless

$$p > \frac{c^-_{\pi,Q}}{1 + c^-_{\pi,Q} - \sigma_{\pi,Q}} \geq \frac{c^-_{\pi,Q}}{1 + c^-_{\pi,Q}}.$$

By (4.10), this threshold is $\Omega(1/(d+1))$ uniformly over scale distributions when $Q/k \to \infty$. These are extremal bounds on subset distributions; they do not assert succinct circuit representations of the extremizers.

A budget $k \leq L_0(\ell)^{a_{\exp}}$, for fixed $a_{\exp} > 0$, permits at most $d = O(\log L_0(\ell))$. Optimizing these bounds within that budget gives success scale $\Theta(1/\log L_0(\ell))$. To certify a vanishing guarantee p, the same argument needs $d = \Omega(1/p)$ and therefore $k = 2^{\Omega(1/p)}$, matching the construction's $2^{O(1/p)}$ budget in order. Thus faster evaluation alone would not improve this bound at a fixed number of inspected points: the worst-case acceptance estimate itself has the logarithmic limitation.

4.3 Indistinguishability on the entire planted region

Evaluating the output on all of K reveals the entire restriction $D \cap K$, not just whether its size is one. Could a different statistic of these same observations improve the test? The next theorem matches the full restriction law between the two source cases, which rules out such an improvement at the stated success scale. The matching laws are abstract subset distributions, and we make no claim that they can be implemented efficiently.

The density mixture. For $0 < \tau < 1$, set

$$z_{k,\tau} = \ln(k/\tau), \qquad \beta_{k,\tau} = \frac{(1-\tau/k)^k}{z_{k,\tau}}. \tag{4.14}$$

Choose a Bernoulli inclusion probability $\vartheta \in [\tau/k, 1]$ with density $1/(\vartheta z_{k,\tau})$. For a fixed $j = 2^m \leq k$, the resulting random subset R satisfies

$$\begin{aligned} \mathbb{E}_R\big[\Pr_{K_m}[|R \cap K_m| = 1]\big] &= \frac{1}{z_{k,\tau}} \int_{\tau/k}^1 j(1-\vartheta)^{j-1}\, d\vartheta \\ &= \frac{(1-\tau/k)^j}{z_{k,\tau}} \geq \beta_{k,\tau}. \end{aligned} \tag{4.15}$$

This elementary calculation is the only density estimate needed for the restriction theorem. The refinements in Appendix A are independent of the computational argument.

Theorem 4.3 (Matching distributions on the planted region)**.** *Fix $d \geq 1$, $k = 2^d < Q$, $0 < \tau < 1$, and*

$$0 < p \leq \frac{\beta_{k,\tau}}{1 + \beta_{k,\tau}}, \tag{4.16}$$

where $\beta_{k,\tau}$ is defined in (4.14)*. For each Q-point set S, there is a presentation-independent pruning distribution μ_S with singleton probability at least p. For every $K \subseteq S$ of size 2^m, $m \leq d$, there is a pruning distribution μ_K on subsets of K, also with singleton probability at least p, such that*

$$\mathrm{Law}(D_S \cap K) = \mathrm{Law}(D_K), \qquad D_S \sim \mu_S, \quad D_K \sim \mu_K.$$

Thus every randomized observation determined by the restriction to K has the same distribution in the two experiments. This includes the full count and arbitrary joint or adaptive probes of subsets of the same K against the same output.

Proof. We first choose one output law on the large set S and then define the laws on smaller sets by restriction. With probability p, return a uniform singleton of S. Otherwise draw ϑ from the density $1/(\vartheta z_{k,\tau})$ on $[\tau/k, 1]$ and include each point independently with probability ϑ. This law, denoted by μ_S, has singleton probability at least p.

For fixed K of size $j = 2^m$, the intersection count has distribution

$$\nu_m^{\mathrm{cnt}} = p\big((1 - j/Q)\delta_0 + (j/Q)\delta_1\big) + (1-p)\int_{\tau/k}^{1} \mathrm{Bin}(j,\vartheta)\frac{d\vartheta}{\vartheta z_{k,\tau}}, \tag{4.17}$$

where δ_r is a point mass at r. By (4.15),

$$\nu_m^{\mathrm{cnt}}(1) = p\frac{j}{Q} + (1-p)\frac{(1-\tau/k)^j}{z_{k,\tau}} \geq (1-p)\beta_{k,\tau} \geq p. \tag{4.18}$$

On the smaller set K, sample $X \sim \nu_m^{\mathrm{cnt}}$ and return a uniform X-element subset. Call this law μ_K. It is pruning and meets the singleton guarantee by (4.18). The restriction of μ_S to K is invariant under every permutation of K. Conditional on its size, it is therefore uniform among subsets of that size. Since its size distribution is ν_m^{cnt}, its full law is exactly μ_K.

The equality holds separately for every m and K, so mixing dimensions preserves it. For adaptive observations, couple the complete restrictions identically and use the same randomness for the observer. The first query inside K has the same answer in both experiments; hence so does the choice of the next query. Induction gives identical complete transcripts, even for overlapping subsets queried on the same output.

These distributions can be included in one presentation-independent set rule: use μ_S at cardinality Q, the constructed restriction rule at cardinalities 2^m for $m \leq d$, and uniform singleton selection at all other nonempty cardinalities. The cases are disjoint because $k < Q$. □

With $\tau = 1/(d+1)$, the threshold in (4.16) is $(1 + o(1))/\ln k$. A discrete density mixture already gives the simpler finite bound $p \leq 1/(1 + e(d+1))$: choose r uniformly from $\{0, \ldots, d\}$ and use density 2^{-r} in the second branch. At scale m, the term $r = m$ gives singleton mass at least $(1-p)\beta_{2^m}/(d+1)$.

So no statistic of the retained subset inside K can tell these matching laws apart. This remains true if each trial uses a fresh independent output and the next dimension is chosen from earlier observations, because the conditional restriction laws agree at every dimension. The limitation therefore applies to every observation within the known region, not only to the singleton test we started with.

Information from the circuit description, or from observations outside K, is not covered by the matching. Proposition B.3 in Appendix B.2 gives a separate bound for exterior membership queries when the observer has no information about the hidden extension beyond K.

5 Isolation with adaptive membership queries

The affine comparison also gives an unconditional bound when the target is accessible only through membership queries. In this model, the output is a random predicate whose value at each point can query the target adaptively. The query budget is charged *per evaluation of this predicate*, not once for the entire surviving set. Theorem 5.2 gives one explicit distribution of affine targets on which every such procedure has small average success. The resulting query–success tradeoff is tight up to absolute constants.

5.1 The membership-query model

Here $n \geq 1$ is the ambient dimension, independent of the source-description length ℓ used earlier. The target is a nonempty set $W \subseteq \mathbb{F}_2^n$. A membership query at y returns $\mathbf{1}_{\{y \in W\}}$. The notation q_{mem} distinguishes this budget from the number of output points inspected in Section 3.

Definition 5.1 (Membership-query pruning procedure)**.** A procedure with query bound q_{mem} consists of a random string ρ, whose distribution is independent of W, and an oracle algorithm A_{mem}. On input (x, ρ), the algorithm outputs a bit after at most q_{mem} adaptive membership queries to W. Its surviving set is

$$D(W; \rho) = \{x \in W : A_{\mathrm{mem}}^{W}(x; \rho) = 1\}. \tag{5.1}$$

The same string ρ is used for all x. Each evaluation starts afresh; there is no target-dependent state inherited from other evaluations. Computation between queries and the length of ρ are unrestricted. The algorithm and the distribution of ρ may depend on n and on a fixed target class, but not on the particular target. Its isolation guarantee is the minimum of $\Pr_\rho[|D(W; \rho)| = 1]$ over that class.

The intersection with W in (5.1) enforces pruning. Thus $q_{\mathrm{mem}} = 0$ means that the *filter* makes no queries; explicitly checking $x \in W$ adds at most one query. Oracle-dependent preprocessing can be included by rerunning it on each evaluation and charging its queries. All shared randomness and advice are independent of the target. In particular, neither a circuit description nor the target's exact dimension or density is supplied.

Ignoring the oracle gives DKMW's target-oblivious model [1, Definition 4.1]. Adaptive querying also permits candidate-first selection, as in Appendix B. Every rule in Definition 5.1 depends only on the target set, but a presentation-invariant circuit transformation need not have an implementation with few membership queries. The model is also different from witness finding with set-intersection queries [10]: here the primitive query is point membership, and the output is a pruning predicate rather than an explicitly returned witness.

5.2 An average-case bound and inverse-linear tightness

For $0 \leq d < n$, let $\mathsf{Aff}(n, d)$ be the following distribution: choose m uniformly from $\{0, \ldots, d\}$, then choose a uniform affine m-subspace W of $\mathbb{F}_2^n$. This is a uniform distribution of *dimensions*, not a uniform distribution over all subspaces in their union.

Theorem 5.2 (Average success under adaptive membership access)**.** *For every procedure in Definition 5.1 and every integer $0 \leq d < n$,*

$$\mathbb{E}_{W \sim \mathsf{Aff}(n,d)}\big[\Pr_\rho[|D(W; \rho)| = 1]\big] \leq \frac{2 + q_{\mathrm{mem}}\, 2^{2d-n}}{d+1}. \tag{5.2}$$

Consequently the same expression bounds its worst-case isolation guarantee on affine subspaces of dimension at most d. The target distribution is independent of the procedure, its randomness, and its advice.

Proof. Choose S uniformly among affine d-subspaces of $\mathbb{F}_2^n$. For fixed $m \leq d$, choose a uniform affine m-subspace K_m of S. The orbit argument of Lemma 3.1 applies with these dimensions: K_m is marginally uniform in the ambient space, and conditional on K_m, S is uniform among its affine d-dimensional extensions. Use the same independent random string ρ on targets K_m and S, and write $D_{K_m} = D(K_m; \rho)$ and $D_S = D(S; \rho)$.

Couple the evaluations on the smaller target. Fix K_m, ρ and collect, over every input point $x \in K_m$, all queries made by $A_{\mathrm{mem}}^{K_m}(x;\rho)$. Their union $\mathcal{Q}(K_m,\rho)$ has size at most $q_{\mathrm{mem}}\,2^m$ and is independent of S conditional on K_m, ρ. For $m < d$ and any fixed $y \notin K_m$,

$$\Pr[y \in S \mid K_m] = \frac{2^d - 2^m}{2^n - 2^m} \le 2^{d-n}. \tag{5.3}$$

The equality follows by symmetry on the points outside K_m: every extension contains exactly $2^d - 2^m$ of those $2^n - 2^m$ points. A union bound therefore gives

$$\Pr[D_S \cap K_m \neq D_{K_m}] \le q_{\mathrm{mem}}\,2^m \frac{2^d - 2^m}{2^n - 2^m} \qquad (m < d). \tag{5.4}$$

Indeed, unless a queried point outside K_m lies in S, every answer along each smaller-target transcript is unchanged. Induction on the queries gives the same next query and final bit, despite adaptivity. When $m = d$, the targets are identical and the error is zero.

Sum over dimensions. For each m, the isolation probability on K_m is at most the probability that $D_S \cap K_m$ is a singleton, plus the coupling error. Conditional on S, ρ, the set D_S is fixed, and K_m is still a uniform affine m-subspace of S. Lemma 3.3 bounds the sum of these singleton-intersection probabilities by 2. The total coupling error, summed only over $m < d$, is at most

$$q_{\mathrm{mem}}\,2^{d-n} \sum_{m=0}^{d-1} 2^m = q_{\mathrm{mem}}\,2^{d-n}(2^d - 1) \le q_{\mathrm{mem}}\,2^{2d-n}.$$

For $d = 0$ the sum is empty. Dividing by $d+1$ proves (5.2). Notice that the flag step uses no success guarantee on S; success assumptions enter only when replacing the actual target probabilities at dimensions $0, \dots, d$ by a common lower bound. □

The lower bound uses the affine-target specialization of the hashing mechanism in Valiant–Vazirani [3]. The direct calculation below includes singleton targets and the full ambient space.

Lemma 5.3 (Zero-query isolation with a dimension bound)**.** *For every integer $0 \le d \le n$, a procedure making no queries to its target has isolation guarantee at least $1/(4(d+1))$ on all nonempty affine subspaces of dimension at most d.*

Proof. Choose i uniformly from $\{0, \dots, d\}$ and a uniform linear map $h : \mathbb{F}_2^n \to \mathbb{F}_2^i$. The filter accepts precisely when $h(x) = 0$. If $W = v + U$ has dimension m, then, conditional on $i = m$, the restriction $h|_U$ is a uniformly random square binary matrix. When it is invertible, $h(v+u) = 0$ has exactly one solution $u \in U$. This occurs with probability

$$\prod_{j=1}^{m} (1 - 2^{-j}) \ge \frac{1}{4}.$$

For $m \ge 1$, factor out $1/2$ and use $\prod_{j=2}^{m}(1-2^{-j}) \ge 1 - \sum_{j=2}^{m} 2^{-j} \ge 1/2$. For $m = 0$, the empty product is one. The chance of choosing $i = m$ is $1/(d+1)$, giving the claim. □

Corollary 5.4 (Optimal order for subexponential query budgets)**.** *Let $n = 3N$ for an integer $N \ge 1$. A membership-query procedure with isolation guarantee p on all nonempty affine subspaces of dimension at most N satisfies*

$$p \le \frac{2 + q_{\mathrm{mem}}\,2^{-N}}{N+1}. \tag{5.5}$$

For $q_{\mathrm{mem}} = 2^{o(n)}$, the optimal guarantee on this class is $\Theta(1/n)$.

Proof. Use $d = N$ in Theorem 5.2. Its error numerator $q_{\mathrm{mem}}\, 2^{-N}$ tends to zero for a subexponential budget. Lemma 5.3, with $d = N$, supplies the matching-order lower bound without making any target queries. □

5.3 The full query–success tradeoff

Let $\mathrm{Iso}(n, q_{\mathrm{mem}})$ be the supremum, over procedures in Definition 5.1, of their minimum isolation probability on *all* nonempty affine subspaces of $\mathbb{F}_2^n$. The minimum includes every dimension, and no target dimension is supplied to the procedure.

Theorem 5.5 (Query–success tradeoff)**.** *For all integers $n \geq 1$ and $0 \leq q_{\mathrm{mem}} \leq 2^n - 1$, set $r_{\mathrm{qry}} = \log(2^n/(q_{\mathrm{mem}} + 1))$. Then*

$$\frac{1}{4(\lceil r_{\mathrm{qry}} \rceil + 1)} \leq \mathrm{Iso}(n, q_{\mathrm{mem}}) \leq \min\left\{1, \frac{3}{\lfloor r_{\mathrm{qry}}/2 \rfloor + 1}\right\}. \tag{5.6}$$

In particular, uniformly throughout this range,

$$\mathrm{Iso}(n, q_{\mathrm{mem}}) = \Theta\left(\frac{1}{1 + \log(2^n/(q_{\mathrm{mem}} + 1))}\right), \tag{5.7}$$

with absolute implicit constants.

Proof. Upper bound. Take $d = \lfloor r_{\mathrm{qry}}/2 \rfloor$. Since $0 \leq r_{\mathrm{qry}} \leq n$, this is an allowed choice in Theorem 5.2, including when $n = 1$. Moreover,

$$q_{\mathrm{mem}}\, 2^{2d-n} \leq \frac{q_{\mathrm{mem}}}{q_{\mathrm{mem}} + 1} \leq 1.$$

The minimum success over all affine targets is no larger than the average in that theorem. This gives the upper bound in (5.6).

Lower bound. Use shared randomness to choose q_{mem} independent uniform candidate points in $\mathbb{F}_2^n$. Query their memberships, using the same ordered list on every evaluation. If any candidate is in W, accept only the first such point. Otherwise use an independent copy of the zero-query filter from Lemma 5.3 with dimension bound $d_{\mathrm{hash}} = \lceil r_{\mathrm{qry}} \rceil \leq n$. This procedure uses at most q_{mem} queries per evaluation.

If $|W| \leq 2^n/(q_{\mathrm{mem}} + 1)$, then $\dim W \leq r_{\mathrm{qry}}$. The fallback filter succeeds with probability at least $1/(4(d_{\mathrm{hash}} + 1))$, independently of the candidate list. Finding a candidate already guarantees isolation, so it cannot lower this success probability. If $|W| > 2^n/(q_{\mathrm{mem}} + 1)$, then $q_{\mathrm{mem}} \geq 1$, and the probability that a candidate is found is

$$1 - \left(1 - \frac{|W|}{2^n}\right)^{q_{\mathrm{mem}}} \geq 1 - \exp\left(-\frac{q_{\mathrm{mem}}}{q_{\mathrm{mem}} + 1}\right) \geq 1 - e^{-1/2} > \frac{1}{4}.$$

This also exceeds the asserted lower bound. For $q_{\mathrm{mem}} = 0$, every target falls into the first case. The finite bounds imply (5.7) because their denominators are within absolute factors of $1 + r_{\mathrm{qry}}$. □

In particular, for every fixed $\delta \in (0, 1)$, even $q_{\mathrm{mem}} \leq 2^{(1-\delta)n}$ permits only $O_\delta(1/n)$ success. Near the size of the universe, the behavior changes: for fixed $a_{\mathrm{exp}} > 0$ and $q_{\mathrm{mem}} = \lfloor 2^n/n^{a_{\mathrm{exp}}} \rfloor$, the optimum is $\Theta(1/\log n)$. The upper bound imposes no limit on computation or seed length. Nonadaptive candidate queries followed by hashing already attain the matching order; adaptivity does not improve that order.

What changes when a circuit description is available. Both arguments use a random inclusion $K \subseteq S$ and the affine flag bound. In the membership model, the output on S is independent of the choice of K conditional on S, because that choice is not an input. The transcript coupling adds a separate fact: the restriction of the output on S usually agrees with the output on the different target K. Presentation invariance in the circuit model supplies the first independence, but not this cross-target agreement. The latter would be false, for example, for lexicographically least selection.

The membership proof sums over linearly many dimensions without executing an enumeration algorithm. The circuit reduction instead decides whether exactly one of the known points survives, and its present implementation uses 2^m output evaluations. This explains the different scales attained by the two proofs; it does not identify the optimal success probability of presentation-invariant circuit transformations. The matching theorem of Section 4 limits observations inside K, not arbitrary uses of the output description. Finally, exact target dimension is substantive information: if it is supplied, the hashing proof can set $i = \dim W$ and obtain constant success with no filter queries. Such information is excluded by Definition 5.1.

6 Conclusion

Under indistinguishability obfuscation, efficient pruning isolation hits a barrier at inverse-logarithmic success: a guarantee of $a/\log L$, even if it is required only on circuit-presented affine spaces, would imply NP $\subseteq$ P/poly, and with nonuniformly secure one-way functions this becomes the impossibility result of Corollary 1.2. The isolator need not be presentation-invariant, and the obfuscator may make negligible functional errors.

The main results share one mechanism, a multiscale test on a known affine subspace. The distributional preprocessing criterion turns average gaps in this test into nonuniform decision circuits, and its invariant and canonical applications need no cryptographic assumption. Run in the dual direction, the same argument produces common lists of equivalent presentation pairs that stay distinguishable after every preprocessing map in a prescribed bounded family.

With membership access alone the comparison is unconditional. The upper bound comes from coupling adaptive transcripts on a small affine target and on a random extension of it; candidate sampling followed by hashing gives the matching order. Theorem 5.5 pins down the whole query–success tradeoff up to absolute constants. In particular, a subexponential query budget leaves the optimum at $\Theta(1/n)$.

In the circuit setting, a polynomial evaluation budget $k = 2^d$ allows only $d = O(\log \ell)$. This is where the $\log \log L$ scale of the information lower bound comes from, and also the $1/\log k$ scale of the restriction-matching theorem. The matching covers every observation inside the planted region but not arbitrary uses of the circuit description. The membership-query proof also relies on cross-target agreement, and neither invariance nor iO supplies it.

The interval between the $\Omega(1/n)$ guarantee of Valiant–Vazirani and $a/\log L$ is still open; here n counts witness variables and L is the description length. One concrete question is whether an assumption weaker than iO yields the restricted preprocessing of Section 3.6. Remark 3.15 describes what such a transformation has to do. Changing witness-independent coordinates leaves the planted region in place, while exposing a different planted decomposition gives away the witness, so the transformation must hide that structure without first recovering a basis. The extremal problem in Appendix A is a separate matter. It concerns the leading constant of the existing tests, not this computational gap.

References

[1] H. Dell, V. Kabanets, D. van Melkebeek, and O. Watanabe, Is Valiant–Vazirani's isolation probability improvable? *Computational Complexity* 22(2):345–383, 2013. doi:10.1007/s00037-013-0059-7.

[2] H. Dell, V. Kabanets, D. van Melkebeek, and O. Watanabe, Is Valiant–Vazirani's isolation probability improvable? ECCC Report TR11-151, Revision 1, 2012.

[3] L. G. Valiant and V. V. Vazirani, NP is as easy as detecting unique solutions, *Theoretical Computer Science* 47 (1986), 85–93.

[4] K. Mulmuley, U. V. Vazirani, and V. V. Vazirani, Matching is as easy as matrix inversion, *Combinatorica* 7(1):105–113, 1987. doi:10.1007/BF02579206.

[5] S. Chari, P. Rohatgi, and A. Srinivasan, Randomness-optimal unique element isolation with applications to perfect matching and related problems, *SIAM Journal on Computing* 24(5):1036–1050, 1995. doi:10.1137/S0097539793250330.

[6] L. A. Hemaspaandra, A. V. Naik, M. Ogihara, and A. L. Selman, Computing solutions uniquely collapses the polynomial hierarchy, *SIAM Journal on Computing* 25(4):697–708, 1996. doi:10.1137/S0097539794268315.

[7] A. V. Naik, K. W. Regan, and D. Sivakumar, On quasilinear time complexity theory, *Theoretical Computer Science* 148(2):325–349, 1995. doi:10.1016/0304-3975(95)00031-Q.

[8] S. Gupta, Isolating an odd number of elements and applications in complexity theory, *Theory of Computing Systems* 31:27–40, 1998. doi:10.1007/s002240000075.

[9] A. R. Klivans and D. van Melkebeek, Graph nonisomorphism has subexponential size proofs unless the polynomial-time hierarchy collapses, *SIAM Journal on Computing* 31(5):1501–1526, 2002. doi:10.1137/S0097539700389652.

[10] A. Kawachi, B. Rossman, and O. Watanabe, The query complexity of witness finding, *Theory of Computing Systems* 61(2):305–321, 2017. doi:10.1007/s00224-016-9708-y.

[11] V. Krishan and N. Limaye, Isolation lemma for directed reachability and NL vs. L, ECCC Report TR16-155, 2016.

[12] S. Gharibian and J. Kamminga, BQP, meet NP: Search-to-decision reductions and approximate counting, In *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024)*, LIPIcs 297, Article 70, 2024. doi:10.4230/LIPIcs.ICALP.2024.70.

[13] R. J. Lipton and N. E. Young, Simple strategies for large zero-sum games with applications to complexity theory, In *Proceedings of STOC*, pp. 734–740, 1994. doi:10.1145/195058.195447.

[14] P. Flajolet, X. Gourdon, and P. Dumas, Mellin transforms and asymptotics: Harmonic sums, *Theoretical Computer Science* 144(1–2):3–58, 1995. doi:10.1016/0304-3975(95)00002-E.

[15] S. Garg, C. Gentry, S. Halevi, M. Raykova, A. Sahai, and B. Waters, Candidate indistinguishability obfuscation and functional encryption for all circuits, *SIAM Journal on Computing* 45(3):882–929, 2016. doi:10.1137/14095772X.

[16] A. Coladangelo, J. Liu, Q. Liu, and M. Zhandry, Hidden cosets and applications to unclonable cryptography, In *Advances in Cryptology—CRYPTO 2021*, LNCS 12825, pp. 556–584, 2021. doi:10.1007/978-3-030-84242-0_20. Full version: arXiv:2107.05692v3, Sections 3.2–3.4.

[17] D. Wichs and G. Zirdelis, Obfuscating compute-and-compare programs under LWE, In *Proceedings of FOCS*, pp. 600–611, 2017. doi:10.1109/FOCS.2017.61. Full version: Cryptology ePrint Archive, Report 2017/276.

[18] N. Bitansky and V. Vaikuntanathan, A note on perfect correctness by derandomization, *Journal of Cryptology* 35, Article 18, 2022. doi:10.1007/s00145-022-09428-0.

[19] S. Ben Daniel, *On the Advice Complexity of Tournaments*, in *COCOON 2012* doi:10.1007/978-3-642-32241-940

A Leading constants for mixtures of affine dimensions

Here we refine the order bounds of Section 4 and identify a candidate leading constant for the optimal scale distribution. None of the Fourier analysis is needed for the complexity characterization or for the restriction-law theorem.

Symbol	Local meaning
ϑ; τ	Bernoulli inclusion probability; lower-cutoff parameter.
ζ	Positive multiplicative phase parameter (not a list size).
$\Psi_+(\zeta)$; $\Psi(\zeta)$	One-sided and two-sided dyadic sums.
κ_*	Maximum of the two-sided sum over one multiplicative period.
$d_{\mathrm{buf}} = \lceil\sqrt{d}\rceil$	Number of additional boundary dimensions in the prior.
$\omega_{d,\zeta}$; $p_{\min}(d,\zeta)$	Density prior; its least singleton probability across dimensions.
$\mathrm{OPT}_{d,N}$	Extremal unnormalized affine-flag sum.

A.1 Refining the constants with Bernoulli subsets

We now refine the constants in two different quantities. Choosing one Bernoulli density lower-bounds the maximum for uniform scales. Mixing densities gives a lower bound valid against every distribution of scales. Only the second comparison can show how close uniform scales are to the optimal choice.

Proposition A.1 (Two Bernoulli refinements)**.** *Let $k = 2^d < Q$. For every fixed $\zeta > 0$, as $d \to \infty$,*

$$(d+1)c_{\mathrm{unif},Q} \geq \Psi_+(\zeta) - o(1), \qquad \Psi_+(\zeta) = \sum_{i=0}^{\infty} \zeta 2^{-i} \exp(-\zeta 2^{-i}). \tag{A.1}$$

The constants $z_{k,\tau}$ and $\beta_{k,\tau}$ are defined in (4.14)*. Every scale distribution satisfies*

$$c_{\pi,Q} \geq \beta_{k,\tau}, \qquad c^-_{\pi,Q} \geq (1 - k/Q)\beta_{k,\tau}. \tag{A.2}$$

For $\tau = 1/(d+1)$, $\beta_{k,\tau} = (1+o(1))/\ln k$. Consequently uniform mixing is within a factor $2\ln 2 + o(1)$ of the optimal mixture. The same statement holds for failure-only maxima if $Q/k \to \infty$.

Proof. For the first bound, take a Bernoulli set of density ζ/k, with d large enough that $\zeta < k$. Equation (4.11) gives

$$(d+1)c_{\mathrm{unif},Q} \geq \Psi_{+,d}(\zeta) := \sum_{i=0}^{d} \zeta 2^{-i}(1 - \zeta 2^{-d})^{2^{d-i}-1}.$$

For each fixed i, the summand tends to $\zeta 2^{-i} e^{-\zeta 2^{-i}}$ and is bounded by $\zeta 2^{-i}$. This summable bound gives $\Psi_{+,d}(\zeta) \to \Psi_+(\zeta)$ by dominated convergence. The positive tail after the first j_{cut} terms is at most $\zeta 2^{1-j_{\mathrm{cut}}}$, allowing direct numerical evaluation. For example,

$$\Psi_+(1) = 1.096085583059141\ldots, \qquad \Psi_+(16) = 1.442704206673319\ldots.$$

The contribution of globally singleton realizations to $\Psi_{+,d}(\zeta)$ equals $\zeta(2k-1)k^{-1}(1-\zeta/k)^{Q-1}$, which tends to zero if $Q/k \to \infty$. Thus the same constants apply to the failure-only uniform maximum.

For the all-mixture bound, randomize the density itself. Draw ϑ from $[\tau/k, 1]$ with probability density $1/(\vartheta z_{k,\tau})$, then include each point independently with probability ϑ. At every permitted size $j = 2^m \leq k$, Equation (4.15) gives the lower bound $\beta_{k,\tau}$ for the expected singleton-intersection

probability. The same lower bound holds at every size, so it survives averaging against any π. Some fixed set attains at least that average, proving the first part of (A.2). For the failure-only bound, the contribution of globally singleton sets is

$$\sigma_{\pi,Q}\Pr[|R|=1]=\sigma_{\pi,Q}\frac{(1-\tau/k)^Q}{z_{k,\tau}}\le\frac{k}{Q}\beta_{k,\tau}.$$

The empty set contributes zero, proving the second part.

With $\tau=1/(d{+}1)$, the numerator of $\beta_{k,\tau}$ tends to one, whereas its denominator is $\ln k{+}\ln(d{+}1)=(1+o(1))\ln k$. Compare this lower bound with $c_{\mathrm{unif},Q}\le 2/(d+1)$ to obtain the approximation ratio. □

A.2 The dyadic periodic function

The calculation in this subsection identifies a candidate leading constant for the optimal distribution of scales. Its lower bound applies to every scale mixture. A matching upper bound for uniform scales would therefore determine the minimax constant, not merely improve a numerical estimate; see (A.5) and (A.9).

The constants near $1/\ln 2$ arise from two ways of averaging the same kernel. A logarithmic density average smooths over the multiplicative scale; dyadic densities preserve a small periodic dependence on its logarithm. For $f(x)=xe^{-x}$, define

$$\Psi(\zeta)=\sum_{i\in\mathbb{Z}}f(\zeta 2^{-i}),\qquad \kappa_*:=\max_{1\le\zeta\le 2}\Psi(\zeta). \tag{A.3}$$

The series converges locally uniformly on $\zeta>0$ and satisfies $\Psi(2\zeta)=\Psi(\zeta)$. Hence $u\mapsto\Psi(2^u)$ is continuous and one-periodic. We compute its Fourier coefficients explicitly, using the logarithmic periodization method for harmonic sums [14]. Here Γ is Euler's gamma function and $\mathrm{i}^2=-1$.

Proposition A.2 (The periodic function and lower bounds for every mixture)**.** *The function $u\mapsto\Psi(2^u)$ has mean $1/\ln 2$ and the Fourier expansion*

$$\Psi(2^u)=\frac{1}{\ln 2}\sum_{h\in\mathbb{Z}}\Gamma\!\left(1-\frac{2\pi\mathrm{i}h}{\ln 2}\right)e^{2\pi\mathrm{i}hu}. \tag{A.4}$$

It is not constant; hence $\kappa_*>1/\ln 2$. Along sequences with $d\to\infty$, $Q=2^N$, and $d=o(N)$,

$$\kappa_*\le\liminf(d+1)\inf_{\pi\in\Pi_d}c_{\pi,Q}\le\liminf(d+1)c_{\mathrm{unif},Q}\le 2. \tag{A.5}$$

The same lower bound holds with $c^-_{\pi,Q}$ in place of $c_{\pi,Q}$. In this regime uniform mixing is within a factor $2/\kappa_*+o(1)$ of the optimal mixture.

Proof. The Fourier expansion. Put $\varphi(u)=f(2^u)$. Then $\Psi(2^u)=\sum_{i\in\mathbb{Z}}\varphi(u-i)$. Unfolding one period and substituting $x=2^u$ gives

$$\int_0^1\Psi(2^u)\,du=\int_{\mathbb{R}}f(2^u)\,du=\frac{1}{\ln 2}\int_0^\infty e^{-x}\,dx=\frac{1}{\ln 2}.$$

The hth Fourier coefficient is

$$\int_{\mathbb{R}}f(2^u)e^{-2\pi\mathrm{i}hu}\,du=\frac{1}{\ln 2}\int_0^\infty e^{-x}x^{-2\pi\mathrm{i}h/\ln 2}\,dx.$$

This gives (A.4). The identity $|\Gamma(1+\mathrm{i}y)|^2 = \pi y/\sinh(\pi y)$, with its continuous extension at zero, implies absolute convergence of the Fourier series. It also shows that the coefficient at $h = 1$ is nonzero, so Ψ is not constant. Finally, $f(x) \le 2(e^{-x/2} - e^{-x})$; the dyadic sum of this bound telescopes to give $\Psi(\zeta) \le 2$.

The uniform-scale lower bound. Fix $\zeta \in [1, 2)$ and choose a Bernoulli set of density $\vartheta_d = \zeta 2^{-\lfloor d/2 \rfloor}$. Its expected unnormalized singleton-intersection sum is

$$\sum_{m=0}^{d} 2^m \vartheta_d (1-\vartheta_d)^{2^m-1} \longrightarrow \Psi(\zeta).$$

Shift the index by $\lfloor d/2 \rfloor$. Each fixed term then converges to $f(\zeta 2^j)$ and, for large d, is bounded by $2f(\zeta 2^j)$. This bound is summable over $j \in \mathbb{Z}$, which justifies the limit. Since some fixed retained set attains at least the expectation, choosing a maximizing phase proves the uniform-scale lower bound.

A density mixture for all scale distributions. Let $d_{\mathrm{buf}} = \lceil \sqrt{d} \rceil$ and fix $\zeta \in [1, 2)$. We use two groups of densities: weight one at $\zeta 2^{-r}$ for $1 \le r \le d + d_{\mathrm{buf}}$, and weight six at 2^{-r} for $0 \le r < d_{\mathrm{buf}}$. Add the weights if a density occurs in both groups, and normalize by $d + 7d_{\mathrm{buf}}$ to obtain $\omega_{d,\zeta}$. The first group preserves the desired dyadic phase; the second ensures a sufficiently large contribution at the smallest planted dimensions. Set

$$p_{\min}(d, \zeta) := \min_{0 \le m \le d} \mathbb{E}_{\vartheta \sim \omega_{d,\zeta}} \left[2^m \vartheta (1-\vartheta)^{2^m-1} \right]. \tag{A.6}$$

We claim that

$$(d+1) p_{\min}(d, \zeta) \longrightarrow \Psi(\zeta). \tag{A.7}$$

Consider first $m < d_{\mathrm{buf}}$. The extra density 2^{-m} contributes at least $6/e > 2$ before normalization; at $m = 0$ it contributes six. For $m \ge d_{\mathrm{buf}}$, fix an integer $j_{\mathrm{cut}} > 0$ and keep the first-group densities indexed by $r = m - j_{\mathrm{cut}}, \dots, m + j_{\mathrm{cut}}$. When $d_{\mathrm{buf}} > j_{\mathrm{cut}}$, these indices are present for every such m. Their sum converges uniformly to $\sum_{i=-j_{\mathrm{cut}}}^{j_{\mathrm{cut}}} f(\zeta 2^{-i})$. Letting d tend to infinity and then j_{cut} tend to infinity shows that the unnormalized minimum is at least $\Psi(\zeta) - o(1)$.

To obtain the reverse limit for this minimum, evaluate it at $m = d$. The first group converges to $\Psi(\zeta)$ by the same domination argument. The second contributes $o(1)$, because its smallest binomial mean is at least $2^{d - d_{\mathrm{buf}} + 1}$. Normalization changes $d + 1$ to $d + 7d_{\mathrm{buf}}$, whose ratio tends to one. This proves (A.7).

Draw a Bernoulli set with density sampled from $\omega_{d,\zeta}$. Every planted scale has expected singleton intersection at least $p_{\min}(d, \zeta)$. Thus $c_{\pi,Q} \ge p_{\min}(d, \zeta)$ for every π. Use a maximizing phase to obtain (A.5). For the failure-only version, the least positive density is at least $2^{-d - d_{\mathrm{buf}}}$. If $d = o(N)$, the probability that the entire Bernoulli set is a singleton is at most

$$Q \exp(-(Q-1) 2^{-d - d_{\mathrm{buf}}}) = o(1/d).$$

Removing such realizations changes any acceptance average by $o(1/d)$, uniformly over π. □

The numerical values distinguish a phase value from its logarithmic mean:

$$\frac{1}{\ln 2} = 1.4426950409\ldots, \qquad \Psi(1) = 1.4427042067\ldots, \qquad \kappa_* = 1.4427093011\ldots.$$

The difference $\Psi(1) - \Psi_+(16)$ is exactly $\sum_{j \ge 5} 2^j e^{-2^j}$, approximately 4.05253×10^{-13}. The first Fourier harmonic has amplitude approximately $1.42602345195 \times 10^{-5}$, and the sum of the higher amplitudes is less than 1.321×10^{-11}. Together with the Fourier tail estimate, these give $1.4427093011 < \kappa_* < 1.4427093012$. A maximizer occurs numerically near $\zeta = 1.81642717416$; no uniqueness claim is needed. All statements use the exact definition (A.3).

The remaining extremal problem. Define

$$\mathrm{OPT}_{d,N} := \max_{R \subseteq \mathbb{F}_2^N} \sum_{m=0}^{d} \Pr[|R \cap K_m| = 1] = (d+1) c_{\mathrm{unif},2^N}. \tag{A.8}$$

For $d \to \infty$ and $d = o(N)$, the established bounds are $\kappa_* - o(1) \le \mathrm{OPT}_{d,N} \le 2$. Because the periodic fluctuation is nonzero, the lower bound already excludes $\mathrm{OPT}_{d,N} = 1/\ln 2 + o(1)$. The remaining question is whether

$$\mathrm{OPT}_{d,N} = \kappa_* + o(1). \tag{A.9}$$

Equivalently, can structured retained sets asymptotically exceed the Bernoulli phase maximum? A matching upper bound in (A.9), together with (A.5), would determine the leading constant for the optimal distribution of scales as well.

For comparison, let R itself be an affine subspace of codimension $d_R \le d$. For $m \le d_R$,

$$\Pr[|R \cap K_m| = 1] = 2^{m-d_R} \prod_{i=0}^{m-1} \frac{1 - 2^{i-d_R}}{1 - 2^{i-N}},$$

and the probability is zero for $m > d_R$. Indeed, the direction space of K_m must inject into the quotient by the direction space of R; the translate then meets R with probability 2^{m-d_R}. With $p_{\mathrm{inj}}(m) = \prod_{i=0}^{m-1}(1-2^{i-d_R})$, the numerator is $p_{\mathrm{inj}}(m) - p_{\mathrm{inj}}(m+1)$, so its sum is one. The denominator changes this by a factor $1 + O(2^{d_R-N})$. Thus affine retained sets give a sum $1 + o(1)$ in the stated regime, below the Bernoulli lower bound.

Consequence for restriction-law matching. The proof of Theorem 4.3 works with any distribution on Bernoulli densities. Replace $\beta_{k,\tau}$ by the least expected singleton probability across the permitted scales. For the distribution $\omega_{d,\zeta}$ from (A.6), exact matching holds when $p \le p_{\min}(d,\zeta)/(1 + p_{\min}(d,\zeta))$. At a maximizing phase this is $(\kappa_* + o(1))/(d+1)$. Hence every fixed $\kappa_0 < \kappa_*$ gives matching distributions for $p \le \kappa_0/(d+1)$ at all sufficiently large d.

B Further consequences of the affine construction

These results use the same construction but are not needed for Corollary 1.3. We collect the explicit field option, additional invariant examples, finite-order implications, and the single-point information bounds here.

Remark B.1 (An explicit field representation). An explicit alternative is to choose the least $N' = 2 \cdot 3^j \ge 2\ell$ and use $X^{2\cdot 3^j} + X^{3^j} + 1 = \Phi_{3^{j+1}}(X)$. Every root has multiplicative order 3^{j+1}. Since $\mathrm{ord}_{3^{j+1}}(2) = 2 \cdot 3^j$, each root has exactly $2 \cdot 3^j$ distinct Frobenius conjugates. This is the polynomial's degree, so it is irreducible over $\mathbb{F}_2$. Here Φ_r is the rth cyclotomic polynomial. The order identity follows from $v_3(4^{3^j} - 1) = j + 1$, where v_3 is the 3-adic valuation, by induction using the factorization of $x^3 - 1$. Since $2\ell \le N' < 6\ell$, this change preserves all size estimates. It removes the field-representation advice, not the nonuniformity of the assumed isolator.

Examples beyond affine hashing. Random weights provide an invariant pruning construction that does not use affine hashing. For a circuit on $n \ge 1$ variables, choose independent weights $\omega_i \sim \mathrm{Unif}(\{1, \dots, R_{\mathrm{wt}}\})$, where $R_{\mathrm{wt}} = 2^{\lceil \log(2n) \rceil}$, and independently choose $j \sim \mathrm{Unif}(\{0, \dots, R_{\mathrm{sum}} - 1\})$, where $R_{\mathrm{sum}} = 2^{\lceil \log(nR_{\mathrm{wt}}+1) \rceil}$. Output $C(x) \wedge \mathbf{1}_{\{\sum_i \omega_i x_i = j\}}$. The Mulmuley–Vazirani–Vazirani lemma gives a unique minimum-weight satisfying assignment with probability at least $1 - n/R_{\mathrm{wt}} \ge 1/2$ [4].

The independently chosen target equals that minimum weight with probability $1/R_{\mathrm{sum}}$. Hence this construction isolates with probability at least $1/(2R_{\mathrm{sum}}) = \Omega(n^{-2})$. The weights and target depend only on n, so the output-set distribution is presentation-invariant. The random guess avoids having to compute the minimum weight.

A second example depends on the target set. Fix a polynomially enumerable set $\mathrm{Cand}_{n,L} \subseteq \{0,1\}^n$ using only n and L, and evaluate C on that set. If any candidate satisfies C, return the lexicographically least satisfying candidate; otherwise apply Valiant–Vazirani. The rule is invariant and preserves the $\Omega(1/n)$ worst-case guarantee. It is not target-oblivious when $\mathrm{Cand}_{n,L}$ contains two points $x < y$: it retains y with probability one when the satisfying set is $\{y\}$, but never when that set is $\{x, y\}$. No single target-oblivious pruning distribution can have both behaviors. These examples illustrate the class without improving its worst-case isolation guarantee.

B.1 Finite-order presentation invariance

If π is concentrated at dimension m, put $j = 2^m$. A valid upper bound $c_{m,Q}$ on $\Pr[|R \cap K_m| = 1]$ gives, by Proposition 4.1,

$$\Pr[|D \cap K| = 1] \le p\frac{j}{Q} + (1-p)c_{m,Q} \quad (C \in \mathcal{C}_\ell^{(1)}), \qquad p^\star_{m,Q} = \frac{c_{m,Q}}{1 + c_{m,Q} - j/Q}. \tag{B.1}$$

The condition $c_{m,Q} \ge j/Q$ ensures that the upper estimate also covers singleton retained sets. The quantity $p^\star_{m,Q}$ is the success threshold certified by this estimate. This is the existing test specialized to one dimension; no new reduction is involved.

Theorem B.2 (First- and second-order invariance)**.** *Let A be a randomized polynomial-size pruning isolator. Either of the following assumptions, for all sufficiently large description lengths L, implies* $\mathrm{NP} \subseteq \mathrm{P/poly}$*:*

(i) retention invariance on affine-subspace inputs and success $p(L) \ge 1/2 + \varepsilon(L)$;

(ii) two-wise retention invariance on affine-subspace inputs and success $p(L) \ge 1/3 + \varepsilon(L)$,

where $\varepsilon(L) > 0$ is bounded below by an inverse polynomial.

Proof. For $m = 0$, use $c_{0,Q} = 1$. Equation (B.1) gives $p^\star_{0,Q} = Q/(2Q-1) = 1/2 + O(Q^{-1})$. For $m = 1$, an affine line over $\mathbb{F}_2$ is an unordered pair of distinct points. If $|R| = r$, then

$$\Pr[|R \cap K_1| = 1] = \frac{2r(Q-r)}{Q(Q-1)} \le \frac{Q}{2(Q-1)}. \tag{B.2}$$

Thus $c_{1,Q} = Q/(2(Q-1))$ is valid, and for $Q \ge 4$ it is at least $2/Q$. The resulting threshold is

$$p^\star_{1,Q} = \frac{Q^2}{3Q^2 - 6Q + 4} = \frac{1}{3} + O(Q^{-1}).$$

In both cases, $Q = 2^{2\ell}$, while $L_0(\ell)$ is polynomial in ℓ. The finite-Q correction is therefore exponentially smaller than the assumed margin $\varepsilon(L_0(\ell))$. The source cases have an inverse-polynomial acceptance gap, and the test uses at most two evaluations of one output. Amplification and Lemma 2.2 finish the proof. □

Four-point version. The same argument with $m = 2$ gives success $27/91 + 1/\operatorname{poly}(L)$ under four-wise invariance. To obtain the intersection bound, identify S with $\mathbb{F}_2^N$, let $f = \mathbf{1}_R$ and $\vartheta = \mathbb{E}_x[f(x)]$, and choose independent uniform x, u, v. Every three distinct positions in the list $x, x+u, x+v, x+u+v$ are jointly uniform before conditioning on the directions. Write $\mathrm{Corr}_4(f)$ for the four-point correlation defined below. Expanding the event of exactly one membership bit equal to one gives

$$4(\vartheta - 3\vartheta^2 + 3\vartheta^3 - \mathrm{Corr}_4(f)), \qquad \mathrm{Corr}_4(f) = \mathbb{E}_u\big[(\mathbb{E}_x[f(x)f(x+u)])^2\big] \geq \vartheta^4.$$

The expression is at most $4\vartheta(1-\vartheta)^3 \leq 27/64$. The directions are linearly independent with probability $(1-1/Q)(1-2/Q)$. Otherwise repeated positions make the exactly-one event impossible. Conditioning on independence gives a uniform affine plane, so

$$c_{2,Q} = \frac{27}{64(1-1/Q)(1-2/Q)}, \qquad p^\star_{2,Q} = \frac{27}{91} + O(Q^{-1}).$$

For $Q \geq 8$, $c_{2,Q} \geq 4/Q$, and the same inverse-polynomial gap argument applies.

B.2 Queries outside the planted subspace

The matching in Theorem 4.3 does not cover a query outside K. Such a query always returns zero in the NO-source case, while a positive YES-source answer would distinguish the cases. We next bound exterior queries under an explicit additional restriction: the observer knows K but has no further information about the larger satisfying space.

Proposition B.3 (Exterior queries with no information about the extension)**.** *Fix an affine m-subspace $K \subseteq \mathbb{F}_2^{3N}$, with $m \leq d < N$. Conditional on K, let S be uniform among affine N-subspaces containing K, and use the matching distributions of Theorem 4.3, with either density mixture. An observer is given K, arbitrary access to the retained subset inside K, and no other information about S. If it makes at most q_{ext} adaptive membership queries outside K, the total-variation distance between its YES and NO transcripts is at most*

$$\min\left\{1,\ q_{\mathrm{ext}}\frac{2^N - 2^m}{2^{3N} - 2^m}\right\}. \tag{B.3}$$

For polynomially many queries this bound is exponentially small in N.

Proof. Expose the full internal restriction $D_{\mathrm{in}} = D \cap K$. Its law does not depend on which extension S was chosen and equals the NO output law. Couple this restriction and the observer's randomness identically in the two experiments. Now follow the hypothetical transcript in which every exterior answer is zero. The queried points along this path depend only on K, D_{in} and the randomness, so they are independent of the hidden extension S. For any fixed point $x \notin K$, symmetry in the quotient by the direction space of K gives

$$\Pr[x \in S \mid K] = \frac{2^N - 2^m}{2^{3N} - 2^m}.$$

Unless one of the at most q_{ext} candidate points belongs to S, all actual answers are zero because $D \subseteq S$, and the transcripts coincide. A union bound and the coupling inequality prove the result. □

The observer in this proposition does not receive the source circuit C or the affine map g. Either could constrain S in ways not captured by the model. The result therefore does not rule out finding a point of $S \setminus K$ by using the presentation. What the restriction theorem does rule out is obtaining new information merely by subdividing the same known K: every observation inside it is already determined by the matched restriction.

B.3 A single known point and anti-concentration

Symbol	Local meaning
shift_v; V	Translation by v; the randomly translated known point.
Y_H	Unique survivor, conditioned on successful isolation.
$p_{\max,A}(H)$	Largest point probability in that conditional distribution.
Z; $\beta_A(C)$	Singleton output or $\perp$; probability it equals the known point.
$\eta_{\mathrm{unif}}(L)$	Allowed multiplicative slack in almost-uniform selection.

We now return to computational consequences of the isolator. With a single known point, its retention probability can be estimated without inspecting any other witness. Specialize the construction to $m = 0$ and translations only. For $v \in \mathbb{F}_2^{3N}$, let $\mathrm{shift}_v(x) = x + v$ and set

$$G_{C,v} := J_{C,0,\mathrm{shift}_v} = H_{C,0} \circ \mathrm{shift}_v^{-1}. \tag{B.4}$$

Keep the same field representation and padding convention. In the NO-source case the only satisfying assignment is the known point v. In the YES-source case the satisfying set is

$$S = v + W_{0,w}, \qquad W_{0,w} = \{0^N\} \times \mathcal{L}_w, \qquad |S| = Q = 2^N. \tag{B.5}$$

A uniform ambient translation V, conditional on the coset S it produces, is uniform among the points of S. Indeed, precisely those points represent the same coset.

For a satisfiable input H, let Y_H have the distribution of the unique survivor conditional on successful isolation. Write

$$p_{\max,A}(H) = \max_{x \in \mathrm{Sol}(H)} \Pr[Y_H = x], \qquad \mathrm{Ent}_\infty(Y_H) = -\log p_{\max,A}(H).$$

Here Ent_∞ is min-entropy, applied to the distribution already conditioned on success.

Proposition B.4 (Anti-concentration of the isolated witness)**.** *Suppose A is a randomized polynomial-size pruning isolator with $p(L) \geq 1/2 + 1/\mathrm{poly}(L)$. If, for an inverse-polynomial $\gamma(L) > 0$, every sufficiently large source $C \in \mathcal{C}_\ell^{(1)}$ satisfies*

$$p_{\max,A}(G_{C,0}) \leq 2p(L) - 1 - \gamma(L), \qquad L = L_0(\ell), \tag{B.6}$$

then $\mathrm{NP} \subseteq \mathrm{P/poly}$. *No presentation invariance is required.*

Proof. The known origin is the only witness on a NO source input, so it survives with probability at least $p = p(L)$. On a YES source input, let $D = D_A(G_{C,0})$ and $p_{\mathrm{act}} = \Pr[|D| = 1]$. Then

$$\Pr[0 \in D] \leq \Pr[D = \{0\}] + \Pr[|D| \neq 1] \leq p_{\max,A}(G_{C,0}) + 1 - p \leq p - \gamma(L).$$

On successful runs, the origin is selected with conditional probability at most $p_{\max,A}(G_{C,0})$, so $\Pr[D = \{0\}] \leq p_{\mathrm{act}} p_{\max,A}(G_{C,0}) \leq p_{\max,A}(G_{C,0})$. All unsuccessful runs together have probability at most $1 - p$. The displayed bound therefore separates origin retention in the two source cases by $\gamma(L)$. Estimate it by repeated output evaluations at zero and apply Lemma 2.2. □

For $p = 1/2 + \varepsilon$, the hypothesis can be written as $\mathrm{Ent}_\infty(Y_{G_{C,0}}) \geq \log(1/(2\varepsilon - \gamma))$, provided $2\varepsilon - \gamma > 0$. An inverse-polynomial upper bound on the largest point probability thus requires only logarithmic min-entropy. Gharibian–Kamminga's almost-uniform condition bounds each witness's selection probability, conditional on success, by $(1 + \eta_{\mathrm{unif}}(L))/|\mathrm{Sol}(H)|$ [12]. Here the YES satisfying set has size 2^N. For polynomially bounded η_{unif}, the bound $(1 + \eta_{\mathrm{unif}}(L))/2^N$ satisfies (B.6) with a suitable inverse-polynomial gap.

B.4 Selection of the planted origin and information

The preceding argument also has a converse form: if the known origin were rarely the unique survivor on every YES source, its retention would decide the source promise. To make this precise, fix $C \in \mathcal{C}_\ell^{(1)}$, sample a uniform translation V, let $D = D_A(G_{C,V})$, and define

$$Z = \begin{cases} x, & D = \{x\}, \\ \perp, & |D| \neq 1, \end{cases} \qquad \beta_A(C) = \Pr[Z = V]. \tag{B.7}$$

The symbol $\perp$ records a nonsingleton output, including the empty set. The variable Z is used in the analysis; the reduction need not compute it from the output circuit.

Theorem B.5 (Necessary selection of the planted origin)**.** *Let A be a randomized polynomial-size pruning isolator with $p(L) = 1/2 + \varepsilon(L)$ and $\varepsilon(L) \geq 1/\operatorname{poly}(L)$. If $\mathrm{NP} \not\subseteq \mathrm{P/poly}$, then for every fixed $\xi \in (0, 2)$, infinitely many source lengths admit $C \in \mathcal{C}_\ell^{(1)}$ satisfying*

$$\beta_A(C) > (2 - \xi)\varepsilon(L_0(\ell)). \tag{B.8}$$

Proof. The origin survives with probability at least p on a NO source. On a YES source, split retention into singleton and unsuccessful outputs to get $\Pr[V \in D] \leq \beta_A(C) + 1 - p$. If the conclusion failed for a fixed ξ at all sufficiently large lengths, this probability would be at most $p - \xi\varepsilon$ on every YES source of those lengths. Sampling V, running A, and evaluating the output at V would then distinguish the source cases with an inverse-polynomial gap. By Lemma 2.2, this would imply the excluded collapse. □

Proposition B.6 (Information carried by the singleton output)**.** *For a fixed YES source, write $\beta = \Pr[Z = V]$. If $\beta \geq 1/Q$, then*

$$\begin{aligned} I(V; Z \mid S) &\geq N - h_2(\beta) - (1 - \beta)\log(Q - 1) \\ &\geq \beta N - h_2(\beta). \end{aligned} \tag{B.9}$$

Under the noncollapse assumption of Theorem B.5, every fixed constant $\varepsilon > 0$ therefore gives $\Omega(N)$ bits on infinitely many source inputs.

Proof. Conditional on S, the origin is uniform on its $Q = 2^N$ points, so $\mathrm{Ent}(V \mid S) = N$. As an information-theoretic estimator, return Z when $Z \in S$ and a fixed point of S otherwise. This guesses V correctly with probability at least β. If its error probability is e_{err}, encode an error bit and, when the guess is wrong, the correct one of the remaining $Q - 1$ points. Hence

$$\mathrm{Ent}(V \mid Z, S) \leq h_2(e_{\mathrm{err}}) + e_{\mathrm{err}} \log(Q - 1).$$

This upper bound is nondecreasing for $e_{\mathrm{err}} \leq 1 - 1/Q$. Since $e_{\mathrm{err}} \leq 1 - \beta$, subtracting it from N proves (B.9).

The first lower bound is nondecreasing for $\beta \geq 1/Q$. Therefore Theorem B.5 permits substitution of $\lambda_0 = (2 - \xi)\varepsilon$, which is larger than $1/Q$ for large inputs, giving $I(V; Z \mid S) \geq \lambda_0 N - h_2(\lambda_0)$. This is linear in N at fixed ε and $\xi < 2$. □

At $p = 2/3$, a suitable sequence of inputs gives $I(V; Z \mid S) \geq N/3 - o(N)$. Here $N = \log |S|$, while a complete witness in the common construction has $3N$ bits. If $\mu_{S,v}$ is the distribution of Z from presentation $G_{C,v}$ and $\overline{\mu}_S = Q^{-1} \sum_{v \in S} \mu_{S,v}$, then equivalently

$$I(V; Z \mid S) = \mathbb{E}_S \Big[\frac{1}{Q} \sum_{v \in S} D_{\mathrm{KL}}(\mu_{S,v} \| \overline{\mu}_S) \Big].$$

Under full presentation invariance, all the conditional output laws on this coset agree, and the information is zero. First-order retention invariance does not determine the probabilities of singleton outputs, so it need not make this information zero. Its collapse consequence follows instead from the separate $m = 0$ argument in Theorem B.2.

C The fiber-local amplification model

This appendix studies amplification in an abstract oracle model, independently of the circuit transformations used in the main theorem. Each witness has its own set of representations, called its fiber. The oracle may depend on the full queried set and the fiber labels, but an amplifier cannot pass information between hidden fibers. This restriction is different from both presentation invariance and DKMW's target-oblivious model.

C.1 The model and the oracle

Let $\mathcal{F}_x$ be the finite set of tagged or encoded copies of a base witness x. Different fibers are disjoint, as is a separate set of public auxiliary witnesses. A query specifies a portion $U_x \subseteq \mathcal{F}_x$ of each active hidden fiber and a public set K_{pub}. The oracle returns a subset of these queried points. Its random choices are not revealed to the transformation.

Definition C.1 (Fiber-local transformation)**.** After fixing its external random string, a fiber-local transformation has the following properties.

(i) For each hidden witness x, its next query portion U_x and its final retention bit depend only on x, its own fiber transcript, and the public transcript.

(ii) The public controller depends only on the public transcript and the external random string.

(iii) The local transition functions and their initialization are the same whether another hidden witness is present or absent. A controller is not given the list of other active fibers or their transcripts.

(iv) The finite round schedule is fixed or determined by the public controller. Public witnesses are removed before the final output is interpreted as a pruning of the base witnesses.

The definition permits tagging, duplication, Boolean operations within a fiber, projection back to its base witness, and repeated isolation of local outputs. It excludes using one hidden transcript to control another fiber or identifying points from different hidden fibers. Inactive rounds can be padded to give the compared experiments a common schedule.

Fix $0 < p \leq 2/3$ and a canonical point $\mathrm{rep}(U_x)$ for every nonempty fiber portion. The oracle $\mathcal{O}_p$ is memoryless and uses fresh independent randomness on each call. An empty query receives the empty answer. If a public witness is present, the oracle returns the canonical public point with probability p and the empty set otherwise; it ignores the hidden fibers on that call.

When no public witness is present, one nonempty hidden fiber receives its canonical point with probability p and the empty set otherwise. For two nonempty hidden fibers with canonical points a, b, use

output	$\varnothing$	$\{a\}$	$\{b\}$	$\{a, b\}$
probability	$1 - 3p/2$	$p/2$	$p/2$	$p/2.$

(C.1)

For completeness, if three or more hidden fibers are present, return one fixed canonical queried point with probability p and the empty set otherwise.

Every answer is a subset of the query, and every nonempty query has singleton probability exactly p. The two-fiber probabilities are nonnegative precisely for $p \le 2/3$. The key identity is marginal: either fiber receives its canonical point with probability $p/2 + p/2 = p$ and nothing otherwise, exactly as it would if it were the only hidden fiber.

Lemma C.2 (Equality of adaptive local transcripts)**.** *Against $\mathcal{O}_p$, the joint distribution of a fixed witness x's local transcript and the public transcript is identical on base sets $\{x\}$ and $\{x, y\}$. The analogous assertion holds for y.*

Proof. Fix the external random string and induct on the rounds. Equal local and public views determine the same next U_x and public query by locality. If the query contains a public point, its answer is Bernoulli-p and the hidden answer is empty in both experiments. Without a public point, an empty U_x again receives an empty answer. Otherwise its canonical point survives with probability p, regardless of whether U_y is empty.

This last conditional law does not depend on the competing transcript, even if that transcript is correlated with the local history. Fresh oracle randomness therefore preserves equality of the joint local/public laws at the next round. Averaging over the external random string proves the lemma. □

C.2 The two-witness inequality

Use DKMW's parameters: p_1 lower-bounds retention on singleton inputs, and p_2 lower-bounds the probability that an output on a two-witness input is not satisfied by both witnesses. In particular, an empty output counts toward p_2; this is not the same event as successful isolation. Their collapse criterion is $p_1 + p_2/2 > 1 + 1/\operatorname{poly}(L)$ [1, 2].

Theorem C.3 (Fiber-local amplification bound)**.** *For every $0 < p \le 2/3$, a fiber-local transformation using finitely many calls to $\mathcal{O}_p$ has parameters satisfying*

$$p_1 + \frac{1}{2} p_2 \le 1. \tag{C.2}$$

In particular, it cannot guarantee isolation probability greater than $2/3$.

Proof. On base set $\{x, y\}$, let $\mathsf{Keep}_x, \mathsf{Keep}_y$ be the two final retention bits. By Lemma C.2 and locality of the output rule, $\Pr[\mathsf{Keep}_x = 1] \ge p_1$ and $\Pr[\mathsf{Keep}_y = 1] \ge p_1$. Inclusion–exclusion gives

$$\Pr[\mathsf{Keep}_x = \mathsf{Keep}_y = 1] \ge \Pr[\mathsf{Keep}_x = 1] + \Pr[\mathsf{Keep}_y = 1] - 1 \ge 2p_1 - 1.$$

Therefore

$$p_2 \le 1 - \Pr[\mathsf{Keep}_x = \mathsf{Keep}_y = 1] \le 2 - 2p_1,$$

which is (C.2). A transformation with isolation guarantee p_{out} has $p_1, p_2 \ge p_{\text{out}}$, and hence $3p_{\text{out}}/2 \le 1$. □

At $p = 2/3$, the oracle itself attains equality. The endpoint is exactly the nonnegativity condition $1 - 3p/2 \ge 0$ in the two-fiber distribution. Independent repetition followed by assignment-wise Boolean combination is one special case of the model. The oracle is an abstract set rule; the bound does not assert that it has a polynomial-size circuit implementation.

C.3 Nonlocal selection and pointwise consensus

Choosing the first nonempty circuit is not a fiber-local operation. Its answer at one point can depend on the existence of a different witness. The following proposition shows that an efficient implementation of this operation would itself have a complexity consequence.

Proposition C.4 (Selecting the first nonempty circuit)**.** *Suppose a randomized polynomial-size transformation takes two circuits, each with at most one satisfying assignment, and outputs a circuit computing their first nonempty member with probability at least* $1/2 + 1/\operatorname{poly}(L)$*, where* L *is the joint input length. Then* $\mathrm{NP} \subseteq \mathrm{P/poly}$*. An input-oblivious randomized pointwise Boolean combiner has worst-case success probability at most* $1/2$ *for this task.*

Proof. Given a source USAT circuit $C(z)$ on n variables, define

$$B_1(b, z) = (\neg b) \wedge C(z), \qquad B_2(b, z) = b \wedge \mathbf{1}_{\{z=0^n\}}.$$

At $a = (1, 0^n)$, the first nonempty circuit accepts if C is unsatisfiable and rejects if C is uniquely satisfiable. Evaluating the output at a therefore distinguishes the promise cases with an inverse-polynomial advantage. Amplification and Lemma 2.2 apply.

For the pointwise assertion, the combiner samples a Boolean function Comb independently of the input descriptions and returns $\mathrm{Comb}(B_1(x), B_2(x))$. Compare the pairs of satisfying sets $(\varnothing, \{v\})$ and $(\{u\}, \{v\})$, with $u \neq v$. At v the combiner sees the same bits $(0, 1)$ in both cases, but the correct outputs are opposite. Its answers at that point have the same distribution, so one instance has success probability at most $1/2$. □

Proposition C.5 (Pointwise majority on the unique-satisfiability promise)**.** *Let* A *be a randomized polynomial-size pruning isolator with* $p(L) \geq 1/2 + 1/\operatorname{poly}(L)$*. At every length there is a fixed polynomial-size list of random strings* $\rho_1, \ldots, \rho_t$ *such that a strict majority of the circuits* $B_i = A(C; \rho_i)$ *are uniquely satisfiable for every satisfiable* C*. On the promise that* C *has at most one satisfying assignment,*

$$\mathrm{Maj}_{i=1}^{t} B_i \equiv C. \tag{C.3}$$

Proof. A polynomial number of independent random strings, a concentration bound, and a union bound over all length-L circuits give the fixed list. If C is unsatisfiable, pruning makes every B_i unsatisfiable. If C has unique witness w, every B_i computes either zero or the point predicate $P_w(x) = \mathbf{1}_{\{x=w\}}$. Successful isolation is then exactly the second case. A strict majority compute P_w, so their pointwise majority agrees with C on every assignment. □

More generally, a pointwise combiner with $\mathrm{Comb}(0^t) = 0$ produces either zero or P_w on such a list, because every membership bit vanishes away from w. If the combination preserves satisfiability, it computes P_w again. It may change the representation, but leaves the original promise problem to be decided. Taking the OR of circuits that are empty or share one possible witness has the same limitation.

D The retention polytope

The main proof uses joint distributions of retention bits. This appendix instead characterizes the possible marginal probabilities for one random pruning. The result is independent of the preceding arguments and separates what follows from first-order constraints from what requires a joint law.

Let $W = \{w_1, \ldots, w_s\}$ and let $D \subseteq W$ be random, with $\Pr[|D| = 1] \geq p$. Write $r_i = \Pr[w_i \in D]$, and let $\mathcal{R}_{s,p}$ be the set of feasible vectors $r \in [0,1]^s$. The simplex is

$$\Delta_{s-1} = \{\boldsymbol{\lambda} \in [0,1]^s : \textstyle\sum_i \lambda_i = 1\}.$$

Here the subscript is its dimension. We use s for the number of witnesses in this appendix, reserving k for the query budget in the preceding results.

Theorem D.1 (The feasible retention vectors)**.** *For $s \geq 1$ and $0 < p \leq 1$,*

$$\mathcal{R}_{s,p} = p\Delta_{s-1} + (1-p)[0,1]^s, \tag{D.1}$$

where the right-hand side is a Minkowski sum. Equivalently, $r \in [0,1]^s$ is feasible if and only if

$$\sum_{i=1}^{s} r_i \geq p, \qquad \sum_{i \in J} r_i \leq p + (1-p)|J| \quad \text{for every } J \subseteq \{1, \ldots, s\}. \tag{D.2}$$

Proof. Decompose a feasible distribution. Let $\alpha_i = \Pr[D = \{w_i\}]$. Since $\sum_i \alpha_i \geq p$, choose $0 \leq \widetilde{\alpha}_i \leq \alpha_i$ with $\sum_i \widetilde{\alpha}_i = p$ and set $\lambda_i = \widetilde{\alpha}_i / p$. Remove mass $\widetilde{\alpha}_i$ from the singleton outcome $\{w_i\}$. Exactly $1-p$ probability mass remains. For $p < 1$, normalize that residual distribution and let x_i be its retention marginals. Then $x \in [0,1]^s$ and $r = p\boldsymbol{\lambda} + (1-p)x$. At $p = 1$, no residual mass remains.

Realize every vector in the sum. Given $r = p\boldsymbol{\lambda} + (1-p)x$, with $\boldsymbol{\lambda} \in \Delta_{s-1}$ and $x \in [0,1]^s$, output a singleton drawn from $\boldsymbol{\lambda}$ with probability p. Otherwise include each point independently with its marginal from x. This realizes r and has singleton probability at least p, proving (D.1). The two summands immediately give the inequalities in (D.2).

Recover the decomposition from the inequalities. For $p < 1$, define

$$\underline{\lambda}_i = \max\{0, (r_i - (1-p))/p\}, \qquad \overline{\lambda}_i = \min\{1, r_i/p\}.$$

Each interval is nonempty. Apply the subset inequality to $J = \{i : r_i > 1-p\}$ to obtain $\sum_i \underline{\lambda}_i \leq 1$. Also $\sum_i \overline{\lambda}_i \geq 1$: either one $r_i \geq p$ already gives $\overline{\lambda}_i = 1$, or the claim follows from $\sum_i r_i \geq p$. We can therefore choose $\lambda_i \in [\underline{\lambda}_i, \overline{\lambda}_i]$ with $\sum_i \lambda_i = 1$. The interval constraints imply $x_i = (r_i - p\lambda_i)/(1-p) \in [0,1]$, producing the desired decomposition. For $p = 1$, the inequalities directly force $\sum_i r_i = 1$. □

Corollary D.2 (Ordered retention probabilities)**.** *If $r_{(1)} \geq \cdots \geq r_{(s)}$ are the ordered coordinates, then, for $1 \leq m \leq s$,*

$$\sum_{j=1}^{m} r_{(j)} \leq p + m(1-p), \qquad r_{(m)} \leq 1 - p + \frac{p}{m}. \tag{D.3}$$

All prefix inequalities are attained simultaneously by one feasible vector. Each individual order-statistic bound is attained as well.

Proof. Apply the subset inequality to the largest m coordinates, and divide by m for the second assertion. Outputting $\{w_1\}$ with probability p and W otherwise gives $(1, 1-p, \ldots, 1-p)$, attaining all prefix bounds. For a fixed m-element subset J, output a uniform singleton of J with probability p and all of J otherwise. The coordinates in J equal $1 - p + p/m$, and all other coordinates are zero. □

Corollary D.3 (The number of low-retention witnesses)**.** *For $1 \le j \le s$, at least j witnesses have retention strictly below p whenever*

$$p > \frac{s-j+1}{2s-2j+1}. \tag{D.4}$$

At equality there is a feasible distribution with exactly $j-1$ such witnesses. If $p > 1/2$, the number of witnesses with retention at least p is at most

$$\min\left\{s, \left\lfloor \frac{p}{2p-1} \right\rfloor\right\}. \tag{D.5}$$

Proof. Set $m = s - j + 1$. If fewer than j witnesses had retention below p, then $r_{(m)} \ge p$. Equation (D.3) would imply $p \le 1 - p + p/m$, or $p \le m/(2m-1)$. At equality, the m-element construction from Corollary D.2 gives exactly m coordinates equal to p and the rest zero.

If m coordinates are at least p, the prefix bound gives $mp \le p + m(1-p)$, and hence $m \le p/(2p-1)$. The same construction attains the integer bound when it is at most s; taking $J = W$ covers the other case. □

Thus at $p = 1/2 + \varepsilon$, at most $\lfloor 1/(4\varepsilon) + 1/2 \rfloor$ witnesses can have retention at least p, regardless of the total number s. This gives many low-retention witnesses when s is large, but does not identify one efficiently. The main affine test avoids that search: it examines a known subspace whose position is random relative to an independent reference output.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il