Electronic Colloquium on Computational Complexity, Report No. 38 (2026)

Hardness Amplification Beyond Boolean Functions

Nobutaka Shimizu
Institute of Science Tokyo
shimizu.n.ah@m.titech.ac.jp

Kenji Yasunaga
Institute of Science Tokyo
yasunaga@c.titech.ac.jp

Abstract

A central goal in average-case complexity is to understand how average-case hardness can be amplified to near-optimal hardness. Classical results such as Yao's XOR lemma establish this principle for Boolean functions, but these techniques typically apply only to artificially constructed functions, rather than to natural computational problems. In this work, we extend hardness amplification beyond the Boolean setting and extend the XOR Lemma to the sum of functions over the finite field $\mathbb{F}_p$, where p is a prime. Specifically, we show that if a function $f\colon \{0,1\}^n \to \mathbb{F}_p$ fails to be computed on at least a δ-fraction of inputs, then the k-wise sum $f^{+k}(x_1,\ldots,x_k) = f(x_1)+\cdots+f(x_k)$ becomes almost optimally unpredictable: no efficient algorithm can compute it with success probability exceeding $\frac{1+\varepsilon}{p}$ for suitable parameters k,δ,ε. Our proof is based on the pseudo-average-min entropy characterization of unpredictability due to Zheng (2014) and Vadhan and Zheng (2012), which we simplify and quantitatively refine to make the dependence of the circuit blow-up on all parameters fully explicit.

As an application, we obtain the first *error-tolerant random self-reduction* for a natural subgraph counting problem. Specifically, we show that any circuit that correctly counts triangles in an Erdős–Rényi random graph with noticeable probability can be transformed into a worst-case circuit with only a quasi-linear overhead.

We further extend the query lower bound framework of Shaltiel and Viola (2010) to the $\mathbb{F}_p$-valued setting, proving that any (possibly adaptive) black-box hardness amplification over $\mathbb{F}_p$ must make at least $\Omega(p\log(1/\delta)/\varepsilon^2)$ oracle queries. Our proof substantially simplifies the core *fixed-set lemma* underlying previous analyses, offering a more modular and entropy-based argument.

Contents

ISSN 1433-8092

1 Introduction

A central pursuit in computational complexity is to understand how *mild* average-case hardness can be amplified into *strong* hardness, which is referred to as *hardness amplification*. Hardness amplification plays a foundational role in diverging areas of theoretical computer science, including derandomization [IW97; STV01], randomness extractors [DT09], one-way functions [Yao82], and error-correcting codes [IJKW10; DHKNT21; Tre03].

However, in these traditional formulations, the amplified problems are usually *artificial* functions constructed from the original one, rather than natural computational tasks. As a result, applying these hardness amplification techniques directly to *natural* problems—such as those arising in linear algebra or graph theory—remains challenging. Recently, there has been growing interest in bridging this gap by extending hardness amplification to more structured or algebraic settings, such as matrix multiplication [AGGS22; HS23; GSS24; HS25], linear-algebraic tasks [AGGSS24; VZ25], planted clique problem [HS24; NR25], and optimization problems [GK20]. These developments highlight the potential of hardness amplification as a unifying principle connecting pseudorandomness, average-case complexity, high-dimensional statistics, and structured computational problems.

The most classical and influential results of hardness amplification are *Yao's XOR lemma* and the *direct product theorem* [Yao82; GNW11]. The former asserts that if a Boolean function f is mildly hard to compute on average, then the XOR of several independent copies of f, i.e., $f(x_1) + \cdots + f(x_k) \bmod 2$, becomes exponentially harder with respect to k. The latter asserts that if a function f (not necessarily Boolean-valued) is mildly hard to compute on average, then the direct product of several independent copies of f, i.e., $(f(x_1), \cdots, f(x_k))$, becomes exponentially harder.

Many previous works towards extending hardness amplification to more natural computational problems essentially rely on the direct product theorem [HS23; HS24; GK20] or deep theory from additive combinatorics [AGGS22; AGGSS24]. For example, the reduction in [HS23] for matrix multiplication exploits the direct product structure inside the matrix multiplication problem. The framework based on additive combinatorics, such as the one in [AGGS22; AGGSS24], is also tailored to the problem-specific structure of linear-algebraic tasks. More importantly, both of these reductions critically rely on the fact that these problems are *efficiently verifiable*: for example, in the case of matrix multiplication, one can efficiently verify whether $AB = C$ holds for given matrices A, B, C. This verifiability limits the applicability of existing amplification methods to broader classes of problems where efficient verification is not available, such as counting problems.

On the other hand, the algebraic nature of Yao's XOR lemma underlies a wide range of applications in error-correcting codes [DHKNT21; IJKW10; STV01; Tre03]. Yet, despite their ubiquity, essentially all known formulations apply only to *Boolean-valued* functions since the reduction in known proofs of the XOR lemma relies on taking the *majority vote* of weak average-case solvers. In contrast, many natural computational problems of combinatorial flavor—such as counting specific subgraphs in an Erdős–Rényi random graph—are *arithmetic* in nature. Extending hardness amplification principles to this arithmetic setting has remained an open conceptual challenge.

1.1 Our Results

This subsection presents our main contributions. At a high level, we extend Yao's XOR lemma from Boolean-valued functions to $\mathbb{F}_p$-valued functions, thereby establishing a general principle of hardness amplification in arithmetic settings. As an application, we obtain the first error-tolerant random self-reduction for the triangle counting problem over Erdős–Rényi random graphs.

Arithmetic XOR Lemma over Prime Finite Fields. Yao's classical XOR lemma applies only to Boolean functions, relying crucially on the majority vote of weak solvers. We overcome this limitation by developing an additive analogue of XOR lemma over $\mathbb{F}_p$. We say that a function $f : \{0,1\}^n \to \mathbb{F}_p$ is *δ-hard for size s* if for any size-s circuit C, it holds that $\Pr_{x\sim\{0,1\}^n}[C(x) \neq f(x)] \geq \delta$.

Roughly speaking, we show that if a function $f : \{0,1\}^n \to \mathbb{F}_p$ is δ-hard for some $\delta > 0$, then the sum of several independent copies of f,

$$f^{+k}(x_1, \ldots, x_k) = f(x_1) + f(x_2) + \cdots + f(x_k) \bmod p,$$

is $(1 - (1+\varepsilon)/p)$-hard for slightly smaller circuits.

Theorem 1.1 (XOR Lemma over $\mathbb{F}_p$, informal; see Theorem 3.4 for the formal statement)**.** *Let $\mathbb{F}_p$ be a finite field such that $p = |\mathbb{F}_p|$ is a prime. Let $f\colon \{0,1\}^n \to \mathbb{F}_p$ be a function that is δ-hard for size s. Then, for any $\varepsilon > 0$ and for some $k = O\left(\frac{p^2}{\delta}\log(p/\varepsilon)\right)$, the function $f^{+k}(x_1, \ldots, x_k) := f(x_1) + \cdots + f(x_k) \bmod p$ is $(1 - \frac{1+\varepsilon}{p})$-hard for size $s' = \Omega\left(s \cdot \frac{\varepsilon^2}{pk^2 \log(k/\varepsilon)}\right)$.*

Remark 1.2. *Our proof crucially relies on the assumption that the field size p is prime. Intuitively, this is because the additive group $\mathbb{Z}_p$ has no nontrivial proper subgroups other than $\{0\}$, which ensures that sums over independent copies of a function mix sufficiently to amplify hardness.*

To see that this assumption is necessary, consider the following simple counterexample over a composite modulus. Let $f : \{0,1\}^n \to \mathbb{Z}_4$ be a random function such that for each x, $f(x)$ independently takes the value 0 or 2 with probability $1/2$. A straightforward counting argument shows that f is $1/2$-hard for any circuit of size $s = 2^{o(n)}$. However, for any $k \in \mathbb{N}$, the k-wise sum function

$$f^{+k}(x_1, \ldots, x_k) := f(x_1) + \cdots + f(x_k) \mod 4$$

still takes values only in $\{0,2\}$ with probability $1/2$, and therefore remains $1/2$-hard. In other words, no hardness amplification occurs over this composite modulus.

Our arithmetic XOR lemma[1] is closely related to the recent work of Li and Vasudevan [LV25], who proved an XOR lemma for integer-valued functions without taking values modulo a prime. However, since their result does not involve taking values modulo a prime, it does not yield the optimal $(1-(1+\varepsilon)/p)$-hardness that we obtain for $\mathbb{F}_p$-valued functions in Theorem 1.1. Interestingly, Theorem 1.1 can be used to derive their integer-valued XOR lemma as a special case. A familiar reader might wonder if Theorem 1.1 follows by combining the direct product theorem [IJKW10] and local list-decoding algorithms for Hadamard codes over $\mathbb{F}_p$ [GRS00]. Unfortunately, this approach fails; see Remark 1.5 for details.

Application: Triangle Counting over Random Graphs. As an important application of the $\mathbb{F}_p$-XOR lemma, we obtain an *error-tolerant random self-reduction* for the triangle counting problem on Erdős–Rényi random graphs, i.e., a random self-reduction that works even if the average-case solver makes errors for a large fraction of inputs.

Theorem 1.3 (informal; see Theorem 4.1 for the formal statement)**.** *Let $\theta \in (0,1)$ be any constant and $\alpha > 0$ be any parameter. Suppose that there exists a size-s circuit C such that, for every sufficiently large n,*

$$\Pr_{G\sim\mathcal{G}(n,\theta)}[C(G) = \mathrm{Tri}(G)] \geq \alpha,$$

where $\mathrm{Tri}(G)$ is the number of triangles in G. Then, there exists a randomized circuit C' of size $\widetilde{O}\big((s+n^2)\cdot \log(1/\alpha)/\alpha^2\big)$ such that, for every sufficiently large n and for every n-vertex graph G, it holds that

$$\Pr_{C'}\big[C'(G) = \mathrm{Tri}(G)\big] \geq 2/3,$$

where the probability is taken over the internal randomness of C'.

Previously, Theorem 1.3 is known to hold for $\alpha \geq 1 - 1/(\log n)^7$ [BBB21] or $\alpha \geq 1 - c$ for a sufficiently small constant $c > 0$ if $\theta = 1/2$ [ABPSS25]. Theorem 1.3 gives the first random self-reduction for a counting problem that works under *weak correctness guarantees* (for any constant $\alpha > 0$), in contrast to previous reductions that required strong correctness guarantees (e.g., $1 - 1/\mathrm{polylog}(n)$ [BBB21; DLW20]). See Section 1.2 for more details.

We note that, unlike the reductions in [BBB21; ABPSS25], our reduction is *non-uniform* (non-constructive): we show that the existence of an average-case solver implies the existence of a worst-case solver, but we do not provide an efficient transformation from one to the other. In contrast, the previous reductions are *uniform* (constructive) in the sense that, given an explicit description of an average-case solver, they output (in nearly the same running time as the average-case solver) a description of a worst-case solver. We leave it as an open problem to prove Theorem 1.3 using a uniform reduction.

Another interesting direction is to detect the optimal success probability α for circuits of size $n^{2+o(1)}$ (under the worst-case hardness of triangle counting for such circuits). The constant circuit that outputs $\lfloor \binom{n}{3}\theta^3 \rfloor$ (the expected number of triangles rounded to integers) achieves $\alpha = \Theta(1/n^2)$ from the local central limit theorem [GK16]. Is this α optimal for circuits of size $n^{2+o(1)}$? Theorem 4.1 enables us to deal with the range $\alpha \geq n^{-o(1)}$ due to the blow-up factor of $\log(1/\alpha)/\alpha^2$.

[1] Formally, the operation considered here is addition modulo p, not XOR over $\{0,1\}$, so the term "sum lemma" would be more precise. We nevertheless use the name "XOR lemma" to emphasize the conceptual connection to Yao's classical hardness amplification result, as this terminology is well established in computational complexity theory.

Query Lower Bound for Black-Box Hardness Amplification. In addition to establishing the arithmetic XOR lemma itself, we investigate the quantitative limitations of black-box hardness amplification. Roughly speaking, a $(1-\delta,\varepsilon)$-black-box hardness amplification consists of a construction Con that transforms a base function $f\colon\{0,1\}^n\to\mathbb{F}_p$ into a more complex function $\mathsf{Con}_f\colon\{0,1\}^m\to\mathbb{F}_p$ (such as f^{+k}) that is intended to be harder to compute on average, together with a reduction $\mathsf{Red}^{(\cdot)}$ that, given oracle access to an algorithm h computing Con_f with advantage ε (i.e., $\Pr_y[h(y)=\mathsf{Con}_f(y)]\geq(1+\varepsilon)/p$), uses a limited number of oracle queries to h to recover the original function f on at least a $(1-\delta)$ fraction of inputs. In other words, the reduction Red converts a solver for the amplified problem into a solver for the base problem while making only a small number of oracle queries. Such frameworks capture all known black-box proofs of XOR lemmas and direct product theorems [SV10; GSV18; Sha23].

A fundamental quantitative question is how many oracle queries Red must make in order to succeed in such a reduction. In Theorem 1.1, the circuit size (or query complexity) of the reduction grows polynomially with the field size p. It is natural to ask whether this dependence can be improved to, say, logarithmic in p.

To address this, we extend the query lower bound framework of Shaltiel and Viola [SV10] and Grinberg, Shaltiel, and Viola [GSV18], originally developed for Boolean functions, to the setting of $\mathbb{F}_p$-valued functions. Our analysis yields a lower bound showing that the number of oracle queries in any black-box hardness amplification over $\mathbb{F}_p$ must necessarily grow at least linearly with p. Informally, we show the following:

Theorem 1.4 (informal; see Theorem 5.2)**.** *Let* $(\mathsf{Con},\mathsf{Red})$ *be any* $(1-\delta,\varepsilon)$ *black-box hardness amplification for* $\mathbb{F}_p$*-valued functions, where* Red *makes at most* q *oracle queries (possibly adaptively) to its oracle. Then, it holds that*

$$q\geq\Omega\left(\frac{p\log(1/\delta)}{\varepsilon^2}\right).$$

This result generalizes the Boolean lower bound of Shaltiel and Viola [SV10] and Grinberg, Shaltiel, and Viola [GSV18] and demonstrates that the polynomial dependence on p in our $\mathbb{F}_p$-XOR lemma is in fact *unavoidable*.

Our proof substantially simplifies the core argument underlying the previous query lower bounds of Shaltiel and Viola [SV10] and Grinberg, Shaltiel, and Viola [GSV18], particularly the so-called *fixed-set lemma* that lies at the heart of their analysis. Specifically, we reformulate the recent min-entropy–based iterative proof introduced by Shaltiel [Sha23] in a more modular and conceptual manner using the chain rule for Kullback–Leibler (KL) divergence. This perspective allows us to view the entire argument as a clean entropy decomposition process, avoiding several of the delicate combinatorial steps present in earlier proofs.

In addition, while the prior works established their lower bounds via the information-theoretic analysis of a *Boolean coin problem*, we extend this argument to the *$\mathbb{F}_p$-valued coin problem*. Our analysis yields a simpler and more general proof based on the χ^2-divergence, which captures the same quantitative behavior as in the Boolean case but applies uniformly to arbitrary finite fields.

1.2 Related Work

Worst-case-to-average-case reductions for subgraph counting problems. There is a rich line of work on worst-case-to-average-case reductions for subgraph counting problems. Goldreich and Rothblum [GR18] proved counting k-clique subgraphs in a random graph drawn from a suitably

chosen distribution is strongly hard assuming the worst-case hardness of it. However, although the distribution is efficiently sampleable, it is far from the Erdős–Rényi random graph.

This issue was addressed by Boix-Adserà, Brennan, and Bresler [BBB21], who presented a worst-case-to-average-case reduction for the clique counting problem. Specifically, they showed that if one can compute the number of k-clique subgraphs in an Erdős–Rényi random graph $G \sim \mathcal{G}(n, \theta)$ (for any constant θ) with probability at least $1 - \frac{1}{(\log n)^{\Omega(k^2)}}$ in time $T(n)$, then one can compute the number of cliques of size k in an arbitrary graph G with probability at least $2/3$ in time $O(T(n) \cdot \mathrm{polylog}(n))$. This result was later generalized to the problem of counting an arbitrary subgraph problems by Dalirrooyfard, Lincoln, and Williams [DLW20]. The main drawback of these results is that they establish only weak average-case hardness. Specifically, assuming the worst-case hardness of subgraph counting problems, they only show that the same problem is δ-hard on Erdős–Rényi random graphs for $\delta = 1/\mathrm{polylog}(n)$.

For the problem of computing the *parity* of k-clique subgraphs, Boix-Adserà, Brennan, and Bresler [BBB21] and Goldreich [Gol20] proved the same result for $\delta = 2^{-O(k^2)}$. This was further improved by Hirahara and Shimizu [HS23], who proved that computing the parity of triangle subgraphs over an Erdős–Rényi random graph is $(1/2 - \varepsilon)$-hard for any constant $\varepsilon > 0$ assuming the worst-case hardness of it. However, they left open the question of whether one can establish the optimal $(1 - (1 + \varepsilon)/p)$-hardness for subgraph counting problem modulo p for any prime p under the corresponding worst-case hardness assumption.

XOR lemma for non-binary functions. For non-binary functions, there has been limited progress. Li and Vasudevan [LV25] recently established Yao's XOR lemma for functions over integers. Specifically, they showed that if a function $f\colon \{0,1\}^n \to \{0, \dots, m\}$ is δ-hard to compute, then the sum $f(x_1) + \cdots + f(x_k)$ of k independent copies becomes $(1 - \varepsilon)$-hard to compute for appropriate parameters k and ε. Hirahara and Shimizu [HS25, Theorem 2.1] established an XOR lemma for multi-output functions $f\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$, which is specialized for a particular problem of matrix multiplication. Hence, their setting differs from that of our XOR lemma.

Remark 1.5. *A familiar reader might wonder if Theorem 1.1 follows by combining the direct product theorem [IJKW10] and the list-decoding algorithm for Hadamard codes over $\mathbb{F}_p$ [GRS00]: For a weakly hard function f, we can apply the direct product theorem to prove that the direct product function*

$$(x_1, \dots, x_k) \mapsto (f(x_1), \dots, f(x_k))$$

is $(1 - \gamma)$-hard if $k = \Omega(\log(1/\gamma)/\delta)$, where γ is a parameter to be specified later. Then, the list-decoding algorithm for Hadamard codes [GL89; GRS00] implies that the function

$$(x_1, \dots, x_k, r_1, \dots, r_k) \mapsto \sum_{i=1}^{k} r_i f(x_i) \tag{1}$$

is $(1/p - O(\gamma))$-hard, where $r_1, \dots, r_k \sim \mathbb{F}_p$.

If $p = 2$, then with probability $\Omega(1/\sqrt{k})$ over the choice of $r_1, \dots, r_k \sim \mathbb{F}_2^k$, exactly $k/2$ of them takes value 1, in which case the function $\sum_i r_i f(x_i)$ can be seen as the $(k/2)$-wise XOR function. This implies that the $(k/2)$-wise XOR function is $(1/2 - O(\sqrt{k} \cdot \gamma))$-hard (see [IJKW10, Section 5] for more detail). Setting $\gamma = \varepsilon/\sqrt{k}$ (a sufficiently large k satisfies $k = \Omega(\log(1/\gamma)/\delta) = \Omega(\log(k/\varepsilon)/\delta)$) yields Yao's XOR lemma.

However, if $p \geq 3$, this argument fails because the probability that the right hand side of Eq. (1) equals to the ℓ-wise sum function for some $\ell = \Theta(k)$ is $\binom{k}{\ell}p^{-k} \leq \binom{k}{k/2}p^{-k} = (2/p)^k \cdot O(1/\sqrt{k})$. Thus, we need to set $\gamma = \varepsilon \cdot (p/2)^k \cdot \Omega(\sqrt{k})$ but we cannot choose k satisfying $k = \Omega(\log(1/\gamma)/\delta) = \Omega(k \log(k/\varepsilon)/\delta)$.

1.3 Proof Overview

In this subsection, we give an overview of the proof of the arithmetic XOR lemma (Theorem 1.1) and worst-case-to-average-case reduction for triangle counting (Theorem 1.3).

Arithmetic XOR lemma (Section 3). Our proof of Theorem 1.1 builds on the characterization of average-case hardness in terms of the *pseudo average-min entropy* introduced by Zheng [Zhe14] and Vadhan and Zheng [VZ12]. Roughly speaking, this characterization (Theorem 2.11) provides an entropic interpretation of average-case hardness: for a function $f : \{0,1\}^n \to \mathbb{F}_p$, let X be uniformly distributed over $\{0,1\}^n$, and consider the joint distribution $(X, f(X))$. Then, f is δ-hard if and only if there exists a pair of random variables (X, Z) satisfying

1. $\widetilde{H}_\infty(Z \mid X) \geq \log_2 \frac{1}{1-\delta}$, and
2. $(X, f(X)) \approx_c (X, Z)$,

where $\widetilde{H}_\infty(Z \mid X)$ denotes the *average conditional min-entropy* of Z given X and $\approx_c$ denotes computational indistinguishability (see Section 2.1 for details).

Given such a pair (X, Z) guaranteed by the above characterization, we consider independent copies $(X_1, Z_1), \ldots, (X_k, Z_k)$. Since taking direct product preserves computational indistinguishability (see, e.g., [HVV06, Lemma 3.2]), we have $(X_1, \ldots, X_k, f(X_1) + \cdots + f(X_k)) \approx_c (X_1, \ldots, X_k, Z_1 + \cdots + Z_k)$. Furthermore, using the fact that $\widetilde{H}_\infty(Z_i \mid X_i) \geq \log_2 \frac{1}{1-\delta}$ for each i together with Fourier analysis over $\mathbb{F}_p$, we show that the sum $Z_1 + \cdots + Z_k$ becomes close to uniform over $\mathbb{F}_p$ when k is sufficiently large. Formally, letting U denote the uniform distribution over $\mathbb{F}_p$, we obtain that $(X_1, \ldots, X_k, f(X_1) + \cdots + f(X_k)) \approx_c (X_1, \ldots, X_k, U)$. Applying the above characterization in the reverse direction, we conclude that $f^{+k}(X_1, \ldots, X_k) := f(X_1) + \cdots + f(X_k)$ is $(1 - (1+\varepsilon)/p)$-hard, establishing the desired hardness amplification.

The results of Zheng [Zhe14] and Vadhan and Zheng [VZ12] were originally developed in the cryptographic setting, where one typically works with polynomial-size circuits and negligible advantage. In our application, however, we require a more precise formulation that explicitly accounts for the circuit-size blowup incurred by the reduction as well as the resulting change in advantage. To the best of our knowledge, this work is the first to apply their pseudo average-min entropy characterization theorem to the analysis of hardness amplification, and in particular, to give a proof of an XOR lemma based on this framework.

Remark 1.6. *The characterization theorem of Zheng [Zhe14] and Vadhan and Zheng [VZ12] can also be proved in the uniform computational model by invoking their uniform minimax theorem. However, in our proof of the XOR lemma, the step showing that computational indistinguishability is preserved under taking direct products (Lemma 3.3) crucially relies on non-uniform advice. Thus, it remains open whether our XOR lemma can be established in a fully uniform setting. We note that Impagliazzo, Jaiswal, Kabanets, and Wigderson [IJKW10, Theorem 1.11] proved a uniform XOR lemma for Boolean functions via a different approach. Moreover, even if our XOR lemma*

could be made uniform, the application to the triangle counting problem involves another step that essentially uses non-uniform advice, so a fully uniform reduction would not follow immediately.

As an additional application, we demonstrate that the pseudo average-min entropy framework also provides a concise derivation of the classical direct product theorem [IJKW10], further illustrating its utility in hardness amplification contexts. See Section 3.1 for details.

Finally, by revisiting the minimax-based proof underlying the pseudo average-min entropy characterization, we show that the same argument extends naturally to search-type problems. In particular, we obtain a *hardcore lemma for efficiently verifiable problems with unique solutions*, which can be viewed as a search analogue of the classical hardcore lemma of Impagliazzo [Imp95]. This class encompasses several central search problems such as the *planted clique problem* [Jer92; Kuč95] and the *Learning Parity with Noise (LPN)* problem [BKW03; Reg09]. This result is conceptually interesting because it demonstrates that the minimax structure underlying our analysis is not limited to prediction problems, but also captures the essence of hardness amplification in search settings. See Section 3.3 for details.

Triangle Counting over Random Graphs (Section 4). The proof of Theorem 1.3 proceeds through several reductions. At a high level, we reduce the problem of computing the exact triangle count to computing the triangle count modulo small primes, and then apply the arithmetic XOR lemma to amplify the average-case hardness. The argument builds on the Chinese Remainder Theorem and on a structural reduction that embeds multiple independent random graphs into a single larger instance.

We start with a circuit C_0 that computes $\mathrm{Tri}(G)$ on a random graph $G \sim \mathcal{G}(N, \theta)$ with success probability at least α. The reduction proceeds in five steps.

1. For any $k \leq N$ and $n \leq N/k$, we construct a circuit C_1 that computes $\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k)$ for $G_1, \ldots, G_k \sim \mathcal{G}(n, \theta)$ with success probability at least α, using C_0 as a subroutine (Lemma 4.2). In particular, for any prime p, taking the output of C_1 modulo p gives a circuit that computes $\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) \bmod p$. This is the only step that relies on the specific structure of the *triangle* counting problem. All subsequent arguments apply more generally to any function that satisfies such an additive property.

2. We then apply the arithmetic XOR lemma (Theorem 1.1) to obtain a circuit C_2 that computes $\mathrm{Tri}(G) \bmod p$ with success probability at least $1-\delta$, where $p \geq 2/\alpha$ is a suitably chosen prime and δ is a small error parameter.

3. Let $2/\alpha \leq p_1 < p_2 < \cdots < p_m$ be primes such that $\prod_{i=1}^{m} p_i \geq n^3$; it suffices to take $m = O(\log n)$. By a union bound, we can compute $\mathrm{Tri}(G) \bmod p_i$ correctly for all $i \in [m]$ with probability at least $1 - m \cdot \delta$.

4. By the Chinese Remainder Theorem, the true value $\mathrm{Tri}(G)$ can be recovered from $(\mathrm{Tri}(G) \bmod p_i)_{i=1}^{m}$ with the same success probability $1 - m \cdot \delta$.

5. Setting $\delta = (\log n)^{-c}$ for a sufficiently large constant c (e.g., $c = 10$) and invoking the worst-case-to-average-case reduction of Boix-Adserà, Brennan, and Bresler [BBB21] (Theorem 4.6), we obtain a circuit that computes $\mathrm{Tri}(G)$ for every n-vertex graph G with success probability at least $2/3$ and size $\tilde{O}((s + N^2)\log(1/\alpha)/\alpha^2)$.

The key technical step lies in reducing the computation of the sum $\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k)$ over k independent random graphs to the triangle counting of a single random graph. This is achieved by embedding the graphs $G_1, \ldots, G_k \sim \mathcal{G}(n, \theta)$ as disjoint induced subgraphs within an nk-vertex random graph, so that the total number of triangles in the larger graph encodes the desired sum.

2 Preliminaries

For a random variable Z, we write $z \sim Z$ to denote that z is drawn from the distribution of Z. The support of Z is denoted by $\mathrm{supp}(Z) = \{z \colon \Pr[Z = z] > 0\}$. Throughout this paper, we often consider a pair of random variables (X, Y) over $U \times V$ for finite sets U and V. Typically, we consider $U = \{0,1\}^n$ and X is uniformly distributed over U. We sometimes distinguish the distribution of a random variable Z and the random variable Z itself. We write $\mathcal{L}_Z$ to denote the distribution of Z, i.e., $\mathcal{L}_Z(z) = \Pr[Z = z]$.

We will use the following standard concentration inequalities.

Lemma 2.1 (Chernoff bound)**.** *Let $X_1, \ldots, X_n$ be independent $[0,1]$-valued random variables such that $\Pr[X_i = 1] = p_i$. Then, for any $\delta > 0$,*

$$\Pr\left[\sum_{i=1}^{n} X_i \le (1-\delta)\,\mathbb{E}[X]\right] \le \exp\left(-\frac{\delta^2}{2}\,\mathbb{E}[X]\right),$$
$$\Pr\left[\sum_{i=1}^{n} X_i \ge (1+\delta)\,\mathbb{E}[X]\right] \le \exp\left(-\frac{\delta^2}{2+2\delta/3}\,\mathbb{E}[X]\right).$$

Lemma 2.2 (Hoeffding bound)**.** *Let $X_1, \ldots, X_n$ be independent $[0,1]$-valued random variables such that $\Pr[X_i = 1] = p_i$. Then, for any $t > 0$,*

$$\Pr\left[\sum_{i=1}^{n} X_i \le \mathbb{E}[X] - t\right] \le \exp\left(-\frac{2t^2}{n}\right),$$
$$\Pr\left[\sum_{i=1}^{n} X_i \ge \mathbb{E}[X] + t\right] \le \exp\left(-\frac{2t^2}{n}\right).$$

2.1 Min-Entropy and Average Min-Entropy

We introduce the notion of min-entropy and average-min entropy [DORS08]. The *min-entropy* of a random variable Z, denoted by $H_\infty(Z)$, is defined by

$$H_\infty(Z) = \log_2\left(\frac{1}{\max_z \Pr[Z = z]}\right).$$

Let (X, Y) be a jointly distributed random variable over $U \times V$ for a finite sets U and V. The *average-min entropy of Y given X*, denoted by $\widetilde{H}_\infty(Y|X)$, is defined by

$$\widetilde{H}_\infty(Y|X) = \log_2\left(\frac{1}{\mathbb{E}_{x\sim X}[\max_{y\in V} \Pr[Y = y|X = x]]}\right).$$

Roughly speaking, $\widetilde{H}_\infty(Y|X)$ quantifies the maximum amount of information that can be extracted from X about Y, which is formally stated as follows:

Proposition 2.3 ([DORS08, Proposition 4.10]). *Let (X, Y) be a jointly distributed random variable. Then, $\widetilde{H}_\infty(Y|X) \geq \ell$ if and only if for every algorithm S, it holds that $\Pr[S(X) = Y] \leq 2^{-\ell}$.*

We shall need the following basic properties of min-entropy and average-min entropy.

Lemma 2.4 ([DORS08, Lemma 2.2]). *If (X, Y) satisfies $\widetilde{H}_\infty(Y|X) \geq \ell$, then for any $\alpha > 0$,*

$$\Pr_{x \sim X}\left[H_\infty(Y|X = x) \geq \ell - \log(1/\alpha)\right] \geq 1 - \alpha.$$

In the following, we present a well-known characterization of min-entropy in terms of convex combinations. We say that a random variable Z is *written as a convex combination of random variables* $W_1, \ldots, W_m$ if there exist non-negative weights $\lambda_1, \ldots, \lambda_m \geq 0$ such that the distribution of Z is given by $\mathcal{L}_Z = \sum_{i=1}^m \lambda_i \mathcal{L}_{W_i}$ and $\sum_{i=1}^m \lambda_i = 1$.

Lemma 2.5. *Let $\ell = \log_2 \frac{1}{1-\delta}$ for some $0 < \delta < 1/2$. A random variable Z has min-entropy ℓ if and only if its distribution Z is written as a convex combination of a finite number of random variables $W^{(1)}, \ldots, W^{(m)}$ such that, for every $i \in [m]$, the distribution of $W^{(i)}$ is obtained as follows: Choose two distinct elements x, y from $\mathrm{supp}(Z)$ and set*

$$\mathcal{L}_{W^{(i)}}(x) = \delta,$$
$$\mathcal{L}_{W^{(i)}}(y) = 1 - \delta.$$

This lemma follows from a simple geometric property of probability simplices: the set of distributions with min-entropy $\log_2(1/(1-\delta))$ coincides with the convex hull of all two-point distributions supported on x, y with probabilities $(\delta, 1-\delta)$. In particular, if the random variable Z has min-entropy $\ell \in \mathbb{N}$, it is widely known (e.g., [Vad12, Lemma 6.10]) that its distribution $\mathcal{L}_Z$ is written as a convex combination of a finite number of random variables $W^{(1)}, \ldots, W^{(m)}$ where the distribution of $W^{(i)}$ is the uniform distribution on a set of size 2^ℓ. However, for the case of $\ell < 1$, we are not aware of a direct reference stating this fact explicitly, so we provide a proof for completeness.

Proof of Lemma 2.5. Write $S = \mathrm{supp}(Z)$. Recall that

$$H_\infty(Z) = -\log_2\big(\max_{z \in S} \mathcal{L}_Z(z)\big).$$

Throughout the proof set $\ell = \log_2 \frac{1}{1-\delta}$, so that $2^{-\ell} = 1 - \delta$. For $x \neq y \in S$, let $\mathcal{L}_{x,y}$ denote the two-point distribution that is supported on $\{x, y\}$ with probabilities $\mathcal{L}_{x,y}(x) = \delta$ and $\mathcal{L}_{x,y}(y) = 1-\delta$.

($\Rightarrow$) We construct finitely many two–point distributions of the form $\mathcal{L}_{x,y}$ whose convex combination equals $\mathcal{L}_Z$.

Let $\mu_0 = \mathcal{L}_Z$ and set $t = 0$. While $|\mathrm{supp}(\mu_t)| > 2$, choose two distinct elements

$$i = \arg\max_{z \in S} \mu_t(z), \qquad j = \arg\max_{z \in S \setminus \{i\}} \mu_t(z),$$

and set

$$\alpha_t \;=\; \min\left\{\frac{\mu_t(i)}{1-\delta}, \frac{\mu_t(j)}{\delta}\right\}.$$

Define the two–point distribution $\nu_t = \mathcal{L}_{i,j}$ and update the residual subprobability measure by

$$\mu_{t+1} \;=\; \mu_t - \alpha_t \nu_t.$$

By construction, μ_{t+1} is nonnegative (since α_t is chosen to be the maximum α such that $\mu_t - \alpha\mathcal{L}_{i,j}$ is nonnegative), its total mass decreases by exactly α_t, and every point mass remains at most $1-\delta$. Moreover, at least one of $\mu_{t+1}(i)$ or $\mu_{t+1}(j)$ becomes zero, so the support size strictly decreases at each step. Since initially $|\mathrm{supp}(\mu_0)| \le |S|$, the process terminates after at most $|S|-3$ iterations, yielding a subprobability measure μ_T supported on at most two points.

If $\mathrm{supp}(\mu_T) = \{x, y\}$ and $\mu_T(x) = a$, $\mu_T(y) = 1-a$ with $a \in [\delta, 1-\delta]$, then

$$\mu_T = \lambda\,\mathcal{L}_{x,y} + (1-\lambda)\,\mathcal{L}_{y,x}, \qquad \lambda = \frac{1-a-\delta}{1-2\delta} \in [0,1].$$

Collecting all extracted pairs (α_t, ν_t) together with this final step, and noting that the total mass decreases by α_t at each iteration, we have $\sum_t \alpha_t = 1$ when $\mu_T = 0$. Therefore, $\mathcal{L}_Z = \sum_t \alpha_t \nu_t$, which expresses $\mathcal{L}_Z$ as a convex combination of finitely many two-point distributions $\mathcal{L}_{x,y}$.

($\Leftarrow$) Conversely, suppose $\mathcal{L}_Z = \sum_{i=1}^m \lambda_i\,\mathcal{L}_{x_i,y_i}$ where each $\mathcal{L}_{x_i,y_i}$ is supported on $\{x_i, y_i\}$ with probabilities $(\delta,\, 1-\delta)$ and $\lambda_i \ge 0$, $\sum_i \lambda_i = 1$. For any $z \in S$,

$$\mathcal{L}_Z(z) = \sum_{i=1}^m \lambda_i\,\mathcal{L}_{x_i,y_i}(z) \le \sum_{i=1}^m \lambda_i\,(1-\delta) = 1-\delta,$$

hence $\max_z \Pr[Z=z] \le 1-\delta$ and therefore $H_\infty(Z) \ge \ell$. □

The proof of the direction ($\Leftarrow$) above immediately implies the following lemma.

Lemma 2.6. *Let Z be a random variable that is written as a convex combination of random variables $W_1, \ldots, W_m$. Then, $H_\infty(Z) \ge \min_{i\in[m]} H_\infty(W_i)$.*

In the following, we show if X and Y are written as convex combinations of random variables $W^{(1)}, \ldots, W^{(m)}$, then the sum $X+Y$ is written as a convex combination of random variables $W^{(i)} + W^{(j)}$ over $i, j \in [m]$.

Lemma 2.7. *Let G be an Abelian group. Let X and Y be random variables over G that are written as convex combinations of G-valued random variables $W^{(1)}, \ldots, W^{(m)}$. Then, the sum $X+Y$ is written as a convex combination of random variables $W^{(i)} + W^{(j)}$ over $i, j \in [m]$.*

Proof. Write

$$\mathcal{L}_X = \sum_{i=1}^m \lambda_i \mathcal{L}_{W^{(i)}},$$
$$\mathcal{L}_Y = \sum_{j=1}^m \mu_j \mathcal{L}_{W^{(j)}},$$

where $\lambda_i, \mu_j \ge 0$ and $\sum_{i=1}^m \lambda_i = 1$ and $\sum_{j=1}^m \mu_j = 1$. Then, for any $a \in G$,

$$\begin{aligned}
\mathcal{L}_{X+Y}(a) &= \sum_{b\in G} \mathcal{L}_X(b)\mathcal{L}_Y(a-b) \\
&= \sum_{i,j} \lambda_i\mu_j \sum_b \mathcal{L}_{W^{(i)}}(b)\mathcal{L}_{W^{(j)}}(a-b) \\
&= \sum_{i,j} \lambda_i\mu_j \mathcal{L}_{W^{(i)}+W^{(j)}}(a).
\end{aligned}$$

Therefore, the sum $X+Y$ is written as a convex combination of random variables $W^{(i)}+W^{(j)}$ over $i,j\in[m]$. □

2.2 Prediction Hardness and Indistinguishability

We introduce the notion of prediction hardness. Roughly speaking, a jointly distributed random variable (X,Y) is weakly hard to predict if any efficient algorithm fails to predict Y given X with high probability, which is formalized as follows:

Definition 2.8. *For a pair of random variables (X,Y), we say that Y is* δ-hard to predict given X for size s *if any size-s circuit C satisfies* $\Pr[C(X)\neq Y]\geq\delta$.

For example, if $Y=f(X)$ for some function $f\colon U\to V$, the prediction hardness of Y given X coincides with the average-case hardness of f in Theorem 1.1.

Remark 2.9. *The formulation in terms of a general pair (X,Y) rather than $(X,f(X))$ is essential for capturing important* search-type *problems where the output is not a deterministic function of the input. For instance, the planted clique and Learning Parity with Noise (LPN) problems (Examples 3.9 and 3.10) can naturally be represented as such pairs, where X encodes a random instance and Y corresponds to the hidden structure to be recovered. This viewpoint allows our framework to encompass both function evaluation problems and efficiently verifiable search problems within a unified hardness-amplification setting.*

Next, we introduce the notion of indistinguishability. Two random variables A and B are computationally indistinguishable if any efficient algorithm cannot distinguish them with high probability. This is formalized as follows:

Definition 2.10. *Two random variables A and B are* γ-indistinguishable for size s *if any size-s circuit C cannot distinguish them with advantage γ, i.e.,*

$$|\mathbb{E}[C(A)]-\mathbb{E}[C(B)]|\leq\gamma.$$

Zheng [Zhe14] and Vadhan and Zheng [VZ12] characterized pseudo-average-min entropy in terms of unpredictability. Roughly speaking, a jointly distributed random variable (X,Y) has pseudo-average-min entropy ℓ if there exists a random variable Z jointly distributed with X such that $\widetilde{H}_\infty(Z|X)\geq\ell$ and the distributions (X,Y) and (X,Z) are computationally indistinguishable. This is formally stated as follows:

Theorem 2.11 (unpredictabiltiy ⇒ pseudo-average-min entropy, nonuniform)**.** *Let (X,Y) be a jointly distributed random variable over $U\times V$ for finite sets U and V. Let $\delta,\gamma>0$ be any parameters. Suppose that Y is 2δ-hard to predict given X for size s. Then, there exists a random variable Z over V jointly distributed with X that has average-min entropy $\widetilde{H}_\infty(Z|X)\geq\log_2\frac{1}{1-\delta}$ such that (X,Y) and (X,Z) are γ-indistinguishable for size $s'=\Omega\Big(s\cdot\frac{\gamma^2}{|V|\log(|V|/\gamma)}\Big)$.*

Zheng [Zhe14] originally proved Theorem 2.11 in the *uniform* setting, using a technically involved argument with several auxiliary parameters. In this work, we present a significantly simplified and modular proof of the non-uniform version. Our proof removes many of the technical artifacts in Zheng's argument and makes the dependence of the circuit blow-up on the parameters $(s,\gamma,|V|)$ explicit. This quantitative clarity is important for our later applications, such as

the triangle-counting reduction in Theorem 1.3, where the overall circuit size must grow by at most $n^{o(1)}$. The resulting statement cleanly separates the information-theoretic and computational components of the argument, which will play a key role in our hardness amplification framework.

The opposite direction of Theorem 2.11 is rather straightforward, and the proof can be found in Section A. Although we state it in the non-uniform setting, the proof works in the uniform setting as well.

Proposition 2.12 (pseudo-average-min entropy $\Rightarrow$ unpredictability)**.** *Let (X, Y) be a jointly distributed random variable over $U \times V$ for finite sets U and V. Let $0 < \alpha \leq \beta$ be any parameters. Suppose that there exists a V-valued random variable Z that is jointly distributed with X that has average-min entropy $\widetilde{H}_\infty(Z|X) \geq \log_2(1/\alpha)$ and such that (X, Y) and (X, Z) are β-indistinguishable for size s. Then, for a sufficiently large constant $c > 0$, Y is $(1 - \alpha - \beta)$-hard to predict given X for size $s - c \log |V|$.*

3 Hardness Amplification from Pseudo-Average-Min Entropy

3.1 Direct Product Theorem

To illustrate the usefulness of the pseudo average-min entropy framework, we show how it naturally yields a clean proof of the classical direct product theorem [IJKW10]. This result is not new—the theorem itself is well known—but our goal here is to demonstrate that the pseudo average-min entropy provides a simple and conceptually transparent route to its proof. To the best of our knowledge, this perspective has not been made explicit before.

Intuitively, suppose that the best possible way to predict Y from X succeeds with probability at most $1 - \delta$. Then, for k independent copies $(X_1, Y_1), \ldots, (X_k, Y_k)$, the best possible strategy to predict $(Y_1, \ldots, Y_k)$ from $(X_1, \ldots, X_k)$ succeeds with probability at most $(1 - \delta)^k$. Information-theoretically, this phenomenon can be formalized as follows:

Lemma 3.1. *Let $(X_1, Y_1), \ldots, (X_k, Y_k)$ be independent pairs of random variables. Then, it holds that*

$$\widetilde{H}_\infty(Y_1, \ldots, Y_k | X_1, \ldots, X_k) = \sum_{i=1}^{k} \widetilde{H}_\infty(Y_i | X_i).$$

Proof. The proof is straightforward. For any fixed $\vec{x} = (x_1, \ldots, x_k) \in U^k$, we have

$$\begin{aligned} \max_{\vec{y} \in V^k} \Pr\left[\vec{Y} = \vec{y} \,\middle|\, \vec{X} = \vec{x}\right] &= \max_{\vec{y} \in V^k} \prod_{i=1}^{k} \Pr\left[Y_i = y_i \,|\, X_i = x_i\right] \\ &= \prod_{i=1}^{k} \max_{y_i \in V} \Pr\left[Y_i = y_i \,|\, X_i = x_i\right] \end{aligned}$$

Taking the expectation over $\vec{x}$ and then taking the logarithm, we obtain the claim. □

The direct product theorem is a computational-complexity analogue of Lemma 3.1 in terms of unpredictability.

Theorem 3.2. *Let $\delta, \varepsilon > 0$ be any parameters. Let (X, Y) be a jointly distributed random variable over $U \times V$ for a finite sets U and V that is δ-hard to predict given X for size s. Let $(X_1, Y_1), \dots, (X_k, Y_k)$ be independent copies of (X, Y). Then, if $k \geq c \cdot \log(1/\varepsilon)/\delta$ for a sufficiently large constant $c > 0$, it holds that $(X_1, \dots, X_k, Y_1, \dots, Y_k)$ is ε-hard to predict given $(X_1, \dots, X_k)$ for size $s' = \Omega\left(s \cdot \frac{\varepsilon^2}{k^2 |V| \log(k|V|/\varepsilon)}\right)$.*

To prove Theorem 3.2, we need a well-known fact that taking the direct product preserves indistinguishability (see, e.g., [HVV06, Lemma 3.2]), which can be shown by a standard hybrid argument.

Lemma 3.3. *Let A, B be random variables that are γ-indistinguishable for size s. Let $A_1, \dots, A_k$ and $B_1, \dots, B_k$ be independent copies of A and B, respectively. Then, $(A_1, \dots, A_k)$ and $(B_1, \dots, B_k)$ are $k\gamma$-indistinguishable for size s.*

Proof. Write $\vec{A} = (A_1, \dots, A_k)$ and $\vec{B} = (B_1, \dots, B_k)$ for simplicity. Suppose that there exists a size-s' circuit D that can distinguish $\vec{A}$ and $\vec{B}$ with advantage strictly greater than $k\gamma$, i.e., $\mathbb{E}[D(\vec{B})] - \mathbb{E}[D(\vec{A})] > k\gamma$. For $i = 0, \dots, k$, let $\vec{H}_i = (A_1, \dots, A_i, B_{i+1}, \dots, B_k)$ be the hybrid random variable. Since $\vec{H}_0 = \vec{B}$ and $\vec{H}_k = \vec{A}$, the distinguishability of D implies

$$\sum_{i=0}^{k} \mathbb{E}[D(\vec{H}_i)] - \mathbb{E}[D(\vec{H}_{i+1})] = \mathbb{E}[D(\vec{B})] - \mathbb{E}[D(\vec{A})] > k\gamma.$$

Therefore, there exists $i \in [k]$ such that

$$\begin{aligned}
\gamma &< \mathbb{E}[D(\vec{H}_i)] - \mathbb{E}[D(\vec{H}_{i+1})] \\
&= \mathop{\mathbb{E}}_{\substack{a_1,\dots,a_{i-1} \sim A \\ b_{i+1},\dots,b_k \sim B}} \left[\mathop{\mathbb{E}}_{a \sim A}[D(a_1, \dots, a_{i-1}, a, b_{i+1}, \dots, b_k)] - \mathop{\mathbb{E}}_{b \sim B}[D(a_1, \dots, a_{i-1}, b, b_{i+1}, \dots, b_k)] \right]
\end{aligned}$$

Therefore, by fixing i, $a_1, \dots, a_{i-1}$, and $b_{i+1}, \dots, b_k$, we obtain a circuit of size s that can distinguish A and B with advantage strictly greater than γ, contradicting the indistinguishability of A and B. □

Now we are ready to prove Theorem 3.2.

Proof of Theorem 3.2. Let $\gamma > 0$ be a parameter to be specified later. From Theorem 2.11 for this γ, we know that there exists a random variable Z that is jointly distributed with X and has average-min entropy $\widetilde{H}_\infty(Z|X) \geq \log_2 \frac{1}{1-\delta/2}$ such that (X, Y) and (X, Z) are γ-indistinguishable for size $s' = \Omega\left(s \cdot \frac{\gamma^2}{|V| \log(|V|/\gamma)}\right)$.

Let $(X_1, Y_1), \dots, (X_k, Y_k)$ be independent copies of (X, Y) and $(X_1, Z_1), \dots, (X_k, Z_k)$ be independent copies of (X, Z). By Lemma 3.3, $(X_1, \dots, X_k, Y_1, \dots, Y_k)$ and $(X_1, \dots, X_k, Z_1, \dots, Z_k)$ are $k\gamma$-indistinguishable for size s'. Moreover, by Lemma 3.1, $\widetilde{H}_\infty(Z_1, \dots, Z_k | X_1, \dots, X_k) = k \log_2 \frac{1}{1-\delta/2}$. Therefore, by Proposition 2.12 for $\alpha = (1 - \delta/2)^k$ and $\beta = k\gamma$, $(Y_1, \dots, Y_k)$ is $(1 - (1-\delta/2)^k - k\gamma)$-hard to predict given $X_1, \dots, X_k$ for size s'. Thus, for some $k = O(\log(1/\varepsilon)/\delta)$, setting $\gamma = \frac{\varepsilon}{2k}$, we obtain that $(Y_1, \dots, Y_k)$ is $(1 - \varepsilon)$-hard to predict given $X_1, \dots, X_k$ for size $s' = \Omega\left(s \cdot \frac{\varepsilon^2}{k^2 |V| \log(k|V|/\varepsilon)}\right)$.

□

3.2 Arithmetic XOR Lemma

In this subsection, we present an arithmetic XOR lemma for any prime p.

Theorem 3.4. *Let $\delta, \varepsilon > 0$ be any parameters and p be a prime. Let (X, Y) be a jointly distributed random variable over $U \times \mathbb{F}_p$ such that Y is δ-hard to predict given X for size s. For $k \in \mathbb{N}$, let $(X_1, Y_1), \dots, (X_k, Y_k)$ be independent copies of (X, Y) and write $\vec{X} = (X_1, \dots, X_k)$. Then, for some $k = O\Big(\frac{p^2}{\delta} \log(p/\varepsilon)\Big)$, it holds that $Y_1 + \cdots + Y_k$ is $(1 - (1+\varepsilon)/p)$-hard to predict given $\vec{X}$ for size $s' = \Omega\Big(s \cdot \frac{\varepsilon^2}{p^3 k^2 \log(kp/\varepsilon)}\Big) - O(k \log p)$.*

Remark 3.5. *In Theorem 3.4, we can set $\varepsilon > 1$. For example, for $2/p \le \alpha < 1$, if we set $\varepsilon = \alpha p - 1 \ge \alpha p/2$, then $Y_1 + \cdots + Y_k$ is $(1-\alpha)$-hard to predict given $\vec{X}$ for size $s' = \Omega\Big(s \cdot \frac{\alpha^2}{pk^2 \log(k/\alpha)}\Big) - O(k \log p)$ and $k = O\Big(\frac{p^2}{\delta} \log(1/\alpha)\Big)$.*

The key to the proof of Theorem 3.4 is to show that if $Z_1, \dots, Z_k$ are independent $\mathbb{F}_p$-valued random variables that has a non-negligible min-entropy, then the distribution of the sum $Z_1 + \cdots + Z_k$ is close to the uniform distribution over $\mathbb{F}_p$, meaning that its min-entropy is close to $\log p$ for a sufficiently large k. This is shown by a Fourier analysis argument.

Lemma 3.6. *Let $\gamma, \delta > 0$ be any parameters and p be a prime. Let $Z_1, \dots, Z_\ell$ be independent $\mathbb{F}_p$-valued random variables such that every Z_i satisfies $H_\infty(Z_i) \ge \log_2 \frac{1}{1-\delta}$ for $0 < \delta < 1/2$. Then, it holds that*

$$H_\infty(Z_1 + \cdots + Z_\ell) \ge \log_2 \frac{p}{1+\gamma}$$

for some $\ell = O\Big(\frac{p^2}{\delta} \log(p/\gamma)\Big)$.

Fourier Analysis for Functions on $\mathbb{F}_p$. Before proving Lemma 3.6, we introduce notation for the Fourier analysis for functions on $\mathbb{F}_p$. For the set of all functions $\{f \colon \mathbb{F}_p \to \mathbb{C}\}$, we associate it with the inner product defined by $\langle f, g \rangle = \sum_{x \in \mathbb{F}_p} f(x) \cdot \overline{g(x)}$, which induces the ℓ^2 norm $\|f\|_2 = \sqrt{\langle f, f \rangle}$. For $a \in \mathbb{F}_p$, we define the character function $\chi_a \colon \mathbb{F}_p \to \mathbb{C}$ by $\chi_a(x) = \omega^{x \cdot a}$, where $\omega = \exp(2\pi i/p)$ is a primitive p-th root of unity. The Fourier transform of f is defined by $\widehat{f}(a) = \langle f, \chi_a \rangle = \sum_{x \in \mathbb{F}_p} f(x) \cdot \overline{\chi_a(x)}$. Then, we can rewrite f as $f = \sum_{a \in \mathbb{F}_p} \widehat{f}(a) \cdot \chi_a$. Note that $f \mapsto \widehat{f}$ is a linear operator. Moreover, it is known that the operator $f \mapsto \widehat{f}$ is unitary, i.e., $\|f\|_2 = \big\|\widehat{f}\big\|_2$ (Parseval's identity).

The convolution of two functions f and g is defined by $(f * g)(x) = \sum_{a \in \mathbb{F}_p} f(a) \cdot g(x - a)$. It is known that the convolution is equivalent to the product of the Fourier transforms, i.e., $\widehat{f * g}(x) = \widehat{f}(x) \cdot \widehat{g}(x)$ for every $x \in \mathbb{F}_p$.

Proof of Lemma 3.6. From Lemmas 2.5 and 2.7, each Z_i can be expressed as a convex combination of random variables $W_i^{(1)}, \dots, W_i^{(m)}$ such that, for every $j \in [m]$, $W_i^{(j)}$ is supported on a two-point set and assigns probability δ to one element and $1 - \delta$ to the other. By Lemma 2.7, the sum $Z_1 + \cdots + Z_\ell$ can therefore be written as a convex combination of random variables of the form

$W_1^{(i_1)} + \cdots + W_\ell^{(i_\ell)}$ for $(i_1, \ldots, i_\ell) \in [m]^\ell$. For every such combination, each $W_t^{(i_t)}$ is two-point supported with probabilities δ and $1-\delta$.

Note that each $W_t^{(i_t)}$ has min-entropy $\log_2 \frac{1}{1-\delta}$. Hence, by Lemma 2.6, it suffices to show that

$$H_\infty\left(W_1^{(i_1)} + \cdots + W_\ell^{(i_\ell)}\right) \geq \log_2 \frac{p}{1+\gamma}$$

for any $(i_1, \ldots, i_\ell) \in [m]^\ell$. Fix such a collection $W_1^{(i_1)}, \ldots, W_\ell^{(i_\ell)}$ and let $W = W_1^{(i_1)} + \cdots + W_\ell^{(i_\ell)}$. Our goal is to show that $\max_{w \in \mathbb{F}_p} \Pr[W = w] \leq \frac{1+\gamma}{p}$.

By translation, we may assume without loss of generality that for every $j \in [\ell]$, each $W_j^{(i_j)}$ satisfies $\Pr\left[W_j^{(i_j)} = 0\right] = 1 - \delta$ and $\Pr\left[W_j^{(i_j)} = w_j\right] = \delta$ for some nonzero $w_j \in \mathbb{F}_p$.

Recall that $\mathcal{L}_X$ denotes the probability mass function of a random variable X. For each $j \in [\ell]$, the Fourier transform of $\mathcal{L}_{W_j^{(i_j)}}$ satisfies

$$\begin{aligned}
\left|\widehat{\mathcal{L}_{W_j^{(i_j)}}}(a)\right|^2 &= \left|\sum_{x \in \mathbb{F}_p} \mathcal{L}_{W_j^{(i_j)}}(x)\, \overline{\chi_a(x)}\right|^2 \\
&= \left|(1-\delta) + \delta\, \overline{\chi_a(w_j)}\right|^2 \\
&= 1 - 2\delta(1-\delta)\left(1 - \cos\frac{2\pi a w_j}{p}\right) \\
&= \begin{cases} 1 & \text{if } a = 0, \\ \leq 1 - \Omega\left(\dfrac{\delta}{p^2}\right) & \text{otherwise.} \end{cases}
\end{aligned}$$

Since $\mathcal{L}_W = \mathcal{L}_{W_1^{(i_1)}} * \cdots * \mathcal{L}_{W_\ell^{(i_\ell)}}$, we have

$$\begin{aligned}
\left|\widehat{\mathcal{L}_W}(a)\right|^2 &= \prod_{j=1}^{\ell} \left|\widehat{\mathcal{L}_{W_j^{(i_j)}}}(a)\right|^2 \\
&= \begin{cases} 1 & \text{if } a = 0, \\ \leq \exp\left(-\Omega\left(\dfrac{\ell\delta}{p^2}\right)\right) & \text{otherwise.} \end{cases}
\end{aligned}$$

Let U denote the uniform distribution over $\mathbb{F}_p$. Then $\widehat{\mathcal{L}_U}(a) = 1$ if $a = 0$ and $\widehat{\mathcal{L}_U}(a) = 0$ otherwise. By Parseval's identity,

$$\begin{aligned}
\max_{w \in \mathbb{F}_p} \Pr[W = w] - \frac{1}{p} &\leq \|\mathcal{L}_W - \mathcal{L}_U\|_2 \\
&= \|\widehat{\mathcal{L}_W} - \widehat{\mathcal{L}_U}\|_2 \\
&= \sqrt{\sum_{a \neq 0} \left|\widehat{\mathcal{L}_W}(a)\right|^2} && \text{(since } \widehat{\mathcal{L}_W}(0) = \widehat{\mathcal{L}_U}(0) = 1\text{)} \\
&\leq p^{1/2} \cdot \exp\left(-\Omega\left(\frac{\ell\delta}{p^2}\right)\right).
\end{aligned}$$

Choosing $\ell = O\left(\frac{p^2}{\delta}\log(p/\gamma)\right)$ ensures that

$$\max_{w\in\mathbb{F}_p}\Pr[W=w]-\frac{1}{p}\le\frac{\gamma}{p},$$

which completes the proof.

□

Utilizing Lemma 2.4, we can extend Lemma 3.6 to the case of average-min entropy as follows:

Lemma 3.7. *Let $\gamma > 0$, $0 < \delta < 1$ be any parameters and p be a prime. There exists some $k = O\left(\frac{p^2}{\delta^2}\log(p/\gamma)\right)$ such that if $(X_1, Z_1), \dots, (X_k, Z_k)$ are independent random variables over $\{0,1\}^n \times \mathbb{F}_p$ such that $\widetilde{H}_\infty(Z_i|X_i) \ge \log_2 \frac{1}{1-\delta}$ for all $i \in [k]$, then it holds that*

$$\widetilde{H}_\infty(Z_1+\cdots+Z_k|X_1,\dots,X_k)\ge\log_2\frac{|V|}{1+\gamma}.$$

Proof. For each $i \in [k]$, let $S_i \subseteq \mathrm{supp}(X_i)$ be the set of x such that $H_\infty(Z_i|_{X_i=x}) \ge \log\frac{1}{1-\delta/2}$. By Lemma 2.4 for $\alpha = 1-\frac{\delta}{2-\delta}$, we have $\Pr[X_i \in S_i] \ge \frac{\delta}{2-\delta} \ge \frac{\delta}{2}$. Let K be the random variable denoting the number of $i \in [k]$ such that $X_i \in S_i$. Then, $\mathbb{E}_{X_1,\dots,X_k}[K] \ge \frac{\delta k}{2}$. Since $X_1, \dots, X_k$ are independent, by the Chernoff bound, we have

$$\Pr\left[K\le\frac{\delta k}{4}\right]\le\exp\left(-\frac{\delta k}{16}\right)\le\frac{\gamma/2}{p}$$

if $k \ge \frac{16\log(2p/\gamma)}{\delta}$.

Write $\vec{X} = (X_1, \dots, X_k)$ and denote a realization of $\vec{X}$ by $\vec{x}$. For any $\vec{x}$ such that $K \ge \frac{\delta k}{4}$, from Lemma 3.6 for $\ell = \delta k/4$, we have $\max_z \Pr\left[Z=z \,\middle|\, \vec{X}=\vec{x}\right] \le \frac{1+\gamma/2}{p}$ if $k \ge c\cdot\frac{p^2}{\delta^2}\log(p/\gamma)$ for some sufficiently large constant $c > 0$. Taking the expectation over $\vec{x}$, we have

$$\mathop{\mathbb{E}}_{\vec{x}\sim\vec{X}}\left[\max_z\Pr\left[Z=z\,\middle|\,\vec{X}=\vec{x}\right]\right]\le\frac{1+\gamma/2}{p}+\frac{\gamma/2}{p}=\frac{1+\gamma}{p}.$$

This completes the proof. □

Proof of Theorem 3.4. Apply Theorem 2.11 for $V = \mathbb{F}_p$ and $\gamma = \frac{\varepsilon}{2kp}$. Then, there exists an $\mathbb{F}_p$-valued random variable Z that is jointly distributed with X and such that $\widetilde{H}_\infty(Z|X) \ge \log_2\frac{1}{1-\delta/2}$ and (X, Z) is γ-indistinguishable for size $s_1 = \Omega\left(s\cdot\frac{\gamma^2}{p\log(p/\gamma)}\right)$.

Let $(X_1, Y_1), \dots, (X_k, Y_k)$ and $(X_1, Z_1), \dots, (X_k, Z_k)$ be independent copies of (X, Y) and (X, Z), respectively. Write $\vec{X} = (X_1, \dots, X_k)$, $\vec{Y} = (Y_1, \dots, Y_k)$, and $\vec{Z} = (Z_1, \dots, Z_k)$. Then, from Lemma 3.3, $(\vec{X}, \vec{Y})$ and $(\vec{X}, \vec{Z})$ are $(\varepsilon/2p)$-indistinguishable for size s_1. In particular, $(\vec{X}, \sum_i Y_i)$ and $(\vec{X}, \sum_i Z_i)$ are $(\varepsilon/2p)$-indistinguishable for size $s_1 - O(k\log p)$. To see this, if there exists a circuit $D(\vec{x}, a)$ that can distinguish $(\vec{X}, \sum_i Y_i)$ and $(\vec{X}, \sum_i Z_i)$ with advantage $\varepsilon/(2p)$, we can distinguish $(\vec{X}, \vec{Y})$ and $(\vec{X}, \vec{Z})$ with the same advantage by computing $D(\vec{x}, \sum_i a_i)$ for given $\vec{x}$ and $\vec{a}$. Moreover, by Lemma 3.7, $\widetilde{H}_\infty(Z_1+\cdots+Z_k|X_1,\dots,X_k) \ge \log_2\frac{|V|}{1+\varepsilon/2}$. Therefore, by Proposition 2.12 for $\alpha = \frac{1+\varepsilon/2}{p}$ and $\beta = \frac{\varepsilon}{2p}$, $Y_1+\cdots+Y_k$ is $(1-\frac{1+\varepsilon/2}{p}-\frac{\varepsilon}{2p})$-hard to predict given $\vec{X}$ for size $s_1 - O(k\log p) = s'$. □

3.3 Hardcore Lemma for Efficiently Verifiable Problems

In this section, we show that the proof technique underlying the pseudo average-min entropy characterization (Theorem 2.11)—based on a minimax theorem argument[Imp95; Zhe14; VZ12]—can be extended to obtain a *hardcore lemma* for efficiently verifiable search problems with unique solutions. Rather than applying the pseudo average-min entropy framework as a black box, we modify its proof (see Section A) to handle search-type unpredictability, where the goal is to recover a hidden witness Y from a given input X. Intuitively, such a pair (X, Y) represents a search problem whose correctness can be efficiently verified by a polynomial-time algorithm.

This class of problems encompasses several fundamental search problems studied in average-case complexity and cryptography, including the *planted clique problem* [Jer92; Kuč95] and the *Learning Parity with Noise (LPN)* problem [BKW03; Reg09]. Both are efficiently verifiable in the sense that, given a candidate solution, one can efficiently check its correctness, yet no efficient algorithm is known to find such a solution under standard hardness assumptions. These examples illustrate that the search-version hardcore lemma captures a broad class of computationally natural and cryptographically relevant problems.

Formally, we consider pairs of random variables (X, Y) equipped with a verifier circuit $F(x, y)$ that accepts (x, Y) and rejects all incorrect candidates $y' \neq Y$ for almost all samples (X, Y) drawn from the underlying distribution. We refer to such pairs as *efficiently verifiable problems with unique solutions*.

Definition 3.8. *A pair of random variables (X, Y) is an (s, γ)-*verifiable problem with unique solution *if there exists a size-s circuit $F\colon U \times V \to \{0,1\}$ such that, for every $(x, y) \in U \times V$,*

$$\Pr_{(X,Y)}\left[F(X,Y) = 1 \text{ and } \forall y \neq Y,\ F(X,y) = 0\right] \geq 1 - \gamma.$$

Example 3.9 (Planted Clique Problem)**.** *The* planted clique problem *[Jer92; Kuč95] can be represented as $(G \cup K_C, C)$, where $G \sim \mathcal{G}(n, 1/2)$ is an Erdős–Rényi random graph and K_C is a k-clique planted on a random vertex set $C \sim \binom{[n]}{k}$. The task is to recover the planted clique C from the combined graph $G \cup K_C$. This problem is* efficiently verifiable with a unique solution*: given a graph G' and a candidate vertex set $C' \subseteq [n]$, one can efficiently verify whether C' induces a k-clique in G'; moreover, if $k \geq 2.01 \log_2 n$, then with high probability over the randomness of $G \cup K_C$, the planted clique C is unique.*

Example 3.10 (Learning Parity with Noise (LPN))**.** *The* Learning Parity with Noise (LPN) *problem [BKW03; Reg09] can be formalized as $((A, As + e), s)$, where $A \sim \mathbb{F}_2^{m \times n}$ is a random matrix, $s \sim \mathbb{F}_2^n$ is a random secret vector, and $e \sim \mathrm{Ber}(p)^{\otimes m}$ is a random noise vector with independent Bernoulli noise of rate $p < 1/2$. The goal is to recover s from the noisy linear system $(A, As + e)$. This problem is also* efficiently verifiable with a unique solution*: given a candidate $s' \in \mathbb{F}_2^n$, one can efficiently check whether $\langle A_i, s' \rangle = (As + e)_i$ holds for a $(1 - p - o(1))$-fraction of rows A_i of A; under the standard noise model, the true secret s is the unique vector satisfying this condition with high probability.*

Now, we state our hardcore lemma. We say that a random variable Z' is *δ-dense in Z* if $\Pr[Z = z] \geq \delta \cdot \Pr[Z' = z]$ for every $z \in \mathrm{supp}(Z)$.

Theorem 3.11. *Let $\gamma, \delta, \varepsilon > 0$ be any parameters. Let (X, Y) be a pair of random variables over $U \times V$ for finite sets U and V that is (s, γ)-uniquely verifiable. If Y is $(1.01\delta + \gamma)$-hard to predict*

given X for size s, then there exists a pair of random variables (X', Y') that is δ-dense in (X, Y) and such that Y' is $(1-\varepsilon)$-hard to predict given X' for size $s' = s \cdot \Omega\left(\frac{\varepsilon}{\log(1/\delta)}\right)$.

Proof. We prove the contrapositive statement. Suppose that for any pair of random variables (X', Y') that is δ-dense in (X, Y), there exists a size-s' circuit C' that can predict Y' given X' with probability at least ε. Consider the following two-player game: Player 1 picks (X', Y') that is δ-dense in (X, Y) and Player 2 picks a size-s' circuit C'. The payoff is given by $\Pr[C'(X') = Y']$. By the assumption, for any strategy of Player 1, there exists a strategy of Player 2 that can achieve a payoff of at least ε. Therefore, by von Neumann's minimax theorem (Theorem A.1), there exists a distribution $\mathcal{C}$ over size-s' circuits such that the expected payoff is at least ε for any strategy of Player 1, i.e.,

$$\forall (X', Y') \text{ such that } X' \text{ is } \delta\text{-dense in } X, \quad \mathop{\mathbb{E}}_{\substack{C' \sim \mathcal{C} \\ (X', Y')}} [C'(X') = Y'] \geq \varepsilon. \tag{2}$$

Fix such a distribution $\mathcal{C}$. Let $p(x, y) = \Pr_{C' \sim \mathcal{C}}[C'(x) = y]$ and $S = \{(x, y) \in U \times V : p(x, y) \geq \varepsilon\}$. Then, we have $\Pr[(X, Y) \in S] \geq 1 - \delta$ since otherwise, the pair $(X', Y') = (X, Y)|_{(X,Y) \notin S}$ is δ-dense in (X, Y) and we can apply Eq. (2) but this contradicts to the definition of S.

Define a randomized circuit C^* as follows: Sample t circuits $C'_1, \ldots, C'_t \sim \mathcal{C}$ and set $y_i = C'_i(x)$, where t is a parameter to be specified later. Run the verification algorithm F on every (x, y_i) and output the first y_i such that $F(x, y_i) = 1$. If no such y_i exists, output an arbitrary $y \in V$. Since $\Pr_{C' \sim \mathcal{C}}[C'(x) = y] \geq \varepsilon$ for every $(x, y) \in S$, if we choose $C'_1, \ldots, C'_t \sim \mathcal{C}$ independently, then with probability at least $1 - (1 - \varepsilon)^t$ over $C'_1, \ldots, C'_t$, it holds that $y \in \{C'_1(x), \ldots, C'_t(x)\}$.

Moreover, since (X, Y) is (s, γ)-uniquely verifiable, for some size-s circuit F, for a $(1-\gamma)$-fraction of $(x, y) \sim (X, Y)$, it holds that $F(x, y) = 1$ and $F(x, y') = 0$ for every $y' \neq y$. Let $T \subseteq U \times V$ be the set of such (x, y). Then, for every $(x, y) \in S \cap T$, we can detect which $y_i = C'_i(x)$ is equal to y by running F on every (x, y_i). That is, $\Pr_{C^*}[C^*(x) = y] \geq 1 - (1 - \varepsilon)^t$ for every $(x, y) \in S \cap T$. Therefore, for some $t = O(\log(1/\delta)/\varepsilon)$, we obtain

$$\Pr_{\substack{(x,y) \sim (X,Y) \\ C^*}} [C^*(x) = y] \geq (1 - \delta - \gamma) \cdot (1 - (1 - \varepsilon)^t) \geq 1 - 1.01\delta - \gamma.$$

By averaging over C^*, there exists a deterministic circuit C of size $O(s' \cdot \log(1/\delta)/\varepsilon)$ such that $\Pr[C(X) = Y] \geq 1 - 1.01\delta - \gamma$. This contradicts the unpredictability assumption of Y given X. □

4 Triangle Counting

For an undirected graph $G = (V, E)$, let $\mathrm{Tri}(G)$ be the number of triangles in G. More formally, $\mathrm{Tri}(G)$ counts the number of unordered triples of distinct vertices $\{u, v, w\} \subseteq V$ such that $\{u, v\}, \{v, w\}, \{w, u\} \in E$. For a constant $\theta \in (0, 1)$, we consider the problem of computing $\mathrm{Tri}(G)$ given $G \sim \mathcal{G}(n, \theta)$, where $\mathcal{G}(n, \theta)$ is the Erdős-Rényi graph with n vertices and edge probability θ (i.e., each edge is present independently with probability θ).

We prove a nonuniform *error-tolerant* worst-case-to-average-case reduction for the triangle counting problem.

Theorem 4.1. *Let $\theta \in (0,1)$ be any constant. Suppose that there exists a size-s circuit C such that*

$$\Pr_{G\sim\mathcal{G}(N,\theta)}[C(G) = \mathrm{Tri}(G)] \geq \alpha.$$

Then, for some $k = (\log N)^{O(1)}$ and for every $n \leq N/k$, there exists a randomized circuit C' of size $\widetilde{O}\big((s+N^2)\cdot\log(1/\alpha)/\alpha^2\big)$ such that for every n-vertex graph G, it holds that

$$\Pr_{C'}\big[C'(G) = \mathrm{Tri}(G)\big] \geq 2/3,$$

where the probability is taken over the internal randomness of C'.

First, we reduce computing the sum $\mathrm{Tri}(G_1)+\cdots+\mathrm{Tri}(G_k)$ for independent random graphs $G_1,\ldots,G_k \sim \mathcal{G}(n,\theta)$ to computing $\mathrm{Tri}(G)$ for a single giant random graph $G \sim \mathcal{G}(N,\theta)$, where $N \geq kn$.

Lemma 4.2. *If there exists a size-s circuit C such that*

$$\Pr_{G\sim\mathcal{G}(N,\theta)}[C(G) = \mathrm{Tri}(G)] \geq \alpha.$$

Then, for any $k \leq N$ and $n \leq N/k$, there exists a size-$O(s+N^2)$ circuit C' such that

$$\Pr_{G_1,\ldots,G_k\sim\mathcal{G}(n,\theta)}[C'(G_1,\ldots,G_k) = \mathrm{Tri}(G_1)+\cdots+\mathrm{Tri}(G_k)] \geq \alpha.$$

To prove Lemma 4.2, we prove that, for a graph $\overline{G} = (\overline{V},\overline{E})$ and disjoint vertex subsets $V_1,\ldots,V_k \subseteq \overline{V}$, we can compute $\mathrm{Tri}(\overline{G}[V_1])+\cdots+\mathrm{Tri}(\overline{G}[V_k])$ from $\mathrm{Tri}(\overline{G})$ and information about edges outside of $V_1,\ldots,V_k$. The special case of $k=2$ is proved in [HS23, Lemma 7.4 of the full version]. We note that this is the only part where we need to rely on the special structure of triangles.

Lemma 4.3. *For an N-vertex graph $\overline{G} = (\overline{V},\overline{E})$ and nonempty disjoint vertex subsets $V_1,\ldots,V_k \subseteq \overline{V}$, let $G = \overline{G}[V_i] = (V_i,E_i)$ be the induced subgraph of $\overline{G}$ on V_i and let $G' = \big(\overline{V},\overline{E}\setminus(E_1\cup\cdots\cup E_k)\big)$ be the graph obtained by removing all edges of $G_1,\ldots,G_k$ from $\overline{G}$. Let A' be the adjacency matrix of G'. Then, it holds that*

$$\mathrm{Tri}(\overline{G}) = \mathrm{Tri}(G_1)+\cdots+\mathrm{Tri}(G_k)+\mathrm{Tri}(G')+\sum_{\{u,v\}\in E_1\cup\cdots\cup E_k}(A'^2)_{u,v}.$$

Proof. Note that the right-hand side counts some triangles in $\overline{G}$ and no triangles in $\overline{G}$ are counted more than once. Thus, it suffices to show that each triangle T of $\overline{G}$ is counted exactly once by the right-hand side. Let $T = \{a,b,c\} \subseteq \overline{V}$ be a triangle in $\overline{G}$. Let $E' = \overline{E}\setminus(E_1\cup\cdots\cup E_k)$ be the edge set of G'.

Suppose $a \notin V_1\cup\cdots\cup V_k$. Then, both $\{a,b\}$ and $\{a,c\}$ are edges in G'. Thus, if $\{b,c\} \in E_1\cup\cdots\cup E_k$, then $\sum_{\{u,v\}\in E_1\cup\cdots\cup E_k}(A'^2)_{u,v}$ counts T once. Otherwise (if $\{b,c\}\in E'$), then T is counted once by $\mathrm{Tri}(G')$.

Suppose that $u \in V_i, v \in V_j, w \in V_\ell$. If i,j,ℓ are distinct, then the three edges of T are in E'; thus T is counted once by $\mathrm{Tri}(G')$. If $i=j=\ell$, then we have $T \subseteq E_i$; thus T is counted

once by $\mathrm{Tri}(G_i)$. Otherwise, without loss of generality, suppose that $a, b \in V_i$ and $c \in V_\ell$ with $\ell \neq i$. Then, $\{u, v\} \in E_i$ and both $\{u, w\}$ and $\{v, w\}$ are edges in E'; Thus, T is counted once by $\sum_{\{a,b\}\in E_i} (A'^2)_{a,b}$.

Therefore, each triangle T of $\overline{G}$ is counted exactly once by $\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) + \mathrm{Tri}(G') + \sum_{\{u,v\}\in E_1\cup\cdots\cup E_k} (A'^2)_{u,v}$, which completes the proof. □

Proof of Lemma 4.2. Let $\overline{V} = \{1, \ldots, N\}$ and $V_1 = \{1, \ldots, n\}$, $V_2 = \{n+1, \ldots, 2n\}$, $\ldots$, $V_k = \{k(n-1)+1, \ldots, kn\}$. Take $\overline{G} = (\overline{V}, \overline{E}) \sim \mathcal{G}(N, \theta)$ and let $G_i = (V_i, E_i) = \overline{G}[V_i]$ be the induced subgraph of $\overline{G}$ on V_i. Then, the distribution of each G_i is $\mathcal{G}(n, \theta)$ and $G_1, \ldots, G_k$ are independent.

Let $G' = \big(\overline{V}, \overline{E} \setminus (E_1 \cup \cdots \cup E_k)\big)$ be the graph obtained by removing all edges of $G_1, \ldots, G_k$ from $\overline{G}$ and let A' be the adjacency matrix of G'. Note that $\overline{G}$ is decomposed into independent parts $G_1, \ldots, G_k$ and G'. Let C be the circuit that computes $\mathrm{Tri}(\overline{G})$ with probability at least α over $\overline{G}$. Then, from Lemma 4.3, we have

$$\begin{aligned}
\alpha &\le \Pr_{\overline{G}}[C(\overline{G}) = \mathrm{Tri}(\overline{G})] \\
&= \Pr_{G_1,\ldots,G_k,G'}\left[C(\overline{G}) = \mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) + \mathrm{Tri}(G') + \sum_{\{u,v\}\in E_1\cup\cdots\cup E_k} (A'^2)_{u,v}\right] \\
&= \mathop{\mathbb{E}}_{G'}\left[\Pr_{G_1,\ldots,G_k}\left[\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) = C(\overline{G}) - \mathrm{Tri}(G') - \sum_{\{u,v\}\in E_1\cup\cdots\cup E_k} (A'^2)_{u,v}\right]\right].
\end{aligned}$$

Therefore, by averaging, there exists a fixed graph G' such that, it holds with probability at least α over $G_1, \ldots, G_k \sim \mathcal{G}(n, \theta)$, that

$$\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) = C(\overline{G}) - \mathrm{Tri}(G') - \sum_{\{u,v\}\in E_1\cup\cdots\cup E_k} (A'^2)_{u,v}.$$

If we fix such G', then $\mathrm{Tri}(G')$ and A'^2 can be given as nonuniform advice. Thus, we can compute $\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k)$ with probability at least α using a circuit of size $O(s + N^2)$. This completes the proof. □

Now, we apply our Mod_p SUM lemma to get a circuit that computes $\mathrm{Tri}(G) \bmod p$ with probability at least $1 - \delta$ using C_1, where $\delta > 0$ is a parameter to be specified later.

Lemma 4.4. *Let $\delta, \alpha > 0$ be any parameters and p be a prime such that $p \ge 2/\alpha$. Suppose that there exists a size-s circuit C such that*

$$\Pr_{G_1,\ldots,G_k\sim\mathcal{G}(n,\theta)}[C(G_1, \ldots, G_k) = \mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) \bmod p] \ge \alpha.$$

Then, for some $k = O\left(\frac{p^2}{\delta}\log(1/\alpha)\right)$, there exists a size-$O\left(s \cdot \frac{pk^2\log(k/\alpha)}{\alpha^2}\right)$ circuit C' such that

$$\Pr_{G\sim\mathcal{G}(n,\theta)}[C(G) = \mathrm{Tri}(G) \bmod p] \ge 1 - \delta.$$

Proof. By assumption, the pair of random variables $(G_1, \ldots, G_k, \mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) \bmod p)$ is not $(1-\alpha)$-hard to predict given $(G_1, \ldots, G_k)$ for size s. Thus, from Theorem 3.4 for $X = \mathcal{G}(n, \theta)$ and $Y = \mathrm{Tri}(G) \bmod p$, the pair $(G, \mathrm{Tri}(G) \bmod p)$ for $G \sim \mathcal{G}(n, \theta)$ is not $(1-\delta)$-hard to predict given G for size $s' = O\left(s \cdot \frac{\alpha^2}{pk^2\log(k/\alpha)}\right)$. This completes the proof. □

To complete the remaining part of the proof of Theorem 4.1, we need two auxiliary results. First, we recall the Chinese Remainder Theorem.

Lemma 4.5 (Chinese Remainder Theorem; See [Sho12, Theorem 4.6]). *Let $M \in \mathbb{N}$ and $p_1, \dots, p_m$ be distinct primes such that $\prod_{i=1}^m p_i \geq M$. Let $a_1, \dots, a_m$ be integers such that $0 \leq a_i < p_i$ for all $i = 1, \dots, m$. Then, there exists a unique integer $0 \leq x < M$ such that*

$$x \equiv a_i \pmod{p_i} \quad \text{for all } i = 1, \dots, m.$$

Moreover, given $p_1, \dots, p_m$ and $a_1, \dots, a_m$ as input, one can compute x in time $O\big((\log M)^2\big)$.

Second, we need a worst-case-to-average-case reduction for the triangle counting problem by Boix-Adserà, Brennan, and Bresler [BBB21]. Indeed, they presented a worst-case-to-average-case reduction for the clique counting problem. In the following, we state their result in the special case of triangle counting.

Theorem 4.6 ([BBB21, Theorem 2.8]). *Let $\theta \in (0,1)$ be any constant. Suppose that there exists a size-s circuit C such that*

$$\Pr_{G \sim \mathcal{G}(N,\theta)}[C(G) = \mathrm{Tri}(G)] \geq 1 - \frac{1}{(\log n)^7}.$$

Then, there exists a randomized circuit C' of size $\widetilde{O}(s)$ such that for every n-vertex graph G, it holds that

$$\Pr_{C'}\big[C'(G) = \mathrm{Tri}(G)\big] \geq \frac{2}{3},$$

where the probability is taken over the internal randomness of C'.

Now we are ready to complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Let C be a size-s circuit that computes $\mathrm{Tri}(G)$ with probability at least α over $G \sim \mathcal{G}(N,\theta)$. Note that we may assume that $\alpha \geq 1/n^3$ since otherwise, we can compute $\mathrm{Tri}(G)$ in time $O(n^3) = O(1/\alpha)$.

Let $2/\alpha \leq p_1 < p_2 \leq \cdots \leq p_m$ be primes such that $\prod_{i=1}^m p_i \geq n^3$. Note that we can take $m = O(\log n / \log(1/\alpha)) = O(\log n)$ and by the prime number theorem, we can take $p_m = (\log n)^{O(1)}$. Set $\delta = (\log n)^{-10}$ and $k = c \cdot \frac{p_m^2 \log(1/\alpha)}{\delta} = (\log n)^{O(1)}$, where c is a sufficiently large constant.

By Lemma 4.2, there exists a size-$O(s + N^2)$ circuit that computes $\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k)$ with probability at least α over $G_1, \dots, G_k \sim \mathcal{G}(n,\theta)$ for any $n \leq N/k$. By taking modulo p of the output of this circuit, we can compute $\mathrm{Tri}(G_1) + \cdots + \mathrm{Tri}(G_k) \bmod p$ with probability at least α over $G_1, \dots, G_k \sim \mathcal{G}(n,\theta)$ for every prime p.

Then, from Lemma 4.4 for every $i \in [m]$, there exists a size-$O\Big((s+N^2) \cdot \frac{p_i k^2 \log(k/\alpha)}{\alpha^2}\Big)$ circuit $C^{(i)}$ that computes $\mathrm{Tri}(G) \bmod p_i$ with probability at least $1 - \delta$ over $G \sim \mathcal{G}(n,\theta)$.

By combining $C^{(1)}, \dots, C^{(m)}$, we obtain a circuit C^* of size $O\Big((s+N^2) \cdot m \cdot \frac{p_m k^2 \log(k/\alpha)}{\alpha^2}\Big)$ that outputs $(\mathrm{Tri}(G) \bmod p_i)_{i \in [m]}$ with probability $1 - m \cdot \delta \geq 1 - (\log n)^{-7}$ over $G \sim \mathcal{G}(n,\theta)$ (here, we apply the union bound over the circuits $C^{(1)}, \dots, C^{(m)}$ to bound the success probability). Then, by the Chinese Remainder Theorem (for $M = n^3$), we can compute $\mathrm{Tri}(G)$ from $(\mathrm{Tri}(G) \bmod p_i)_{i \in [m]}$.

Finally, by the worst-case-to-average-case reduction of Boix-Adserà, Brennan, and Bresler [BBB21], we can compute $\mathrm{Tri}(G)$ for any G with probability at least $2/3$. The resulting circuit has size

$$O\left((s+N^2)\cdot m\cdot\frac{p_m k^2\log(k/\alpha)}{\alpha^2}+(\log n)^2\right)\cdot(\log n)^{O(1)}=(s+N^2)\cdot(\log n)^{O(1)}\cdot\log(1/\alpha)/\alpha^2.$$

This completes the proof. □

5 Query Lower Bound

In this section, we establish a lower bound on the number of oracle queries required for any black-box hardness amplification of $\mathbb{F}_p$-valued functions (Theorem 5.2). Our result extends the framework of Shaltiel and Viola [SV10] and Grinberg, Shaltiel, and Viola [GSV18], which were originally developed for Boolean functions, to functions taking values in an arbitrary finite group G of size p.[2]

At a high level, a $(1-\delta,\varepsilon)$-black-box hardness amplification is a generic reduction transforming a function that is only mildly hard to compute into one that is significantly harder, using oracle access to a candidate solver.

Definition 5.1. *A $(1-\delta,\varepsilon)$-black-box hardness amplification of functions over G of size p is a pair $(\mathsf{Con},\mathsf{Red})$ such that*

- *A construction Con maps a function $f\colon\{0,1\}^n\to G$ to the function $\mathsf{Con}_f\colon\{0,1\}^m\to G$. Without loss of generality, we assume that $n\le m$;*
- *A reduction Red is an oracle circuit with two input $x\in\{0,1\}^m$ and $\alpha\in\{0,1\}^a$ such that for every $f\colon\{0,1\}^n\to G$ and $h\colon\{0,1\}^m\to G$ with*

$$\Pr_{y\sim\{0,1\}^m}[h(y)=\mathsf{Con}_f(y)]\ge\frac{1+\varepsilon}{p},$$

there is an advice $\alpha\in\{0,1\}^a$ such that

$$\Pr_{x\sim\{0,1\}^n}\left[\mathsf{Red}^h(x,\alpha)=f(x)\right]\ge 1-\delta.$$

We note that the queries made by the reduction Red^h can be adaptive, i.e., the i-th query q_i can depend on the answers $h(q_1),\dots,h(q_{i-1})$ of foregoing queries. We show that any reduction of a black-box hardness amplification must necessarily make at least $\Omega(p\log(1/\delta)/\varepsilon^2)$ oracle queries. Our proof follows the general outline of the previous works, but we substantially simplify the technical core— the *fixed-set lemma*—by replacing the intricate combinatorial argument with a short, modular, and information-theoretic proof based on the chain rule of KL divergence (see Section 5.3 for the proof).

Throughout this section, we consider a finite group G of size $p\ge 2$.

[2]The group structure is assumed only for notational convenience. Our proof applies more generally to functions taking values in any finite alphabet of size p, since the argument relies solely on the cardinality of the range rather than the group operation.

Theorem 5.2. *Let* $(\mathsf{Con}, \mathsf{Red})$ *be a* $(1-\delta, \varepsilon)$*-black-box hardness amplification of functions over* G *of size* p *such that* Red *makes at most* q *(adaptive) oracle queries. Assume there exist constants* $\nu > 0$ *and* $\delta_0 > 0$ *such that* $\max\{p, q, a\} \leq 2^{\nu n}$ *and* $\delta \leq \delta_0$*. Then, there exists a constant* $c > 0$ *such that*

$$q \geq c \cdot \min\left\{\frac{2^m}{p}, \frac{p \log(1/\delta)}{\varepsilon^2}\right\}.$$

5.1 Zoom Lemma

We prove a zoom lemma for hardness amplification of functions over G. In the proof, we identify strings $Z \in G^{\{0,1\}^m}$ with functions $Z\colon \{0,1\}^m \to G$, writing $Z(y)$ for the value of Z at index y. For $\beta \in [0,1]$, define a β-biased distribution Θ_β over $\{0,1\}^{2^m}$ as follows: choose string $x_0 \in G^{\{0,1\}^m}$ with exactly $\beta 2^m$ zeros and with the remaining $(1-\beta)2^m$ entries equally distributed among the other $p-1$ elements of G (assume that $\beta 2^m$ and $(1-\beta)2^m/(p-1)$ are integers), then apply a uniformly random permutation of the coordinates. Thus, Θ_β does not depend on the particular choice of x_0. Similarly, we identify functions $f\colon \{0,1\}^n \to G$ with strings $f \in G^{\{0,1\}^n}$.

Lemma 5.3 (Zoom lemma for black-box hardness amplification of functions over finite groups)**.** *There exists* $\nu > 0$ *and* $\delta_0 > 0$ *such that the following holds. Let* $(\mathsf{Con}, \mathsf{Red})$ *be a* $(1-\delta, \varepsilon)$*-black-box hardness amplification of functions over* G *of size* p *such that* Red *makes at most* q *(adaptive) oracle queries. Assume that* $\max\{p, q, a\} \leq 2^{\nu n}$ *and* $\delta \leq \delta_0$*. Then, there exist* $f\colon \{0,1\}^n \to G$*,* $x \in \{0,1\}^n$*,* $I \subseteq \{0,1\}^m$ *of size* $O(qa/\delta^2)$*, and* $v \in G^I$ *such that*

$$\Pr_{Z \sim \Theta_{(1+\varepsilon)/p}}\left[\mathsf{Red}^{\mathsf{Con}_f + Z}(x, \alpha) = f(x) \,\middle|\, Z(I) = v\right] \geq 1 - \sqrt{2\delta};$$

$$\Pr_{Z' \sim \Theta_{1/p}}\left[\mathsf{Red}^{\mathsf{Con}_f + Z'}(x, \alpha) = f(x) \,\middle|\, Z'(I) = v\right] \leq \frac{1.01}{p}.$$

Proof. By definition, the fraction of 0's in $Z \sim \Theta_{(1+\varepsilon)/p}$ is exactly $(1+\varepsilon)/p$. Thus, we have

$$\Pr_{Z \sim \Theta_{(1+\varepsilon)/p}}[(\mathsf{Con}_f + Z)(x) = \mathsf{Con}_f(x)] = \frac{1+\varepsilon}{p}$$

for every $f\colon \{0,1\}^n \to G$ and $x \in \{0,1\}^m$. By the property of a $(1-\delta, \varepsilon)$-black box hardness amplification, for every f and $Z \sim \Theta_{(1+\varepsilon)/p}$, it holds that

$$\Pr_Z\left[\exists \alpha \in \{0,1\}^a : \Pr_{x \sim \{0,1\}^n}\left[\mathsf{Red}^{\mathsf{Con}_f + Z}(x, \alpha) = f(x)\right] \geq 1 - \delta\right] = 1.$$

Let F be a uniformly random function $F\colon \{0,1\}^n \to G$. By choosing $\alpha \in \{0,1\}^a$ uniformly at random, we have

$$\Pr_{F, Z, \alpha}\left[\Pr_x\left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x)\right] \geq 1 - \delta\right] \geq 2^{-a}.$$

By an averaging argument, there exists $\alpha \in \{0,1\}^a$ such that

$$\Pr_{F, Z}\left[\Pr_x\left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x)\right] \geq 1 - \delta\right] \geq 2^{-a}.$$

Fix such an α.

We apply the fixed-set lemma (Lemma 5.6) by letting $Y := (F, Z) \in G^{\{0,1\}^n} \times G^{\{0,1\}^m}$, $X := Y|_{\mathcal{E}}$, and $\mathcal{E}$ be the event such that

$$\Pr_x \left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x) \right] \geq 1 - \delta. \tag{3}$$

Then, $\Pr[Y \in \mathcal{E}] \geq 2^{-a}$, which implies $\mathrm{KL}(X \parallel Y) \leq a$, where $\mathrm{KL}(X \parallel Y)$ denotes the KL divergence of X and Y (see Section 5.3 for the definition). By choosing $\gamma := \delta$, Lemma 5.6 implies there exist $I = (I_1, I_2) \subseteq \{0,1\}^n \times \{0,1\}^m$ of size $O(qa/\delta^2)$ and $v = (v_1, v_2) \in G^{I_1} \times G^{I_2}$ such that

$$\left| \Pr_{F,Z,x} \left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z(I_2) = v_2 \right] \right.$$
$$\left. - \Pr_{F,Z,x} \left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z(I_2) = v_2, \mathcal{E} \right] \right| \leq \delta.$$

By definition of $\mathcal{E}$ (see Eq. (3)), $\Pr_{F,Z,x} \left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z(I_2) = v_2, \mathcal{E} \right] \geq 1 - \delta$. Thus,

$$\Pr_{F,Z,x} \left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z(I_2) = v_2 \right] \geq 1 - 2\delta,$$

which implies, by a Markov argument, that

$$\Pr_{F,x} \left[\Pr_Z \left[\mathsf{Red}^{\mathsf{Con}_F + Z}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z(I_2) = v_2 \right] \geq 1 - \sqrt{2\delta} \right] \geq 1 - \sqrt{2\delta}. \tag{4}$$

In Lemma 5.4, we prove that

$$\Pr_{F,x,Z' \sim \Theta_{1/p}} \left[\mathsf{Red}^{\mathsf{Con}_F + Z'}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z'(I_2) = v_2 \right] \leq \frac{1}{p} \left(1 + \frac{1}{200} \right).$$

By a Markov argument, the above implies that

$$\Pr_{F,x} \left[\Pr_{Z' \sim \Theta_{1/p}} \left[\mathsf{Red}^{\mathsf{Con}_F + Z'}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z'(I_2) = v_2 \right] \leq \frac{1}{p} \left(1 + \frac{1}{100} \right) \right] \geq \frac{1}{202}. \tag{5}$$

By (4) and (5), there exist f and $x \in \{0,1\}^n$ such that

$$\Pr_{Z \sim \Theta_{(1+\varepsilon)/p}} \left[\mathsf{Red}^{\mathsf{Con}_f + Z}(x, \alpha) = f(x) \,\middle|\, Z(I_2) = v_2 \right] \geq 1 - \sqrt{2\delta};$$
$$\Pr_{Z' \sim \Theta_{1/p}} \left[\mathsf{Red}^{\mathsf{Con}_f + Z'}(x, \alpha) = f(x) \,\middle|\, Z'(I_2) = v_2 \right] \leq \frac{1}{p} \left(1 + \frac{1}{100} \right)$$

for sufficiently small δ_0. Hence, the statement follows. □

Lemma 5.4.

$$\Pr_{F,x,Z' \sim \Theta_{1/p}} \left[\mathsf{Red}^{\mathsf{Con}_F + Z'}(x, \alpha) = F(x) \,\middle|\, F(I_1) = v_1, Z'(I_2) = v_2 \right] \leq \frac{1}{p} \left(1 + \frac{1}{200} \right)$$

Proof. Let Z_2' be a random variable sampled according to $\Theta_{1/p}$ conditioned on $Z_2'(I_2) = v_2$. Let $\mathcal{F}_1$ be the set of all functions $f\colon \{0,1\}^n \to G$ satisfying $f(I_1) = v_1$. Define the set B_1 such that

$$B_1 := \left\{ f \in \mathcal{F}_1 \colon \Pr_{Z_2'} \left[\mathsf{Adv}^f_{\mathsf{Red}^{\mathsf{Con}_f + Z_2'}(\alpha)} \geq \frac{\gamma}{4} \right] \geq \frac{\gamma}{4} \right\},$$

where

$$\mathsf{Adv}^f_{\mathsf{Red}^O(\alpha)} := \Pr_{x \sim \{0,1\}^n} \left[\mathsf{Red}^O(x, \alpha) = f(x) \right] - \frac{1}{p} \quad \text{and} \quad \gamma := \frac{1}{200p}.$$

To analyze $\mathsf{Red}^{\mathsf{Con}_f + Z_2'}$, we will replace the oracle with a simpler distribution *independent* from f. Let U_2 be a uniformly random function $U_2\colon \{0,1\}^m \to G$ conditioned on $U_2(I_2) = v_2$. On the unfixed coordinates $\{0,1\}^m \setminus I_2$, the distribution Z_2' (sampled from $\Theta_{1/p}$ and conditioned on $Z_2'(I_2) = v_2$), is distributed as sampling *without replacement* from a multiset that contains, for each $a \in G$, exactly $2^m/p - n_a$ copies of a, where $n_a := |\{i \in I_2 \colon v_2(i) = a\}|$. In contrast, the distribution U_2 corresponds to *independent sampling with replacement* from the uniform distribution over G. Since both Z_2' and U_2 are conditioned on the same fixed values over I_2, their behavior on the remaining coordinates can be compared by the standard coupling argument for sampling with and without replacement. By a classical comparison (e.g., [DF80]), the output distribution of any q-query (adaptive) oracle algorithm A satisfies

$$d_{\mathrm{TV}}\left(A^{Z_2'}, A^{U_2}\right) \leq \frac{\binom{q}{2}}{2^m - |I_2| - 1} \leq \frac{q^2}{2^m} \tag{6}$$

for any f.[3] Since $|I_2| = O(qa/\delta^2)$, we have $|I_2| \leq 2^{m-1}$ for sufficiently small ν, and hence the second inequality holds. Since (6) implies $d_{\mathrm{TV}}\left(A^{\mathsf{Con}_f + Z_2'}, A^{\mathsf{Con}_f + U_2}\right) \leq q^2/2^m$, we have for every t,

$$\Pr_{U_2}\left[\mathsf{Adv}^f_{\mathsf{Red}^{\mathsf{Con}_f + U_2}(\alpha)} \geq t \right] \geq \Pr_{Z_2'}\left[\mathsf{Adv}^f_{\mathsf{Red}^{\mathsf{Con}_f + Z_2'}(\alpha)} \geq t \right] - \frac{q^2}{2^m}.$$

We can define the set $B_1^{(1)}$ such that

$$B_1 \subseteq B_1^{(1)} := \left\{ f \in \mathcal{F}_1 : \Pr_{U_2}\left[\mathsf{Adv}^f_{\mathsf{Red}^{\mathsf{Con}_f + U_2}(\alpha)} \geq \frac{\gamma}{4} \right] \geq \frac{\gamma}{4} - \frac{q^2}{2^m} \right\}.$$

Let $\varepsilon_2 := \frac{\gamma}{4} - \frac{q^2}{2^m}$. We assume $\varepsilon_2 \geq \frac{\gamma}{8}$, which holds for sufficiently small ν.

Let U be a uniformly random function $U\colon \{0,1\}^m \to G$. Notice that $U(y)$ is distributed identically to $(\mathsf{Con}_f + U_2)(y)$ when $y \notin I_2$, and is equal to $(\mathsf{Con}_f + U_2)(y)$ with probability $1/p$ when $y \in I_2$. Thus, we have

$$\Pr_{U}\left[\mathsf{Adv}^f_{\mathsf{Red}^{U}(\alpha)} \geq t \right] \geq \Pr_{U_2}\left[\mathsf{Adv}^f_{\mathsf{Red}^{\mathsf{Con}_f + U_2}(\alpha)} \geq t \right] \cdot p^{-|I_2|}$$

[3] Since both Z_2' and U_2 fix the same values on I_2, they induce distributions over $\{0,1\}^m \setminus I_2$ supported on the same multiset of size $N := 2^m - |I_2|$: Z_2' samples without replacement, while U_2 samples with replacement. The Diaconis–Freedman coupling generates two length-q sequences $(C_{\mathsf{wo}}, C_{\mathsf{wr}})$ that remain identical until a collision occurs in the with-replacement process, which happens with probability at most $\frac{i-1}{N-1}$ at step i. Hence, $\Pr[C_{\mathsf{wo}} \neq C_{\mathsf{wr}}] \leq \sum_{i=1}^{q} \frac{i-1}{N-1} = \binom{q}{2}/(N-1)$, and thus $d_{\mathrm{TV}}(C_{\mathsf{wo}}, C_{\mathsf{wr}}) \leq \binom{q}{2}/(N-1)$. The first inequality in (6) follows since A only post-processes oracle answers. This argument does not rely on global uniformity and remains valid under the above conditioning.

for any t. Hence, we can define the set $B_1^{(2)}$ such that

$$B_1^{(1)} \subseteq B_1^{(2)} := \left\{ f \in \mathcal{F}_1 : \Pr_U \left[\mathsf{Adv}^f_{\mathsf{Red}^U(\alpha)} \geq \varepsilon_2 \right] \geq \varepsilon_2 \cdot p^{-|I_2|} \right\}.$$

By averaging over all $f \in B_1^{(2)}$, there exists a function $u\colon \{0,1\}^m \to G$ such that an $\varepsilon_2 \cdot p^{-|I_2|}$-fraction of $f \in B_1^{(2)}$ lies in

$$B_1^{(3)} := \left\{ f \in \mathcal{F}_1 \,\middle|\, \mathsf{Adv}^f_{\mathsf{Red}^u(\alpha)} \geq \varepsilon_2 \right\}.$$

We fix such a u. Now, $\mathsf{Red}^u(\cdot, \alpha)$ is a fixed function. The probability that a random function F_1 from $\mathcal{F}_1$ satisfies $\mathsf{Adv}^{F_1}_{\mathsf{Red}^u(\alpha)} \geq \varepsilon_2$ is bounded above as follows:

$$\begin{aligned}
\Pr_{F_1 \sim \mathcal{F}_1} \left[\mathsf{Adv}^{F_1}_{\mathsf{Red}^u(\alpha)} \geq \varepsilon_2 \right] &\leq \Pr_{F_1} \left[\sum_{x \in \{0,1\}^n} \frac{1}{2^n} \cdot \mathbb{1} \left\{ \mathsf{Red}^U(x, \alpha) = F_1(x) \right\} \geq \frac{1}{p} + \frac{\gamma}{8} \right] \\
&\leq \Pr_{F_1} \left[\sum_{x \notin I_1} \mathbb{1} \left\{ \mathsf{Red}^u(x, \alpha) = F_1(x) \right\} \geq 2^n \left(\frac{1}{p} + \frac{\gamma}{8} \right) - |I_1| \right] \\
&\leq \Pr_{F_1} \left[\sum_{x \notin I_1} \mathbb{1} \left\{ \mathsf{Red}^u(x, \alpha) = F_1(x) \right\} \geq 2^n \left(\frac{1}{p} + \frac{\gamma}{16} \right) \right] \qquad (7)
\end{aligned}$$

where the last inequality uses $|I_1|/2^n \leq \gamma/16$, which holds for sufficiently small ν. Let $S = \sum_{x \notin I_1} \mathbb{1} \left\{ \mathsf{Red}^u(x, \alpha) = F_1(x) \right\}$, which is the sum of $N := 2^n - |I_1|$ independent indicators. Since the marginal distribution of $F_1(x)$ is uniform over G for each $x \in \{0,1\}^n \setminus I_1$, we have $\mathbb{E}_{F_1}[S] = \frac{N}{p} \leq \frac{2^n}{p}$. By the Hoeffding bound (Lemma 2.2), Eq. (7) satisfies

$$\begin{aligned}
(7) &\leq \Pr \left[S \geq \mathbb{E}[S] + 2^n \cdot \frac{\gamma}{16} \right] \\
&\leq \exp \left(-\frac{2}{N} \cdot \left(2^n \cdot \frac{\gamma}{16} \right)^2 \right) \\
&\leq \exp \left(-c \, \gamma^2 \, 2^n \right) \\
&= 2^{-\Omega(2^n/p^2)}
\end{aligned}$$

for some absolute constant $c > 0$ (recall $\gamma = 1/(200p)$). Hence,

$$\Pr \left[F_1 \in B_1^{(3)} \right] = \Pr_{F_1 \sim \mathcal{F}_1} \left[\mathsf{Adv}^{F_1}_{\mathsf{Red}^u(\alpha)} \geq \varepsilon_2 \right] \leq 2^{-\Omega(2^n/p^2)}.$$

Therefore, we have

$$\begin{aligned}
\Pr[F_1 \in B_1] &\leq \Pr_{F_1}[F_1 \in B_1^{(2)}] \\
&\leq \frac{p^{|I_2|}}{\varepsilon_2} \cdot \Pr_{F_1}[F_1 \in B_1^{(3)}] && \because \Pr_{f \sim B_1^{(2)}}[f \in B_1^{(3)}] \geq \frac{\varepsilon_2}{p^{|I_2|}} \\
&\leq \frac{p^{O(qa/\delta^2)}}{\gamma/8} \cdot 2^{-\Omega(2^n/p^2)},
\end{aligned}$$

which is less than $\gamma/2$ for sufficiently small ν and large n.

For any $f \notin B_1$,

$$\mathop{\mathbb{E}}_{Z'_2}\left[\mathsf{Adv}^{f}_{\mathsf{Con}_f+Z'_2(\alpha)}\right] \le \frac{\gamma}{4}+\frac{\gamma}{4}=\frac{\gamma}{2}.$$

Thus, since $\mathsf{Adv}^{f}_{\mathsf{Red}^{\mathsf{Con}_f+Z'_2}(\alpha)} \le 1$, we have that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{F_1,Z'_2}\left[\mathsf{Adv}^{F_1}_{\mathsf{Red}^{\mathsf{Con}_{F_1}+Z'_2}(\alpha)}\right] \\
&= \Pr[F_1 \in B_1]\cdot \mathop{\mathbb{E}}_{F_1,Z'_2}\left[\mathsf{Adv}^{F_1}_{\mathsf{Red}^{\mathsf{Con}_{F_1}+Z'_2}(\alpha)} \,\middle|\, F_1 \in B_1\right] + \Pr[F_1 \notin B_1]\cdot \mathop{\mathbb{E}}_{F_1,Z'_2}\left[\mathsf{Adv}^{F_1}_{\mathsf{Red}^{\mathsf{Con}_{F_1}+Z'_2}(\alpha)} \,\middle|\, F_1 \notin B_1\right] \\
&< \frac{\gamma}{2}\cdot 1 + 1\cdot\frac{\gamma}{2} \le \gamma,
\end{aligned}$$

which implies the statement of the lemma. □

5.2 Proof of Query Lower Bound

Using a Zoom lemma, we prove a lower bound on the number of queries made by a black-box hardness amplification of functions over G.

Proof of Theorem 5.2. We use the zoom lemma (Lemma 5.3) to construct a circuit C^{Z^*} that simulates $\mathsf{Red}^{\mathsf{Con}_f+Z^*}(x,\alpha)$ conditioned on $Z^*(I)=v$, where $Z^* \in \{Z,Z'\}$. Let $f\colon \{0,1\}^n \to G, x \in \{0,1\}^n, I \subseteq G^{\{0,1\}^m}$, and $v \in G^I$ as in Lemma 5.3. By using the oracle $Z^* \in \{Z,Z'\}$, C^{Z^*} runs $\mathsf{Red}^{O}(x,\alpha)$, where for each oracle query $y \in \{0,1\}^m$ by Red^O, we simulate the oracle O by setting the answer to be $\mathsf{Con}_f(y)+v_y$ if $y \in I$, and $\mathsf{Con}_f(y)+Z^*(y)$ otherwise. Finally, C outputs 1 if the output by $\mathsf{Red}^{O}(x,\alpha)$ equals $f(x)$, and 0 otherwise.

We claim that C *almost* simulates $\mathsf{Red}^{\mathsf{Con}_f}(x,\alpha)$ conditioned on $Z^*(I)=v$ in the following sense:

Claim 5.5. *Let $I \subseteq G^{\{0,1\}^m}, v \in G^I, q \in \mathbb{N}$ as in Lemma 5.3. Let C^Z be any adaptive q-query decision tree that makes oracle queries to Z. Let $\beta \in \left\{\frac{1}{p},\frac{1+\varepsilon}{p}\right\}$. Let $Z^{(1)}$ be a random string obtained by sampling $Z \sim \Theta_\beta$ and then setting $Z(I) \leftarrow v$. Let $Z^{(2)}=Z$ conditioned on $Z(I)=v$.*

Then, it holds that

$$\left|\Pr\left[C^{Z^{(1)}}=1\right]-\Pr\left[C^{Z^{(2)}}=1\right]\right| \le \frac{q|I|}{2(2^m-|I|-q)} = o(1).$$

We will prove Claim 5.5 later. Since C^Z makes at most q oracle queries to Z, from Claim 5.5, we have

$$\Pr_{Z\sim\Theta_{(1+\varepsilon)/p}}\left[C^Z=1\right] \ge 1-\sqrt{2\delta}-o(1) \quad \text{and} \quad \Pr_{Z'\sim\Theta_{1/p}}\left[C^{Z'}=1\right] \le \frac{1.01}{p}+o(1).$$

Note that C makes at most q oracle queries. Thus, by Lemma B.1, for any $p \ge 2$, there exists a $c>0$ such that $q \ge c\cdot\min\left\{\frac{2^m}{p},\frac{p\log(1/\delta)}{\varepsilon^2}\right\}$. □

Finally, we prove Claim 5.5.

Proof of Claim 5.5. For notational convenience, let $M = 2^m$. Recall that Θ_β is the distribution over G^M obtained by uniformly permuting a multiset containing βM copies of 0 and an equal number of copies of each element of $G \setminus \{0\}$. For $g \in G$, let M_g and ℓ_g denote the numbers of g appearing in $Z \sim \Theta_\beta$ and v, respectively. Note that $\sum_g M_g = M$, $\sum_g \ell_g = |I|$, and

$$M_g = \begin{cases} \beta M & \text{if } g = 0, \\ \frac{1-\beta}{p-1} M & \text{if } g \neq 0. \end{cases}$$

Suppose C^Z makes q oracle queries to $Z \sim \Theta_\beta$ and receives q responses $Z_1, \dots, Z_q \in G$. Since $Z \sim \Theta_\beta$, the responses can be seen as the outcome of q samples from the follwing process: Consider a bin containing M_g balls with label g for each $g \in G$. Then, sample q balls from the bin without replacement. Then, Z_i is the label of the i-th ball sampled.

With this in mind, we shall consider $C^{Z^{(j)}}$ for $j = 1, 2$. Let $Z_1^{(j)}, \dots, Z_q^{(j)}$ be the responses of $C^{Z^{(j)}}$ to the oracle queries to $Z^{(j)}$. Then, $Z_1^{(j)}, \dots, Z_q^{(j)}$ are the outcome of q samples from the follwing process:

- If the oracle is $Z^{(1)}$, then consider a bin containing M_g balls with label g for each $g \in G$. For $i = 1, \dots, q$, sample a ball from the bin and remove it from the bin. If the i-th query q is in I, set $Z_i^{(1)} = v_q$. Otherwise, set $Z_i^{(1)}$ to be the label of the ball sampled.
- If the oracle is $Z^{(2)}$, then consider a bin containing $M_g - \ell_g$ balls with label g for each $g \in G$. If the i-th query q is in I, then set $Z_i^{(1)} = v_q$. Otherwise, sample a ball from the bin and set $Z_i^{(1)}$ to be the label of the ball sampled. Then, remove the ball from the bin.

Now we prove that we can couple $(Z_1^{(1)}, \dots, Z_q^{(1)})$ and $(Z_1^{(2)}, \dots, Z_q^{(2)})$ such that

$$(Z_1^{(1)}, \dots, Z_q^{(1)}) = (Z_1^{(2)}, \dots, Z_q^{(2)})$$

with probability at least $1 - o(1)$, which implies the claim. We prove this by induction on $i = 1, \dots, q$. For notational convenience, define $Z_0^{(1)} = Z_0^{(2)} = 0$. Suppose that $Z_j^{(1)} = Z_j^{(2)}$ for all $j < i$ (if $i = 1$, this is always true). Note that the i-th query q is identical for $C^{Z^{(1)}}$ and $C^{Z^{(2)}}$. If $q \in I$, then $Z_i^{(1)} = Z_i^{(2)} = v_q$ with probability 1. Otherwise, we compare the distribution of $Z_i^{(1)}$ and $Z_i^{(2)}$.

- For $Z_i^{(1)}$, at the beginnig of the i-th round of the process, the bin contains $M - (i-1)$ balls in total and the number of balls with label g lies between $M_g - (i-1)$ and M_g.
- For $Z_i^{(2)}$, at the beginnig of the i-th round of the process, the bin contains $M - |I| - (i-1)$ balls in total and the number of balls with label g lies between $M_g - \ell_g - (i-1)$ and $M_g - \ell_g$.

Therefore, the total variation distance between $Z_i^{(1)}$ and $Z_i^{(2)}$ satisfies

$$d_{\mathrm{TV}}\left(Z_i^{(1)}, Z_i^{(2)}\right) \leq \frac{1}{2} \cdot \frac{\sum_g \ell_g}{M - |I| - (i-1)} \leq \frac{|I|}{2(M - |I| - q)}.$$

Thus, by the maximal coupling inequality (see, e.g., [LP17, Proposition 4.7]), there exists a coupling of $Z_i^{(1)}$ and $Z_i^{(2)}$ such that

$$\Pr\left[Z_i^{(1)} \neq Z_i^{(2)}\right] \leq \frac{|I|}{2(M - |I| - q)}. \tag{8}$$

Repeating this for $i = 1, \ldots, q$, we can couple $(Z_1^{(1)}, \ldots, Z_q^{(1)})$ and $(Z_1^{(2)}, \ldots, Z_q^{(2)})$ such that

$$\Pr\left[(Z_1^{(1)}, \ldots, Z_q^{(1)}) \neq (Z_1^{(2)}, \ldots, Z_q^{(2)})\right] \leq \sum_{i=1}^{q} \Pr\left[Z_i^{(1)} \neq Z_i^{(2)}\right] \leq \frac{q|I|}{2(M - |I| - q)}.$$

By the assumption on $|I|, q, M = 2^m$ (and $n \leq m$), by taking the constant ν of Lemma 5.3 to be sufficiently small, the right hand side is at most $o(1)$. □

5.3 A Simple Proof of the Fixed-Set Lemma

The fixed-set lemma is originally proved by Grinberg, Shaltiel, and Viola [GSV18] which was later generalized by Shaltiel [Sha23, Lemma 2.12]. In this section, we prove a simple proof of the generalized version based on the chain rule of KL divergence. Before proving the lemma, we recall some basic facts of information theory.

Notation of Information Theory. Let X and Y be two random variables such that $\mathrm{supp}(X) \subseteq \mathrm{supp}(Y)$. The KL divergence of X and Y is defined by $\mathrm{KL}(X \parallel Y) = \mathbb{E}_{x \sim X}\left[\log \frac{\Pr[X=x]}{\Pr[Y=x]}\right]$. Unless otherwise specified, we always use the natural logarithm. Let (X, W) and (Y, Z) be jointly distributed random variables. The conditional KL divergence of X given Z and Y given W is defined by $\mathrm{KL}(X|W \parallel Y|Z) = \mathbb{E}_{w \sim W}\left[\mathrm{KL}(X|_{W=w} \parallel Y|_{Z=w})\right]$. It is known that the chain rule of KL divergence holds:

$$\mathrm{KL}(X \parallel Y) = \mathrm{KL}(W \parallel Z) + \mathrm{KL}(X|W \parallel Y|Z).$$

We recall Pinsker's inequality, which states that

$$\mathrm{KL}(X \parallel Y) \geq 2 d_{\mathrm{TV}}(X, Y)^2.$$

The data-processing inequality states that for any function f, we have

$$\mathrm{KL}(X \parallel Y) \geq \mathrm{KL}(f(X) \parallel f(Y)).$$

Consider a random vector $X = (X_1, \ldots, X_n)$ that takes value in Ω^n. For $I = \{i_1 < \cdots < i_q\} \subseteq [n]$, we denote by $X_I := (X_{i_1}, \ldots, X_{i_q})$ the projection of X onto I. For $x = (x_1, \ldots, x_q)$, we denote by $X|_{X_I = x}$ be the random vector X conditioned on $X_I = x$.

Now we state the fixed-set lemma.

Lemma 5.6 (Fixed-Set Lemma)**.** *Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ be two random vectors that takes value in Ω^n such that $\mathrm{supp}(X) \subseteq \mathrm{supp}(Y)$. Suppose that $\mathrm{KL}(X \parallel Y) \leq a$. Let $q \in \mathbb{N}$ and $\gamma > 0$ be any parameters. Then, there exist a subset $I \subseteq [n]$ of size at most $O(qa/\gamma^2)$ and $x \in \Omega^I$ such that, for any adaptive q-query decision tree $T^{(\cdot)}$ that makes q adaptive queries, it holds that*

$$\left|\Pr\left[T^X = 1 \mid X_I = x\right] - \Pr\left[T^Y = 1 \mid Y_I = x\right]\right| \leq \gamma.$$

The original lemma [Sha23, Lemma 2.12] is stated for X being $X = Y|_{\mathcal{E}}$ for some event $\mathcal{E}$ of Y such that $\Pr[Y \in \mathcal{E}] \geq e^{-a}$. Our version of the lemma can also be applied to this setting as $\mathrm{KL}(Y|_{\mathcal{E}} \parallel Y) \leq a$.

Remark 5.7. *Grinberg, Shaltiel, and Viola [GSV18] proved a special case of Lemma 5.6 for Y being uniformly random over $\{0,1\}^n$. Shaltiel [Sha23] proved that if there exists a q-query decision tree that distinguishes X and Y, then we can fix q coordinates of X and Y such that the resulting random vectors X' and Y' satisfies $\Pr[Y' \in \mathcal{E}] \geq e^{\Omega(\gamma^2)} \cdot \Pr[Y \in \mathcal{E}]$. Since we can apply this result for $O(a/\gamma^2)$ times and each application fixes q coordinates, we can fix $O(qa/\gamma^2)$ coordinates such the resulting random vectors are indistinguishable. In this paper, we provide an alternative proof by considering the KL divergence of X and Y. A benefit of this approach is that we can exploit the chain rule of KL divergence, which makes the proof simpler. In the following, we prove the iterative approach for KL divergence.*

Lemma 5.8. *Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ be two random variables that takes value in Ω^n such that $\mathrm{supp}(X) \subseteq \mathrm{supp}(Y)$. Suppose that there exists a decision tree $T^{(\cdot)}$ that makes q adaptive queries such that*

$$\left| \Pr_X \left[T^X = 1\right] - \Pr_Y \left[T^Y = 1\right] \right| > \gamma.$$

Then, there exists an index subset $I \subseteq [n]$ of size q and $w \in \Omega^q$ such that

$$\mathrm{KL}(X|_{X_I=w} \parallel Y|_{Y_I=w}) \leq \mathrm{KL}(X \parallel Y) - 2\gamma^2.$$

Proof. Let $T^{(\cdot)}$ be an adaptive q-query decision tree that satisfies the requirement. Let $W, Z \in \Omega^q$ be the random vectors that $T^{(\cdot)}$ receives from oracles X and Y, respectively. That is, in the i-th query, T^X receives W_i and T^Y receives Z_i. By the assumption on $T^{(\cdot)}$, from Pinsker's inequality, we have

$$\mathrm{KL}(W \parallel Z) \geq 2d_{\mathrm{TV}}(W, Z)^2 \geq 2\gamma^2.$$

By the chain rule, we have

$$\mathrm{KL}(X \parallel Y) = \mathrm{KL}(W \parallel Z) + \mathrm{KL}(X|W \parallel Y|Z) \geq 2\gamma^2 + \mathop{\mathbb{E}}_{w \sim W} \mathrm{KL}(X|_{W=w} \parallel Y|_{Z=w}).$$

Thus, by averaging, there exists a $w^* \in \mathrm{supp}(W)$ such that

$$\mathrm{KL}(X \parallel Y) \geq 2\gamma^2 + \mathrm{KL}(X|_{W=w^*} \parallel Y|_{Z=w^*}).$$

Note that $X|_{W=w^*}$ and $Y|_{Z=w^*}$ can be written as $X|_{X_I=w^*}$ and $Y|_{Y_I=w^*}$ for some $I \subseteq [n]$ of size q, respectively.

□

Now we are ready to prove the fixed-set lemma.

Proof of Lemma 5.6. Consider the following process that produces a sequence of random vectors $(X^{(0)}, Y^{(0)}), (X^{(1)}, Y^{(1)}), \ldots, (X^{(t)}, Y^{(t)})$:

1. Initialize $X^{(0)} := X$ and $Y^{(0)} := Y$.
2. For $i = 0, 1, 2, \ldots,$ do the following:

(a) If there exists a q-query decision tree $T^{(\cdot)}$ that makes q adaptive queries such that

$$\left|\Pr_{X^{(i)}}[T^{X^{(i)}} = 1] - \Pr_{Y^{(i)}}[T^{Y^{(i)}} = 1]\right| > \gamma,$$

then, by Lemma 5.8, there exists a index subset $I^{(i)} \subseteq [n]$ of size q and $w^{(i)} \in \Omega^q$ such that

$$\mathrm{KL}\Big(X^{(i)}|_{X_{I^{(i)}}=w^{(i)}} \,\Big\|\, Y^{(i)}|_{Y_{I^{(i)}}=w^{(i)}}\Big) \le \mathrm{KL}\Big(X^{(i)} \,\Big\|\, Y^{(i)}\Big) - 2\gamma^2.$$

Let $X^{(i+1)} := X^{(i)}|_{X_{I^{(i)}}=w^{(i)}}$ and $Y^{(i+1)} := Y^{(i)}|_{Y_{I^{(i)}}=w^{(i)}}$.

(b) Otherwise, output $X^{(i)}$ and $Y^{(i)}$.

By the non-negativity of KL divergence, we have that

$$0 \le \mathrm{KL}\Big(X^{(t)} \,\Big\|\, Y^{(t)}\Big) \le \mathrm{KL}(X \parallel Y) - 2t\gamma^2 \le a - 2t\gamma^2.$$

Therefore, we have $t \le \frac{a}{2\gamma^2}$. Moreover, $X^{(t)}$ and $Y^{(t)}$ are obtained from X and Y by fixing coordinates in $\bigcup_{i=0}^{t} I^{(i)}$, which has size at most $qt \le \frac{aq}{2\gamma^2}$. Finally, since the procedure stops at most t times, any adaptive q-query decision tree $T^{(\cdot)}$ cannot distinguish $X^{(t)}$ and $Y^{(t)}$, i.e.,

$$\left|\Pr_{X^{(t)}}[T^{X^{(t)}} = 1] - \Pr_{Y^{(t)}}[T^{Y^{(t)}} = 1]\right| \le \gamma.$$

This completes the proof. □

References

[ABPSS25] Prashanth Amireddy, Amik Raj Behera, Manaswi Paraashar, Srikanth Srinivasan, and Madhu Sudan. "Low degree local correction over the Boolean cube". en. In: *Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, 2025, pp. 5504–5511. DOI: 10.1137/1.9781611978322.187 (cit. on p. 4).

[AGGS22] Vahid R Asadi, Alexander Golovnev, Tom Gur, and Igor Shinkar. "Worst-case to average-case reductions via additive combinatorics". In: *Symposium on Theory of Computing (STOC)*. 2022, pp. 1566–1574. DOI: 10.1145/3519935.3520041 (cit. on p. 2).

[AGGSS24] Vahid R Asadi, Alexander Golovnev, Tom Gur, Igor Shinkar, and Sathyawageeswar Subramanian. "Quantum worst-case to average-case reductions for all linear problems". en. In: *Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, 2024, pp. 2535–2567. DOI: 10.1137/1.9781611977912.90 (cit. on p. 2).

[BBB21] Enric Boix-Adserà, Matthew Brennan, and Guy Bresler. "The Average-Case Complexity of Counting Cliques in Erdös–Rényi Hypergraphs". In: *SIAM Journal on Computing* (2021), FOCS19–39–FOCS19–80. DOI: 10.1137/20M1316044 (cit. on pp. 4, 6, 8, 22, 23).

[BKW03] Avrim Blum, Adam Kalai, and Hal Wasserman. "Noise-tolerant learning, the parity problem, and the statistical query model". en. In: *Journal of the ACM* 50 (4 2003), pp. 506–519. DOI: 10.1145/792538.792543 (cit. on pp. 8, 18).

[DF80] P. Diaconis and D. Freedman. "Finite Exchangeable Sequences". In: *The Annals of Probability* 8 (1980), pp. 745–64 (cit. on p. 26).

[DHKNT21] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. "List-Decoding with Double Samplers". In: *SIAM Journal on Computing* 50 (2 2021), pp. 301–349. DOI: 10.1137/19M1276650 (cit. on pp. 2, 3).

[DLW20] Mina Dalirrooyfard, Andrea Lincoln, and Virginia Vassilevska Williams. "New Techniques for Proving Fine-Grained Average-Case Hardness". In: *Symposium on Foundations of Computer Science (FOCS)*. 2020, pp. 774–785. DOI: 10.1109/FOCS46700.2020.00077 (cit. on pp. 4, 6).

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data". In: *SIAM Journal on Computing* 38 (1 2008), pp. 97–139. DOI: 10.1137/060651380 (cit. on pp. 9, 10).

[DT09] Anindya De and Luca Trevisan. "Extractors Using Hardness Amplification". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. 2009, pp. 462–475. DOI: 10.1007/978-3-642-03685-9_35 (cit. on p. 2).

[GK16] Justin Gilmer and Swastik Kopparty. "A local central limit theorem for triangles in a random graph". en. In: *Random Structures & Algorithms* 48 (4 2016), pp. 732–750. DOI: 10.1002/rsa.20604. URL: http://dx.doi.org/10.1002/rsa.20604 (visited on 02/13/2026) (cit. on p. 4).

[GK20] Elazar Goldenberg and Karthik C. S. "Hardness Amplification of Optimization Problems". In: *Innovations in Theoretical Computer Science Conference (ITCS)*. Ed. by Thomas Vidick. ITCS. 2020, 1:1–1:13. DOI: 10.4230/LIPIcs.ITCS.2020.1 (cit. on p. 2).

[GL89] O Goldreich and L A Levin. "A hard-core predicate for all one-way functions". In: *Proceedings of Symposium on Theory of Computing (STOC)* (1989), pp. 25–32. DOI: 10.1145/73007.73010. URL: http://dx.doi.org/10.1145/73007.73010 (cit. on p. 6).

[GNW11] Oded Goldreich, Noam Nisan, and Avi Wigderson. "On Yao's XOR-Lemma". In: *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation* (2011), pp. 273–301. DOI: 10.1007/978-3-642-22670-0_23 (cit. on p. 2).

[Gol20] Oded Goldreich. "On Counting t-Cliques Mod 2". In: *ECCC TR20-104* (2020) (cit. on p. 6).

[GR18] Oded Goldreich and Guy Rothblum. "Counting t-Cliques: Worst-Case to Average-Case Reductions and Direct Interactive Proof Systems". In: *Symposium on Foundations of Computer Science (FOCS)*. 2018, pp. 77–88. DOI: 10.1109/FOCS.2018.00017 (cit. on p. 5).

[GRS00] Oded Goldreich, Ronitt Rubinfeld, and Madhu Sudan. "Learning Polynomials with Queries: The Highly Noisy Case". In: *SIAM Journal on Discrete Mathematics* 13 (4 2000), pp. 535–570. DOI: 10.1137/S0895480198344540 (cit. on pp. 4, 6).

[GSS24] Ashish Gola, Igor Shinkar, and Harsimran Singh. *Matrix Multiplication Reductions*. en. 2024. DOI: 10.4230/LIPICS.APPROX/RANDOM.2024.34 (cit. on p. 2).

[GSV18] Aryeh Grinberg, Ronen Shaltiel, and Emanuele Viola. "Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs". en. In: *Symposium on Foundations of Computer Science (FOCS)*. 2018. DOI: 10.1109/focs.2018.00094 (cit. on pp. 5, 23, 30, 31).

[HS23] Shuichi Hirahara and Nobutaka Shimizu. "Hardness Self-Amplification: Simplified, Optimized, and Unified". In: *Symposium on Theory of Computing (STOC)*. STOC 2023. 2023. DOI: 10.1145/3564246.3585189 (cit. on pp. 2, 6, 20).

[HS24] Shuichi Hirahara and Nobutaka Shimizu. "Planted Clique Conjectures Are Equivalent". In: *Symposium on Theory of Computing (STOC)*. STOC 2024. 2024, pp. 358–366. DOI: 10.1145/3618260.3649751 (cit. on p. 2).

[HS25] Shuichi Hirahara and Nobutaka Shimizu. "Error-correction of matrix multiplication algorithms". In: *Symposium on Theory of Computing (STOC)*. 2025. DOI: 10.1145/3717823.3718244 (cit. on pp. 2, 6).

[HVV06] Alexander Healy, Salil Vadhan, and Emanuele Viola. "Using Nondeterminism to Amplify Hardness". In: *SIAM Journal on Computing* 35 (4 2006), pp. 903–931. DOI: 10.1137/S0097539705447281 (cit. on pp. 7, 14).

[IJKW10] Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. "Uniform Direct Product Theorems: Simplified, Optimized, and Derandomized". In: *SIAM Journal on Computing* 39 (4 2010), pp. 1637–1665. DOI: 10.1137/080734030 (cit. on pp. 2–4, 6–8, 13).

[Imp95] R Impagliazzo. "Hard-core distributions for somewhat hard problems". In: *Foundations of Computer Science (FOCS)*. 1995, pp. 538–545. DOI: 10.1109/SFCS.1995.492584 (cit. on pp. 8, 18).

[IW97] Russell Impagliazzo and Avi Wigderson. "P = BPP if E requires exponential circuits: derandomizing the XOR lemma". In: *Symposium on Theory of Computing (STOC)*. 1997. DOI: 10.1145/258533.258590 (cit. on p. 2).

[Jer92] Mark Jerrum. "Large cliques elude the metropolis process". en. In: *Random Structures & Algorithms* 3 (4 1992), pp. 347–359. DOI: 10.1002/rsa.3240030402 (cit. on pp. 8, 18).

[Kuč95] Luděk Kučera. "Expected complexity of graph partitioning problems". In: *Discrete Applied Mathematics* 57 (2 1995), pp. 193–212. DOI: 10.1016/0166-218X(94)00103-K (cit. on pp. 8, 18).

[LP17] David Levin and Yuval Peres. *Markov Chains and Mixing Times*. 2nd. Providence, Rhode Island: American Mathematical Society, 2017. DOI: 10.1090/mbk/107. URL: http://dx.doi.org/10.1090/mbk/107 (cit. on p. 29).

[LV25] Yunqi Li and Prashant Nalini Vasudevan. "Hardness amplification for real-valued functions". en. In: *Computational Complexity Conference (CCC)*. 2025, 2:1–2:25. DOI: 10.4230/LIPICS.CCC.2025.2 (cit. on pp. 4, 6).

[NR25] Ansh Nagda and Prasad Raghavendra. "On optimal distinguishers for Planted Clique". In: *Symposium on Foundations of Computer Science (2025), to appear* (2025) (cit. on p. 2).

[Reg09] Oded Regev. "On lattices, learning with errors, random linear codes, and cryptography". en. In: *Journal of the ACM* 56 (6 2009), pp. 1–40. DOI: 10.1145/1568318.1568324 (cit. on pp. 8, 18).

[Sha23] Ronen Shaltiel. "Is it possible to improve Yao's XOR lemma using reductions that exploit the efficiency of their oracle?" en. In: *Computational complexity* 32 (1 2023), pp. 1–47. DOI: 10.1007/s00037-023-00238-9 (cit. on pp. 5, 30, 31).

[Sho12] Victor Shoup. *A computational introduction to number theory and algebra*. Cambridge University Press, 2012. DOI: 10.1017/cbo9781139165464 (cit. on p. 22).

[STV01] Madhu Sudan, Luca Trevisan, and Salil Vadhan. "Pseudorandom Generators without the XOR Lemma". In: *Journal of Computer and System Sciences* 62 (2 2001), pp. 236–266. DOI: 10.1006/jcss.2000.1730 (cit. on pp. 2, 3).

[SV10] Ronen Shaltiel and Emanuele Viola. "Hardness Amplification Proofs Require Majority". In: *SIAM Journal on Computing* 39 (7 2010), pp. 3122–3154. DOI: 10.1137/080735096 (cit. on pp. 5, 23).

[Tre03] L Trevisan. "List-decoding using the XOR lemma". In: *Symposium on Foundations of Computer Science (FOCS)*. 2003, pp. 126–135. DOI: 10.1109/SFCS.2003.1238187 (cit. on pp. 2, 3).

[Vad12] Salil Vadhan. "Pseudorandomness". In: *Foundations and Trends in Theoretical Computer Science* 7 (1-3 2012), pp. 1–336. DOI: 10.1561/0400000010 (cit. on p. 10).

[VZ12] Salil Vadhan and Colin Jia Zheng. "Characterizing pseudoentropy and simplifying pseudorandom generator constructions". In: *Symposium on Theory of Computing (STOC)*. 2012. DOI: 10.1145/2213977.2214051 (cit. on pp. 7, 12, 18).

[VZ25] Vinod Vaikuntanathan and Or Zamir. "Improving Algorithmic Efficiency using Cryptography". In: *Symposium on Discrete Algorithms (SODA), to appear* (2025) (cit. on p. 2).

[Yao82] Andrew C Yao. "Theory and application of trapdoor functions". In: *Symposium on Foundations of Computer Science (SFCS)*. 1982, pp. 80–91. DOI: 10.1109/SFCS.1982.45 (cit. on p. 2).

[Zhe14] Jia Zheng. "A Uniform Min-Max Theorem and Characterizations of Computational Randomness". en. 2014 (cit. on pp. 7, 12, 18, 35).

A Unpredictability and Pseudo Average-Min Entropy

In this section, we revisit the characterization of pseudo-average-min entropy in terms of unpredictability given by Zheng [Zhe14], which is a complexity-theoretic analogue of Proposition 2.3.

To prove Theorem 2.11, we recall von Neumann's minimax theorem.

Theorem A.1 (von Neumann's minimax theorem)**.** *Let A, B be two finite sets and Δ_A, Δ_B be sets of all probability distributions over A and B, respectively. Let* $\mathrm{val}\colon A \times B \to \mathbb{R}$ *be a function. Then,*

$$\min_{p_B \in \Delta_B} \max_{a \in \Delta_A} \mathop{\mathbb{E}}_{\substack{b \sim p_B \\ a \sim p_A}} [\mathrm{val}(a,b)] = \max_{p_A \in \Delta_A} \min_{b \in \Delta_B} \mathop{\mathbb{E}}_{\substack{a \sim p_A \\ b \sim p_B}} [\mathrm{val}(a,b)].$$

Proof of Theorem 2.11. We prove the contrapositive statement. Suppose that for any jointly distributed random variable (X, Z) with $\widetilde{H}_\infty(Z|X) \geq \log_2 \frac{1}{1-\delta}$, there exists a size-$s'$ circuit D that can distinguish (X, Y) and (X, Z) with advantage γ, i.e.,

$$\mathbb{E}[D(X,Y)] - \mathbb{E}[D(X,Z)] \geq \gamma.$$

Here, we use the fact that, without loss of generality, we can remove the absolute value by considering the negation of D. We will construct a circuit C of size $s = O(s' \cdot |V| \log(|V|/\gamma)/\gamma^2)$ that predicts Y given X with probability at least $1 - 2\delta$, contradicting the assumption of Y.

Consider the following two-player game: Player 1 picks a distribution (X, Z) with $\widetilde{H}_\infty(Z|X) \geq \log_2 \frac{1}{1-\delta}$ and Player 2 picks a distinguisher D of size s'. The payoff is given by $\mathbb{E}[D(X,Y)] - \mathbb{E}[D(X,Z)]$. By the assumption, for any strategy of Player 1, there exists a strategy of Player 2 that can achieve a payoff of at least γ. Therefore, by von Neumann's minimax theorem, there exists a distribution $\mathcal{D}$ over size-s' circuits such that the payoff is at least γ for any strategy of Player 1, i.e.,

$$\forall (X,Z) \text{ such that } \widetilde{H}_\infty(Z|X) \geq \log_2 \frac{1}{1-\delta}, \quad \mathop{\mathbb{E}}_{\substack{D \sim \mathcal{D} \\ (X,Y)}} [D(X,Y)] - \mathop{\mathbb{E}}_{\substack{D \sim \mathcal{D} \\ (X,Z)}} [D(X,Z)] \geq \gamma. \tag{9}$$

Fix such distribution $\mathcal{D}$ and define a function $p\colon U \times V \to [0,1]$ by

$$p(x,y) = \mathop{\mathbb{E}}_{D \sim \mathcal{D}} [D(x,y)].$$

Although the distribution $\mathcal{D}$ is not necessarily efficiently sampleable, we can approximate $p(x,y)$ for given (x,y) by sampling as follows.

Claim A.2. *Let $\alpha, \beta > 0$ be any parameters. There exists a deterministic circuit $\widetilde{P}(x,y)$ of size $O\left(s' \cdot \frac{\log(|V|/\beta)}{\alpha^2}\right)$ such that*

$$\Pr_{x \sim X} \left[\forall y \in V,\ \left| p(x,y) - \widetilde{P}(x,y) \right| \leq \alpha \right] \geq 1 - \beta.$$

Proof. Consider a randomized circuit $P(x,y)$ that samples $t = O\left(\frac{\log(|V|/\beta)}{\alpha^2}\right)$ circuits $D_1, \ldots, D_t$ from $\mathcal{D}$ and outputs the average $\frac{1}{t}\sum_{i \in [t]} D_i(x,y)$. By the Hoeffding's inequality, we have

$$\forall (x,y) \in U \times V, \quad \Pr_P [|p(x,y) - P(x,y)| \leq \alpha] \geq 1 - \exp\left(-2t\alpha^2\right) \geq 1 - \frac{\beta}{|V|}.$$

By the union bound over $y \in V$, we have

$$\forall x \in U, \quad \Pr_P [\forall y \in V,\ |p(x,y) - P(x,y)| \leq \alpha] \geq 1 - \beta.$$

In particular, this yields

$$\Pr_{x \sim X, P} [\forall y \in V,\ |p(x,y) - P(x,y)| \leq \alpha] \geq 1 - \beta.$$

By averaging, there exist $D_1, \ldots, D_t$ such that $\widetilde{P}(x,y) := \frac{1}{t}\sum_{i \in [t]} D_i(x,y)$ satisfies the claim. □

Now we back to the proof of Theorem 2.11. Let $\alpha, \beta > 0$ be parameters that will be specified later. Let $\widetilde{P}(x, y)$ be the circuit guaranteed by Claim A.2. For given $x \in V$, our circuit C outputs an arbitrary $y \in V$ that maximizes $\widetilde{P}(x, y)$. Note that $\widetilde{P}(x, y)$ is a deterministic circuit and thus can be implemented as a size-$O\left(s' \cdot \frac{\log(|V|/\beta)}{\alpha^2}\right)$ circuit.

We prove the correctness of C, that is, $\Pr[C(X) = Y] \geq 1 - 2\delta$. Call an input x *good* if $\left|p(x, y) - \widetilde{P}(x, y)\right| \leq \alpha$ for all $y \in V$. Note that by Claim A.2, we have $\Pr[X \text{ is good}] \geq 1 - \beta$. Let $C^*(x)$ be the "ideal" circuit that outputs an arbitrary $y \in V$ that maximizes $p(x, y)$ for given $x \in U$. Then, for every good x, we have $p(x, C(x)) \geq p(x, C^*(x)) - \alpha$.

Suppose $\Pr[C(X) = Y] \leq 1 - 2\delta$ for contradiction. Let Z be a random variable over V jointly distributed with X defined by

$$Z|_{X=x} = \begin{cases} C(x) & \text{with probability } 1/2, \\ Y|_{X=x} & \text{with probability } 1/2. \end{cases}$$

In other words, given $X = x$, the random variable $Z|_{X=x}$ outputs $C(x)$ with probability $1/2$ and outputs a random value drawn from $Y|_{X=x}$ with probability $1/2$. Using our assumption, we can claim that Z has average-min entropy $\widetilde{H}_\infty(Z|X) \geq \log_2 \frac{1}{1-\delta}$. To see this, for every x, the probability $\Pr[Z = y|X = x]$ is maximized when $y = C(x)$, in which case we have

$$\Pr[Z = C(x)|X = x] = \frac{1}{2} + \frac{1}{2} \cdot \Pr[Y = C(x)|X = x].$$

(note that for every $y \neq C(x)$, it holds that $\Pr[Z = y|X = x] \leq 1/2$). Taking expectation over $x \sim X$, we have

$$\mathop{\mathbb{E}}_{x \sim X}\left[\Pr[Z|_{X=x} = C(x)]\right] = \frac{1}{2} + \frac{1}{2} \cdot \Pr[Y = C(X)] \leq 1 - \delta.$$

Here, we used our assumption that $\Pr[C(X) = Y] \leq 1 - 2\delta$. Thus, the average-min entropy of Z given X satisfies

$$\widetilde{H}_\infty(Z|X) = \log_2 \frac{1}{\mathbb{E}_{x \sim X}[\Pr[Z|_{X=x} = C(x)]]} \geq \log_2 \frac{1}{1 - \delta},$$

which certifies that $\widetilde{H}_\infty(Z|X) \geq \log_2 \frac{1}{1-\delta}$.

Therefore, Eq. (9) applies to this (X, Z) and yields

$$\begin{aligned} \gamma &\leq \mathbb{E}[p(X, Y)] - \mathbb{E}[p(X, Z)] \\ &= \mathbb{E}[p(X, Y)] - \left(\frac{1}{2}\mathbb{E}[p(X, C(X))] + \frac{1}{2}\mathbb{E}[p(X, Y)]\right) \\ &= \frac{1}{2}\mathbb{E}[p(X, Y)] - \frac{1}{2}\mathbb{E}[p(X, C(X))] \\ &\leq \frac{1}{2}(\mathbb{E}[p(X, C^*(X))] - \mathbb{E}[p(X, C(X))]). \end{aligned} \tag{10}$$

In the last inequality, we used the definition of C^*. For every good x, it holds that $p(x, C^*(x)) \leq p(x, C(x)) + \alpha$. Since $\Pr[X \text{ is good}] \geq 1 - \beta$, Eq. (10) yields

$$0 < \gamma \leq (10) \leq \frac{1}{2}((1 - \beta) \cdot \alpha + \beta \cdot 1) \leq \frac{\alpha + \beta}{2}.$$

Setting $\alpha = \beta = \gamma$, we obtain a contradiction. For this choice of α and β, the size of C is at most $O\left(s' \cdot \frac{\log(|V|/\beta)}{\alpha^2}\right) = O\left(s' \cdot \frac{\gamma^2}{|V|\log(|V|/\gamma)}\right)$, which completes the proof. □

We prove Proposition 2.12, the opposite direction of Theorem 2.11.

Proof of Proposition 2.12. We prove the contrapositive statement. Suppose that there exists a size-s' circuit C such that

$$\Pr[C(X) = Y] > \alpha + \beta.$$

Let D be the circuit that is given (x, y) and outputs 1 if $C(x) = y$ and 0 otherwise. Note that the size of D is at most $s' + O(\log |V|) \le s$.

Let Z be an arbitrary random variable over V jointly distributed with X such that $\widetilde{H}_\infty(Z|X) \ge \log_2 \frac{1}{\alpha}$. From Proposition 2.3, we have $\Pr[C(X) = Z] \le 2^{-\log_2 \frac{1}{\alpha}} = \alpha$. Then, we have

$$\begin{aligned}\Pr[D(X,Y) = 1] - \Pr[D(X,Z) = 1] &= \underbrace{\Pr[C(X) = Y]}_{>\alpha+\beta} - \underbrace{\Pr[C(X) = Z]}_{\le\alpha} \\ &\ge \beta.\end{aligned}$$

That is, the circuit D distinguishes (X, Y) and (X, Z) with advantage at least β, □

B Coin Problem over $\mathbb{F}_p$

In this section, we prove a query lower bound for any algorithm that distinguishes between a $(1+\varepsilon)/p$-biased distribution and a $1/p$-biased distribution over $\mathbb{F}_p^N$. Fix a parameter $N \in \mathbb{N}$. Recall that for a parameter $\beta \in [0, 1]$, Θ_β denotes the distribution over $\mathbb{F}_p^N$ obtained by uniformly permuting a multiset containing βN copies of 0 and an equal number of copies of each element of $\{1, \dots, p-1\}$ (in Section 5, we set $N = 2^m$).

Lemma B.1. *Let $A^{(\cdot)}$ be an oracle algorithm that satisfies the following:*

$$\begin{aligned}\Pr_{Z\sim\Theta_{(1+\varepsilon)/p}}[A^Z = 1] &\ge 1 - \delta, \\ \Pr_{Z'\sim\Theta_{1/p}}[A^{Z'} = 1] &\le 0.51.\end{aligned}$$

Then, there exists a universal constant $c > 0$ such that the number of adaptive queries made by A is at least $q \ge c \cdot \min\left\{N/p, p\log(1/\delta)/\varepsilon^2\right\}$.

Proof. We borrow the notation of information theory from Section 5.3. For two random variables X and Y, the χ^2-*divergence* of X and Y is defined by $\chi^2(X \parallel Y) = \sum_{y\in\mathrm{supp}(Y)} \frac{(\Pr[X=x]-\Pr[Y=y])^2}{\Pr[Y=y]}$. It is known that $\mathrm{KL}(X \parallel Y) \le \chi^2(X \parallel Y)$.

Let $0 < c \le 1/2$ be a sufficiently small constant and assume for contradiction that $q \le c \cdot \min\left\{N/p, p\log(1/\delta)/\varepsilon^2\right\}$. Let W and W' be the random vectors that A receives from oracles Z and Z', respectively. Let $\mathrm{Ber}(\gamma)$ denote the Bernoulli distribution with parameter γ. By the

data-processing inequality, we have

$$\begin{aligned}\mathrm{KL}\big(W' \,\big\|\, W\big) &\geq \mathrm{KL}\Big(A^{W'} \,\Big\|\, A^{W}\Big)\\ &\geq \mathrm{KL}(\mathrm{Ber}(0.51) \,\|\, \mathrm{Ber}(1-\delta))\\ &= 0.51 \log \frac{0.51}{1-\delta} + 0.49 \log \frac{0.49}{\delta}\\ &= \Omega(\log(1/\delta)). \end{aligned} \tag{11}$$

On the other hand, by the chain rule of KL divergence, we have

$$\begin{aligned}\mathrm{KL}\big(W' \,\big\|\, W\big) &= \sum_{i=1}^{q} \mathrm{KL}\big(W'_{\leq i} | W'_{\leq i-1} \,\big\|\, W_{\leq i} | W_{\leq i-1}\big)\\ &= \sum_{i=1}^{q} \mathrm{KL}\big(W'_{i} | W'_{\leq i-1} \,\big\|\, W_{i} | W_{\leq i-1}\big)\\ &= \sum_{i=1}^{q} \mathop{\mathbb{E}}_{w' \sim W'_{\leq i-1}} \Big[\mathrm{KL}\Big(W'_i|_{W'_{\leq i-1}=w'} \,\Big\|\, W_i|_{W_{\leq i-1}=w'}\Big)\Big]\\ &\leq \sum_{i=1}^{q} \mathop{\mathbb{E}}_{w' \sim W'_{\leq i-1}} \Big[\chi^2\Big(W'_i|_{W'_{\leq i-1}=w'} \,\Big\|\, W_i|_{W_{\leq i-1}=w'}\Big)\Big] \end{aligned} \tag{12}$$

In the last inequality, we use the fact that $\mathrm{KL}(X \,\|\, Y) \leq \chi^2(X \,\|\, Y)$.

Now, we bound the χ^2-divergence of $W'_i|_{W'_{\leq i-1}=w'}$ and $W_i|_{W_{\leq i-1}=w'}$ from above for any fixed w'. Recall that $W'_{\leq i}$ is the string that the oracle algorithm $A^{Z'}$ receives from the first i queries. Let $j_1, \ldots, j_i$ be the indices of the queries that $A^{Z'}$ makes, i.e., $W'_1 = Z'_{j_1}, \ldots, W'_i = Z'_{j_i}$. Although $j_1, \ldots, j_i$ are random since $A^{(\cdot)}$ makes adaptive queries, conditioned on $W'_{\leq i-1} = w'$, the i-th query location j_i is fixed. Since the original string Z' is obtained by a random shuffle, the distribution of W'_{j_i} is obtained by choosing random index $j \sim [N] \setminus \{j_1, \ldots, j_{i-1}\}$ and setting $W'_{j_i} = Z'_j$.

For $a \in \mathbb{F}_p$, let $N_a(w)$ be the number of occurrences of a in w. Then, we have

$$\begin{aligned}\Pr\left[W'_i = a \,\middle|\, W'_{\leq i-1} = w'\right] &= \frac{\text{number of a's in } W' - \text{number of a's in } w'}{N-i+1},\\ &= \frac{N/p - N_a(w')}{N-i+1}. \end{aligned}$$

Similarly, we have

$$\Pr\left[W_i = a \,\middle|\, W_{\leq i-1} = w'\right] = \begin{cases} \frac{(1+\varepsilon)N/p - N_0(w')}{N-i+1}, & \text{if } a = 0,\\ \frac{(1-\varepsilon/(p-1))N/p - N_a(w')}{N-i+1}, & \text{otherwise.} \end{cases}$$

In particular, since $q \leq c \cdot \frac{N}{p}$, if we take $c \leq 1/2$, we have

$$\Pr\left[W'_i = a \,\middle|\, W'_{\leq i-1} = w'\right] \geq \frac{N/p - q}{N} \geq \frac{1}{2p}$$

and for $a = 0$,

$$\begin{aligned}\Big(\Pr\Big[W'_i|_{W'_{\le i-1}=w'} = 0\Big] - \Pr\big[W_i|_{W_{\le i-1}=w} = 0\big]\Big)^2 &\le \left(\frac{\varepsilon N/p}{N-i+1}\right)^2 \\ &\le \left(\frac{\varepsilon}{(1-c)p}\right)^2 \\ &\le \left(\frac{2\varepsilon}{p}\right)^2\end{aligned}$$

and for $a \neq 0$,

$$\begin{aligned}\Big(\Pr\Big[W'_i|_{W'_{\le i-1}=w'} = a\Big] - \Pr\big[W_i|_{W_{\le i-1}=w} = a\big]\Big)^2 &\le \left(\frac{\varepsilon N/(p(p-1))}{N-i+1}\right)^2 \\ &\le \left(\frac{\varepsilon}{(1-c)p(p-1)}\right)^2 \\ &\le \left(\frac{4\varepsilon}{p^2}\right)^2.\end{aligned}$$

Therefore, we can bound the χ^2-divergence as follows:

$$\begin{aligned}\chi^2\Big(W'_i|_{W'_{\le i-1}=w'} \,\Big\|\, W_i|_{W_{\le i-1}=w'}\Big) &\le 2p \cdot \sum_{a\in\mathbb{F}_p}\Big(\Pr\Big[W'_i|_{W'_{\le i-1}=w'} = a\Big] - \Pr\big[W_i|_{W_{\le i-1}=w} = a\big]\Big)^2 \\ &\le 2p \cdot \left(\frac{2\varepsilon}{p}\right)^2 + 2p \cdot \left(\frac{4\varepsilon}{p^2}\right)^2 \\ &\le \frac{8\varepsilon^2}{p} + \frac{32\varepsilon^2}{p^3} \\ &\le \frac{40\varepsilon^2}{p}. \qquad (13)\end{aligned}$$

Combining Eqs. (12) and (13), we have

$$\Omega(\log(1/\delta)) \le \mathrm{KL}\big(W' \,\big\|\, W\big) \le 40q \cdot \frac{\varepsilon^2}{p}.$$

This yields a contradiction since we have assumed that $q \le c \cdot p\log(1/\delta)/\varepsilon^2$ for a sufficiently small constant c. □

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