Tighter Bounds for the Oblivious Bit-Fixing Inner Product Extractor on Biased Seeds

Jack Doerner
University of Virginia, Charlottesville, VA, USA

Lawrence Roy
Aarhus University, Denmark

Abstract

The *Inner Product Extractor* (IPE) of Impagliazzo, Levin, and Luby (STOC'89) takes a seed $h \in \mathbb{F}^\gamma$ and a source $x \in \{0,1\}^\gamma$ for some $\gamma \in \mathbb{N}$ and produces $\langle h, x \rangle$ with error $\varepsilon = \mathsf{SD}((\langle \mathcal{H}, \mathcal{X} \rangle, \mathcal{H}), (\mathcal{Y}, \mathcal{H}))$ such that

$$\varepsilon \leq \frac{1}{2}\sqrt{|\mathbb{F}|^\gamma / 2^{H_\infty(\mathcal{H})}}\ \sqrt{|\mathbb{F}| / 2^{H_\infty(\mathcal{X})}}$$

where $\mathcal{Y}$ is the uniform distribution over $\mathbb{F}$, and $\mathcal{H}$ and $\mathcal{X}$ are the independent but possibly *non-uniform* distributions from which h and x are drawn, respectively. In other words, the IPE's error grows with the square root of seed bias, at most. This square root arises because prior works bound the squared error using the 2-universality of the IPE. The analysis requires an even power of the error, and the IPE is not 4-universal.

Motivated by applications to multiparty computation, we revisit the problem of the IPE with biased seeds and prove far tighter bounds on the influence of seed bias by bypassing universal hashing. We first prove an *Elevated* General Leftover Hash Lemma, which yields an n^{th} root bound for functions that are *almost* n-universal. Bounding number of inputs on which the IPE is not 4-universal yields $\varepsilon = \mathsf{SD}((\langle \mathcal{H}, \mathcal{W} \rangle, \mathcal{H}), (\mathcal{Y}, \mathcal{H}))$ where

$$\varepsilon \lesssim \frac{2.1}{2}\left(|\mathbb{F}|^\gamma / 2^{H_\infty(\mathcal{H})}\right)^{\frac{1}{4}} \sqrt{|\mathbb{F}| / 2^{H_\infty(\mathcal{W})}}$$

for any *oblivious bit-fixing* source $\mathcal{W}$ with $2^{0.585 H_\infty(\mathcal{W})} \leq |\mathbb{F}| \leq 2^{H_\infty(\mathcal{W})}$. Next, we use matroid theory to *directly* analyze the n-way multicollision probability of the IPE, yielding an asymptotic bound for any even n. For $n \geq 4$, $0 < \epsilon \leq 0.83/(n-2)$, and $|\mathbb{F}| \leq 2^{(1-\epsilon)\cdot H_\infty(\mathcal{W})}$, as $|\mathbb{F}| \to \infty$,

$$\varepsilon \leq \frac{(n-1)}{2}\left(|\mathbb{F}|^\gamma / 2^{H_\infty(\mathcal{H})}\right)^{\frac{1}{n}} \sqrt{2^{-\epsilon \cdot H_\infty(\mathcal{W})}}\ (1 + o(1)).$$

Computing a *concrete* version of this bound requires time exponential in n. We compute concrete $\{4, 6, 8\}^{\text{th}}$-root bounds and demonstrate that no one choice of n is optimal. Finally, we introduce a new class of *seed-adaptive* oblivious bit-fixing sources, extend our results to such sources, and use this extension to fix a bug that we identify in the proof of the oblivious linear evaluation protocol of Doerner et al. (SP'24).

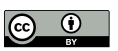
2012 ACM Subject Classification Security and privacy → Cryptography; Security and privacy → Information-theoretic techniques

Keywords and phrases Leftover hash lemma, Inner product extractor, Randomness extraction, Oblivious linear evaluation

Digital Object Identifier 10.4230/LIPIcs.ITC.2026.1

Related Version *Full Version*: `https://eprint.iacr.org/2026/654` [12]

Funding *Jack Doerner*: The work of Jack Doerner was supported in part by the Brown University Data Science Institute.
Lawrence Roy: The work of Lawrence Roy was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement number 101124977 (DECRYPSIS).



Acknowledgements The authors thank Daniel Wichs for a helpful discussion, and abhi shelat and Ethan Rathbun for the use of their computers. Part of this work was conducted while Lawrence Roy was visiting the Simons Institute for the Theory of Computing.

1 Introduction

A *randomness extractor* (often simply an *extractor*) is a function that produces a nearly uniformly-distributed output using a weak source of entropy. As such, they are a foundational primitive with a wide variety of applications in cryptography and connections to many other important concepts in theoretical computer science. *Seeded* extractors, first introduced by Nisan and Zuckerman [22], take a uniformly sampled seed in addition to their weak entropy source. *Strong* seeded extractors guarantee that the statistical distance between the uniform distribution and the distribution of the extractor's output (usually known as the *error* of the extractor) be bounded even given knowledge of the seed, which ensures that the seed can be reused, while the source is sampled afresh at each invocation. Intuitively, a seeded extractor can be thought of as (re)using a small amount of high-quality randomness to turn a large amount of low-quality randomness into an output that is close to uniform.

Perhaps the best known specific randomness extractor construction is the *Inner Product Extractor* (IPE) of Chor and Goldreich [2], which takes as its seed a vector h of γ elements of some field $\mathbb{F}$, and as its source a γ-bit vector x of sufficient min-entropy, and produces an output $y = \langle h, x \rangle$. It is a classic exercise in many introductory cryptography courses to prove that the error of the inner product extractor is bounded by the following equation:

$$\mathsf{SD}((\langle \mathcal{H}, \mathcal{X} \rangle, \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \leq \sqrt{|\mathbb{F}|/2^{H_\infty(\mathcal{X})}} \tag{1}$$

where $\mathcal{X}$ is the source distribution, $\mathcal{Y}$ is the uniform distribution over the output domain $\mathbb{F}$, and $\mathcal{H}$ is the distribution from which h is drawn (i.e. the uniform distributions over $\mathbb{F}^\gamma$), and where $H_\infty(\cdot)$ is the min-entropy function and $\mathsf{SD}(\cdot, \cdot)$ is statistical distance. This proof typically proceeds by first observing that any two distinct inputs differ in at least one bit, and since the seed element corresponding to this bit is uniformly sampled and contributes to the output for only one of the two inputs, the probability of a collision on any two distinct inputs (over the choice of the seed) is $1/|\mathbb{F}|$. Functions with this collision probability on distinct inputs are said to be *2-Universal Hash Functions*, and we can apply the celebrated *Leftover Hash Lemma* (LHL) of Impagliazzo, Levin, and Luby [17] to show that the output of such a function on a uniform seed and a source from $\mathcal{X}$ has a statistical distance bounded as per Equation (1).

It is not always the case that a high-quality seed is available in practice, which motivates the study of *multi-source* extractors that take two or more weak entropy sources. If a multi-source extractor attains its statistical guarantee even under the disclosure of one or more of its sources, it is said to be *strong*.[1] Generalizing the LHL to account for biased seeds yields the implication that the IPE is a strong seed-biased extractor with

$$\mathsf{SD}((\langle \mathcal{G}, \mathcal{X} \rangle, \mathcal{G}), (\mathcal{Y}, \mathcal{G})) \leq \sqrt{|\mathbb{F}|^{\gamma+1}/2^{H_\infty(\mathcal{G})+H_\infty(\mathcal{X})}}$$

for any *arbitrarily biased* seed distribution $\mathcal{G}$ over $|\mathbb{F}|^\gamma$ that is independent of $\mathcal{X}$ [25, 3, 20].

[1] Hereinafter, we often refer to strong multi-source extractors as *seed-biased* extractors.

Why Study the IPE?

In the 40 years since the IPE was first introduced, extractors have advanced a great deal. In particular, a recent work of Doran and Fridman [13] proposes an extractor for oblivious bit-fixing sources (the same variety we consider) that outputs almost all of the entropy in its source, requires very little source entropy (relative to the total number of source bits), and achieves low error, all *without* the use of a seed, which eliminates the possibility that the seed can be biased, or that the bias of the source can depend upon the seed. Yet we claim that the IPE retains an important advantage that makes it worthy of continuing study: it can be computed by a depth-1 circuit over $\mathbb{F}$, and it is linear in both the source and seed. These properties make it ideal in the context of *multiparty computation* (MPC) based upon linear secret sharing over $\mathbb{F}$, where round count usually depends upon circuit depth and the IPE fits natively into the computation paradigm. Because the IPE also achieves negligible error, it is suitable for use in MPC protocols that compute cryptographic functions. To the best of our knowledge, the only other extractors that combine these properties are also based upon universal hashing.[2]

Why Focus on Oblivious Bit-Fixing Sources?

Each sample from an *oblivious bit-fixing* (OBF) distribution $\mathcal{W}$ over $\mathbb{1}^\gamma$ contains $H_\infty(\mathcal{W})$ uniform bits (at fixed locations) and $\gamma - H_\infty(\mathcal{W})$ bits that are constant. Oblivious bit-fixing sources naturally model a common situation in MPC protocol analysis: selective failure attacks (by which the adversary conditions protocol aborts on the values of specific secrets) can sometimes effectively fix secret random bits. This pattern appears in particular in a number of works that construct Oblivious Linear Evaluation (OLE) in the Oblivious Transfer (OT) hybrid model [19, 8, 9, 16, 10, 5]. This work focuses exclusively on oblivious bit-fixing sources. We do not know how to extend our results to sources with arbitrary bias,[3] and we consider such an extension to be an interesting open problem for follow-up work.

The Impact of Public Seeds

In the MPC context, it would be ideal to sample and fix a seed as a public *common random string*, and thereafter use the IPE as a quasi-seedless extractor. This would eliminate the need to use expensive seed-sampling protocols later on, which are likely to incur additional rounds in the best case, and may not be possible at all in the worst case [4]. Fixing the seed in advance introduces a new issue, however: bias in the source can now depend upon the value of the seed! This violates the source/seed independence upon which the generalized LHL (and the analyses of most two-source extractors) relies. In order to use the IPE in this way, we must bound the effects of source/seed dependence upon the IPE's error. We show in this work that in the context of OBF sources and the IPE, this kind of seed-dependent source bias can be modeled as (additional) bias in the seed alone: thus we can focus our attention on minimizing the impact of seed bias on the IPE's error.

[2] For example, universal hashing from polynomial evaluation achieves these properties *if* online secure computation of the powers of the seed can be avoided. See e.g. Degabriele et al. [6] and the citations therein.

[3] See Section 3.2 for a discussion of the barriers that prevent them from being extended in the same way as the LHL.

Why can We Hope for Improvement?

Both the classic LHL and the General LHL upper-bound the statistical distance of the IPE's output distribution with a square-root. This is because, within the proof of both lemmas, the *square* of the statistical distance is transformed into a statement about 2-way collision probability, and this is bounded using the 2-universality of the IPE. The choice of the exponent used is motivated by specific constraints: the n^{th} power of statistical distance can be transformed into a statement about n-way collision probability, and a bound on the latter would imply an n^{th}-root bound on statistical distance. Moreover, due to the use of Jensen's lemma, the bias term $|\mathbb{F}|^\gamma/2^{H_\infty(\mathcal{G})}$ appears *without* an exponent, which means that it appears under an n^{th}-root in the final bound, dramatically reducing its impact (and improving the performance of the extractor under seed bias). However, Jensen's inequality only applies to convex functions, which implies that we can only choose even values of n. The IPE is *not* 4-universal, and thus the classic proof breaks down for any $n > 2$. This leaves a pair of open questions:

> Can we bound the n-way collision probability of the IPE for $n \geq 4$?
>
> Is the impact of seed bias on the error of the IPE bounded by the n^{th} root of the bias, for any $n > 2$?

In this work, we answer both questions in the affirmative. We demonstrate that the square-root GLHL is not tight for the IPE on bit-fixing sources by introducing concrete n^{th}-root bounds for $n \in \{4, 6, 8\}$, which we derive via multiple techniques, and an asymptotic bound for any fixed even n as $H_\infty(\mathcal{W}), |\mathbb{F}| \to \infty$.

1.1 Our Contributions

1. We introduce a new class of oblivious bit-fixing sources in which the locations of the fixed bits are chosen *adaptively* by an adversary with knowledge of the seed, and show that the error due to this kind of source bias is upper-bounded by the error due to a certain magnitude of seed bias alone. This notation is similar to that of seedless extraction, in that it *non-constructively* implies a seedless extractor by taking the best seed (the seed with the least worst-case bias).
2. We bound the number of inputs on which the inner product extractor is not 4-universal, and use this to produce 4^{th}-root bound for the error of the IPE via a new generalization of the LHL that does not require the extractor to be a perfectly universal hash. This result is stated in full as Theorem 32 in Section 5, but its simplest and clearest statement is via the following:

   ▶ **Corollary 1** (of Theorem 32). *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 13) from $W = \mathbb{1}^\gamma$ to $Y = \mathbb{F}_m$. If $\gamma \geq \xi \geq \log_2 m \geq 0.585\xi$, $m > 3$, $2^\xi \gg 4$, and $\eta > 0$, then for any arbitrary seed distribution $\mathcal{H}$ on H such that $H_\infty(\mathcal{H}) \geq \eta$ and any oblivious bit-fixing (Definition 8) source distribution $\mathcal{W}$ on W with $H_\infty(\mathcal{W}) \geq \xi$, we have*

$$\mathsf{SD}((\langle \mathcal{H}, \mathcal{W} \rangle, \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \lesssim \frac{2.1}{2} \sqrt[4]{\frac{m^\gamma}{2^\eta}} \sqrt{\frac{m}{2^\xi}}$$

   where $\mathcal{Y}$ is the uniform distribution on Y. For any seed-adaptive *oblivious bit-fixing (Definition 10) distribution $\mathcal{W}'$ on W with $H_\infty(\mathcal{W}') \geq \xi$,*

$$\mathsf{SD}((\langle \mathcal{H}, \mathcal{W}' \rangle, \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \lesssim \frac{2.1}{2} \sqrt[4]{\frac{m^\gamma \binom{\gamma}{\xi}}{2^\eta}} \sqrt{\frac{m}{2^\xi}}$$

3. Taking a different approach, we analyze the overall i-way collision probability of the inner product extractor using techniques from *matroid* theory. Specifically, we describe an equation with exponentially many terms that exactly computes the probability we need, and we show how this equation yields a concrete n^{th}-root bound on the error of the IPE for any n. Using a $2^{O(n^2)}$ algorithm, we compute concrete n^{th}-root bounds for the IPE at $n \in \{4, 6, 8\}$. The 4^{th}-root bound we derive via this method is actually both tighter and simpler to express than the one we derived manually.[4] On the other hand, the 8^{th}-root bound we derive via this method has 89 terms under the radical and takes nearly one page to print. Using these concrete bounds, we show that the optimal n depends on $|\mathbb{F}|$, γ, and $H_\infty(\mathcal{W})$. This result is reported in the full version of this paper [12].
4. In the full version of this paper [12], we asymptotically bound our concrete n^{th}-root error bound, as both the output length and the min-entropy of the source go to ∞. This result is stated in full as Theorem 38 in Section 6, but its simplest and clearest statement is via the following:

   ▶ **Corollary 2** (of Theorem 38). *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 13) from $W = \mathbb{1}^\gamma$ to $Y = \mathbb{F}_m$. If $\gamma \geq \xi > 0$, $\eta > 0$, and $m \leq 2^{(1-\epsilon)\cdot\xi}$ for some $0 < \epsilon \leq 0.83/(n-2)$, then for any even constant integer $n \geq 4$, any arbitrary seed distribution $\mathcal{H}$ on H such that $H_\infty(\mathcal{H}) \geq \eta$, and any oblivious bit-fixing (Definition 8) source distribution $\mathcal{W}$ on W with $H_\infty(\mathcal{W}) \geq \xi$, as $m \to \infty$ we have*

$$\mathsf{SD}((\langle \mathcal{H}, \mathcal{W} \rangle, \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \leq \frac{n-1}{2} \sqrt[n]{\frac{m^\gamma}{2^\eta}} \sqrt{2^{-\epsilon\cdot\xi}} \, (1 + o(1))$$

   where $\mathcal{Y}$ is the uniform distribution on Y. For any seed-adaptive *oblivious bit-fixing (Definition 10) distribution $\mathcal{W}'$ on W with $H_\infty(\mathcal{W}') \geq \xi$, as $m \to \infty$ we have*

$$\mathsf{SD}((\langle \mathcal{H}, \mathcal{W}' \rangle, \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \leq \frac{n-1}{2} \sqrt[n]{\frac{m^\gamma \binom{\gamma}{\xi}}{2^\eta}} \sqrt{2^{-\epsilon\cdot\xi}} \, (1 + o(1))$$

5. We describe a bug in the security proof of a widely-deployed OLE protocol [8, 9, 10] that stems from treating a particular IPE source as if it were oblivious bit-fixing, when in actuality it is a *seed-adaptive* oblivious bit-fixing source. We compute the exact number of additional source bits required in order for this protocol to attain the security level originally claimed, under each of the bounds we have introduced. In some *deployed* parameter regimes, our bounds improve over the GLHL of Lee et al. [20] by a factor of nearly 150, in terms of the number of extra source bits.

1.2 Abbreviated Technical Overview

We set the stage in Section 3 by showing that reasoning about the performance of the IPE on seed-adaptive OBF sources and *arbitrarily* biased seeds can be reduced to the problem of reasoning about the performance of the IPE on uniform sources and seeds drawn uniformly from an arbitrary subset of the overall seed set. This simplifies the problem considerably.

As we previously discussed, within the proof of the original LHL, the *square* of the statistical distance is transformed into a statement about 2-way collision probability, and this is bounded using the 2-universality of the IPE. Most of the proof is compatible with any

[4] The value of our EGLHL-derived 4^{th}-root bound comes from the fact that we prove it without using a computer to enumerate exponentially-large sets.

even power n of statistical distance (which will be transformed into a statement about n-way collision probability), but the IPE is not n-universal for any even $n > 2$. We develop two distinct means to work around this limitation.

First, we observe that even if the IPE is not n-universal for some even n, there may be only a negligible number of inputs that witness this fact, and thus the IPE behaves as though it were n-universal with overwhelming probability over the coins of the source. Motivated by this observation, we develop a general n^{th}-root *Elevated General Hash Lemma* (EGLHL) for any hash function on uniform sources and seeds drawn uniformly from an arbitrary subset of the total seed space. Our lemma does not require n-universality, but instead requires a bound on j-way *distinct-input collision probability* (CPDI) of the hash function, for every $j \in [n]$. This result is restated and proven in Section 4.1:

▶ **Lemma 3** (Elevated General Leftover Hash Lemma). *Let H be a family of hash functions from X to Y. Let $G \subseteq H$ and let $\mathcal{G}$, $\mathcal{X}$, and $\mathcal{Y}$ denote the uniform distributions over G, X, and Y respectively. Let* CPDI *be the* distinct-input collision probability *of H (see Definition 16), and let $\varepsilon = \mathsf{SD}((\mathcal{G}(\mathcal{X}), \mathcal{G}), (\mathcal{Y}, \mathcal{G}))$. If $n \in \mathbb{N}^+$ is even, then*

$$\varepsilon \leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left(1 + \sum_{j=1}^{n} \mathsf{CPDI}(j) \sum_{i=j}^{n} \binom{n}{i} \left\{ {i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{i-1}}{(-1)^{n-i} |X|^{i}}\right)}$$

For the IPE in particular, we can bound the j-way distinct-input collision probability using the *exceptional input fraction* (EIF) for sets of j inputs; that is, the fraction of all sets of j inputs that are witnesses to the fact that the IPE is not j-universal. In Section 5, we bound the EIF for every $j \leq 4$ by writing each possible set of inputs as the columns of a matrix, and then counting the possibilities for each row. We exhaustively categorize combinations of rows that imply j-universality; in particular, we need sufficiently-many linearly independent rows, and the fraction of possibilities that have this property is enlarged by the requirement that columns be distinct. Categorizing possible row combinations yields a bound on EIF, and by implication the 4^{th} root bound reported in Corollary 1. However, we are unable to devise a similar bound on EIF for sets of $j > 4$ inputs. We thus introduce a second approach that allows us to *exactly* compute n-way collision probability for $n > 4$.

Our second approach essentially generalizes the first, using tools from combinatorics. In particular, there is a matroid corresponding to the set of all possible $\{0, 1\}$-matrices with n distinct columns; this is the linear matroid of all vertices in an $(n-1)$-cube. Using this matroid, we can sample rows *independently*, and compute the probability that the corresponding matrix has a particular rank (which implies a particular collision probability). Properties related to sampling randomly from the matroid and computing ranks are all described by a polynomial invariant of the matroid called the Tutte polynomial (and the closely-related *coboundary* polynomial that can be derived from it), and we can compute the exact collision probability of the IPE by evaluating a particular derivative of this polynomial at a constant. Using our exact formula for collision probability, we give an n^{th}-root bound for the error of the IPE on uniform sources and seeds drawn uniformly from an arbitrary subset of the total seed space. This result is restated and proven in the full version [12]:

▶ **Theorem 4.** *Let $H = \mathbb{F}_m^{\xi}$ be an IPE family (Definition 13) from $X = \mathbb{1}^{\xi}$ to $Y = \mathbb{F}_m$. Let $G \subseteq H$ and let $\mathcal{G}$, $\mathcal{X}$, and $\mathcal{Y}$ denote the uniform distributions over G, X, and Y respectively. Let $\overline{\chi}_{M_{i-1}^{\square}}$ be the coboundary polynomial of the cube matroid (Definition 36), and let $\varepsilon = \mathsf{SD}((\langle \mathcal{G}, \mathcal{X} \rangle, \mathcal{G}), (\mathcal{Y}, \mathcal{G}))$. If $n \in \mathbb{N}^+$ is even, then*

$$\varepsilon \leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left(1 + \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{n-i}}{2^{\xi \cdot (i-1)}} \left(\frac{d}{d\alpha}\right)^{\xi} \overline{\chi}_{M_{i-1}^{\square}}(m, e^{\alpha})\Big|_{\alpha=0}\right)}$$

Unfortunately, the most efficient algorithm that we are aware of for computing Tutte polynomials runs in $2^{O(n^2)}$ time. For small $n \leq 8$ this is feasible and we do it. The resulting polynomials and the error bounds that they imply include dozens of terms, but evaluating them on specific sets of parameters is efficient, once they are computed. We have included these error bounds in the full version [12]. Towards understanding the behavior of the IPE's error when n is larger, we perform an asymptotic analysis of the Tutte polynomial of the cube matroid. We begin by decomposing it into an exponential number of terms, each of which relates to a specific *flat* of the matroid,[5] and then we observe that the largest terms asymptotically are those corresponding flats that have the largest size relative to their rank. We classify the terms according to this property, with one class of maximal terms per rank, plus one class of non-maximal terms. We then bound each class individually; although our bounds are relatively course, they give us sufficient information to identify and concretely upper-bound the asymptotically dominant term in the overall error, and to asymptotically upper-bound the other terms. Thus we arrive at Corollary 2.

1.3 Applications to Malicious OLE-from-OT Protocols via Seed-Adaptive Oblivious Bit Fixing

Oblivious Transfer (OT) [23, 14] is a two-party protocol that allows a *receiver* to select one of two messages supplied by a *sender*. *Oblivious Linear Evaluation* (OLE) [21] is a natural algebraic generalization of OT, in which the sender supplies the coefficients of a line, and the receiver learns the evaluation of this line over some finite ring on a chosen secret point. We can view OT as producing secret shares of the product of boolean secrets, and OLE as producing secret shares of the product of secrets in an arbitary finite ring. Both OT and OLE have been the subject of intense study, and both have served as building blocks in a very large variety of applications.

It is natural to ask whether OLE can be derived from OT, and if so, how efficiently? One of the earliest and best known constructions answering this question was Gilboa's semi-honest multiplication protocol [15], which can perform a single OLE over $\mathbb{F}_m$ using $\mu = \log_2(m)$ OT instances. In the *malicious* setting, generic approaches can be applied to enhance the security of Gilboa's protocol [18], but these attain efficiency only under amortization. The question of an explicit, purely OT-based maliciously secure OLE protocol with concrete efficiency in the non-amortized setting remained open until it was implicitly closed by Keller, Orsini, and Scholl [19], whose protocol required a few additional rounds and a communication overhead factor of roughly 6 relative to Gilboa's baseline. Later, purely OT-based maliciously secure OLE protocols with overhead factors between 2 and 4 and no added rounds (in the random oracle model) were given by Doerner et al. [8, 9, 10]. The last of these remains the state of the art. Haitner et al. [16] proposed a protocol with even lower overhead, but it realizes a *relaxed* notion of OLE or else relies upon on the DDH assumption. Later, Doerner et al. [7] took a different approach based upon the Chinese remainder theorem which yields a purely OT-based OLE protocol with communication cost that is *subquadratic* in the modulus length. Their protocol is the first to outperform Gilboa's semi-honest baseline in the non-amortized setting, but in order to achieve malicious security it requires at least five rounds.

[5] A flat can be thought of as a collection of rows to which no additional rows can be added without increasing the rank, i.e., a linear subspace.

Fixing A Bug in the DKLs OLE Protocols

The DKLs OLE protocols [8, 9, 10] work by securely computing the IPE using OT instances. Specifically, OT is used to compute a sharing of $\alpha \cdot \langle h, x \rangle$, where $x \in \mathbb{1}^\gamma$ is the extractor source, supplied by the receiver, $\alpha \in \mathbb{F}_m$ is the sender's secret input to the multiplication, and $h \in \mathbb{F}_m^\gamma$ is uniformly random *public* extractor seed. This secret sharing is then converted into a secret sharing of $\alpha \cdot \beta$ for an arbitrary secret input $\beta \in \mathbb{F}_m$ supplied by the receiver.

Part of the security analysis of Doerner et al. deals with selective failure attacks, by which a malicious sender can cause the protocol to detectably abort if it guesses a particular subset of the bits of x correctly. If such an attack occurs, and the protocol *does not abort*, then the corrupt sender learns those bits, and they thus become fixed. The security proof relies upon the extractor's output being indistinguishable from uniform, conditioned on the fixing of these source bits. Specifically, Doerner et al. *exclude* the seed elements corresponding to the fixed bits from the analysis, and argue about the distribution of the inner product of the remaining source bits and seed elements.

It is exactly at this point that there is a bug in the versions of Doerner et al.'s proof dated December 2023 and earlier. Specifically, in their proof of Claim 6.10, they analyze the extractor's statistical distance from uniform using the classic LHL [17], which requires a uniform seed. However, the adversary can adaptively choose which bits to fix (and thus which seed elements to exclude) with full knowledge of the seed, which was sampled and published in advance. The non-excluded seed elements are therefore adversarially biased, and the classic LHL does not apply.

We can fix the bug in Doerner et al. by characterizing the seed bias induced by seed-adaptive oblivious bit fixing, and then analyzing the extractor under this bias. In Section 3 we prove that if c out of γ source bits are fixed, then the min-entropy of the seed is reduced by $\log_2(\binom{\gamma}{c})$ and the source min-entropy is reduced by c. This leads finally to a key question:

> How many *extra* source bits are needed to attain a particular level of security under seed-adaptive oblivious bit-fixing, relative to non-adaptive oblivious bit-fixing.

In fixing the bug, we incur a performance penalty that hinges on the answer to this question. Specifically, we must perform a number of additional OT instances proportionate to the number of extra source bits, which means that the cost of the patched protocol will grow linearly with this number. We answer the question and report the number of extras concretely for a variety of parameter regimes in the full version [12]. Doerner et al. [10] suggest to deploy their protocol with $|m| = 256$ and to set the statistical parameter $s = 80$. In this parameter regime, the techniques we introduce in this paper reduce the number of extra OT instances that we must add by a factor of 11, from 271 to 22. The improvement is even greater if $|m| = 384$ is used (as might be required when working with pairing curves [11], for example): in this case the number of extra OTs reduces from 295 to 2!

2 Preliminaries

Notation

In this document, we use $=$ for equality, $\equiv$ for congruence, and $\leftarrow$ for sampling. We typically use capital italic letters (e.g. X) to denote sets, capital caligraphic letters (e.g. $\mathcal{X}$) to denote distributions, and lower case italic letters (e.g. x) to denote matrices, vectors, or scalars. Set literals are denoted with $\{\ldots\}$, ordered tuple literals with $(\ldots)$, and $[a, b]$ generates the ordered tuple of integers from a to b inclusive (or from 1 to b if a is omitted). $|X|$ denotes

the size of X. Function names are typeset either in roman font for standard mathematical functions, or sans-serif font for functions that we define. Some mathematical functions (such as rank and dim) use a subscript to disambiguate the field or matroid relative to which they are defined. $\mathbb{N}^+$ denotes the set of all natural numbers greater than zero, $\mathbb{F}_m$ a prime field with order m, and $\mathbb{1} = \{0, 1\}$. The variable e always refers to Euler's number. $H_\infty(\cdot)$ denotes the min-entropy function and $\mathsf{SD}(\cdot, \cdot)$ denotes statistical distance. For additional notation related to matroids, we direct the reader to the full version [12].

2.1 Useful Combinatorial Notions

▶ **Definition 5** (The Falling Factorial Function)**.**

$$x^{\underline{n}} = \overbrace{x(x-1)\dots(x-n+1)}^{n \text{ factors}}$$

▶ **Definition 6** (Stirling Numbers of the Second Kind)**.** *The second-kind Stirling number* $\left\{ {i \atop j} \right\}$ *counts the number of ways to partition* i *elements into exactly* j *subsets. It is given by the following sum:*

$$\left\{ {i \atop j} \right\} = \frac{1}{j!} \sum_{k=0}^{j} (-1)^{j-k} \binom{j}{k} k^i$$

2.2 Randomness Extractors

A strong extractor is an efficient function which takes a random source and a random seed, and produces uniform-looking bits, even to a distinguisher that can see the seed.[6] The strongest notion we will consider takes input from an arbitrary distribution of sufficient min-entropy.

▶ **Definition 7** ((Strong) Randomness Extractors)**.** *Given* $\varepsilon > 0$*, a source distribution* $\mathcal{X}$ *over source set* X*, and a seed distribution* $\mathcal{H}$ *over seed set* H*, a* strong $(\mathcal{X}, \mathcal{H}, \varepsilon)$-extractor *is an efficient function* $\mathrm{Ext} : H \times X \to Y$*, usually written* $h(x)$ *for* $h \in H$[7] *and* $x \in X$*, such that* $\mathsf{SD}((\mathcal{H}(\mathcal{X}), \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \leq \varepsilon$,[8] *where* $\mathcal{Y}$ *is the uniform distributions on the output set* Y*.*

Overloading notation, we say that an $(\xi, \mathcal{H}, \varepsilon)$-extractor *is an* $(\mathcal{X}, \mathcal{H}, \varepsilon)$*-extractor over every source* $\mathcal{X}$ *such that* $H_\infty(\mathcal{X}) \geq \xi$*.*

However, we will mostly stick to a weaker class of extractors, which are only guaranteed to work on oblivious bit-fixing sources.

▶ **Definition 8** (Oblivious Bit-Fixing (OBF) Sources)**.** *A* (ξ, γ)-OBF source *is a distribution* $\mathcal{W}$ *on* $W = \mathbb{1}^\gamma$ *wherein the individual bits are independently sampled, each bit is either uniform or constant, and there are exactly* ξ *uniform bits, in fixed locations. Equivalently,* $\mathcal{W}$ *is a fixed permutation applied to the uniform distribution on* $\mathbb{1}^\xi \times \{f\}$*, for some fixed bit vector* $f \in \mathbb{1}^{\gamma - \xi}$*.*

[6] *Weak* extractors provide guarantees only when the seed is hidden. We do not consider any weak extractors in this work, and therefore all extractors mentioned herein should be assumed to be strong.
[7] We often refer to H as an *extractor family* and think of $h \in H$ as a function $h : X \to Y$.
[8] The $\mathcal{H}$'s in this equation are the same random variable, not independent samples from $\mathcal{H}$.

We note that the literature contains a variation on bit-fixing sources wherein an adversary can choose the values of the fixed bits as a function of the uniform ones [1]. These are known as *non-oblivious* bit-fixing (NOBF) sources. We do not consider such sources in this work, and unless otherwise specified, all bit-fixing sources used in this work should be presumed to be oblivious.

▶ **Definition 9** (Oblivious Bit-Fixing (OBF) Extractors). *A* strong $(\xi, \mathcal{H}, \varepsilon)$-OBF extractor *is a strong* $(\mathcal{W}, \mathcal{H}, \varepsilon)$*-extractor per Definition 7 for every source* $\mathcal{W}$ *that is* (ξ, γ)-oblivious-bit-fixing *for some* $\gamma \geq \xi$*, per Definition 8.*

We also introduce a new variation on bit-fixing sources, in which the indices of the fixed bits are chosen adaptively by an adversary as a function of the seed.

▶ **Definition 10** (Seed-Adaptive OBF Sources). *A* (ξ, γ)-SAOBF source *is a distribution* $\mathcal{W}$ *on* $W = \mathbb{1}^\gamma$ *wherein the individual bits are independently sampled, each bit is either uniform or constant, and there are exactly* ξ *uniform bits, in locations chosen by the adversary as a function of the extractor's seed. Equivalently,* $\mathcal{W}$ *is a permutation that is* adversarially *derived from the seed and applied to the uniform distribution on* $\mathbb{1}^\xi \times \{f\}$*, for some fixed* $f \in \mathbb{1}^{\gamma-\xi}$*.*

▶ **Definition 11** (Seed-Adaptive OBF Extractors). *A* strong $(\xi, \mathcal{H}, \varepsilon)$-SAOBF extractor *is a strong* $(\mathcal{W}, \mathcal{H}, \varepsilon)$*-extractor per Definition 7 for every source* $\mathcal{W}$ *that is* (ξ, γ)-seed-adaptive-oblivious-bit-fixing *for some* $\gamma \geq \xi$*, per Definition 10.*

Extractors as defined above are often flexible enough to tolerate sources with significant unknown (and even adversarial) bias, so long as those sources have sufficient min-entropy. Allowing this same kind of flexibility over the seed distribution defines a biased-seed extractor; our main concern in this paper is analyzing extractors with biased seeds.

▶ **Definition 12** (Biased-Seed Extractors). *A* strong $(\mathcal{X}, \eta, \varepsilon)$-biased-seed extractor *is an strong* $(\mathcal{X}, \mathcal{H}, \varepsilon)$*-extractor for all seed distributions* $\mathcal{H}$ *such that* $H_\infty(\mathcal{H}) \geq \eta$*.*

The particular extractor we will analyze is the inner product extractor.

▶ **Definition 13** (Inner Product Extractor). *The* $\mathbb{F}_m$*-inner-product extractor has source set* $X = \mathbb{1}^\gamma$*, seed set* $H = \mathbb{F}_m^\gamma$*, and evaluation* $h(x) = \langle h, x \rangle$*.*

2.3 Universal Hashing and Exceptional Inputs

▶ **Definition 14** (Multicollision Probability). *For a family of hash functions* $H : X \to Y$ *and an integer* i*, let* CP *be a function that computes the* i*-way collision probability of a uniformly-sampled member of* H*. That is, if we let* $\mathcal{H}$ *and* $\mathcal{X}$ *be the uniform distributions over* H *and* X *respectively, then*

$$\mathsf{CP}(i) = \Pr_{\substack{h \leftarrow \mathcal{H} \\ (x_1, \ldots, x_i) \leftarrow \mathcal{X}^i}} \left[h(x_{i_1}) = h(x_{i_2}) \forall i_1, i_2 \right].$$

▶ **Definition 15** (Distinct Inputs). *For an input domain* X *and an integer* j *let* DI *be a function computing the* j*-pairwise-distinct input set, i.e.*

$$\mathsf{DI}(j) = \left\{ (x_1, \ldots, x_j) \in X^j : \nexists\, i_1, i_2 \text{ s.t. } i_1 \neq i_2 \wedge x_{i_1} = x_{i_2} \right\}.$$

We denote the uniform distribution over this set as $\mathcal{U}_{\mathsf{DI}(j)}$*. Notice,* $|\mathsf{DI}(j)| = |X|^{\underline{j}}$*.*

▶ **Definition 16** (Distinct-Input Multicollision Probability). *For a family of hash functions $H : X \to Y$ and an integer j, let* CPDI *be a function that computes the j-way collision probability of a uniformly-sampled member of H, conditioned on pairwise-distinct inputs. That is,*

$$\mathsf{CPDI}(j) = \Pr_{\substack{h \leftarrow \mathcal{H} \\ (x_1,\ldots,x_j) \leftarrow \mathcal{U}_{\mathsf{DI}(j)}}} \left[h(x_{i_1}) = h(x_{i_2}) \forall i_1, i_2\right].$$

For convenience we also assert that $\mathsf{CPDI}(1) = 1$ *for any X.*

▶ **Definition 17** (Universal Hash Function). *A hash function family $H : X \to Y$ is ℓ-universal if for every $(x_1, \ldots, x_\ell) \in \mathsf{DI}(\ell)$,*

$$\Pr_{h \leftarrow \mathcal{H}}\left[h(x_{i_1}) = h(x_{i_2}) \forall i_1, i_2\right] \leq |Y|^{1-\ell}.$$

Note that the above condition trivially implies $\mathsf{CPDI}(\ell) \leq |Y|^{1-\ell}$.

▶ **Definition 18** (Exceptional Inputs). *For a family of hash functions $H : X \to Y$ and integers j and k, let* EI *be a function computing the set of j-pairwise-distinct input tuples on which H is not k-universal. That is,*

$$\mathsf{EI}(j,k) = \left\{ \begin{array}{l} (x_1, \ldots, x_j) \in \mathsf{DI}(j) : \\ \quad \Pr_{h \in \mathcal{H}}\left[h(x_{i_1}) = h(x_{i_2}) \forall i_1, i_2\right] > |Y|^{1-k} \end{array} \right\}.$$

Abusing notation, we will often write $\mathsf{EI}(j) = \mathsf{EI}(j,j)$.

▶ **Definition 19** (Exceptional Input Fraction). *For a family of hash functions $H : X \to Y$ and integers j and k, let* EIF *be a function computing the* fraction *of j-pairwise-distinct input tuples on which H is not k-universal. That is,*

$$\mathsf{EIF}(j,k) = \frac{|\mathsf{EI}(j,k)|}{|\mathsf{DI}(j)|}.$$

Abusing notation, we will often write $\mathsf{EIF}(j) = \mathsf{EIF}(j,j)$.

▶ **Proposition 20.** *For any ℓ-universal hash function family, we have the following generic equalities and bounds on* EIF*:*

$$\begin{aligned} &\mathsf{EIF}(j,k) \leq \mathsf{EIF}(j,k+1) \\ &\mathsf{EIF}(j,k) = 0 && \text{if } k \leq \ell \\ &\mathsf{EIF}(j) = 0 && \text{if } j \leq \ell \end{aligned}$$

▶ **Proposition 21.** *For any ℓ-universal hash function family $H : X \to Y$, we have the following generic bounds on* CPDI*:*

$$\begin{aligned} \mathsf{CPDI}(j) &\leq |Y|^{1-j} + \sum_{k=\ell+1}^{j} \mathsf{EIF}(j,k)\left(|Y|^{2-k} - |Y|^{1-k}\right) && (2) \\ &\leq |Y|^{1-j} + \mathsf{EIF}(j)|Y|^{1-\ell} && (3) \\ \mathsf{CPDI}(j) &\leq \mathsf{CPDI}(j-1) \\ \mathsf{CPDI}(j) &\leq |Y|^{1-j} && \text{if } j \leq \ell \end{aligned}$$

▶ **Proposition 22.** *For any function family $H : X \to Y$, we have the following generic equalities for* CP*:*

$$\mathsf{CP}(i) = \sum_{y \in Y} \Pr_{\substack{x \leftarrow \mathcal{X} \\ h \leftarrow \mathcal{H}}} [h(x) = y]^i$$

$$\mathsf{CP}(0) = |Y|$$

$$\mathsf{CP}(1) = 1$$

$$\mathsf{CP}(i) = \sum_{j=1}^{i} \left\{ {i \atop j} \right\} \frac{|X|^{\underline{j}}}{|X|^i} \mathsf{CPDI}(j) \qquad \text{if } i > 0 \tag{4}$$

Proofs of Equations (2) and (4) are given in the full version [12]. The full version also contains additional preliminaries relating to matroids, Tutte polynomials, and coboundary polynomials, and a simple proof that the rank of a matrix is only sensitive to the field over which the entries of that matrix is defined when the order of that field is small relative to the length of the short side of the matrix.

3 Restricting the Problem

In this section, we will prove a few useful lemmas that allow us to restrict our attention to only *uniform* sources and *uniform-over-subset* seeds for the remainder of the document.

3.1 The Sufficiency of Uniform-Over-Subset Seed Distributions

Randomness extractors typically assume a uniform seed distribution and tolerate a biased source distribution with sufficient min-entropy. A standard lemma about randomness extractors [24, Lemma 6.10] asserts that in order to reason about all sources with a certain min-entropy, it is sufficient to reason about sources that are uniformly distributed over arbitrary correspondingly-sized subsets of the source space. We adapt this idea to arbitrarily-biased *seeds* using following lemma (proven in the full version [12]). Hereafter in this document, we will restrict our focus to seed distributions $\mathcal{G}$ that are uniform over subsets $G \subseteq H$.

▶ **Lemma 23** (Convex Combination of Uniform Seeds)**.** *Let ε be a concave function and let H be an extractor family. If H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon(|G|^{-1}))$-extractor per Definition 7 for every $G \subseteq H$ and $\mathcal{G}$ uniformly distributed over G, then then H is also a strong $(\mathcal{X}, \eta, \varepsilon(2^{-\eta}))$-*biased-seed *extractor for all η, per Definition 12.*

3.2 The (In)Sufficiency of Uniform Source Distributions

Lemma 23 showed that in order to reason about any arbitrily-biased seed distributions $\mathcal{H}$, we need only prove that an extractor is good on the set of all seed distributions $\mathcal{G}$ that are uniform over some $G \subseteq H$ where $|G| = 2^{H_\infty(\mathcal{H})}$. It would be tempting to use a similar line of reasoning to restrict our attention to uniform-over-subset source distributions, and claim a generalization to all sources with sufficient min-entropy, and indeed this is often done in other works (in particular, this technique is used in the classic LHL proof, and in Lee et al.'s proof of the GLHL [20]). Unfortunately, our analysis in the remaining sections of this paper does *not* permit this line of reasoning. We are able to show that IPE is a good extractor on the source distribution that is uniform over the entire source set X, but there is a specific point in each of our proofs where the chain of implication fails if the source distribution is uniform over only a *subset* of X:

- In Definition 19 and the proof of Lemma 31 we take $\mathsf{EIF}(j)$ to be the number of inputs on which the extractor is not j-universal, over the number of pairwise-distinct input tuples. If the source distribution is uniform over a subset of X (and has probability 0 outside that subset), then the concrete bound we calculate on $\mathsf{EIF}(4)$ is false, because we must assume all "bad" inputs to be within the subset, meaning that the denominator shrinks with the subset, but the numerator does not. This invalidates Theorem 32 for arbitrary non-uniform source distributions. Our analysis could potentially be adjusted to account for this, but we do not expect such an adjustment to produce useful bounds for signficantly biased distributions over X.
- An additional example relates to a proof that appears in the full version of this paper [12].

(Static) Oblivious Bit-Fixing Sources

Although we cannot generalize from our results for uniform sources to results for *arbitrary* source distributions, we *can* generalize to the more restricted classes of *oblivious bit-fixing* (OBF) sources given in Definition 8. We begin with the simpler case of bits that are fixed *independently* of the seed.[9]

▶ **Lemma 24** (Generalization to Oblivious Bit-Fixing Sources). *Let $\varepsilon, \eta \geq 0$ and let ξ and γ be integers such that $\gamma \geq \xi$. Let $H_\xi = \mathbb{F}_m^\xi$ be the IPE family (Definition 13) from $X = \mathbb{1}^\xi$ to $Y = \mathbb{F}_m$, and let $H_\gamma = \mathbb{F}_m^\gamma$ be the IPE family from $W = \mathbb{1}^\gamma$ to Y. If H_ξ is a strong $(\mathcal{X}, \eta - (\gamma - \xi)\log_2(m), \varepsilon)$-biased-seed extractor per Definition 12, where $\mathcal{X}$ is the uniform distribution over X, then H_γ is a strong (ξ, η, ε)-biased-seed-OBF extractor per Definitions 9 and 12.*

Proof. By Lemma 23, we can restrict our attention to the set of seed distributions $\mathcal{G}$ that are uniform over some $G \subseteq H$ such that $H_\infty(\mathcal{G}) \geq \eta$. Recall from Definition 8 that any (ξ, γ)-OBF source $\mathcal{W}$ on $W = \mathbb{1}^\gamma$ is a fixed permutation applied to the uniform distribution on $\mathbb{1}^\xi \times \{f\}$, for some fixed bit vector $f \in \mathbb{1}^{\gamma-\xi}$. In other words, every $w \in W$ can be partitioned bitwise into f and some $x \in X$, and sampling $w \leftarrow \mathcal{W}$ corresponds to sampling $x \leftarrow \mathcal{X}$ and interleaving the bits of x with the bits of f. We can also partition any $h \in H_\gamma$ into $h_x \in H_\xi$ (the elements of h corresponding to bits of w that are partitioned into x) and h_f (the elements of h corresponding to bits of w that are partitioned into f). It is easy to see that $\langle h, w\rangle = \langle h_x, x\rangle + \langle h_f, f\rangle$, where $\langle h_x, x\rangle$ is an instance of the inner-product extractor on the uniform source $\mathcal{X}$, and thus the statistical distance between $\langle h, w\rangle$ and the uniform distribution is no greater than the distance between $\langle h_x, x\rangle$ and the uniform distribution.

It remains only to reason about the distribution of h_x, which we will denote $\mathcal{G}_x$. In particular, the partitioning process maps *at least* one and *at most* $m^{\gamma-\xi}$ values $h \in H_\gamma$ onto each $h_x \in H_\xi$. It therefore follows that $H_\infty(\mathcal{G}_x) \geq H_\infty(\mathcal{G}) - (\gamma - \xi)\log_2(m)$. Since H_ξ is a $(\mathcal{X}, \mathcal{G}_x, \varepsilon)$-extractor by assumption, it follows that H_γ is a $(\mathcal{W}, \mathcal{G}, \varepsilon)$-extractor. This holds for any (ξ, γ)-OBF source $\mathcal{W}$, any biased seed $\mathcal{G}$ such that $H_\infty(\mathcal{G}) \geq \eta$, and any $\mathcal{G}_x$ induced by the pair. Thus we have the lemma. ◀

3.3 Seed-Adaptive Oblivious Bit-Fixing Sources

We now turn our attention to *seed-adaptive* OBF sources, as described in Definition 10. As above, we can partition the source bit vector w into a sub-vector f of fixed bits and a sub-vector x of uniformly distributed bits, and we can partition the corresponding elements

[9] Such an analysis was already performed by Doerner et al. [8, 9, 10]. Their analysis is correct for the non-seed-adaptive case, but as we discussed in Section 1.3, they erred in that their proof actually required a seed-adaptive analysis.

of h into h_f and h_x. When $\mathcal{W}$ is *seed-adaptive*, however, the latter partitioning induces (additional) bias into the seed. We quantify that bias by defining an augmented seed distribution $\mathcal{D}^{\mathsf{Adv}}(\mathcal{G}, c)$ as a function of the (pre-bit-fixing) distribution $\mathcal{G}$ and the number c of fixed bits, after which we lower bound the min-entropy of $\mathcal{D}^{\mathsf{Adv}}_{\mathcal{G},c}$ relative to the min-entropy of $\mathcal{G}$.

Algorithm 1 $\mathcal{D}^{\mathsf{Adv}}_{\mathcal{G},\xi}$.

$\mathcal{D}$ `samples` $h = (h_1, \ldots, h_\gamma) \leftarrow \mathcal{G}$
`The adversary produces` $I \leftarrow \mathsf{Adv}(h, \xi)$ `such that` $I \subset [\gamma]$ `and` $|I| = \gamma - \xi$
$\mathcal{D}$ `outputs` $h^* = (h_i)_{i \notin I}$

▶ **Lemma 25.** *For all integers $\gamma \geq \xi > 0$ and all (not necessarily efficient) adversaries* Adv*, and any distibution $\mathcal{G}$ on $\mathbb{F}_m^\gamma$*

$$H_\infty\left(\mathcal{D}^{\mathsf{Adv}}_{\mathcal{G},\xi}\right) \geq H_\infty(\mathcal{G}) - (\gamma - \xi)\log_2(m) - \log_2\binom{\gamma}{\xi}$$

Proof. The min-entropy $H_\infty(\mathcal{D}^{\mathsf{Adv}}_{\mathcal{G},\xi})$ is defined as the negative logarithm of the probability of the most likely event in the distribution $\mathcal{D}^{\mathsf{Adv}}_{\mathcal{G},\xi}$. The adversary's influence is restricted to deleting $(\gamma - \xi)$ entries from h. For any fixed output h^*, the number of pre-image vectors h from which $(\gamma - \xi)$ elements can be deleted to yield h^* is $\binom{\gamma}{\gamma-\xi} m^{\gamma-\xi}$. The probability of the most likely value of h^* is upper-bounded by the probability that one of its pre-images is sampled, which is $\binom{\gamma}{\gamma-\xi} m^{\gamma-\xi} / 2^{H_\infty(\mathcal{G})}$, and the lemma follows. ◀

We therefore have the following seed-adaptive counterpart to Lemma 24:

▶ **Lemma 26** (Generalization to SAOBF Sources)**.** *Let $\varepsilon, \eta \geq 0$ and let ξ and γ be integers such that $\gamma \geq \xi$. Let $H_\xi = \mathbb{F}_m^\xi$ be the IPE family (Definition 13) from $X = \mathbb{1}^\xi$ to $Y = \mathbb{F}_m$, and let $H_\gamma = \mathbb{F}_m^\gamma$ be the IPE family from $W = \mathbb{1}^\gamma$ to Y. If H_ξ is a strong $(\mathcal{X}, \eta - (\gamma - \xi)\log_2(m) - \log_2\binom{\gamma}{\xi}, \varepsilon)$-biased-seed extractor per Definition 12, where $\mathcal{X}$ is the uniform distribution over X, then H_γ is a strong (ξ, η, ε)-biased-seed-SAOBF extractor per Definitions 11 and 12.*

Proof Sketch. The proof is similar to the proof of Lemma 24, except that the distribution of h_x is $\mathcal{D}^{\mathsf{Adv}}_{\mathcal{G},\xi}$ instead of $\mathcal{G}_x$. ◀

Hereafter in this document, we will restrict our focus to the uniform source distribution $\mathcal{X}$ over $\mathbb{1}^\xi$.

4 The Elevated General Leftover Hash Lemma

In this section, we develop an *Elevated* General Leftover Hash Lemma for *Almost-n*-Universal Hash functions. By *elevated*, we mean that our hash function is derived via the n^{th} power of the norm (as opposed to the squared norm, in the case of the standard GLHL) and it therefore establishes an n-root bound on statistical distance. By *almost-n*-universal, we mean that it does not require the hash function on which it acts to be n-universal, but instead requires only that the Distinct-Input Collision Probability be bounded generally. This is crucial: it allows us to produce a tighter analysis of a hash function that is ℓ-universal and *almost but not quite* n-universal than we could using ℓ-universality alone. Apart from this bound on collision probability, we require little information about the hash function, and work over general, unstructured domains (H, G, X, Y). Our proof proceeds very similarly to the proof of the GLHL of Lee et al. [20] and to the original LHL proof of Impagliazzo, Levin, and Luby [17].

In Section 4.2, we use additional structural information about the IPE to specialize our lemma and replace the bound on distinct-input collision probability with a bound on the number of *exceptional* inputs on which the IPE is not n-universal[10]. Later, in Section 5, we finally bound the number of exceptional inputs for 4-universality, allowing us to calculate an explicit 4^{th}-root upper bound on statistical distance for the IPE.

4.1 A Family-Agnostic n^{th} GLHL

▶ **Lemma 3.** *Let H be a family of hash functions from X to Y. Let $G \subseteq H$ and let $\mathcal{G}$ and $\mathcal{X}$ denote the uniform distributions over G and X. Let* CPDI *be the distinct-input collision probability of H (Definition 16). If $n \in \mathbb{N}^+$ is even, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 7, where*

$$\varepsilon \leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left(1+\sum_{j=1}^{n}\mathsf{CPDI}(j)\sum_{i=j}^{n}\binom{n}{i}\left\{{i \atop j}\right\}\frac{|X|^{\underline{j}}|Y|^{i-1}}{(-1)^{n-i}|X|^{i}}\right)}$$

Proof. For all even n we have:

$$\begin{aligned}
&2^n \cdot \mathsf{SD}((\langle\mathcal{G},\mathcal{X}\rangle,\mathcal{G}),(\mathcal{Y},\mathcal{G}))^n \\
&= \left(\sum_{h\in G}\sum_{y\in Y}\frac{1}{|G|}\left|\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right|\right)^n \\
&= |Y|^n\left(\frac{\sum_{h\in G}\sum_{y\in Y}\left|\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right|}{\sum_{h\in G}\sum_{y\in Y}1}\right)^n \\
&\leq |Y|^n\frac{\sum_{h\in G}\sum_{y\in Y}\left(\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right)^n}{\sum_{h\in G}\sum_{y\in Y}1} \qquad (5)\\
&= \frac{|Y|^{n-1}}{|G|}\sum_{h\in G}\sum_{y\in Y}\left(\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right)^n \\
&\leq \frac{|Y|^{n-1}}{|G|}\sum_{h\in H}\sum_{y\in Y}\left(\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right)^n \qquad (6)\\
&= \frac{|Y|^{n-1}}{|G|}\sum_{h\in H}\sum_{y\in Y}\sum_{i=0}^{n}\binom{n}{i}\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]^i\frac{1}{(-|Y|)^{n-i}} \qquad (7)\\
&= \frac{|H|}{|G|}\sum_{i=0}^{n}\binom{n}{i}\sum_{y\in Y}\Pr_{\substack{x\leftarrow\mathcal{X}\\ h\leftarrow\mathcal{H}}}[h(x)=y]^i\frac{|Y|^{i-1}}{(-1)^{n-i}} \\
&= \frac{|H|}{|G|}\sum_{i=0}^{n}\binom{n}{i}\mathsf{CP}(i)\frac{|Y|^{i-1}}{(-1)^{n-i}} \qquad (8)\\
&= \frac{|H|}{|G|}\left(1+\sum_{i=1}^{n}\binom{n}{i}\sum_{j=1}^{i}\left\{{i \atop j}\right\}\frac{|X|^{\underline{j}}}{|X|^i}\mathsf{CPDI}(j)\frac{|Y|^{i-1}}{(-1)^{n-i}}\right)
\end{aligned}$$

[10] A hash function is not n universal on a tuple of n distinct inputs if it has a probability greater than $|Y|^{1-n}$ of producing an n-way collision on those *specific* inputs, over the choice of its seed. See also Definition 18.

$$= \frac{|H|}{|G|}\left(1 + \sum_{j=1}^{n} \mathsf{CPDI}(j) \sum_{i=j}^{n} \binom{n}{i} \left\{ {i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{i-1}}{(-1)^{n-i}|X|^i}\right)$$

where

Equation (5) follows from Jensen's inequality and the fact that n is even
Equation (6) follows from the fact that $G \subseteq H$
Equation (7) follows from the binomial theorem
The lemma follows from taking the n^{th} root. ◀

4.2 Applying Lemma 3 to the IPE

▶ **Proposition 27.** *For the Inner Product Extractor we have* $\mathsf{CPDI}(j) \geq |Y|^{1-j}$.

Proof. For any tuple of inputs $(x_1, \ldots, x_j) \in \mathsf{DI}(j)$, the probability of a collision under $(x_1, \ldots, x_j)$ is exactly

$$\mathsf{CPDI}(j) = \sum_{y \in \mathbb{F}_m} \Pr_{h \leftarrow \mathbb{F}_m^{\xi}} \left[\begin{bmatrix} h^T & y \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_j \\ 1 & \cdots & 1 \end{bmatrix} = \vec{0} \right].$$

Since $\vec{0}$ is in the span of any matrix, the system of linear equations implicitly defined by matrix multiplication within the probability statement is always consistent. It has $\xi + 1$ unknowns (h, y) and a coefficient matrix with rank no greater than j (since this is the number of its columns), and thus by the Rouché-Capelli Theorem the number of solutions (each of which corresponds to a collision) is no less than $m^{\xi+1-j}$. These solutions are partitioned among the various fixed values of y over which the sum is taken, with a consistent denominator comprising the total number of possible assignments of h, which is m^{ξ}. Thus the final probability of a collision must be greater than $m^{\xi+1-j}/m^{\xi} = |Y|^{1-j}$ ◀

▶ **Lemma 28.** *The Inner Product Extractor is* 3*-universal if the source is at least 2 bits long.*

Proof. Per Proposition 20, we can equivalently prove that $\mathsf{EIF}(2) = \mathsf{EIF}(3) = 0$. The proof that $\mathsf{EIF}(2) = 0$ is traditional, we omit it. To see that $\mathsf{EIF}(3) = 0$, we need only a slight extension of the argument that $\mathsf{EIF}(2) = 0$. Observe that if we interpret any three inputs as the columns in a boolean matrix, then any row where the first two columns are distinct must have a value in the third column equal to the value in either the first or second. Without loss of generality, let it be the first. Because 3-universality quantifies only over input sets with three distinct inputs (i.e. three distinct columns in the corresponding matrix), there must exist another row in which the third column is *not* equal to the first, and using this row, we can repeat the argument for $\mathsf{EIF}(2) = 0$ to conclude that $\mathsf{EIF}(3) = 0$. ◀

▶ **Lemma 29.** *Let* H *be an IPE family (Definition 13) from* X *to* Y*. Let* $G \subseteq H$ *and let* $\mathcal{G}$ *and* $\mathcal{X}$ *denote the uniform distributions over* G *and* X*. If* $n \in \mathbb{N}^+$ *is even, then* H *is a strong* $(\mathcal{X}, \mathcal{G}, \varepsilon)$*-extractor per Definition 7, where*

$$\varepsilon \leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left(\begin{array}{l} 1 + \displaystyle\sum_{j=1}^{n}\sum_{i=j}^{n} \binom{n}{i}\left\{ {i \atop j} \right\} \frac{|X|^{\underline{j}}|Y|^{i-j}}{(-1)^{n-i}|X|^i} \\ \quad + \displaystyle\sum_{j=4}^{n}\sum_{k=4}^{j} \mathsf{EIF}(j,k) \sum_{i=\lceil j/2 \rceil}^{n/2} \binom{n}{2i}\left\{ {2i \atop j} \right\} \frac{|X|^{\underline{j}}|Y|^{2i-k+1}}{|X|^{2i}} \end{array}\right)}$$

Proof. Beginning from Lemma 3, we have

$$
\begin{aligned}
&2^n \cdot \mathsf{SD}((\langle \mathcal{G}, \mathcal{X} \rangle, \mathcal{G}), (\mathcal{Y}, \mathcal{G}))^n \\
&\leq \frac{|H|}{|G|} \left(1 + \sum_{j=1}^{n} \mathsf{CPDI}(j) \sum_{i=j}^{n} \binom{n}{i} \left\{ {i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{i-1}}{(-1)^{n-i} |X|^i} \right) \\
&= \frac{|H|}{|G|} \left(1 + \sum_{j=1}^{n} \mathsf{CPDI}(j) \left(\begin{array}{l} \displaystyle\sum_{i=\lceil j/2 \rceil}^{n/2} \binom{n}{2i} \left\{ {2i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{2i-1}}{|X|^{2i}} \\ \displaystyle - \sum_{i=\lfloor j/2 \rfloor}^{n/2} \binom{n}{2i+1} \left\{ {2i+1 \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{2i}}{|X|^{2i+1}} \end{array} \right) \right) \\
&\leq \frac{|H|}{|G|} \left(\begin{array}{l} \displaystyle 1 + \sum_{j=1}^{n} |Y|^{1-j} \sum_{i=\lceil j/2 \rceil}^{n/2} \binom{n}{2i} \left\{ {2i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{2i-1}}{|X|^{2i}} \\ \displaystyle + \sum_{j=1}^{n} \sum_{k=\ell+1}^{j} \mathsf{EIF}(j,k) |Y|^{2-k} \sum_{i=\lceil j/2 \rceil}^{n/2} \binom{n}{2i} \left\{ {2i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{2i-1}}{|X|^{2i}} \\ \displaystyle - \sum_{j=1}^{n} |Y|^{1-j} \sum_{i=\lfloor j/2 \rfloor}^{n/2} \binom{n}{2i+1} \left\{ {2i+1 \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{2i}}{|X|^{2i+1}} \end{array} \right)
\end{aligned}
$$

Where the last step follows from Proposition 27 and Equation (2) of Proposition 21, and the conclusion follows from Lemma 28 ◀

▶ **Lemma 30** (Simplified Version of Lemma 29)**.** *Let H be an IPE family (Definition 13) from X to Y. Let $G \subseteq H$ and let $\mathcal{G}$ and $\mathcal{X}$ denote the uniform distributions over G and X. If $n \in \mathbb{N}^+$ is even, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 7, where*

$$
\varepsilon \leq \frac{1}{2} \sqrt[n]{\frac{|H|}{|G|} \left(\begin{array}{l} \displaystyle 1 + \sum_{j=1}^{n} \sum_{i=j}^{n} \binom{n}{i} \left\{ {i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{i-j}}{(-1)^{n-i} |X|^i} \\ \displaystyle + \sum_{j=4}^{n} \mathsf{EIF}(j) \sum_{i=\lceil j/2 \rceil}^{n/2} \binom{n}{2i} \left\{ {2i \atop j} \right\} \frac{|X|^{\underline{j}} |Y|^{2i-3}}{|X|^{2i}} \end{array} \right)}
$$

Proof. The proof is identical to the proof for Lemma 29, except that we use Equation (3) from Proposition 21 instead of Equation (2). ◀

5 A Simple Bound for the IPE at $n = 4$

In this section we bound the number of exceptional inputs to 4-universality for the IPE using techniques similar to those employed by prior works to show that it is 2-universal. We then plug this bound into Lemma 29 to derive our first explicit 4^{th}-root bound on the statistical distance of the IPE.

5.1 Bounding $\mathsf{EIF}(4)$

▶ **Lemma 31.** *If $\xi \geq 5$ and $m > 3$, then the IPE family (Definition 13) from $X = \mathbb{1}^\xi$ to $Y = \mathbb{F}_m$ has $\mathsf{EIF}(4) \leq 15 \cdot 6^{\log_2|X|}/|X|^{\underline{4}}$*

Proof. Let $W(\xi, 4) = \mathbb{1}^{\xi \times 4}$ and let $\widetilde{W}(\xi, 4) \subset W(\xi, 4)$ be the set of all matrices in $W(\xi, 4)$ that have four pairwise-distinct columns. We can interpret any set of 4 inputs $(x_1, x_2, x_3, x_4) \in \mathsf{DI}(4)$ as a unique matrix $w \in \widetilde{W}(\xi, 4)$ and vice versa. We will divide our argument into two mutually-exclusive cases based upon the value w, and then bound the probability that each of these cases occurs when w is sampled.

1. w contains a row such that the value in one column of that row is distinct from the values in the others; i.e. there exists a row in w with parity 1. In this case, we can apply the argument that $\mathsf{EIF}(3) = 0$ (Lemma 28) to the columns that are identical to see that the hash function is 4-universal on w.
2. w does *not* contain a row such that the value in one column of that row is distinct from the values in the others; i.e. all rows in w have parity 0. Let us now consider the *sub*-case that there are five distinct rows with Hamming weight 2 (i.e. the values in any two columns of these rows are 0 and the values in the other two columns are 1). We first observe that among any five distinct rows with Hamming weight 2, four must be linearly independent.[11][12] Without loss of generality, let us number these four linearly independent rows 1 through 4. Let $(a_1, \ldots, a_\xi) = h$ be the individual elements of the seed of the IPE, and suppose that there exists some y such that

$$\langle a_{[4]}, w_{[4],k} \rangle + y'_k = y \qquad \text{and} \qquad y'_k = \langle a_{[5,\xi]}, w_{[5,\xi],k} \rangle$$

for $k \in [4]$. Let us fix any $a_{[5,\xi]}$. This in turn fixes y'_1, y'_2, y'_3, y'_4. For each of the m possible values of y, we have a system of four equations with four unknowns (i.e. $a_{[4]}$) and a coefficient matrix of rank four (i.e. $w_{[4],k}$). By the Rouché–Capelli theorem, such a system has at most one solution. That is, for every w, $a_{[5,\xi]}$, there is at most one assignment of $a_{[4]}$ that causes a four-way collision with output value y. Summing over the m possible values of y, and dividing by the m^4 possible assignments of $a_{[4]}$ gives us an upper bound on the overall collision probability of m^{-3}. In other words, the hash function is 4-universal in this sub-case.

Putting our cases together, we find that the set $\overline{W}(\xi, 4)$ of matrices on which the inner product extractor is *not* 4-universal is a subset of those whose rows assume the values of at most four distinct vectors of hamming weight 2, or the 0 or 1 vector. In other words $|\overline{W}(\xi, 4)| \leq \binom{6}{4} 6^\xi$. We have

$$\mathsf{EIF}(4) \leq \frac{|\overline{W}(\xi, 4) \cap \widetilde{W}(\xi, 4)|}{|\widetilde{W}(\xi, 4)|} \leq \frac{|\overline{W}(\xi, 4)|}{|\widetilde{W}(\xi, 4)|} \leq 15 \frac{6^\xi}{|X|^{\underline{4}}}$$ ◀

5.2 Putting it Together

Now we can derive a concrete upper bound on distinguishability.

[11] Up to permutations on rows, there are 6 boolean matrices of dimension $(5, 4)$ that have pairwise-distinct rows, each of hamming weight 2. All of them have rank 4 over the real numbers and therefore rank 4 over $\mathbb{F}_m$ for $m > 3$, and rank is invariant under permutation of rows.

[12] Most of the matrices with exactly four distinct rows of hamming weight 2 are also 4 universal, which means it's easy to make this analysis slightly tighter, but it's uglier and it doesn't appear to lead to a concrete improvement.

▶ **Theorem 32.** *Let $H = \mathbb{F}_m^{\xi}$ be the IPE family (Definition 13) from $X = \mathbb{1}^{\xi}$ to $Y = \mathbb{F}_m$. Let $G \subseteq H$ and let $\mathcal{G}$ and $\mathcal{X}$ denote the uniform distributions over G and X. If $|Y| > 3$, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 7, where*

$$\varepsilon \leq \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}\left(\frac{|X|^{\underline{4}}}{|X|^4} - 4\frac{|X|^{\underline{3}}}{|X|^3} + 6\frac{|X|^{\underline{2}}}{|X|^2} - 3\right) \\ + \frac{|Y|}{|X|}\left(6\frac{|X|^{\underline{3}}}{|X|^3} - 12\frac{|X|^{\underline{2}}}{|X|^2} + 6\right) \\ + \frac{|Y|^2}{|X|^2}\left(7\frac{|X|^{\underline{2}}}{|X|^2} - 4\right) + \frac{|Y|^3}{|X|^3} + 15\frac{|Y|3^{\log_2|X|}}{|X|^3}\end{array}\right)}$$

Proof. Beginning from Lemma 30, we have

$$\begin{aligned}
&\mathsf{SD}((\langle\mathcal{G},\mathcal{X}\rangle,\mathcal{G}),(\mathcal{Y},\mathcal{G})) \\
&\leq \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}1 + \sum_{j=1}^{4}\sum_{i=j}^{4}\binom{4}{i}\left\{{i \atop j}\right\}\frac{|X|^{\underline{j}}|Y|^{i-j}}{(-1)^{4-i}|X|^i} \\ + \mathsf{EIF}(4)\binom{4}{4}\left\{{4 \atop 4}\right\}\frac{|X|^{\underline{4}}|Y|}{|X|^4}\end{array}\right)} \\
&\leq \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}1 + \sum_{j=1}^{4}\sum_{i=j}^{4}\binom{4}{i}\left\{{i \atop j}\right\}\frac{|X|^{\underline{j}}|Y|^{i-j}}{(-1)^{4-i}|X|^i} \\ + 15\frac{|Y|6^{\log_2|X|}}{|X|^4}\end{array}\right)} \qquad (9) \\
&= \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}1 + \sum_{i=1}^{4}\binom{4}{i}\frac{|X||Y|^{i-1}}{(-1)^{4-i}|X|^i} + \sum_{i=2}^{4}\binom{4}{i}\left\{{i \atop 2}\right\}\frac{|X|^{\underline{2}}|Y|^{i-2}}{(-1)^{4-i}|X|^i} \\ + \sum_{i=3}^{4}\binom{4}{i}\left\{{i \atop 3}\right\}\frac{|X|^{\underline{3}}|Y|^{i-3}}{(-1)^{4-i}|X|^i} + \frac{|X|^{\underline{4}}}{|X|^4} + 15\frac{|Y|6^{\log_2|X|}}{|X|^4}\end{array}\right)} \\
&= \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}-3 + 6\frac{|Y|}{|X|} - 4\frac{|Y|^2}{|X|^2} + \frac{|Y|^3}{|X|^3} + 6\frac{|X|^{\underline{2}}}{|X|^2} - 12\frac{|X|^{\underline{2}}|Y|}{|X|^3} \\ + 7\frac{|X|^{\underline{2}}|Y|^2}{|X|^4} - 4\frac{|X|^{\underline{3}}}{|X|^3} + 6\frac{|X|^{\underline{3}}|Y|}{|X|^4} + \frac{|X|^{\underline{4}}}{|X|^4} + 15\frac{|Y|6^{\log_2|X|}}{|X|^4}\end{array}\right)}
\end{aligned}$$

where Equation (9) follows from Lemma 31. ◀

We simplify the above theorem statement slightly by observing that for any constant c, $x^{\underline{c}}/x^c \approx 1$ when $x \gg c$.

▶ **Corollary 33.** *Let $H = \mathbb{F}_m^{\xi}$ be the IPE family (Definition 13) from $X = \mathbb{1}^{\xi}$ to $Y = \mathbb{F}_m$. Let $G \subseteq H$ and let $\mathcal{G}$ and $\mathcal{X}$ denote the uniform distributions over G and X. If $|Y| > 3$ and $|X| \gg 4$, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 7, where*

$$\varepsilon \lesssim \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(3\frac{|Y|^2}{|X|^2} + \frac{|Y|^3}{|X|^3} + 15\frac{|Y|3^{\log_2|X|}}{|X|^3}\right)}$$

Proof. Follows directly from Theorem 32. ◀

▶ **Corollary 34.** *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 13) from $W = \mathbb{1}^\gamma$ to $Y = \mathbb{F}_m$. If $m > 3$, $\gamma \geq \xi$, $2^\xi \gg 4$, and $\eta > 0$, then H is a strong (ξ, η, ε)-biased-seed-OBF extractor per Definitions 9 and 12, where*

$$\varepsilon \lesssim \frac{1}{2}\sqrt[4]{\frac{m^\gamma}{2^\eta}\left(3\frac{m^2}{2^{2\xi}} + \frac{m^3}{2^{3\xi}} + 15\frac{m\cdot 3^\xi}{2^{3\xi}}\right)}$$

Furthermore, H is a strong $(\xi, \eta, \varepsilon')$-biased-seed-SAOBF extractor per Definitions 11 and 12, where

$$\varepsilon' \lesssim \frac{1}{2}\sqrt[4]{\frac{m^\gamma\binom{\gamma}{\xi}}{2^\eta}\left(3\frac{m^2}{2^{2\xi}} + \frac{m^3}{2^{3\xi}} + 15\frac{m\cdot 3^\xi}{2^{3\xi}}\right)}$$

Proof. We first observe that the function bounding ε in Corollary 33 is a concave function of $|G|^{-1}$, and thus we can apply Lemma 23. The corollary then follows from Lemmas 24 and 26. ◀

Finally, restricting the relationship between $|X|$ and $|Y|$ gives us the clean statement of the 4^{th}-root bound featured in the abstract.

▶ **Corollary 35.** *Let $H = \mathbb{F}_m^\xi$ be the IPE family (Definition 13) from $X = \mathbb{1}^\xi$ to $Y = \mathbb{F}_m$. Let $G \subseteq H$ and let $\mathcal{G}$ and $\mathcal{X}$ denote the uniform distributions over G and X. If $|X| \geq |Y| \geq |X|^{0.585}$, $|Y| > 3$, and $|X| \gg 4$, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 7, where*

$$\varepsilon \lesssim \frac{2.1}{2}\sqrt[4]{\frac{|H|}{|G|}}\sqrt{\frac{|Y|}{|X|}}$$

Proof. Follows from Theorem 32 via the fact that $|Y|^3/|X|^3 \leq |Y|^2/|X|^2$ when $|X| \geq |Y| > 0$ and $|Y|3^{\log_2|X|}/|X|^3 \leq |Y|^2/|X|^2$ when $|Y| \geq |X|^{\log_2 3 - 1} > 0$. ◀

▶ **Corollary 1.** *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 13) from $W = \mathbb{1}^\gamma$ to $Y = \mathbb{F}_m$. If $\gamma \geq \xi \geq \log_2 m \geq 0.585\xi$, $m > 3$, $2^\xi \gg 4$, and $\eta > 0$, then H is a strong (ξ, η, ε)-biased-seed-OBF extractor per Definitions 9 and 12, where*

$$\varepsilon \lesssim \frac{2.1}{2}\sqrt[4]{\frac{m^\gamma}{2^\eta}}\sqrt{\frac{m}{2^\xi}}$$

Furthermore, H is a strong $(\xi, \eta, \varepsilon')$-biased-seed-SAOBF extractor per Definitions 11 and 12, where

$$\varepsilon' \lesssim \frac{2.1}{2}\sqrt[4]{\frac{m^\gamma\binom{\gamma}{\xi}}{2^\eta}}\sqrt{\frac{m}{2^\xi}}$$

Proof. We first observe that the function bounding ε in Corollary 35 is a concave function of $|G|^{-1}$, and thus we can apply Lemma 23. The corollary then follows from Lemmas 24 and 26. ◀

6 Further Results from Matroid Theory in the Full Version

In the full version of this paper [12], we relate the *exact* multicollision probability of the IPE to the coboundary polynomial of the cube matroid.

▶ **Definition 36** (The Cube Matroid). *Let $M_n^{\square}$ be the matroid on ground set $\mathbb{1}^n$ with rank function $\mathrm{rank}_{M_n^{\square}}(A) = \dim(A)$. In other words, the rank function of $M_n^{\square}$ is defined identically to the rank function for matrices. Note that as with matrices, the rank function can be defined over any field, not necessarily $\mathbb{F}_2$. We leave the field over which rank is defined implicit unless it is relevant.*

▶ **Lemma 37.** *Let $H = \mathbb{F}_m^{\xi}$ be an IPE family (Definition 13) from $X = \mathbb{1}^{\xi}$ to $Y = \mathbb{F}_m$. Let $\overline{\chi}_{M_{i-1}^{\square}}$ be the coboundary polynomial of the cube matroid (Definition 36). The i-way multicollision probability of H for $i \in \mathbb{N}^+$ is exactly given by*

$$\mathsf{CP}(i) = \frac{m^{1-i}}{2^{\xi\cdot(i-1)}} \left(\frac{d}{d\alpha}\right)^{\xi} \overline{\chi}_{M_{i-1}^{\square}}(m, e^{\alpha})\Bigg|_{\alpha=0}$$

Then, we use the above result to compute a *second* version of the Elevated General Leftover Hash Lemma for the IPE.

▶ **Theorem 4.** *Let $H = \mathbb{F}_m^{\xi}$ be an IPE family (Definition 13) from $X = \mathbb{1}^{\xi}$ to $Y = \mathbb{F}_m$. Let $G \subseteq H$ and let $\mathcal{G}$ and $\mathcal{X}$ denote the uniform distributions over G and X. Let $\overline{\chi}_{M_{i-1}^{\square}}$ be the coboundary polynomial of the cube matroid (Definition 36). If $n \in \mathbb{N}^+$ is even, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 7, where*

$$\varepsilon \leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left(1 + \sum_{i=1}^{n}\binom{n}{i}\frac{(-1)^{n-i}}{2^{\xi\cdot(i-1)}}\left(\frac{d}{d\alpha}\right)^{\xi}\overline{\chi}_{M_{i-1}^{\square}}(m, e^{\alpha})\Bigg|_{\alpha=0}\right)} \tag{10}$$

The first version of an IPE-specific EGLHL given in Lemma 29 was limited in its usefulness because it requires us to prove a concrete bound on $\mathsf{EIF}(j)$ for every $j \in [4, n]$, and we do not know how to prove such bounds for any $j > 4$. On the other hand, we know how to concretely compute the second version for any n, but only in $2^{O(n^2)}$ time. In the full version of this paper, we perform this exponential computation for $n \in \{4, 6, 8\}$ and report the concrete bounds that result. Additionally, we asymptotically upper-bound the right hand side of Equation (10). This yields the following theorem and its corollary:

▶ **Theorem 38.** *Let $H = \mathbb{F}_m^{\xi}, X = \mathbb{1}^{\xi}, Y = \mathbb{F}_m$, and $G \subseteq H$ be as in Theorem 4, let $\mathcal{G}$ and $\mathcal{X}$ be the uniform distributions over G and X, and assume that $m \leq 2^{(1-\epsilon)\cdot\xi}$ for some $0 < \epsilon < 1$. For any even constant $n \in \mathbb{N}^+$, as $m \to \infty$, H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 7, where*

$$\varepsilon \leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left((n-1)!!\cdot 2^{-n\cdot\epsilon\cdot\xi/2} + O\left(2^{-(n/2+1)\cdot\epsilon\cdot\xi} + 2^{-\xi} + \left(\frac{3}{4}\right)^{\xi} 2^{-\epsilon\cdot\xi}\right)\right)}.$$

▶ **Corollary 2.** *Let $H = \mathbb{F}_m^{\gamma}$ be the IPE family (Definition 13) from $W = \mathbb{1}^{\gamma}$ to $Y = \mathbb{F}_m$. If $\gamma \geq \xi > 0$, $\eta > 0$, and $m \leq 2^{(1-\epsilon)\cdot\xi}$ for some $0 < \epsilon \leq 0.83/(n-2)$, then for any even constant integer $n \geq 4$, as $m \to \infty$, H is a strong (ξ, η, ε)-biased-seed-OBF extractor per Definitions 9 and 12, where*

$$\varepsilon \leq \frac{n-1}{2}\sqrt[n]{\frac{m^{\gamma}}{2^{\eta}}}\sqrt{2^{-\epsilon\cdot\xi}}\,(1 + o(1))$$

Furthermore, H is a strong $(\xi, \eta, \varepsilon')$-biased-seed-SAOBF extractor per Definitions 11 and 12, where

$$\varepsilon' \leq \frac{n-1}{2}\sqrt[n]{\frac{m^{\gamma}\binom{\gamma}{\xi}}{2^{\eta}}}\sqrt{2^{-\epsilon\cdot\xi}}\,(1 + o(1))$$

References

1 Michael Ben-Or and Nathan Linial. Collective coin flipping. *Adv. Comput. Res.*, 5:91–115, 1989.

2 Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 429–442, 1985. doi:10.1109/SFCS.1985.62.

3 Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988. doi:10.1137/0217015.

4 Richard Cleve. Limits on the security of coin flips when half the processors are faulty (extended abstract). In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, pages 364–369, 1986. doi:10.1145/12130.12168.

5 Ran Cohen, Jack Doerner, Yashvanth Kondi, and abhi shelat. Secure multiparty computation with identifiable abort via vindicating release. In *Advances in Cryptology - CRYPTO 2024 - 44th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2024, Proceedings, Part VIII*, 2024. doi:10.1007/978-3-031-68397-8_2.

6 Jean Paul Degabriele, Jan Gilcher, Jérôme Govinden, and Kenneth G. Paterson. Sok: Efficient design and implementation of polynomial hash functions over prime fields. In *IEEE Symposium on Security and Privacy, SP 2024, San Francisco, CA, USA, May 19-23, 2024*, pages 3128–3146. IEEE, 2024. doi:10.1109/SP54263.2024.00132.

7 Jack Doerner, Iftach Haitner, Yuval Ishai, and Nikolaos Makriyannis. From OT to OLE with subquadratic communication. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security, CCS 2025, Taipei, Taiwan, October 13-17, 2025*, pages 2249–2263. ACM, 2025. doi:10.1145/3719027.3765225.

8 Jack Doerner, Yashvanth Kondi, Eysa Lee, and abhi shelat. Secure two-party threshold ECDSA from ECDSA assumptions. In *Proceedings of the 39th IEEE Symposium on Security and Privacy (S&P)*, 2018.

9 Jack Doerner, Yashvanth Kondi, Eysa Lee, and abhi shelat. Threshold ECDSA from ECDSA assumptions: The multiparty case. In *2019 IEEE Symposium on Security and Privacy, SP 2019, San Francisco, CA, USA, May 19-23, 2019*, pages 1051–1066. IEEE, 2019. doi:10.1109/SP.2019.00024.

10 Jack Doerner, Yashvanth Kondi, Eysa Lee, and abhi shelat. Threshold ECDSA in three rounds. In *Proceedings of the 45th IEEE Symposium on Security and Privacy (S&P)*, 2024.

11 Jack Doerner, Yashvanth Kondi, Eysa Lee, abhi shelat, and LaKyah Tyner. Threshold BBS+ signatures for distributed anonymous credential issuance. In *Proceedings of the 44th IEEE Symposium on Security and Privacy (S&P)*, 2023.

12 Jack Doerner and Lawrence Roy. Tighter bounds for the oblivious bit-fixing inner product extractor on biased seeds. Cryptology ePrint Archive, Paper 2026/654, 2026. URL: https://eprint.iacr.org/2026/654.

13 Dean Doron and Ori Fridman. Bit-Fixing Extractors for Almost-Logarithmic Entropy. In Alina Ene and Eshan Chattopadhyay, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2025)*, volume 353 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:19, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.APPROX/RANDOM.2025.33.

14 Shimon Even, Oded Goldreich, and Abraham Lempel. A randomized protocol for signing contracts. *Communications of the ACM*, 28(6):637–647, 1985. doi:10.1145/3812.3818.

15 Niv Gilboa. Two party RSA key generation. In *Annual International Cryptology Conference (CRYPTO)*, pages 116–129, 1999. doi:10.1007/3-540-48405-1_8.

16 Iftach Haitner, Nikolaos Makriyannis, Samuel Ranellucci, and Eliad Tsfadia. Highly efficient ot-based multiplication protocols. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 180–209, 2022. doi:10.1007/978-3-031-06944-4_7.

17 Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Pseudo-random generation from one-way functions (extended abstracts). In David S. Johnson, editor, *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washington, USA*, pages 12–24. ACM, 1989. doi:10.1145/73007.73009.

18 Yuval Ishai, Manoj Prabhakaran, and Amit Sahai. Founding cryptography on oblivious transfer - efficiently. In David A. Wagner, editor, *Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings*, volume 5157 of *Lecture Notes in Computer Science*, pages 572–591. Springer, 2008. doi:10.1007/978-3-540-85174-5_32.

19 Marcel Keller, Emmanuela Orsini, and Peter Scholl. MASCOT: faster malicious arithmetic secure computation with oblivious transfer. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 830–842, 2016. doi:10.1145/2976749.2978357.

20 Chia-Jung Lee, Chi-Jen Lu, Shi-Chun Tsai, and Wen-Guey Tzeng. Extracting randomness from multiple independent sources. *IEEE Trans. Inf. Theory*, 51(6):2224–2227, 2005. doi:10.1109/TIT.2005.847746.

21 Moni Naor and Benny Pinkas. Oblivious polynomial evaluation. *SIAM Journal on Computing*, 35(5):1254–1281, 2006. doi:10.1137/S0097539704383633.

22 Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. Syst. Sci.*, 52(1):43–52, 1996. doi:10.1006/JCSS.1996.0004.

23 M. O. Rabin. How to exchange secrets by oblivious transfer. TR-81, Harvard, 1981.

24 Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. doi:10.1561/0400000010.

25 Umesh Vazirani. Towards a strong communication complexity theory or generating quasi-random sequences from two communicating slightly-random sources. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, STOC '85, pages 366–378, New York, NY, USA, 1985. Association for Computing Machinery. doi:10.1145/22145.22186.