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

Jack Doerner
j@ckdoerner.net
University of Virginia

Lawrence Roy
ldr709@gmail.com
Aarhus University

June 15, 2026

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

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

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).

Contents

1 Introduction **1**
- 1.1 Our Contributions . . . 3
- 1.2 Abbreviated Technical Overview . . . 5
- 1.3 Applications to Malicious OLE-from-OT Protocols via Seed-Adaptive Oblivious Bit Fixing . . . 7
- 1.4 Concrete Performance Improvements for the IPE under Seed-Adaptive Oblivious Bit-Fixing . . . 9
- 1.5 TLDR: The Bottom Line for the Security of DKLs OLE . . . 11

2 Preliminaries **11**
- 2.1 Useful Combinatorial Notions . . . 12
- 2.2 Randomness Extractors . . . 12
- 2.3 Universal Hashing and Exceptional Inputs . . . 14
- 2.4 Matroids . . . 16
- 2.5 The Tutte and Coboundary Polynomials . . . 18
- 2.6 On the Sensitivity of Dimension and Rank to Field Order . . . 19

3 Restricting the Problem **19**
- 3.1 The Sufficiency of Uniform-Over-Subset Seed Distributions . . . 20
- 3.2 The (In)Sufficiency of Uniform Source Distributions . . . 21
- 3.3 Seed-Adaptive Oblivious Bit-Fixing Sources . . . 23

4 The Elevated General Leftover Hash Lemma **24**
- 4.1 A Family-Agnostic n^{th} GLHL . . . 24
- 4.2 Applying Lemma 1.3 to the IPE . . . 25

5 A Simple Bound for the IPE at $n = 4$ **27**
- 5.1 Bounding $\mathsf{EIF}(4)$ . . . 27
- 5.2 Putting it Together . . . 28

6 The Exact Multicollision Probability of the IPE **31**
- 6.1 An Exponential Generator for the Accumulation of Exponentiated Ranks . . . 32
- 6.2 Multicollision Probability in Terms of Coboundary Polynomials . . . 36

7 Tighter Bounds for the IPE at $n \geq 4$ **38**
- 7.1 Bounding Statistical Distance for any Even n . . . 38
- 7.2 Computing Coboundary Polynomials in $2^{O(n^2)}$ Time . . . 39
- 7.3 Asymptotic Bounds for any Fixed Even n . . . 40

A Concrete Bounds at $n \in \{4, 6, 8\}$ **51**

B Some Coboundary Polynomials **52**

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 [NZ96], 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 [CG85], 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:

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

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 [ILL89] 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

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

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

for any *arbitrarily biased* seed distribution $\mathcal{G}$ over $|\mathbb{F}|^\gamma$ that is independent of $\mathcal{X}$ [Vaz85, CG88, LLTT05].

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 [DF25] 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 [KOS16, DKLs18, DKLs19, HMRT22, DKLs24, CDKs24]. 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 [Cle86]. 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

[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. [DGGP24] 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.

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.

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 5.2 in Section 5, but its simplest and clearest statement is via the following:

Corollary 1.1 (of Theorem 5.2)**.** *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 2.11) 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 2.6) 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 2.8) 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. We therefore report our computer-assisted bounds in Appendix A. Using these concrete bounds, we show that the optimal n depends on $|\mathbb{F}|$, γ, and $H_\infty(\mathcal{W})$.

4. 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 7.7 in Section 7.3, but its simplest and clearest statement is via the following:

Corollary 1.2 (of Theorem 7.7)**.** *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 2.11) 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 2.6) source distribution $\mathcal{W}$ on W with $H_\infty(\mathcal{W}) \geq \xi$, as $m \to \infty$ we have*

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

[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.

where $\mathcal{Y}$ is the uniform distribution on Y. For any seed-adaptive *oblivious bit-fixing (Definition 2.8) 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 [DKLs18, DKLs19, DKLs24] 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. [LLTT05] by a factor of more than 25, 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 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 1.3 (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 2.14), 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 1.1. 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 Section 7.1:

Theorem 1.4. *Let $H = \mathbb{F}_m^{\xi}$ be an IPE family (Definition 2.11) 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 (Definition 2.30) of the cube matroid (Definition 6.10), 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 Appendix A. 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 1.2.

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

Oblivious Transfer (OT) [Rab81, EGL85] is a two-party protocol that allows a *receiver* to select one of two messages supplied by a *sender*. *Oblivious Linear Evaluation* (OLE) [NP06] 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 [Gil99], 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 [IPS08], 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 [KOS16], 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. [DKLs18, DKLs19, DKLs24]. The last of these remains the state of the art. Haitner et al. [HMRT22] 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. [DHIM25] took

[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.

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.

Fixing A Bug in the DKLs OLE Protocols. The DKLs OLE protocols [DKLs18, DKLs19, DKLs24] 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 [ILL89], 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 Section 1.4. Doerner et al. [DKLs24] 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 7, from 268 to 38. The improvement is even greater if $|m|= 384$ is used (as might be required when working with pairing curves [DKL+23], for example): in this case the number of extra OTs reduces from 295 to 11, a factor of more than 25!

Removing a Round from the HMRT weak-OLE Protocol. The highly communication-efficient weak-OLE protocol of Haitner et al. [HMRT22, Protocol 4.10] makes use of the inner product extractor in a similar manner to the DKLs protocols. Unlike the Doerner et al., Haitner et al. specify that the seed is sampled by the OT receiver and transmitted to the sender *after* the sender is committed to its OT inputs. This adds one round, but ensures that the sender cannot condition its inputs (which it can use to execute a selective failure attack) on the value of the seed. In the "Almost-monochromatic d" case of their proof, they exclude seed and source elements corresponding to locations where the sender has provided incorrect OT inputs, just as Doerner et al. do, and then apply their Lemma 3.8 (derived from the LHL) which insists that the pruned seed and the pruned source be independent and uniform. Because of the extra round in their protocol, this independence is enforced, and unlike Doerner et al. they do not have a bug. However, we suggest that our results in this work could be used to derive a version of their Lemma 3.8 which permits bias in the (pruned) seed, and that their protocol could then be modified to sample the seed *before* the sender becomes committed to its OT inputs, saving a round. We state this as a suggestion because a proof is outside the scope of this work. If Haitner et al.'s protocol were modified as we suggest, it would match the DKLs protocols in terms of both round count and assumptions, revealing a simple trade-off: their protocol realizes a weaker functionality, and thereby saves roughly a factor of two in communication.

1.4 Concrete Performance Improvements for the IPE under Seed-Adaptive Oblivious Bit-Fixing

To demonstrate the concrete impact of the bounds we present in this paper, we use our bounds (along with the bound of Lee et al. [LLTT05], for comparison) to compute how many *extra* source bits are required to fix the bug of Doerner et al. that was described in Section 1.3. In particular, we assume that the seed is uniform, the adversary fixes c bits of the source, and the distinguishing advantage is the extractor's error, scaled by 2^{-c}.[6] To fix the bug we require sufficiently many source bits to guarantee that the advantage is less than 2^{-s} for every $0 \leq c \leq 2s$. That is, if we let ε_c denote the error of the IPE when exactly c bits are fixed, then we insist that $\varepsilon_c \cdot 2^{-c} \leq 2^{-s}$ for every $0 \leq c \leq s$.

[6] In the DKLs proof, each fixed bit induces an abort with probability 1/2 (independently of one another). The extractor's output is hidden from the adversary if an abort occurs, leading to the scaling factor.

As a baseline, the classic LHL [ILL89] tells us that $|m|+2s$ source bits are required to achieve the above criteria if the bits are fixed *statically*, as Doerner et al. erroneously assumed. We report the number of extra bits required to achieve the same criteria under *adaptive* oblivious bit fixing in Table 1.5. For $n = 2$, these values were computed using the bound of Lee et al. [LLTT05]. For $n = \{4, 6, 8\}$, we compute the values using our matroid-based concrete bounds, which are derived from Theorem 1.4 via Lemmas 3.1 and 3.5. For $n = 4$ we additionally compute values using Theorem 5.2 via Lemmas 3.1 and 3.5, which avoids matroids by reasoning about exceptional inputs. These computations were all performed using a `sagemath` script which incremented the source length until the criteria mentioned above was satisfied under each of the bounds.

s	n	Bound	$\mu = 32$	$\mu = 64$	$\mu = 128$	$\mu = 192$	$\mu = 256$	$\mu = 384$	$\mu = 512$	$\mu = 1024$
40	2	[LLTT05]	102	113	130	142	152	167	178	209
40	4	Theorem 5.2	47	41	36	43	49	58	65	82
40	4	Theorem 1.4	**45**	39	36	43	49	58	65	82
40	6	Theorem 1.4	54	42	**18**	15	19	25	30	41
40	8	Theorem 1.4	76	61	32	**5**	**7**	**10**	**12**	**21**
40	best	Theorem 1.4	**45**	**38**	**18**	**5**	**7**	**10**	**12**	**21**
60	2	[LLTT05]	151	164	184	199	212	232	247	290
60	4	Theorem 5.2	73	68	58	55	63	76	85	110
60	4	Theorem 1.4	**71**	**66**	56	55	63	76	85	110
60	6	Theorem 1.4	89	77	52	**28**	21	29	35	52
60	8	Theorem 1.4	122	108	78	49	19	**10**	**13**	**24**
60	best	Theorem 1.4	**71**	**66**	**49**	**28**	**14**	**10**	**13**	**24**
80	2	[LLTT05]	199	213	235	253	268	292	312	365
80	4	Theorem 5.2	100	95	85	74	74	90	102	134
80	4	Theorem 1.4	**98**	**93**	83	72	74	90	102	134
80	6	Theorem 1.4	123	111	87	63	**38**	31	39	61
80	8	Theorem 1.4	169	154	125	95	66	**11**	**14**	**26**
80	best	Theorem 1.4	**98**	**93**	**79**	**60**	**38**	**11**	**14**	**26**

Table 1.5: **Source Length Overhead.**[7] The classic LHL tells use that we need $\mu + 2s$ source bits, where $\mu = |m|$ is the size of the modulus over which the IPE is evaluated, in order to achieve $\varepsilon_c \cdot 2^{-c} \leq 2^{-s}$ for all $0 \leq c \leq s$, where ε_c is the error of the IPE on a source with exactly c statically obliviously fixed bits. This table reports the number of *extra* source bits (i.e. the number beyond $\mu + 2s$) required to achieve the same criteria when the source bits are fixed *seed-adaptively* rather than statically. The "best" row quantifies over all c, and then for each c it quantifies over $n \in \{2, 4, 6, 8\}$, whereas the other rows fix a single value of n first, and then quantify over all c.

[7] In versions of this paper dated June 8 2026 and earlier, the numbers reported in this table were incorrect.

1.5 TLDR: The Bottom Line for the Security of DKLs OLE

For deployers of the DKLs protocols, this work raises two critical questions:

- *Is the DKLs protocol vulnerable to attack if it is implemented as originally described (in the versions of their paper dated December 2023 and earlier)?*

  Although the DKLs protocol is not as secure as originally claimed, the drop in security is so minor that we do *not* recommend modifying existing implementations, particularly if those implementations target 80 bits of statistical security (i.e. $s = 80$) as recommended by DKLs.

  As described in the previous subsection, the parameters originally suggested by the authors of DKLs were chosen to ensure that $\varepsilon_c \cdot 2^{-c} \leq 2^{-s}$ for $0 \leq c \leq s$, where ε_c is the error of the IPE on a source with exactly c seed-adaptively and obliviously fixed bits. They erroneously used the static (not seed-adaptive) bound of $\varepsilon_c \leq 2^{(\mu-\gamma+c)/2}$, and so specified that $\gamma = \mu + 2s$ (i.e. $\xi = \kappa + 2s$ in their notation). When $\mu = 256$ and $s = 80$, as recommended by DKLs, we can use Theorem 1.4 to prove that their parameterization yields $\varepsilon_c \cdot 2^{-c} \leq 2^{-69.8}$ for every $0 \leq c \leq s$ (where $n \in \{2, 4, 6, 8\}$ is chosen as the optimal number for each c). Similarly, when $\mu = 256$ and $s = 40$, we can prove that $\varepsilon_c \cdot 2^{-c} \leq 2^{-36.9}$ for every $0 \leq c \leq s$. Thus their error results in a statistical security loss of no more than 10.2 bits, when targeting 80 bits, or no more than 3.1 bits, when targeting 40 bits.

- *What changes must be made to the DKLs protocol in order to achieve the level of security it originally targeted?*

  When signing over a 256-bit curve and targeting 80 bits of statistical security (i.e. $\mu = 256, s = 80$), we recommend to set $\gamma = 454$ in future implementations. In the notation of DKLs, this means setting $\xi = 454$, which is an increase of 38 relative to the original number.

  When signing over a 256-bit curve and targeting 40 bits of statistical security (i.e. $\mu = 256, s = 40$), we recommend to set $\gamma = 343$ in future implementations. In the notation of DKLs, this means setting $\xi = 343$, which is an increase of 7 relative to the original number.

  For other combinations of parameters, the required increase in γ (i.e. ξ in the notation of DKLs) is given as the *Source Length Overhead* in Table 1.5.

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 Section 2.4.

2.1 Useful Combinatorial Notions

Definition 2.1 (The Falling Factorial Function)**.**

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

Definition 2.2 (Stirling Numbers of the First Kind)**.** *The* unsigned *first-kind Stirling number* $\left[{i \atop j}\right]$ *counts the number of permutations of* i *elements that have exactly* j *disjoint cycles. It is also the absolute value of the coefficient of the* j^{th} *power in the expansion of the* i^{th} *falling factorial. That is, Stirling numbers of the first kind are defined such that*

$$x^{\underline{i}} = \sum_{j=0}^{i} (-1)^{i-j} \cdot \left[{i \atop j}\right] \cdot x^j$$

Definition 2.3 (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:*

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

Definition 2.4 (Exponential Generators)**.** *For any function family* $f_i(x)$ *that is well defined over* $i \in \mathbb{N}$*, an* exponential generator *of the family is a function* $F(\alpha, x)$ *such that*

$$f_i(x) = \left(\frac{d}{d\alpha}\right)^i F(\alpha, x)\Big|_{\alpha=0}$$

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.[8] The strongest notion we will consider takes input from an arbitrary distribution of sufficient min-entropy.

Definition 2.5 ((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 $\textsc{Ext} : H \times X \to Y$, usually written $h(x)$ for $h \in H$[9] and $x \in X$, such that $\mathsf{SD}((\mathcal{H}(\mathcal{X}), \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \leq \varepsilon$,[10] 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 2.6 (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 $f \in \mathbb{1}^{\gamma - \xi}$.*

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 [BL89]. 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 2.7 (Oblivious Bit-Fixing (OBF) Extractors)**.** *A* strong $(\xi, \mathcal{H}, \varepsilon)$-OBF extractor *is a strong $(\mathcal{W}, \mathcal{H}, \varepsilon)$-extractor per Definition 2.5 for every source $\mathcal{W}$ that is (ξ, γ)-*oblivious-bit-fixing *for some $\gamma \geq \xi$, per Definition 2.6.*

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 2.8 (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 2.9 (Seed-Adaptive OBF Extractors)**.** *A* strong $(\xi, \mathcal{H}, \varepsilon)$-SAOBF extractor *is a strong $(\mathcal{W}, \mathcal{H}, \varepsilon)$-extractor per Definition 2.5 for every source $\mathcal{W}$ that is (ξ, γ)-*seed-adaptive-oblivious-bit-fixing *for some $\gamma \geq \xi$, per Definition 2.8.*

[8] *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.

[9] We often refer to H as an *extractor family* and think of $h \in H$ as a function $h : X \to Y$.

[10] The $\mathcal{H}$'s in this equation are the same random variable, not independent samples from $\mathcal{H}$.

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 2.10 (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 2.11 (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 2.12 (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 2.13 (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 2.14 (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 2.15 (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 2.16 (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,*

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

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

Definition 2.17 (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 2.18. *For any ℓ-universal hash function family, we have the following generic equalities and bounds on* EIF*:*

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

Proposition 2.19. *For any ℓ-universal hash function family $H : X \to Y$, we have the following generic bounds on* CPDI*:*

$$\mathsf{CPDI}(j) \le |Y|^{1-j} + \sum_{k=\ell+1}^{j} \mathsf{EIF}(j,k)\left(|Y|^{2-k} - |Y|^{1-k}\right) \tag{2}$$

$$\le |Y|^{1-j} + \mathsf{EIF}(j)|Y|^{1-\ell} \tag{3}$$

$$\begin{aligned} \mathsf{CPDI}(j) &\le \mathsf{CPDI}(j-1) \\ \mathsf{CPDI}(j) &\le |Y|^{1-j} && \text{if } j \le \ell \end{aligned}$$

Proof of Equation (2). Beginning with Definition 2.14, and making use of Definitions 2.15 to 2.17, we have

$$\begin{aligned} \mathsf{CPDI}(j) &= \sum_{(x_1,\ldots,x_j)\in\mathsf{DI}(j)} \Pr_{h\leftarrow\mathcal{H}}\left[h(x_{i_1}) = h(x_{i_2})\forall i_1,i_2\right] \cdot \Pr_{(x'_1,\ldots,x'_j)\leftarrow\mathcal{U}_{\mathsf{DI}(j)}}\left[x'_i = x_i \forall i\right] \\ &= \frac{1}{|\mathsf{DI}(j)|} \sum_{(x_1,\ldots,x_j)\in\mathsf{DI}(j)} \Pr_{h\leftarrow\mathcal{H}}\left[h(x_{i_1}) = h(x_{i_2})\forall i_1,i_2\right] \\ &\le \frac{1}{|\mathsf{DI}(j)|} \sum_{(x_1,\ldots,x_j)\in\mathsf{DI}(j)-\mathsf{EI}(j,j)} |Y|^{1-j} + \frac{1}{|\mathsf{DI}(j)|} \sum_{(x_1,\ldots,x_j)\in\mathsf{EI}(j,\ell+1)} |Y|^{1-\ell} \end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{|\mathsf{DI}(j)|} \sum_{k=\ell+2}^{j} \sum_{(x_1,\ldots,x_j)\in \mathsf{EI}(j,k)-\mathsf{EI}(j,k-1)} |Y|^{2-k} \\
&= \frac{|\mathsf{DI}(j)|-|\mathsf{EI}(j,j)|}{|\mathsf{DI}(j)|}|Y|^{1-j} + \frac{|\mathsf{EI}(j,\ell+1)|}{|\mathsf{DI}(j)|}|Y|^{1-\ell} \\
&\quad + \sum_{k=\ell+2}^{j} \frac{|\mathsf{EI}(j,k)|-|\mathsf{EI}(j,k-1)|}{|\mathsf{DI}(j)|}|Y|^{2-k} \\
&= |Y|^{1-j} + \sum_{k=\ell+1}^{j} \frac{|\mathsf{EI}(j,k)|}{|\mathsf{DI}(j)|}\left(|Y|^{2-k}-|Y|^{1-k}\right)
\end{aligned}
$$

□

Proposition 2.20. *For any function family $H : X \to Y$, we have the following generic equalities for* CP*:*

$$
\begin{aligned}
\mathsf{CP}(i) &= \sum_{y\in Y} \Pr_{h\leftarrow\mathcal{H}}^{x\leftarrow\mathcal{X}}[h(x)=y]^i \\
\mathsf{CP}(0) &= |Y| \\
\mathsf{CP}(1) &= 1 \\
\mathsf{CP}(i) &= \sum_{j=1}^{i} \left\{ {i \atop j} \right\} \frac{|X|^{\underline{j}}}{|X|^i} \mathsf{CPDI}(j) \qquad \text{if } i > 0 \qquad (4)
\end{aligned}
$$

Proof of Equation (4). Beginning with Definition 2.12, we observe that we can partition the event space according to the number of inputs that are distinct. Combining this with Definitions 2.3, 2.13 and 2.14 yields

$$
\begin{aligned}
\mathsf{CP}(i) &= \Pr_{\substack{h\leftarrow\mathcal{H} \\ (x_1,\ldots,x_i)\leftarrow\mathcal{X}^i}} [h(x_{i_1}) = h(x_{i_2}) \forall i_1, i_2] \\
&= \sum_{j=1}^{i} \Pr_{(x_1,\ldots,x_i)\leftarrow\mathcal{X}^i} [(x_1,\ldots,x_i) \text{ contains } j \text{ distinct values}] \cdot \mathsf{CPDI}(j) \\
&= \sum_{j=1}^{i} \frac{\begin{pmatrix} \text{number of ways to partition} \\ i \text{ values into exactly } j \text{ subsets} \end{pmatrix} \cdot (|\mathsf{DI}(j)|)}{|X|^i} \cdot \mathsf{CPDI}(j)
\end{aligned}
$$

□

2.4 Matroids

In this subsection, we give a brief review of basic matroid theory, and in Section 2.5 we introduce several useful polynomials and theorems related to matroids. For a deeper and more comprehensive introduction to the topic, we refer the reader to Oxley [Oxl11]. There are multiple equivalent definitions of a matroid. We choose to root our definition in the rank function.

Definition 2.21 (Matroid)**.** *A matroid M is characterized by a* ground set *of elements, denoted E_M, and a* rank function rank_M *that is well defined on any subset of E_M.* rank_M *must satisfy the following properties:*

1. *For any* $A \subseteq E_M$, $0 \leq \text{rank}_M(A) \leq |A|$.
2. *For any* $A \subset B \subseteq E_M$, $\text{rank}_M(A) \leq \text{rank}_M(B)$. *That is,* rank_M *must be* order-preserving.
3. *For any* $A, B \subseteq E_M$, $\text{rank}_M(A \cup B) + \text{rank}_M(A \cap B) \leq \text{rank}_M(A) + \text{rank}_M(B)$. *That is,* rank_M *must be a* submodular set function.

By convention, $\text{rank}_M(M) = \text{rank}_M(E_M)$.

Definition 2.22 (Independent Sets and Bases of Matroids)**.** *An* independent set *of matroid* M *is any subset* $I \subseteq E_M$ *such that* $\text{rank}_M(I) = |I|$. *We use* $\mathcal{I}_M$ *to denote the set of all independent sets of* M. *A* basis *of* M *is a maximal independent set. That is, a basis is any set* $B \in \mathcal{I}_M$ *such that* $\text{rank}_M(B) = \text{rank}_M(M)$. *We use* $\mathcal{B}_M \subseteq \mathcal{I}_M$ *to denote the set of all bases of* M.

Definition 2.23 (Flats of Matroids)**.** *A* flat *of matroid* M *is any subset* $F \subseteq E_M$ *such that for every* $\eta \in E_M - F$, $\text{rank}_M(F \cup \{\eta\}) > \text{rank}_M(F)$. *In other words, a flat is a set to which no element can be added without increasing its rank. We use* $\mathcal{F}_M$ *to denote the set of all flats of* M.

Definition 2.24 (Restriction, Deletion, Contraction, and Matroid Minors)**.** *A matroid* R *is a* restriction *of matroid* M *to set* S, *denoted* $R = M|S$, *if* $E_R = S$ *and* $\mathcal{I}_R = \{I : I \in \mathcal{I}_M \wedge I \subseteq S\}$. *The rank function of the restricted matroid is defined such that* $\text{rank}_R(A) = \text{rank}_M(A)$ *for any* $A \subseteq E_R$. *If* $D = E_M - S$, *then we can equivalently say that* R *is a* deletion *of* D *from* M, *and denote this operation* $R = M \setminus D$.

A matroid C *is a* contraction *of matroid* M *by set* S, *denoted* $C = M/S$, *if* $E_C = E_M - S$, $\mathcal{I}_C = \{I - S : I \in \mathcal{I}_M\}$, *and* $\text{rank}_C(A) = \text{rank}_M(A \cup S) - \text{rank}_M(S)$.

A matroid N *is a* minor *of matroid* M *if it can be arrived at from* M *by a sequence of restrictions, deletions, and contractions.*

Remark 2.25. *Notice that if for some matroid* M *we have* $S \subseteq E_M$, $R = M|S$, *and* $C = M/S$, *then*

$$\begin{aligned}\text{rank}_R(R) + \text{rank}_C(C) &= \text{rank}_M(E_R) + \text{rank}_M(E_C \cup S) - \text{rank}_M(S) \\ &= \text{rank}_M(M).\end{aligned}$$

Definition 2.26 (Loops and Coloops of Matroids)**.** *An element* $\eta \in E_M$ *is a* loop *of matroid* M *if it is not a member of any independent set of* M. *An element* $\eta \in E_M$ *is a* coloop *of matroid* M *if it is a member of every basis of* M. *In other words,* η *is a loop of* M *if* $\forall I \in \mathcal{I}_M, \eta \notin I$, *and* η *is a coloop of* M *if* $\forall B \in \mathcal{B}_M, \eta \in B$.

Definition 2.27 (Disjointness and Disjoint Unions of Matroids)**.** *Matroids* M *and* N *are said to be* disjoint *if* $E_M \cap E_N = \varnothing$. *The* disjoint union *(or* direct sum*) of matroids* M *and* N, *denoted* $M \oplus N$ *is the matroid defined by the ground set* $E_{M \oplus N} = E_M \cup E_N$ *and independent sets* $\mathcal{I}_{M \oplus N} = \{I \cup J :$

$I \in \mathcal{I}_M, J \in \mathcal{I}_N\}$. *By definition,* $M \oplus N$ *has rank function* $\operatorname{rank}_{M\oplus N}(A) = \operatorname{rank}_M(A \cap E_M) + \operatorname{rank}_N(A \cap E_N)$. *We assert that the component matroids of a disjoint union are indeed disjoint by adding distinct labels to the elements of their ground sets, if necessary.*

Remark 2.28. *Even if it happens to be the case that two matroids* M *and* N *are disjoint, and their rank functions are well-defined one one-another's ground sets, it is not necessarily the case that* $\operatorname{rank}_{M\oplus N}(A) = \operatorname{rank}_M(A) = \operatorname{rank}_N(A)$ *for some* $A \subseteq E_{M\oplus N}$.

2.5 The Tutte and Coboundary Polynomials

Definition 2.29 ([MRIS12, Definitions 2.1 and 2.6])**.** *If* M *is a matroid with a totally ordered ground set* E_M *and a basis set* $\mathcal{B}_M$*, then the* Tutte polynomial *of* M *is defined by*

$$\begin{aligned} T_M(x,y) &= \sum_{A\subseteq E_M} (x-1)^{\operatorname{rank}_M(M)-\operatorname{rank}_M(A)}(y-1)^{|A|-\operatorname{rank}_M(A)} \\ &= \sum_{B\in\mathcal{B}_M} x^{\iota(M,B)}y^{\varepsilon(M,B)} \end{aligned}$$

where $\iota(M,B)$ *and* $\varepsilon(M,B)$ *compute the number of internally and externally active elements in* M *with respect to* B.[11]

Definition 2.30 ([MRIS12, Equations 20–21])**.** *If* M *is a matroid with ground set* E_M*, then the* coboundary polynomial $\overline{\chi}_M(q,t)$ *and the* characteristic polynomial $\chi_M(q)$ *of* M *are defined by*

$$\begin{aligned} \overline{\chi}_M(q,t) &= \sum_{X\in\mathcal{F}_M} t^{|X|}\chi_{M/X}(q) \\ \chi_M(q) &= \sum_{X\subseteq E_M} (-1)^{|X|}q^{\operatorname{rank}_M(M)-\operatorname{rank}_M(X)} \end{aligned}$$

Lemma 2.31 ([MRIS12, Equation 23])**.** *For any matroid* M*, the Tutte polynomial and the coboundary polynomial of* M *are related by:*

$$\overline{\chi}_M(q,t) = (t-1)^{\operatorname{rank}_M(M)}T_M\left(\frac{q+t-1}{t-1}, t\right)$$

Definition 2.32 ([Wel99, Theorem 2.16])**.** *Let* $\mathcal{M}$ *be a class of matroids that is closed under disjoint unions and the taking of minors. A function* f *is said to be a* Tutte–Grothendieck Invariant *of* $\mathcal{M}$ *if it satisfies the following two conditions:*

1. $\forall M \in \mathcal{M}\exists(a,b)$ s.t. $\forall\hat{\eta} \in E_M, f(M) = a \cdot f(M \setminus \{\hat{\eta}\}) + b \cdot f(M/\{\hat{\eta}\})$
2. $\forall M, N \in \mathcal{M}, f(M \oplus N) = f(M) \cdot f(N)$

[11] Since we do not use internal and external activity elsewhere in this work, we do not define them further.

Lemma 2.33 ([Wel99, Theorem 2.16])**.** *Let $\mathcal{M}$ be a class of matroids that is closed under disjoint unions and the taking of minors, and let f be a Tutte–Grothendieck Invariant of $\mathcal{M}$ (see Definition 2.32). For any $M \in \mathcal{M}$,*

$$f(M) = a^{|E_M| - \text{rank}_M(M)} b^{\text{rank}_M(M)} T_M\left(\frac{x_0}{b}, \frac{y_0}{a}\right)$$

where (a, b) are as in Definition 2.32, and x_0 is the value that f assumes on a matroid that has a ground set comprising a single coloop, and y_0 is the values f assumes on a matroid that has a ground set comprising a single loop.

2.6 On the Sensitivity of Dimension and Rank to Field Order

We now introduce a proofs that dim of a set of vectors (and rank of a matrix) is only sensitive to the order of the field in which it is evaluated when the order of that field is small, relative to the number of vectors in the set (or the length of the short side of the matrix). This allows us to (mostly) ignore the field over which dim and rank are calculated for the rest of this work.

Lemma 2.34. *Let $A \in (\mathbb{1}^{\xi})^{i}$ be a set of vectors with boolean entries. If for some $k > 0$, $\dim_{\mathbb{R}}(A) = k$, and if $m > 2^{-k} \cdot (k+1)^{(k+1)/2}$, then $\dim_{\mathbb{F}_m}(A) = k$.*

Proof. Let w be a matrix whose columns are the vectors in A. Because $\dim_{\mathbb{R}}(A) = k$, there must exist some minor w' of matrix w such that $w' \in \mathbb{1}^{k \times k}$ and $\det_{\mathbb{R}}(w') \neq 0$. By Hadamard's inequality, we also know that $|\det(w')| \leq 2^{-k} \cdot (k+1)^{(k+1)/2} < m$, and thus $\det(w') \not\equiv 0 \pmod{m}$, which implies $\text{rank}_{\mathbb{F}_m}(w') = k$. This finally implies that $\text{rank}_{\mathbb{F}_m}(w) = k$ and thus $\dim_{\mathbb{F}_m}(A) = k$. □

Corollary 2.35. *Let $A \in (\mathbb{1}^{\xi})^{i}$ be a set of vectors with boolean entries. If for some $m' \geq 2$ and $k > 0$, $\dim_{\mathbb{F}_{m'}}(A) = k$, and if $m > 2^{-k} \cdot (k+1)^{(k+1)/2}$, then $\dim_{\mathbb{F}_m}(A) \geq k$.*

Remark 2.36. *In this work, we predominantly consider relatively low-dimension vectors (i.e. low-rank matrices, implicitly). In particular, our n^{th}-root bounds depend upon the dimension of sets of vectors of length at most n, or the ranks of matrices whose short sides are of length at most n. The largest value of n for which we develop concrete bounds is 8. For $n \leq 21$, we know the maximal determinant (over the real numbers) among all $n \times n$ boolean matrices [OEI21], which can be considerably better than what is implied by Hadamard's inequality. This implies the concrete lower bounds for dimensional/rank invariance reported in Table 2.37, via reasoning similar to that used in the proof of Lemma 2.34.*

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.

n	4	6	8	10	12	14	16	18	20
$\min(m)$	4	10	57	321	3646	25516	327681	3411969	56640626

Table 2.37: **Minimal Field Order for Dimensional Invariance.** For every $m \geq \min(m)$, every $\xi \in \mathbb{N}^+$, and every $A \in (\mathbb{1}^\xi)^n$, $\dim_{\mathbb{F}_m}(A) = \dim_{\mathbb{R}}(A)$.

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 [Vad12, 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, and include a relatively standard proof for the sake of completeness. Hereafter in this document, we will restrict our focus to seed distributions $\mathcal{G}$ that are uniform over subsets $G \subseteq H$.

Lemma 3.1 (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 2.5 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 2.10.*

Proof. Consider the following inequality, defined over all pairs $(\mathcal{H}, p)$ such that $\mathcal{H}$ is an arbitrary distribution over H and $0 \leq p \leq 1$.

$$\mathsf{SD}((\mathcal{H}(\mathcal{X}), \mathcal{H}), (\mathcal{Y}, \mathcal{H})) \leq \varepsilon(p) \tag{5}$$

The lemma holds if Equation (5) holds for all pairs of the form $(\mathcal{H}, 2^{-H_\infty(\mathcal{H})})$, under the *assumption* that Equation (5) holds for all pairs of the form $(\mathcal{G}, 2^{-H_\infty(\mathcal{G})})$, where $\mathcal{G}$ is the uniform distribution over some subset $G \subseteq H$.

Notice that the left side of Equation (5) is a convex function of the probability distribution $\mathcal{H}$. That is, both the real and random distributions are linear functions of $\mathcal{H}$, and statistical distance is convex. Since ε is concave, we have that $\mathsf{SD}((\mathcal{H}(\mathcal{X}), \mathcal{H}), (\mathcal{Y}, \mathcal{H})) - \varepsilon(p)$ is a convex function of $(\mathcal{H}, p)$. Therefore, the set of pairs $(\mathcal{H}, p)$ that satisfy Equation (5) is a convex set.

To complete our proof, we will show that every pair $(\mathcal{H}, 2^{-H_\infty(\mathcal{H})})$ is in the *convex hull* of pairs of the form $(\mathcal{G}, 2^{-H_\infty(\mathcal{G})})$ where $\mathcal{G}$ is a uniform distribution over some subset $G \subseteq H$. That is, we will represent each $\mathcal{H}$ as as a convex combination of uniform distributions on subsets of H. The lemma then follows from the convexity of Equation (5).

Let us consider the *layer cake representation* of $\mathcal{H}$. Let $G_y \subset H$ be the set of all $h \in H$ such that $\Pr_{\mathcal{H}}(h) \geq y$, let $\mathcal{G}_y$ be the uniform distribution on G_y, and let $p_y = 2^{-H_\infty(\mathcal{G}_y)}$ be the probability of any particular outcome in $\mathcal{G}_y$. Recall

that $2^{-H_\infty(\mathcal{H})}$ is the maximum probability in $\mathcal{H}$. For any fixed $h \in H$, the layer cake formula yields

$$\Pr_{\mathcal{H}}(h) = \int_{y=0}^{y=2^{-H_\infty(\mathcal{H})}} \mathbb{1}_{G_y}(h) \cdot dy = \int_{y=0}^{y=2^{-H_\infty(\mathcal{H})}} \Pr_{h' \leftarrow \mathcal{G}_y}(h = h') \cdot \frac{dy}{p_y},$$

where $\mathbb{1}_{G_y}$ is the indicator function for the set G_y. It follows that

$$\mathcal{H} = \int_{y=0}^{2^{-H_\infty(\mathcal{H})}} \mathcal{G}_y \cdot \frac{dy}{p_y}.$$

The above integral defines a convex combination of the distributions $\mathcal{G}_y$ because

$$\int_{y=0}^{y=2^{-H_\infty(\mathcal{H})}} \frac{dy}{p_y} = \int_{y=0}^{y=2^{-H_\infty(\mathcal{H})}} \sum_{h \in H} \Pr_{h' \leftarrow \mathcal{G}_y}(h = h') \cdot \frac{dy}{p_y} = \sum_{h \in H} \Pr_{\mathcal{H}}(h) = 1.$$

Taking the same convex combination of min-entropies, we have:

$$\int_{y=0}^{y=2^{-H_\infty(\mathcal{H})}} p_y \cdot \frac{dy}{p_y} = \int_{y=0}^{y=2^{-H_\infty(\mathcal{H})}} dy = 2^{-H_\infty(\mathcal{H})}.$$

Therefore, $(\mathcal{H}, 2^{-H_\infty(\mathcal{H})})$ is in the convex hull of the set $\{\mathcal{G}_y, p_y\}_{y \in [0, 2^{-H_\infty(\mathcal{H})}]}$, and the lemma follows. □

3.2 The (In)Sufficiency of Uniform Source Distributions

Lemma 3.1 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 [LLTT05]). 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 2.17 and the proof of Lemma 5.1 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 5.2 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.

- In Equation (26), we transpose a matrix and replace i uniformly sampled length-ξ columns of bits with ξ uniformly sampled length-i columns of bits. This step does *not* hold if each of the i initial columns was not uniformly sampled from its domain. Thus Theorem 1.4 is also invalid for arbitrary non-uniform source distributions.

(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 2.6. We begin with the simpler case of bits that are fixed *independently* of the seed.[12]

Lemma 3.2 (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 2.11) 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 2.10, where $\mathcal{X}$ is the uniform distribution over X, then H_γ is a strong (ξ, η, ε)-biased-seed-OBF extractor per Definitions 2.7 and 2.10.*

Proof. By Lemma 3.1, 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 2.6 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

[12] Such an analysis was already performed by Doerner et al. [DKLs18, DKLs19, DKLs24]. 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.

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 2.8. 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 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 3.3. $\mathcal{D}^{\mathsf{Adv}}_{\mathcal{G},\xi}$

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

Lemma 3.4. *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 3.2:

Lemma 3.5 (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 2.11) 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 2.10, where $\mathcal{X}$ is the uniform distribution over X, then H_γ is a strong (ξ, η, ε)-biased-seed-SAOBF extractor per Definitions 2.9 and 2.10.*

Proof Sketch. The proof is similar to the proof of Lemma 3.2, 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. [LLTT05] and to the original LHL proof of Impagliazzo, Levin, and Luby [ILL89].

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[13]. 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 1.3. *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 2.14). If $n \in \mathbb{N}^+$ is even, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 2.5, 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:

$$\begin{aligned}
&2^n \cdot \mathsf{SD}((\langle \mathcal{G}, \mathcal{X} \rangle, \mathcal{G}), (\mathcal{Y}, \mathcal{G}))^n \\
&= \left(\sum_{h \in G} \sum_{y \in Y} \frac{1}{|G|} \left| \Pr_{x \leftarrow \mathcal{X}} [h(x) = y] - \frac{1}{|Y|} \right| \right)^n \\
&= |Y|^n \left(\frac{\sum_{h \in G} \sum_{y \in Y} \left| \Pr_{x \leftarrow \mathcal{X}} [h(x) = y] - \frac{1}{|Y|} \right|}{\sum_{h \in G} \sum_{y \in Y} 1} \right)^n
\end{aligned}$$

[13] 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 2.16.

$$\leq |Y|^n \frac{\sum_{h\in G}\sum_{y\in Y}\left(\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right)^n}{\sum_{h\in G}\sum_{y\in Y}1} \tag{6}$$

$$= \frac{|Y|^{n-1}}{|G|}\sum_{h\in G}\sum_{y\in Y}\left(\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right)^n$$

$$\leq \frac{|Y|^{n-1}}{|G|}\sum_{h\in H}\sum_{y\in Y}\left(\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]-\frac{1}{|Y|}\right)^n \tag{7}$$

$$= \frac{|Y|^{n-1}}{|G|}\sum_{h\in H}\sum_{y\in Y}\sum_{i=0}^{n}\binom{n}{i}\Pr_{x\leftarrow\mathcal{X}}[h(x)=y]^i\frac{1}{(-|Y|)^{n-i}} \tag{8}$$

$$= \frac{|H|}{|G|}\sum_{i=0}^{n}\binom{n}{i}\sum_{y\in Y}\Pr_{\substack{x\leftarrow\mathcal{X}\\ h\leftarrow\mathcal{H}}}[h(x)=y]^i\frac{|Y|^{i-1}}{(-1)^{n-i}}$$

$$= \frac{|H|}{|G|}\sum_{i=0}^{n}\binom{n}{i}\mathsf{CP}(i)\frac{|Y|^{i-1}}{(-1)^{n-i}} \tag{9}$$

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

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

where

Equation (6) follows from Jensen's inequality and the fact that n is even

Equation (7) follows from the fact that $G \subseteq H$

Equation (8) follows from the binomial theorem

The lemma follows from taking the n^{th} root. □

4.2 Applying Lemma 1.3 to the IPE

Proposition 4.1. *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 4.2. *The Inner Product Extractor is* 3*-universal if the source is at least 2 bits long.*

Proof. Per Proposition 2.18, 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 4.3. *Let H be an IPE family (Definition 2.11) 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 2.5, 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 1.3, we have

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

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

Where the last step follows from Proposition 4.1 and Equation (2) of Proposition 2.19, and the conclusion follows from Lemma 4.2 □

Lemma 4.4 (Simplified Version of Lemma 4.3)**.** *Let H be an IPE family (Definition 2.11) 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 2.5, where*

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

Proof. The proof is identical to the proof for Lemma 4.3, except that we use Equation (3) from Proposition 2.19 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 4.3 to derive our first explicit 4^{th}-root bound on the statistical distance of the IPE.

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

Lemma 5.1. *If $\xi \geq 5$ and $m > 3$, then the IPE family (Definition 2.11) 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 4.2) 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.[14][15] 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

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

5.2 Putting it Together

Now we can derive a concrete upper bound on distinguishability.

[14] 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 $m > 3$ (see Lemma 2.34 and Remark 2.36), and rank is invariant under permutation of rows.

[15] 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 5.2. *Let $H = \mathbb{F}_m^\xi$ be the IPE family (Definition 2.11) 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 2.5, 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 4.4, we have

$$\begin{aligned}
&\mathsf{SD}((\langle\mathcal{G},\mathcal{X}\rangle,\mathcal{G}),(\mathcal{Y},\mathcal{G})) \\
&\leq \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}1 + \sum_{j=1}^{4}\sum_{i=j}^{4}\binom{4}{i}\left\{{i \atop j}\right\}\frac{|X|^{\underline{j}}|Y|^{i-j}}{(-1)^{4-i}|X|^i} \\ + \mathsf{EIF}(4)\binom{4}{4}\left\{{4 \atop 4}\right\}\frac{|X|^{\underline{4}}|Y|}{|X|^4}\end{array}\right)} \\
&\leq \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}1 + \sum_{j=1}^{4}\sum_{i=j}^{4}\binom{4}{i}\left\{{i \atop j}\right\}\frac{|X|^{\underline{j}}|Y|^{i-j}}{(-1)^{4-i}|X|^i} \\ + 15\frac{|Y|6^{\log_2|X|}}{|X|^4}\end{array}\right)} \qquad (10) \\
&= \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}1 + \sum_{i=1}^{4}\binom{4}{i}\frac{|X||Y|^{i-1}}{(-1)^{4-i}|X|^i} + \sum_{i=2}^{4}\binom{4}{i}\left\{{i \atop 2}\right\}\frac{|X|^{\underline{2}}|Y|^{i-2}}{(-1)^{4-i}|X|^i} \\ + \sum_{i=3}^{4}\binom{4}{i}\left\{{i \atop 3}\right\}\frac{|X|^{\underline{3}}|Y|^{i-3}}{(-1)^{4-i}|X|^i} + \frac{|X|^{\underline{4}}}{|X|^4} + 15\frac{|Y|6^{\log_2|X|}}{|X|^4}\end{array}\right)} \\
&= \frac{1}{2}\sqrt[4]{\frac{|H|}{|G|}\left(\begin{array}{l}-3 + 6\frac{|Y|}{|X|} - 4\frac{|Y|^2}{|X|^2} + \frac{|Y|^3}{|X|^3} + 6\frac{|X|^{\underline{2}}}{|X|^2} - 12\frac{|X|^{\underline{2}}|Y|}{|X|^3} \\ + 7\frac{|X|^{\underline{2}}|Y|^2}{|X|^4} - 4\frac{|X|^{\underline{3}}}{|X|^3} + 6\frac{|X|^{\underline{3}}|Y|}{|X|^4} + \frac{|X|^{\underline{4}}}{|X|^4} + 15\frac{|Y|6^{\log_2|X|}}{|X|^4}\end{array}\right)}
\end{aligned}$$

where Equation (10) follows from Lemma 5.1. □

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 5.3. *Let $H = \mathbb{F}_m^\xi$ be the IPE family (Definition 2.11) 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 2.5, 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 5.2. □

Corollary 5.4. *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 2.11) 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 2.7 and 2.10, 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 2.9 and 2.10, 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 5.3 is a concave function of $|G|^{-1}$, and thus we can apply Lemma 3.1. The corollary then follows from Lemmas 3.2 and 3.5. □

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

Corollary 5.5. *Let $H = \mathbb{F}_m^\xi$ be the IPE family (Definition 2.11) 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 2.5, where*

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

Proof. Follows from Theorem 5.2 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 1.1. *Let $H = \mathbb{F}_m^\gamma$ be the IPE family (Definition 2.11) 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 2.7 and 2.10, 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 2.9 and 2.10, 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 5.5 is a concave function of $|G|^{-1}$, and thus we can apply Lemma 3.1. The corollary then follows from Lemmas 3.2 and 3.5. □

6 The Exact Multicollision Probability of the IPE

We now investigate the *exact* multicollision probability of the IPE in order to derive bounds for $n > 4$ and a tighter bound at $n = 4$. By the Rouché-Capelli theorem, the chance of an IPE multicollision on a given set of inputs has to do with the rank of the set of inputs, with probability decreasing with m exponentiated to the rank. We therefore focus on counting the number of matrices of each rank, where the values in the matrices are restricted to $\{0, 1\}$, but the *rank function* is defined over an arbitrary field, and so computing the sum over exponentiated ranks. A classic result in combinatorics [vLW92, Corollary of Theorem 25.2] gives that the number of $\xi \times i$ matrices of rank r as

$$\text{rankcount}_m(\xi, i, r) = \binom{\xi}{r}_m \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k}_m m^{i \cdot k + \binom{r-k}{2}}$$

where $\binom{\xi}{k}_m$ is a Gaussian coefficient,[16] when the matrix entries may assume any value in $\mathbb{F}_m$ and the rank function is *also* defined over $\mathbb{F}_m$. To our knowledge, no analogue of this equation exists when the value range and rank function are mismatched.[17] We must therefore devise one.

Our method recasts the sum over exponentiated matr*ix* ranks[18] as a sum over exponentiated matr*oid* ranks. We begin in Section 6.1 by proving that the latter sum satisfies certain properties that allow it to be computed as the ξ^{th} derivative of a specific evaluation of a *fixed* polynomial that is defined by the matroid. Then, in Section 6.2, we rewrite $\mathsf{CP}(i)$ (see Definition 2.12) in terms of this polynomial. In Section 7 we derive concrete and asymptotic bounds for $n \geq 4$ by plugging in $\mathsf{CP}(i)$ and computing or bounding the polynomials.

[16] Also known as a q-binomial coefficient.

[17] It might be tempting to directly combine the rankcount_2 function with Lemma 2.34 in an attempt to construct bounds on such an analogue. We advise the reader that we have been down this road and did not find it to lead anywhere useful.

[18] See Equation (29) in the proof of Lemma 6.11 for more on why it is necessary to compute this sum.

6.1 An Exponential Generator for the Accumulation of Exponentiated Ranks

Definition 6.1 (Accumulation of Exponentiated Ranks)**.** *For integers ℓ, m and matroid M with ground set E_M, we define*

$$f_{\ell,M}(m) = \sum_{(\eta_1,\ldots,\eta_\ell)\in E_M^\ell} m^{\mathrm{rank}_M(M)-\mathrm{rank}_M(\{\eta_1,\ldots,\eta_\ell\})}.$$

Furthermore, we define an exponential generator (see Definition 2.4) for $f_{\ell,M}(m)$:

$$F_M(\alpha, m) = \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} f_{\ell,M}(m).$$

Lemma 6.2. *For matroids M and N and for $f_{\ell,(\cdot)}(m)$ as given in Definition 6.1 it holds that*

$$f_{\ell,M\oplus N}(m) = \sum_{k=0}^{\ell} \binom{\ell}{k} f_{k,M}(m) \cdot f_{\ell-k,N}(m)$$

Proof. Let $L = M \oplus N$. Starting from the definition of f we have

$$\begin{aligned}
f_{\ell,L}(m) &= \sum_{(\eta_1,\ldots,\eta_\ell)\in E_L^\ell} m^{\mathrm{rank}_L(L)-\mathrm{rank}_L(\{\eta_1,\ldots,\eta_\ell\})} \\
&= \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{\substack{(\eta_1,\ldots,\eta_k)\in E_M^k \\ (\eta_{k+1},\ldots,\eta_\ell)\in E_N^{\ell-k}}} \left(\begin{array}{l} m^{\mathrm{rank}_M(M)-\mathrm{rank}_M(\{\eta_1,\ldots,\eta_k\})} \\ \cdot\, m^{\mathrm{rank}_N(N)-\mathrm{rank}_N(\{\eta_{k+1},\ldots,\eta_\ell\})} \end{array} \right) \qquad (11) \\
&= \sum_{k=0}^{\ell} \binom{\ell}{k} \left(\begin{array}{l} \left(\displaystyle\sum_{(\eta_1,\ldots,\eta_k)\in E_M^k} m^{\mathrm{rank}_M(M)-\mathrm{rank}_M(\{\eta_1,\ldots,\eta_k\})} \right) \\ \cdot \left(\displaystyle\sum_{(\eta_1,\ldots,\eta_{\ell-k})\in E_N^k} m^{\mathrm{rank}_N(N)-\mathrm{rank}_N(\{\eta_1,\ldots,\eta_{\ell-k}\})} \right) \end{array} \right)
\end{aligned}$$

where Equation (11) follows from Definition 2.27 and the fact that there are $\binom{\ell}{k}$ ways to interleave $(\eta_1,\ldots,\eta_k)$ and $(\eta_{k+1},\ldots,\eta_\ell)$ to form a single ordered tuple of ℓ elements. The lemma follows from Definition 6.1. □

Lemma 6.3. *For matroids M and N, and for $F_{(\cdot)}(m)$ as given in Definition 6.1 it holds that*

$$F_{M\oplus N}(\alpha, m) = F_M(\alpha, m) \cdot F_N(\alpha, m)$$

Proof. Let $L = M \oplus N$. Starting from the definition of F we have

$$F_L(\alpha, m) = \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} f_{\ell,L}(m)$$

$$= \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} \sum_{k=0}^{\ell} \binom{\ell}{k} f_{k,M}(m) \cdot f_{\ell-k,N}(m) \tag{12}$$

$$= \sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell} \frac{\alpha^k \alpha^{\ell-k}}{k!\,(\ell-k)!} f_{k,M}(m) \cdot f_{\ell-k,N}(m) \tag{13}$$

$$= \left(\sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} f_{\ell,M}(m) \right) \cdot \left(\sum_{k=0}^{\infty} \frac{\alpha^k}{k!} f_{k,N}(m) \right)$$

where

Equation (12) follows from Lemma 6.2

Equation (13) follows from the fact that $\binom{\ell}{k} = \frac{\ell!}{k!(\ell-k)!}$

and the lemma follows from Definition 6.1. □

Lemma 6.4. *For matroid M and any $\hat{\eta} \in E_M$ such that $\hat{\eta}$ is not a loop, let $\widehat{M} = M \setminus \{\hat{\eta}\}$ and $\widetilde{M} = M/\{\hat{\eta}\}$. For $f_{\ell,(\cdot)}(m)$ as given in Definition 6.1 it holds that*

$$f_{\ell,M}(m) = f_{\ell,\widehat{M}}(m) + \sum_{k=0}^{\ell-1} \binom{\ell}{k} f_{k,\widetilde{M}}(m)$$

Proof. Starting from the definition of f we have

$$\begin{aligned} f_{\ell,M}(m) &= \sum_{(\eta_1,\ldots,\eta_\ell) \in E_M^\ell} m^{\operatorname{rank}_M(M) - \operatorname{rank}_M(\{\eta_1,\ldots,\eta_\ell\})} \\ &= \sum_{k=0}^{\ell-1} \binom{\ell}{k} \sum_{(\eta_1,\ldots,\eta_k) \in (E_M - \{\hat{\eta}\})^k} m^{\operatorname{rank}_M(M) - \operatorname{rank}_M(\{\eta_1,\ldots,\eta_k,\hat{\eta}\})} \\ &\quad + \sum_{(\eta_1,\ldots,\eta_\ell) \in (E_M - \{\hat{\eta}\})^\ell} m^{\operatorname{rank}_M(M) - \operatorname{rank}_M(\{\eta_1,\ldots,\eta_\ell\})} \end{aligned} \tag{14}$$

$$\begin{aligned} &= f_{\ell,\widehat{M}}(m) + \sum_{k=0}^{\ell-1} \binom{\ell}{k} \sum_{(\eta_1,\ldots,\eta_k) \in (E_{\widetilde{M}})^k} m^{\operatorname{rank}_M(M) - \operatorname{rank}_M(\{\eta_1,\ldots,\eta_k,\hat{\eta}\})} \\ &= f_{\ell,\widehat{M}}(m) + \sum_{k=0}^{\ell-1} \binom{\ell}{k} \sum_{(\eta_1,\ldots,\eta_k) \in (E_{\widetilde{M}})^k} m^{\operatorname{rank}_{\widetilde{M}}(\widetilde{M}) - \operatorname{rank}_{\widetilde{M}}(\{\eta_1,\ldots,\eta_k\})} \end{aligned} \tag{15}$$

where

Equation (14) partitions the tuples $(\eta_1, \ldots, \eta_\ell) \in E_M^\ell$ according to the number of times $\hat{\eta}$ appears in each of them, separating out the case that it does not appear at all, and leverages the fact that the rank of a tuple is equal to the rank of the set of values in that tuple (i.e. rank is insensitive to repetition)

Equation (15) uses the fact that $\operatorname{rank}_M(A) = \operatorname{rank}_{\widetilde{M}}(A - \{\eta\}) + \operatorname{rank}_M(\{\eta\})$, per Definition 2.24. □

Lemma 6.5. *For matroid M and any $\hat{\eta} \in E_M$ such that $\hat{\eta}$ is neither a loop nor a coloop, and for $F_{(\cdot)}(m)$ as given in Definition 6.1 it holds that*

$$F_M(\alpha, m) = a \cdot F_{M \setminus \{\eta\}}(\alpha, m) + b \cdot F_{M/\{\eta\}}(\alpha, m)$$

where $a = 1$ and $b = e^\alpha - 1$.

Proof. Let $\widehat{M} = M \setminus \{\hat{\eta}\}$ and $\widetilde{M} = M/\{\hat{\eta}\}$. Starting from the definition of F we have

$$\begin{aligned}
F_M(\alpha, m) &= \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} f_{\ell,M}(m) \\
&= \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} \left(f_{\ell,\widehat{M}}(m) + \sum_{k=0}^{\ell-1} \binom{\ell}{k} f_{k,\widetilde{M}}(m) \right) && (16) \\
&= F_{\widehat{M}}(\alpha, m) + \sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell-1} \frac{\alpha^k \cdot \alpha^{\ell-k}}{k!\,(\ell-k)!} f_{k,\widetilde{M}}(m) && (17) \\
&= F_{\widehat{M}}(\alpha, m) + \sum_{k=0}^{\infty} \sum_{\ell'=1}^{\infty} \frac{\alpha^{\ell'} \cdot \alpha^k}{\ell'!\,k!} f_{k,\widetilde{M}}(m) && (18) \\
&= F_{\widehat{M}}(\alpha, m) + F_{\widetilde{M}}(\alpha, m) \sum_{\ell'=1}^{\infty} \frac{\alpha^{\ell'}}{\ell'!} \\
&= F_{\widehat{M}}(\alpha, m) + (e^\alpha - 1) \cdot F_{\widetilde{M}}(\alpha, m)
\end{aligned}$$

where

Equation (16) follows from Lemma 6.4

Equation (17) follows from the fact that $\binom{\ell}{k} = \frac{\ell!}{k!(\ell-k)!}$

Equation (18) sets $\ell' = (\ell - k)$. □

Lemma 6.6. *$F_{(\cdot)}(\alpha, m)$ (see Definition 6.1) is a Tutte–Grothendieck invariant (see Definition 2.32) on all matroids.*

Proof. We note first of all that computing F_M requires only knowing the rank function and ground set of M. It is therefore well-defined on all matroids. The two necessary conditions hold by Lemmas 6.3 and 6.5. □

Lemma 6.7. *$F_{(\cdot)}(\alpha, m)$ (see Definition 6.1) assumes the value $y_0 = e^\alpha$ on matroids that have ground sets comprising a single loop.*

Proof. Let M be any matroid that contains only a single loop. Beginning from the definition of F we have

$$F_M(\alpha, m) = \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} f_{\ell,M}(m)$$

$$= \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} \sum_{(\eta_1,\ldots,\eta_\ell)\in E_M^\ell} m^{\mathrm{rank}_M(M)-\mathrm{rank}_M(\{\eta_1,\ldots,\eta_\ell\})} \tag{19}$$

$$= \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} = e^\alpha \tag{20}$$

where

Equation (19) follows from Definition 6.1

Equation (20) follows because M contains only one element, which is a loop. Therefore $\mathrm{rank}_M(A) = 0$ for any $A \subseteq E_M$, and $|E_M^\ell| = 1$ for any ℓ. □

Lemma 6.8. *$F_{(\cdot)}(\alpha, m)$ (see Definition 6.1) assumes the value $x_0 = m + e^\alpha - 1$ on matroids that have ground sets comprising a single coloop.*

Proof. Let M be any matroid that contains only a single coloop. Beginning from the definition of F we have

$$F_M(\alpha, m) = \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} f_{\ell,M}(m)$$

$$= \sum_{\ell=0}^{\infty} \frac{\alpha^\ell}{\ell!} \sum_{(\eta_1,\ldots,\eta_\ell)\in E_M^\ell} m^{\mathrm{rank}_M(M)-\mathrm{rank}_M(\{\eta_1,\ldots,\eta_\ell\})} \tag{21}$$

$$= m^{\mathrm{rank}_M(M)} + \sum_{\ell=1}^{\infty} \frac{\alpha^\ell}{\ell!} = m + e^\alpha - 1 \tag{22}$$

where

Equation (21) follows from Definition 6.1

Equation (22) follows because M contains only one element, which is a coloop. It follows that $|E_M^\ell| = 1$ for any ℓ, and every non-empty subset A of E_M is a coloop, with $\mathrm{rank}_M(A) = \mathrm{rank}_M(M) = 1$. On the other hand, the rank of the empty set is 0. □

Lemma 6.9. *If $\overline{\chi}_M$ is the coboundary polynomial for matroid M (see Definition 2.30), then*

$$F_M(\alpha, m) = \overline{\chi}_M(m, e^\alpha)$$

Proof.

$$F_M(\alpha, m) = (e^\alpha - 1)^{\mathrm{rank}_M(M)} T_M\left(\frac{x_0}{e^\alpha - 1}, y_0\right) \tag{23}$$

$$= (e^\alpha - 1)^{\mathrm{rank}_M(M)} T_M\left(\frac{m}{e^\alpha - 1} + 1, e^\alpha\right) \tag{24}$$

where

Equation (23) holds due to Lemma 6.6 with (a, b) as in Lemma 6.5 (see Lemma 2.33)

Equation (24) follows from Lemmas 6.7 and 6.8

and the lemma follows from Lemma 2.31. □

6.2 Multicollision Probability in Terms of Coboundary Polynomials

Definition 6.10 (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 6.11. *Let $H = \mathbb{F}_m^{\xi}$ be an IPE family (Definition 2.11) from $X = \mathbb{1}^{\xi}$ to $Y = \mathbb{F}_m$. Let $\overline{\chi}_{M_{i-1}^{\square}}$ be the coboundary polynomial (Definition 2.30) of the cube matroid (Definition 6.10). The i-way multicollision probability of H for $i \in \mathbb{N}^+$ is exactly given by*

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

Proof.

$$\begin{aligned}
\mathsf{CP}(i) &= \sum_{y \in \mathbb{F}_m} \Pr_{\substack{x \leftarrow \mathbb{1}^{\xi} \\ h \leftarrow \mathbb{F}_m^{\xi}}} \left[\langle h, x \rangle = y\right]^i \\
&= m \cdot \Pr_{\substack{x \leftarrow \mathbb{1}^{\xi} \\ h \leftarrow \mathbb{F}_m^{\xi} \\ y \leftarrow \mathbb{F}_m}} \left[\langle h, x \rangle = y\right]^i \\
&= m \cdot \Pr_{\substack{(x_1, \ldots, x_i) \leftarrow (\mathbb{1}^{\xi})^i \\ h \leftarrow \mathbb{F}_m^{\xi} \\ y \leftarrow \mathbb{F}_m}} \left[\begin{bmatrix} h^{\top} & -y \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_i \\ 1 & \cdots & 1 \end{bmatrix} = \vec{0}\right] \\
&= \frac{m}{2^{\xi \cdot i}} \sum_{(x_1, \ldots, x_i) \in (\mathbb{1}^{\xi})^i} m^{-\mathrm{rank}\begin{bmatrix} x_1 & \cdots & x_i \\ 1 & \cdots & 1 \end{bmatrix}} \qquad (25) \\
&= \frac{m}{2^{\xi \cdot i}} \sum_{(x_1, \ldots, x_{\xi}) \in (\mathbb{1}^i)^{\xi}} m^{-\mathrm{rank}\begin{bmatrix} 1 & & & \\ \vdots & x_1 & \cdots & x_{\xi} \\ 1 & & & \end{bmatrix}} \qquad (26) \\
&= \frac{m}{2^{\xi \cdot i}} \sum_{(x_1, \ldots, x_{\xi}) \in (\mathbb{1}^i)^{\xi}} m^{-\mathrm{rank}\begin{bmatrix} 1 & & & \\ \vdots & x'_1 & \cdots & x'_{\xi} \\ 1 & & & \end{bmatrix}}
\end{aligned}$$

$$\text{where } x_i' = \begin{cases} x_i & \text{if } x_i = \begin{bmatrix} 0 \\ z \end{bmatrix} \text{ for } z \in \mathbb{1}^{i-1} \\ \vec{1} - x_i & \text{otherwise} \end{cases} \tag{27}$$

$$= \frac{m}{2^{\xi \cdot i}} \sum_{(x_1,\ldots,x_\xi) \in (\mathbb{1}^{i-1})^\xi} m^{-\text{rank}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & & & \\ \vdots & x_1 & \cdots & x_\xi \\ 1 & & & \end{bmatrix}} \tag{28}$$

$$= \frac{1}{2^{\xi \cdot (i-1)}} \sum_{(x_1,\ldots,x_\xi) \in (\mathbb{1}^{i-1})^\xi} m^{-\text{rank}\begin{bmatrix} x_1 & \cdots & x_\xi \end{bmatrix}} \tag{29}$$

$$= \frac{1}{2^{\xi \cdot (i-1)}} \sum_{(x_1,\ldots,x_\xi) \in (\mathbb{1}^{i-1})^\xi} m^{-\text{rank}_{M_{i-1}^{\square}}(\{x_1,\ldots,x_\xi\})} \tag{30}$$

$$= \frac{m^{-\text{rank}_{M_{i-1}^{\square}}(M_{i-1}^{\square})}}{2^{\xi \cdot (i-1)}} \left(\frac{d}{d\alpha}\right)^\xi F_{M_{i-1}^{\square}}(\alpha, m)\Big|_{\alpha=0} \tag{31}$$

$$= \frac{m^{-\text{rank}_{M_{i-1}^{\square}}(M_{i-1}^{\square})}}{2^{\xi \cdot (i-1)}} \left(\frac{d}{d\alpha}\right)^\xi \overline{\chi}_{M_{i-1}^{\square}}(m, e^\alpha)\Big|_{\alpha=0} \tag{32}$$

where

Equation (25) follows because $\begin{bmatrix} h^\top & -y \end{bmatrix}$ is uniform over $\mathbb{F}_m^{1 \times (\xi+1)}$. If we let

$$A = \begin{bmatrix} x_1 & \cdots & x_i \\ 1 & \cdots & 1 \end{bmatrix}$$

then we can see that $\begin{bmatrix} h^\top & -y \end{bmatrix} A$ is uniformly distributed in image (A). It is always the case that $\vec{0} \in$ image (A), and $|\text{image}\,(A)| = m^{\text{rank}(A)}$.

Equation (27) follows because rank is invariant to column operations.

Equation (28) follows because there are 2^ξ permutations of leading bits in Equation (27), each of which maps one combination of remaining bits to some fixed assignment of the set $(x_1, \ldots, x_\xi)$ as defined in Equation (28).

Equation (29) follows because $\vec{1}$ cannot be in the span of any set of vectors that all begin with 0.

Equation (30) introduces the matroid specified in Definition 6.10. Note that the rank function of this matroid is implicitly defined over $\mathbb{F}_m$, just as the matrix rank function was previously! This will be important in later sections.

Equation (31) follows from Definition 6.1 via Definition 2.4.

Equation (32) follows from Lemma 6.9.

The lemma follows due to the fact that $\text{rank}_{M_{i-1}^{\square}}(M_{i-1}^{\square}) = i - 1$. □

7 Tighter Bounds for the IPE at $n \geq 4$

In this section, we compute a *second* version of the Elevated General Leftover Hash Lemma for the IPE, using the equation for exact collision probability of the IPE that we introduced in Section 6. The first version of an IPE-specific EGLHL given in Lemma 4.3 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 (given in Theorem 1.4) for any n, but only in $2^{O(n^2)}$ time. In Section 7.2 we perform this exponential computation for small values of n, and in Section 7.3 we prove asymptotic bounds (which do not require exponential time calculations) for any n as $\xi \to \infty$.

7.1 Bounding Statistical Distance for any Even n

Theorem 1.4. *Let $H = \mathbb{F}_m^\xi$ be an IPE family (Definition 2.11) 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 (Definition 2.30) of the cube matroid (Definition 6.10). If $n \in \mathbb{N}^+$ is even, then H is a strong $(\mathcal{X}, \mathcal{G}, \varepsilon)$-extractor per Definition 2.5, where*

$$\varepsilon \leq \frac{1}{2} \sqrt[n]{\frac{|H|}{|G|} \left(1 + \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{n-i}}{2^{\xi \cdot (i-1)}} \left(\frac{d}{d\alpha} \right)^{\xi} \overline{\chi}_{M_{i-1}^{\square}}(m, e^{\alpha}) \Big|_{\alpha=0} \right)} \tag{33}$$

Proof.

$$\begin{aligned}
& 2^n \cdot \mathsf{SD}((\langle \mathcal{G}, \mathcal{X} \rangle, \mathcal{G}), (\mathcal{Y}, \mathcal{G}))^n \\
& \leq \frac{|H|}{|G|} \sum_{i=0}^{n} \binom{n}{i} \mathsf{CP}(i) \frac{m^{i-1}}{(-1)^{n-i}} && (34) \\
& = \frac{|H|}{|G|} \left(1 + \sum_{i=1}^{n} \binom{n}{i} \mathsf{CP}(i) \frac{m^{i-1}}{(-1)^{n-i}} \right) && (35) \\
& = \frac{|H|}{|G|} \left(1 + \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{n-i}}{2^{\xi \cdot (i-1)}} \left(\frac{d}{d\alpha} \right)^{\xi} \overline{\chi}_{M_{i-1}^{\square}}(m, e^{\alpha}) \Big|_{\alpha=0} \right) && (36)
\end{aligned}$$

where

Equation (34) follows as did Equation (9) in the proof of Lemma 1.3.

Equation (35) follows due to Proposition 2.20.

Equation (36) follows due to Lemma 6.11.

The theorem follows from taking the n^{th} root. □

7.2 Computing Coboundary Polynomials in $2^{O(n^2)}$ Time

In order to evaluate the bound in Theorem 1.4 at some parameterization (n, m, ξ), it is necessary to compute the coboundary polynomials $\overline{\chi}_{M_i^\square}$ (see Definition 2.30) of the cube matroid $M_i^\square$ (see Definition 6.10) for i up to $n-1$. We note that these polynomials technically depend upon m, because the $\text{rank}_{M_i^\square}$ function varies with m. However, Lemma 2.34 and Remark 2.36 tell us that for every n, there is a (relatively small) m above which $\overline{\chi}_{M_i^\square}$ defined with respect to m is equivalent to $\overline{\chi}_{M_i^\square}$ defined with respect to the real numbers. We therefore restrict our attention in this work to $\overline{\chi}_{M_i^\square}$ as defined with respect to the reals, and since we only work with integral values in this paper, it is sufficient that we further restrict ourselves to the rationals.

Computing the coboundary polynomial of a matroid is a nontrivial task. Definition 2.30 defines the polynomial as the sum of a *doubly-exponential* number of terms, each of which inolves computing a matroid rank. Lemma 2.31 relates the polynomial to the Tutte polynomial, which as typically defined still requires one to sum over all (exponentially many) subsets of the matroid's ground set. However, Tutte's original formulation, which we give as an alternative in Definition 2.29, involves summing only over all *bases* of the matroid, of which there are no more than $\binom{|E_M|}{\text{rank}_M(M)}$. For $M_i^\square$, the number of bases bounded specifically by $\binom{2^i}{i} \in 2^{O(i^2)}$.[19] It is this approach that we employ.

We wrote a `sagemath` program that computes the Tutte polynomial of $M_i^\square$ (with respect to the rationals) by partitioning $\mathcal{B}_{M_i^\square}$ among the available CPU threads, and then transforms the Tutte polynomial into the coboundary polynomial via Lemma 2.31. We ran our script on 100 threads (50 cores) of an AMD Ryzen Threadripper 7980X. The computations of $\overline{\chi}_{M_i^\square}$ for $i \in [0, 5]$ completed in less than 0.5 second each, $\overline{\chi}_{M_6^\square}$ required 92 seconds, and $\overline{\chi}_{M_7^\square}$ required 59.5 hours. This is sufficient to compute an 8^{th}-root bound using Theorem 1.4. Computing a 10^{th}-root bound would require us to find $\overline{\chi}_{M_9^\square}$, which we think impractical given currently known techniques and currently available computational resources. We have printed $\overline{\chi}_{M_i^\square}$ (with respect to the rationals) for $i \in [0, 7]$ in full in Appendix B.

Since the longest of the coboundary polynomials we have computed has 107 terms, we will not plug them into Theorem 1.4 and simplify by hand. Instead, we used `sagemath` to perform this task and arrived finally at closed-form n^{th}-root bounds for $n \in \{4, 6, 8\}$, which we have printed in full in Appendix A. Per Table 2.37, these bounds hold for $m \geq \{4, 10, 57\}$ respectively.

As we observed in Section 1.1, the 4^{th}-root bound we have computed using Theorem 1.4 and the techniques in this section is actually both tigher and simpler[20] than the one we derived manually in Theorem 5.2. The value of Theorem 5.2 comes from the fact that it can be proven without resorting to the assistance of a computer to enumerate exponentially-large sets.

[19] Thus the algorithm is in the complexity class EXP, but not in the class E.

[20] Simpler in the sense that it has nine terms under the radical instead of eleven, and does not involve any falling factorials.

7.3 Asymptotic Bounds for any Fixed Even n

In this section, our task is to asymptotically upper-bound the right hand side of Equation (33). We can see that this involves a sum over coboundary polynomials and binomial coefficients. Per Definition 2.30, the coboundary polynomial of a matroid is the weighted sum the characteristic polynomials of the contractions of that matroid by each of its flats. We will begin by partitioning the flats of the cube matroid by their codimension, and then find the size of each partition in Lemma 7.4. Subsequently, we will compute an asymptotic upper bound on a particular sum of such partition sizes in Lemma 7.5, and an asymptotic upper bound on the characteristic polynomial of a matroid contracted by one of its flats in Lemma 7.6, and then, finally, we will assemble these lemmas in order to prove the main result for this section (i.e. a bound on the first sum mentioned at the beginning of this paragraph), which is Theorem 7.7.

Lemma 7.1 ([MW19, Theorem 3.1])**.** *For any flat $X \in \mathcal{F}_{M_n^\square}$ of rank k, either $|X| = 2^k$ or $|X| \leq \frac{3}{4} \cdot 2^k$.*

Definition 7.2 (Maximal Flats)**.** *We call a flat X of $M_n^\square$* maximal *if $|X| = 2^{\operatorname{rank}_{M_n^\square}(X)}$. We use $\mathcal{F}^*_{M_n^\square}$ to denote the set of all maximal flats of $M_n^\square$, and $\mathcal{F}^{*,k}_{M_n^\square}$ to denote the subset of maximal flats of codimension k, i.e., of rank $n-k$. We use $\mathcal{F}^\circ_{M_n^\square}$ to denote the set of all* non-*maximal flats; i.e. $\mathcal{F}^\circ_{M_n^\square} = \mathcal{F}_{M_n^\square} - \mathcal{F}^*_{M_n^\square}$.*

Lemma 7.3. *Every maximal flat $X \in \mathcal{F}^{*,k}_{M_n^\square}$ contains an orthogonal basis $e_1, \ldots, e_{n-k} \in \mathbb{1}^n$. This basis is unique up to a permutation of basis vectors.*

Proof. Let $X \in \mathcal{F}^{*,k}_{M_n^\square}$ be a maximal flat of codimension k, and let $u, v \in X$. Also let X_2 be the version of X defined over the finite field $\mathbb{F}_2$ instead of the rationals, and $u_2, v_2 \in \mathbb{F}_2^n$ be the corresponding vectors reduced modulo 2. If $u \in X$ then $u_2 \in X_2$, and the converse also holds because reduction modulo 2 is injective on $\mathbb{1}$-vectors (and thus $|X_2| = 2^{n-k} = |X|$).

Claim. *If $u \in X$ and $v \in X$, then $u \oplus v \in X$ and $u \wedge v \in X$, where the operations $u \oplus v$ and $u \wedge v$ perform bitwise XOR and AND respectively.*

Proof of Claim. We have that $(u \oplus v)_2 = u_2 + v_2 \in X_2$, because X_2 is a linear subspace of $\mathbb{F}_2$. Therefore, $u \oplus v \in X$. We also have $2(u \wedge v) = u + v - (u \oplus v)$, so by linearity $u \wedge v \in X$. □

Next, we build an orthogonal basis greedily: select basis vector $e_i \in X$ as a vector linearly independent from $e_1, \ldots, e_{i-1}$ such that the 1-norm $\|e_i\|_1$ is minimized. Clearly this process produces a basis for X. To see that it will be an orthogonal basis, assume toward contradiction that e_i is the first vector selected by this process that is *not* orthogonal to some e_j for $j < i$. We have $e_i \wedge e_j \in X$ and $e_i \wedge e_j \neq 0$. Furthermore, $e_i \wedge e_j$ must be a vector that is both strictly shorter than e_j and orthogonal to e_ℓ for $\ell < j$, because it is a subset of the nonzero elements of e_j. Therefore, $e_i \wedge e_j$ is linearly independent from e_ℓ for $\ell < j$, and because it is shorter than e_j, it should have been selected instead of e_j (i.e. we have produced a contradiction).

Finally, we must show that the orthogonal basis we have constructed is unique up to permutation. This is equivalent to showing that any two orthogonal bases $e_1, \ldots, e_{n-k}$ and $e'_1, \ldots, e'_{n-k}$ and all $i, j \in [n-k]$, e_i and e'_j are either the same or orthogonal. If we suppose that for some $i, j \in [n-k]$, e_i and e'_j are *not* orthogonal, then we have $e_i \wedge e'_j \neq 0$ and $e_i \wedge e'_j \in X$. This vector has a subset of the nonzero entries of e_i and e'_j, so it is orthogonal to all $e_{i'}$ and $e'_{j'}$ for $i' \neq i$ and $j' \neq j$, which implies it must be a scalar multiple of both e_i and e'_j since it is also in the span of both bases. Because e_i, e'_j, and $e_i \wedge e'_j$ are all $\mathbb{1}$-vectors, and none of them are zero, the aforementioned conditions can only hold if $e_i \wedge e'_j = 1 \cdot e_i$ and $e_i \wedge e'_j = 1 \cdot e'_j$. Therefore, $e_i = e'_j$. □

Lemma 7.4. *The number of maximal flats in $M_n^{\square}$ satisfies*

$$\left|\mathcal{F}^{*,k}_{M_n^{\square}}\right| = \left\{ {n+1 \atop n+1-k} \right\}.$$

Proof. We will count flats by counting orthogonal bases. Lemma 7.3 established that every flat in $\mathcal{F}^{*,k}_{M_n^{\square}}$ contains an orthogonal basis, and that this basis is unique up to permutation. Conversely, every orthogonal basis of vectors $e_1, \ldots, e_{n-k} \in \mathbb{1}^n$ defines a unique flat containing 2^{n-k} vectors. Orthogonality implies that the sum of any subset of $e_1, \ldots, e_{n-k}$ is a $\mathbb{1}$ vector, and thus $e_1, \ldots, e_{n-k}$ defines a linear map from $\mathbb{1}^{n-k}$ into $M_n^{\square}$.

Next, we put these orthogonal bases (ignoring order) into bijection with the partitions of a set of $n+1$ items into $n-k+1$ non-empty subsets, of which there are $\left\{ {n+1 \atop n+1-k} \right\}$. An orthogonal basis already defines a partition of some subset of the n dimensions (some may be left out) into $n-k$ disjoint non-empty subsets: every basis vector's nonzero entries determine the dimensions contained in the corresponding subset, and orthogonality implies disjointness. Now consider a set of $n+1$ items comprising the n dimensions, plus one additional "marker" item. We can partition this set of items into the same $n-k$ disjoint non-empty subsets just described, plus one additional subset that contains the marker item and any dimensions that were *not* used by the basis.[21]

This map can be inverted by constructing one basis vector from each subset in the partition that does not contain the marker item: now disjointness implies orthogonality. Thus, this map is a bijection. □

Later, we will take a sum over the partition sizes that we established in Lemma 7.4. In Lemma 7.5, we bound the size of that sum asymptotically.

Lemma 7.5. *For integers n and k,*

$$\sum_{i=k}^{n} (-1)^{n-i} \binom{n}{i} \left\{ {i \atop i-k} \right\} = \begin{cases} 0 & \text{if } 0 \leq 2k < n \\ (n-1)!! & \text{if } 2k = n \end{cases} \tag{37}$$

[21] The presence of the marker item in this subset indicates that it does not correspond to a vector in the basis, unlike the other subsets.

Proof. Let us define a function u_k such that

$$u_k(i) = \left\{ {i \atop i-k} \right\} = \sum_{j=0}^{k} \binom{k-i}{k+j} \binom{k+i}{i+j} \begin{bmatrix} k+j \\ j \end{bmatrix} = \sum_{j=0}^{k} \binom{k-i}{k+j} \binom{k+i}{k-j} \begin{bmatrix} k+j \\ j \end{bmatrix}$$

where the second equality holds by [GKP94, Equation 6.26]. The only factors in the final expression that depend on i are $\binom{k-i}{k+j}$ and $\binom{k+i}{k-j}$, which are polynomials in i because $\binom{x}{y} = x^{\underline{y}}/y!$. Thus u_k is a polynomial. Asymptotically, Stirling numbers of the second kind satisfy [Hsu48, Equation 2]

$$u_k(i) = \frac{(i-k)^{2k}}{2^k k!}(1 + O((i-k)^{-1})) = \frac{i^{2k}}{2^k k!}(1 + O(i^{-1})).$$

Therefore, $\deg(u_k) = 2k$, and the leading coefficient of u_k is $\frac{1}{2^k k!}$.

These properties of u_k turn out to be enough to compute the binomial sum in Equation (37). By [Rui96, Theorem 1 and Corollary 2], that sum is equal to $n!$ times the coefficient of the degree-n term in u_k; i.e.

$$\sum_{i=k}^{n} (-1)^{n-i} \binom{n}{i} u_k(i) = \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} u_k(i) = \begin{cases} 0 & \text{if } 0 \le 2k < n \\ \frac{n!}{2^k k!} & \text{if } 2k = n \end{cases}$$

where the first equality holds because $u_k(i) = 0$ for $0 \le i < k$. Finally, we have $\frac{n!}{2^k k!} = (n-1)!!$[22] when $n = 2k$ [Wei25]. □

Next, our asymptotic upper bound on the characteristic polynomial of a matroid that has been contracted by one of its flats:

Lemma 7.6. *For any matroid M and flat $X \in \mathcal{F}_M$, as $m \to \infty$,*

$$\chi_{M/X}(m) = m^{\operatorname{rank}_M(E_M) - \operatorname{rank}_M(X)}(1 \pm O(m^{-1}))$$

Proof. By Definition 2.30,

$$\chi_{M/X}(m) = \sum_{Y \subseteq E_{M/X}} (-1)^{|Y|} m^{\operatorname{rank}_{M/X}(M/X) - \operatorname{rank}_{M/X}(Y)}$$

The highest power of m occurs only when $\operatorname{rank}_{M/X}(Y) = 0$. Since X is a flat of M, it contains all loops in M, so there are no remaining loops in M/X. Thus $\operatorname{rank}_{M/X}(Y) = 0$ implies $Y = \emptyset$. Splitting off this leading term, we get

$$\begin{aligned} \chi_{M/X}(m) &= m^{\operatorname{rank}_{M/X}(M/X)} + \sum_{\substack{Y \subseteq E_{M/X} \text{ s.t.} \\ \operatorname{rank}_{M/X}(Y) \ge 1}} (-1)^{|Y|} m^{\operatorname{rank}_{M/X}(M/X) - \operatorname{rank}_{M/X}(Y)} \\ &= m^{\operatorname{rank}_{M/X}(M/X)} \pm O(m^{\operatorname{rank}_{M/X}(M/X)-1}) \\ &= m^{\operatorname{rank}_M(M) - \operatorname{rank}_M(X)}(1 \pm O(m^{-1})) \end{aligned}$$

□

[22] The *double factorial* operator is defined such that $(n-1)!! = (n-1)\cdot(n-3)\cdot(n-5)\cdot\ldots\cdot 3\cdot 1$. It is not to be confused with an iterated factorial.

Finally, we are ready to state our main asymptotic result.

Theorem 7.7. *Let $H = \mathbb{F}_m^\xi, X = \mathbb{1}^\xi, Y = \mathbb{F}_m$, and $G \subseteq H$ be as in Theorem 1.4, 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 2.5, where*

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

Proof. Starting from Theorem 1.4, we have

$$\begin{aligned}
&\frac{|G|}{|H|}(2\cdot \mathsf{SD}((\langle \mathcal{G}, \mathcal{X}\rangle, \mathcal{G}), (\mathcal{Y}, \mathcal{G})))^n \\
&\leq 1 + \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{n-i}}{2^{\xi\cdot(i-1)}} \left(\frac{d}{d\alpha}\right)^\xi \overline{\chi}_{M_{i-1}^\square}(m, e^\alpha)\Big|_{\alpha=0} \\
&= 1 + \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{n-i}}{2^{\xi\cdot(i-1)}} \left(\frac{d}{d\alpha}\right)^\xi \sum_{X \in \mathcal{F}_{M_{i-1}^\square}} e^{|X|\alpha}\, \chi_{M_{i-1}^\square/X}(m)\Big|_{\alpha=0} && (38) \\
&= 1 + \sum_{i=1}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in \mathcal{F}_{M_{i-1}^\square}} \left(\frac{|X|}{2^{i-1}}\right)^\xi \chi_{M_{i-1}^\square/X}(m) && (39) \\
&= \sum_{i=1}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in (\mathcal{F}_{M_{i-1}^\square} - \{M_{i-1}^\square\})} \left(\frac{|X|}{2^{i-1}}\right)^\xi \chi_{M_{i-1}^\square/X}(m). && (40)
\end{aligned}$$

where

Equation (38) follows from Definition 2.30

Equation (39) follows by evaluating the derivative

Equation (40) follows by separating out the case where X is the whole matroid $M_{i-1}^\square$. Per Definition 2.30 and the binomial theorem, we have

$$\left(\frac{|M_{i-1}^\square|}{2^{i-1}}\right)^\xi \chi_{M_{i-1}^\square/M_{i-1}^\square}(m) = 1 \quad \text{and} \quad \sum_{i=1}^{n} \binom{n}{i}(-1)^{n-i} = -1.$$

Before taking the next step, we will develop some intuition by analyzing the behavior of each term in Equation (40) as $\xi, m \to \infty$. By Lemma 7.6, $\chi_{M_{i-1}^\square/X}(m) = m^{i-1-\mathrm{rank}_{M_{i-1}^\square}(X)}(1 \pm O(m^{-1}))$, so the term involving X decays as $\mathrm{rank}_{M_{i-1}^\square}(X)$ increases. On the other hand, increasing $|X|$ increases the base for ξ. Thus, in the case where $m = \Theta(2^\xi)$,[23] the largest terms in the sum are

[23] We note that $m = \Theta(2^\xi)$ contradicts our assumption in the theorem statement that $m \leq 2^{(1-\epsilon)\cdot\xi}$ for some $0 < \epsilon < 1$. For now we are only developing intuition.

those where $|X|$ is maximal for the rank of X. By Lemma 7.1 and Definition 7.2, if X has codimension k (i.e., $\operatorname{rank}_{M_{i-1}^{\square}}(X) = i-1-k$), then the maximal value $|X| = 2^{i-1-k}$ occurs if and only if X is a maximal flat. This completes our intuition that the terms in Equation (40) relating to maximal flats are the largest.

Since the maximal flats seem to correspond to the largest terms in Equation (40), we split the sum into maximal and non-maximal flats. Let $A_k(\xi, m)$ sum over all maximal flats of codimension k, and let $B(\xi, m)$ be the sum over the non-maximal flats. Now we have

$$\begin{aligned}
&\frac{|G|}{|H|}(2 \cdot \mathsf{SD}((\langle \mathcal{G}, \mathcal{X}\rangle, \mathcal{G}), (\mathcal{Y}, \mathcal{G})))^n \\
&\leq \sum_{i=1}^{n} \binom{n}{i} (-1)^{n-i} \sum_{k=1}^{i} \sum_{X \in \mathcal{F}^{*,k}_{M_{i-1}^{\square}}} \left(\frac{|X|}{2^{i-1}}\right)^{\xi} \chi_{M_{i-1}^{\square}/X}(m) \\
&\quad + \sum_{i=1}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in \mathcal{F}^{\circ}_{M_{i-1}^{\square}}} \left(\frac{|X|}{2^{i-1}}\right)^{\xi} \chi_{M_{i-1}^{\square}/X}(m) \\
&= \sum_{k=1}^{n} \sum_{i=k}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in \mathcal{F}^{*,k}_{M_{i-1}^{\square}}} 2^{-k\xi} \chi_{M_{i-1}^{\square}/X}(m) + B(\xi, m) \\
&= \sum_{k=1}^{n} A_k(\xi, m) + B(\xi, m) \qquad (41)
\end{aligned}$$

Next, we bound the above terms as $m \to \infty$, beginning with $A_k(\xi, m)$. We have

$$\begin{aligned}
&A_k(\xi, m) \\
&= \sum_{i=k}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in \mathcal{F}^{*,k}_{M_{i-1}^{\square}}} 2^{-k\cdot\xi} \chi_{M_{i-1}^{\square}/X}(m) \\
&= \sum_{i=k}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in \mathcal{F}^{*,k}_{M_{i-1}^{\square}}} 2^{-k\cdot\xi} m^k \cdot (1 \pm O(m^{-1})) \qquad (42) \\
&= 2^{-k\cdot\xi} m^k \sum_{i=k}^{n} \binom{n}{i} (-1)^{n-i} \left|\mathcal{F}^{*,k}_{M_{i-1}^{\square}}\right| \qquad (43) \\
&\quad + \sum_{i=k}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in \mathcal{F}^{*,k}_{M_{i-1}^{\square}}} \pm O(2^{-k\cdot\xi} m^{k-1}) \\
&\leq 2^{-k\cdot\xi} m^k \sum_{i=k}^{n} \binom{n}{i} (-1)^{n-i} \left|\mathcal{F}^{*,k}_{M_{i-1}^{\square}}\right| + O(2^{-k\cdot\xi} m^{k-1}) \qquad (44)
\end{aligned}$$

$$= 2^{-k\cdot\xi} m^k \sum_{i=k}^{n} \binom{n}{i} (-1)^{n-i} \left\{ {i \atop i-k} \right\} + O(2^{-k\xi} m^{k-1}) \tag{45}$$

$$= \begin{cases} O(2^{-k\xi} m^{k-1}) & \text{if } k < n/2 \\ 2^{-k\cdot\xi} m^k \cdot (n-1)!! + O(2^{-k\xi} m^{k-1}) & \text{if } k = n/2 \\ O(2^{-k\cdot\xi} m^k) & \text{in general} \end{cases} \tag{46}$$

$$\leq \begin{cases} O(2^{-k\xi} m^{k-1}) & \text{if } k < n/2 \\ 2^{-k\cdot\epsilon\cdot\xi} \cdot (n-1)!! + O(2^{-k\cdot\xi} m^{k-1}) & \text{if } k = n/2 \\ O(2^{-k\cdot\epsilon\cdot\xi}) & \text{in general} \end{cases} \tag{47}$$

$$\leq \begin{cases} O(2^{-\xi}) & \text{if } k < n/2 \\ 2^{-k\cdot\epsilon\cdot\xi} \cdot (n-1)!! + O(2^{-\xi}) & \text{if } k = n/2 \\ O(2^{-k\cdot\epsilon\cdot\xi}) & \text{in general} \end{cases} \tag{48}$$

where

Equation (42) follows by Lemma 7.6 and Definition 7.2

Equation (44) follows because n is a constant, which implies that the set $\mathcal{F}^{*,k}_{M^{\square}_{i-1}}$ is constant.

Equation (45) follows by Lemma 7.4

Equation (46) follows in the first and second cases from Lemma 7.5, and in the third case from the fact the value of the sum in Equation (45) depends only upon n, which is a constant per the theorem statement

Equation (47) follows from the fact that $2^{-k\cdot\xi} m^k \leq 2^{-k\cdot\xi} 2^{(1-\epsilon)\cdot k\cdot\xi} = 2^{-k\cdot\epsilon\cdot\xi}$ per the theorem statement, and the fact that n is a constant

Equation (48) follows from the fact that $2^{-k\cdot\xi} m^{k-1} \leq 2^{-k\cdot\xi + (k-1)\cdot(1-\epsilon)\cdot\xi} = 2^{-\xi-\epsilon\cdot(k-1)\cdot\xi} \leq 2^{-\xi}$.

Now we bound $B(\xi, m)$ as $m \to \infty$. We have

$$\begin{aligned} & B(\xi, m) \\ &= \sum_{i=1}^{n} \binom{n}{i} (-1)^{n-i} \sum_{X \in \mathcal{F}^{\circ}_{M^{\square}_{i-1}}} \left(\frac{|X|}{2^{i-1}} \right)^{\xi} \chi_{M^{\square}_{i-1}/X}(m) \\ &\leq \sum_{\substack{i \in [n] \text{ s.t.} \\ i \text{ is even}}} \binom{n}{i} \sum_{X \in \mathcal{F}^{\circ}_{M^{\square}_{i-1}}} \left(\frac{|X|}{2^{i-1}} \right)^{\xi} \chi_{M^{\square}_{i-1}/X}(m) \qquad (49) \\ &\leq \sum_{\substack{i \in [n] \text{ s.t.} \\ i \text{ is even}}} \binom{n}{i} \sum_{X \in \mathcal{F}^{\circ}_{M^{\square}_{i-1}}} \left(\frac{3}{4} \cdot 2^{\operatorname{rank}(X)-i+1} \right)^{\xi} \chi_{M^{\square}_{i-1}/X}(m) \qquad (50) \end{aligned}$$

$$\leq \sum_{\substack{i\in[n] \text{ s.t.} \\ i \text{ is even}}} \binom{n}{i} \sum_{X\in\mathcal{F}^{\circ}_{M^{\square}_{i-1}}} \left(\frac{3}{4}\cdot 2^{\mathrm{rank}(X)-i+1}\right)^{\xi} \cdot O\left(m^{i-1-\mathrm{rank}(X)}\right) \tag{51}$$

$$\leq \sum_{\substack{i\in[n] \text{ s.t.} \\ i \text{ is even}}} \binom{n}{i} \sum_{X\in\mathcal{F}^{\circ}_{M^{\square}_{i-1}}} O\left(\left(\frac{3}{4}\cdot 2^{\epsilon\cdot(\mathrm{rank}(X)-i+1)}\right)^{\xi}\right) \tag{52}$$

$$\leq O\left(\left(\frac{3}{4}\cdot 2^{-\epsilon}\right)^{\xi}\right) \tag{53}$$

where

Equation (49) follows from the fact that for any matroid M, $\chi_M(m)$ is non-negative as $m \to \infty$, per Lemma 7.6

Equation (50) follows from Lemma 7.1 and Definition 7.2

Equation (51) follows from Lemma 7.6

Equation (52) follows from the fact that $m^{i-1-\mathrm{rank}(X)} \leq 2^{(1-\epsilon)(i-1-\mathrm{rank}(X))\cdot\xi}$

Equation (53) follows because n is a constant, and for every $X \in \mathcal{F}^{\circ}_{M^{\square}_{i-1}}$ it holds that $\mathrm{rank}(X) < i-1$. The only flat of $M^{\square}_{i-1}$ that has rank $i-1$ is $M^{\square}_{i-1}$ itself, which is a member of $\mathcal{F}^{*}_{M^{\square}_{i-1}}$.

Finally, plugging Equations (48) and (53) back into Equation (41) yields

$$\begin{aligned}
&\sum_{k=1}^{n} A_k(\xi,m) + B(\xi,m) \\
&\leq 2^{-n\cdot\epsilon\cdot\xi/2}\cdot(n-1)!! + O(2^{-\xi}) + O(2^{-\epsilon\cdot(n/2+1)\cdot\xi}) + O\left(\left(\frac{3}{4}\cdot 2^{-\epsilon}\right)^{\xi}\right) \\
&= 2^{-n\cdot\epsilon\cdot\xi/2}\cdot(n-1)!! + O\left(2^{-\epsilon\cdot(n/2+1)\cdot\xi} + 2^{-\xi} + \left(\frac{3}{4}\right)^{\xi} 2^{-\epsilon\cdot\xi}\right)
\end{aligned}$$

and Theorem 7.7 follows. □

Corollary 7.8. *Let $H = \mathbb{F}_m^{\xi}, X = \mathbb{1}^{\xi}, Y = \mathbb{F}_m, G \subseteq H$ be as in Theorem 1.4, 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 $0 < \epsilon \leq 0.83/(n-2)$. For any even constant integer $n \geq 4$, as $m \to \infty$, H is a strong $(\mathcal{X},\mathcal{G},\varepsilon)$-extractor per Definition 2.5, where*

$$\varepsilon \leq \frac{n-1}{2}\sqrt[n]{\frac{|H|}{|G|}}\sqrt{2^{-\epsilon\cdot\xi}}\,(1+o(1)).$$

Proof. In this case, $\log_2(4/3) > \epsilon \cdot (n/2-1)$ which implies that $\log_2(4/3)+\epsilon > n \cdot \epsilon/2$, which in turn implies that

$$\left(\frac{3}{4}\right)^{\xi} 2^{-\epsilon\cdot\xi} = 2^{-(\log_2(4/3)+\epsilon)\cdot\xi} < o(2^{-n\cdot\epsilon\cdot\xi/2}).$$

Furthermore, we have

$$\epsilon < \frac{\log_2(4/3)}{n/2-1} = \frac{2}{n}\cdot\frac{\log_2(4/3)}{1-2/n} \leq \frac{2}{n}\cdot 2\log_2(4/3) < \frac{2}{n},$$

which implies that $2^{-\xi} < o(2^{-n\cdot\epsilon\cdot\xi/2})$. Finally, $(n-1)!! = (n-1)\cdot(n-3)\cdot\ldots\cdot 3\cdot 1 < (n-1)^{n/2-1}$. Plugging these inequalities into Theorem 7.7 and recalling that n is a constant yields

$$\begin{aligned}
\varepsilon &\leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left((n-1)^{n/2-1}\cdot 2^{-n\cdot\epsilon\cdot\xi/2} + O\left(2^{-(n/2+1)\cdot\epsilon\cdot\xi} + o(2^{-n\cdot\epsilon\cdot\xi/2})\right)\right)} \\
&\leq \frac{1}{2}\sqrt[n]{\frac{|H|}{|G|}\left((n-1)^{n/2-1}\cdot 2^{-n\cdot\epsilon\cdot\xi/2} + o(2^{-n\cdot\epsilon\cdot\xi/2})\right)} \\
&\leq \frac{(n-1)^{1/2-1/n}}{2}\sqrt[n]{\frac{|H|}{|G|}2^{-n\cdot\epsilon\cdot\xi/2}\cdot(1+o(1))}
\end{aligned}$$

and the corollary follows. □

Finally, we arrive at the asymptotic bound we presented in the abstract.

Corollary 1.2. *Let $H = \mathbb{F}_m^{\gamma}$ be the IPE family (Definition 2.11) 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 2.7 and 2.10, 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 (ξ,η,ε')-biased-seed-SAOBF extractor per Definitions 2.9 and 2.10, where

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

Proof. We first observe that the function bounding ε in Corollary 7.8 is a concave function of $|G|^{-1}$, and thus we can apply Lemma 3.1. The corollary then follows from Lemmas 3.2 and 3.5. □

Code Availability

The scripts used to generate these results are available from https://github.com/jackdoerner/OBF-IPE-Error-Bound-Scripts.

Acknowledgments

The authors thank Daniel Wichs for a helpful discussion, and abhi shelat and Ethan Rathbun for the use of their computers. The work of Jack Doerner was supported in part by the Brown University Data Science Institute. 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). Part of this work was conducted while Lawrence Roy was visiting the Simons Institute for the Theory of Computing.

References

[BL89] Michael Ben-Or and Nathan Linial. Collective coin flipping. *Adv. Comput. Res.*, 5:91–115, 1989.

[CDKs24] 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.

[CG85] 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.

[CG88] 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.

[Cle86] 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.

[DF25] 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.

[DGGP24] 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.

[DHIM25] 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.

[DKL+23] 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.

[DKLs18] 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.

[DKLs19] 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.

[DKLs24] 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.

[EGL85] Shimon Even, Oded Goldreich, and Abraham Lempel. A randomized protocol for signing contracts. *Communications of the ACM*, 28(6):637–647, 1985.

[Gil99] Niv Gilboa. Two party RSA key generation. In *Annual International Cryptology Conference (CRYPTO)*, pages 116–129, 1999.

[GKP94] Ronald L. Graham, Donald Ervin Knuth, and Oren. Patashnik. Concrete mathematics: A foundation for computer science, 2nd ed., 1994.

[HMRT22] 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.

[Hsu48] Leetsch C Hsu. Note on an asymptotic expansion of the nth difference of zero. *The Annals of Mathematical Statistics*, 19(2):273–277, 1948.

[ILL89] 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.

[IPS08] 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.

[KOS16] 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.

[LLTT05] 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.

[MRIS12] Criel Merino, Marcelino Ramírez-Ibáñez, and Guadalupe Sanchez. The tutte polynomial of some matroids. *International Journal of Combinatorics*, 2012, 02 2012.

[MW19] Nolmar Melo and Andreas Winter. Intersection patterns of linear subspaces with the hypercube. *Journal of Combinatorial Theory, Series A*, 164:60–71, 2019.

[NP06] Moni Naor and Benny Pinkas. Oblivious polynomial evaluation. *SIAM Journal on Computing*, 35(5):1254–1281, 2006.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. Syst. Sci.*, 52(1):43–52, 1996.

[OEI21] OEIS Foundation Inc. A003432, 2021. Published electronically at https://oeis.org/A003432.

[Oxl11] James Oxley. *Matroid Theory*. Oxford University Press, 02 2011.

[Rab81] M. O. Rabin. How to exchange secrets by oblivious transfer. TR-81, Harvard, 1981.

[Rui96] Sebastián Martín Ruiz. An algebraic identity leading to wilson's theorem. *The Mathematical Gazette*, 80(489):579–582, 1996.

[Vad12] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012.

[Vaz85] 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, page 366–378, New York, NY, USA, 1985. Association for Computing Machinery.

[vLW92] Jacobus H. van Lint and Richard M. Wilson. *A course in combinatorics*. Cambridge University Press, 1992.

[Wei25] Eric W. Weisstein. Double factorial. From MathWorld–A Wolfram Web Resource, 2025. Visited on 2025-05-25.

[Wel99] Dominic Welsh. The tutte polynomial. *Random Struct. Algorithms*, 15(3–4):210–228, October 1999.

A Concrete Bounds at $n \in \{4, 6, 8\}$

In this section we report the $\{4, 6, 8\}^{\text{th}}$-root bounds that we have computed using Theorem 1.4 together with the coboundary polynomials reported in Appendix B. For the details of the method used to compute them, we refer the reader to Section 7.2. Because these bounds have been computed with respect to the rank function over the rational numbers, they hold for $m \geq \{4, 10, 57\}$ respectively, per Table 2.37.

$$\varepsilon \leq \frac{1}{2}\left(\frac{|H|}{|G|}\right)^{\frac{1}{4}} \left(\begin{array}{l} \left(\frac{1}{8}\right)^{\xi} m^3 + 3\left(\frac{1}{4}\right)^{\xi} m^2 - 7\left(\frac{1}{8}\right)^{\xi} m^2 + 3\left(\frac{3}{8}\right)^{\xi} m - 12\left(\frac{1}{4}\right)^{\xi} m + \\ 15\left(\frac{1}{8}\right)^{\xi} m - 3\left(\frac{3}{8}\right)^{\xi} + 9\left(\frac{1}{4}\right)^{\xi} - 9\left(\frac{1}{8}\right)^{\xi} \end{array}\right)^{\frac{1}{4}}$$

$$\varepsilon \leq \frac{1}{2}\left(\frac{|H|}{|G|}\right)^{\frac{1}{6}} \left(\begin{array}{l} \left(\frac{1}{32}\right)^{\xi} m^5 + 25\left(\frac{1}{16}\right)^{\xi} m^4 - 31\left(\frac{1}{32}\right)^{\xi} m^4 + 15\left(\frac{1}{8}\right)^{\xi} m^3 + \\ 195\left(\frac{3}{32}\right)^{\xi} m^3 - 570\left(\frac{1}{16}\right)^{\xi} m^3 + 375\left(\frac{1}{32}\right)^{\xi} m^3 + 45\left(\frac{3}{16}\right)^{\xi} m^2 + \\ 75\left(\frac{5}{32}\right)^{\xi} m^2 + 285\left(\frac{1}{8}\right)^{\xi} m^2 - 2805\left(\frac{3}{32}\right)^{\xi} m^2 + 4510\left(\frac{1}{16}\right)^{\xi} m^2 - \\ 2130\left(\frac{1}{32}\right)^{\xi} m^2 + 10\left(\frac{5}{16}\right)^{\xi} m + 105\left(\frac{7}{32}\right)^{\xi} m - 270\left(\frac{3}{16}\right)^{\xi} m - \\ 120\left(\frac{5}{32}\right)^{\xi} m - 2775\left(\frac{1}{8}\right)^{\xi} m + 11385\left(\frac{3}{32}\right)^{\xi} m - 13590\left(\frac{1}{16}\right)^{\xi} m + \\ 5270\left(\frac{1}{32}\right)^{\xi} m - 10\left(\frac{5}{16}\right)^{\xi} - 105\left(\frac{7}{32}\right)^{\xi} + 225\left(\frac{3}{16}\right)^{\xi} + 45\left(\frac{5}{32}\right)^{\xi} + \\ 2475\left(\frac{1}{8}\right)^{\xi} - 8775\left(\frac{3}{32}\right)^{\xi} + 9625\left(\frac{1}{16}\right)^{\xi} - 3485\left(\frac{1}{32}\right)^{\xi} \end{array}\right)^{\frac{1}{6}}$$

$$\varepsilon \leq \frac{1}{2}\left(\frac{|H|}{|G|}\right)^{\frac{1}{8}} \left(\begin{array}{l}
\left(\frac{1}{128}\right)^{\xi} m^7 + 119\left(\frac{1}{64}\right)^{\xi} m^6 - 127\left(\frac{1}{128}\right)^{\xi} m^6 + 490\left(\frac{1}{32}\right)^{\xi} m^5 + \\
5103\left(\frac{3}{128}\right)^{\xi} m^5 - 12600\left(\frac{1}{64}\right)^{\xi} m^5 + 7035\left(\frac{1}{128}\right)^{\xi} m^5 + 105\left(\frac{1}{16}\right)^{\xi} m^4 + \\
7350\left(\frac{3}{64}\right)^{\xi} m^4 + 19950\left(\frac{5}{128}\right)^{\xi} m^4 + 58730\left(\frac{1}{32}\right)^{\xi} m^4 - \\
419643\left(\frac{3}{128}\right)^{\xi} m^4 + 548891\left(\frac{1}{64}\right)^{\xi} m^4 - 215439\left(\frac{1}{128}\right)^{\xi} m^4 + \\
630\left(\frac{3}{32}\right)^{\xi} m^3 + 4760\left(\frac{5}{64}\right)^{\xi} m^3 + 18585\left(\frac{9}{128}\right)^{\xi} m^3 + 20440\left(\frac{1}{16}\right)^{\xi} m^3 + \\
145530\left(\frac{7}{128}\right)^{\xi} m^3 - 68880\left(\frac{3}{64}\right)^{\xi} m^3 - 332220\left(\frac{5}{128}\right)^{\xi} m^3 - \\
4584860\left(\frac{1}{32}\right)^{\xi} m^3 + 13243230\left(\frac{3}{128}\right)^{\xi} m^3 - 12278980\left(\frac{1}{64}\right)^{\xi} m^3 + \\
3831835\left(\frac{1}{128}\right)^{\xi} m^3 + 280\left(\frac{5}{32}\right)^{\xi} m^2 + 315\left(\frac{9}{64}\right)^{\xi} m^2 + 3360\left(\frac{1}{8}\right)^{\xi} m^2 + \\
2940\left(\frac{15}{128}\right)^{\xi} m^2 + 8820\left(\frac{7}{64}\right)^{\xi} m^2 + 2730\left(\frac{13}{128}\right)^{\xi} m^2 + \\
32970\left(\frac{3}{32}\right)^{\xi} m^2 + 72996\left(\frac{11}{128}\right)^{\xi} m^2 + 74200\left(\frac{5}{64}\right)^{\xi} m^2 - \\
276885\left(\frac{9}{128}\right)^{\xi} m^2 + 206990\left(\frac{1}{16}\right)^{\xi} m^2 - 3185490\left(\frac{7}{128}\right)^{\xi} m^2 - \\
1859844\left(\frac{3}{64}\right)^{\xi} m^2 - 8158920\left(\frac{5}{128}\right)^{\xi} m^2 + 100201220\left(\frac{1}{32}\right)^{\xi} m^2 - \\
191437890\left(\frac{3}{128}\right)^{\xi} m^2 + 142082129\left(\frac{1}{64}\right)^{\xi} m^2 - 37769977\left(\frac{1}{128}\right)^{\xi} m^2 + \\
35\left(\frac{35}{128}\right)^{\xi} m + 728\left(\frac{13}{64}\right)^{\xi} m + 1540\left(\frac{11}{64}\right)^{\xi} m + 756\left(\frac{21}{128}\right)^{\xi} m + \\
1120\left(\frac{5}{32}\right)^{\xi} m + 10080\left(\frac{9}{64}\right)^{\xi} m + 7140\left(\frac{17}{128}\right)^{\xi} m - 22848\left(\frac{1}{8}\right)^{\xi} m + \\
18900\left(\frac{15}{128}\right)^{\xi} m - 64680\left(\frac{7}{64}\right)^{\xi} m + 114660\left(\frac{13}{128}\right)^{\xi} m - \\
330918\left(\frac{3}{32}\right)^{\xi} m - 473704\left(\frac{11}{128}\right)^{\xi} m - 1429960\left(\frac{5}{64}\right)^{\xi} m + \\
1687455\left(\frac{9}{128}\right)^{\xi} m - 6438040\left(\frac{1}{16}\right)^{\xi} m + 16611742\left(\frac{7}{128}\right)^{\xi} m + \\
16160928\left(\frac{3}{64}\right)^{\xi} m + 148295980\left(\frac{5}{128}\right)^{\xi} m - 748697390\left(\frac{1}{32}\right)^{\xi} m + \\
1119690075\left(\frac{3}{128}\right)^{\xi} m - 714967876\left(\frac{1}{64}\right)^{\xi} m + 169824305\left(\frac{1}{128}\right)^{\xi} m - \\
35\left(\frac{35}{128}\right)^{\xi} - 728\left(\frac{13}{64}\right)^{\xi} - 1540\left(\frac{11}{64}\right)^{\xi} - 756\left(\frac{21}{128}\right)^{\xi} - 1400\left(\frac{5}{32}\right)^{\xi} - \\
10395\left(\frac{9}{64}\right)^{\xi} - 7140\left(\frac{17}{128}\right)^{\xi} + 19488\left(\frac{1}{8}\right)^{\xi} - 21840\left(\frac{15}{128}\right)^{\xi} + \\
55860\left(\frac{7}{64}\right)^{\xi} - 117390\left(\frac{13}{128}\right)^{\xi} + 297318\left(\frac{3}{32}\right)^{\xi} + 400708\left(\frac{11}{128}\right)^{\xi} + \\
1351000\left(\frac{5}{64}\right)^{\xi} - 1429155\left(\frac{9}{128}\right)^{\xi} + 6210505\left(\frac{1}{16}\right)^{\xi} - 13571782\left(\frac{7}{128}\right)^{\xi} - \\
14239554\left(\frac{3}{64}\right)^{\xi} - 139824790\left(\frac{5}{128}\right)^{\xi} + 653021810\left(\frac{1}{32}\right)^{\xi} - \\
941080875\left(\frac{3}{128}\right)^{\xi} + 584628317\left(\frac{1}{64}\right)^{\xi} - 135677633\left(\frac{1}{128}\right)^{\xi}
\end{array}\right)^{\frac{1}{8}}$$

B Some Coboundary Polynomials

In this section we report the coboundary polynomials $\overline{\chi}_{M_i^{\square}}$ (see Definition 2.30) that we have computed for the cube matroids $M_i^{\square}$ (see Definition 6.10) for $i \in [0, 7]$ over the rational numbers. For the details of the method used to compute them, we refer the reader to Section 7.2.

$\overline{\chi}_{M_0^{\square}}(q,t) = t$

$\overline{\chi}_{M_1^{\square}}(q,t) = qt + t^2 - t$

$\overline{\chi}_{M_2^{\square}}(q,t) = t^4 + q^2t + 3\,qt^2 - 3\,qt - 3\,t^2 + 2\,t$

$$\overline{\chi}_{M_3^{\square}}(q,t) = t^8 + 6\,qt^4 + q^3t + 7\,q^2t^2 + 3\,qt^3 - 6\,t^4 - 7\,q^2t - 24\,qt^2 - 3\,t^3 + 15\,qt + 17\,t^2 - 9\,t$$

$$\begin{aligned}\overline{\chi}_{M_4^{\square}}(q,t) = {} & t^{16} + 10\,qt^8 - 10\,t^8 + 15\,qt^6 + 25\,q^2t^4 - 15\,t^6 + q^4t + 15\,q^3t^2 + 30\,q^2t^3 - 70\,qt^4 - 15\,q^3t \\ & - 135\,q^2t^2 - 150\,qt^3 + 45\,t^4 + 80\,q^2t + 365\,qt^2 + 120\,t^3 - 170\,qt - 245\,t^2 + 104\,t\end{aligned}$$

$$\begin{aligned}\overline{\chi}_{M_5^{\square}}(q,t) = {} & t^{32} + 15\,qt^{16} - 15\,t^{16} + 45\,qt^{12} - 45\,t^{12} + 10\,qt^{10} + 65\,q^2t^8 - 10\,t^{10} \\ & - 120\,qt^8 + 225\,q^2t^6 + 105\,qt^7 + 55\,t^8 + 90\,q^3t^4 + 75\,q^2t^5 - 1170\,qt^6 \\ & - 105\,t^7 + q^5t + 31\,q^4t^2 + 195\,q^3t^3 - 420\,q^2t^4 - 120\,qt^5 + 945\,t^6 - 31\,q^4t \\ & - 660\,q^3t^2 - 2805\,q^2t^3 - 810\,qt^4 + 45\,t^5 + 375\,q^3t + 4990\,q^2t^2 + 11385\,qt^3 \\ & + 1140\,t^4 - 2130\,q^2t - 14610\,qt^2 - 8775\,t^3 + 5270\,qt + 10249\,t^2 - 3485\,t\end{aligned}$$

$$\begin{aligned}\overline{\chi}_{M_6^{\square}}(q,t) = {} & t^{64} + 21\,qt^{32} - 21\,t^{32} + 105\,qt^{24} - 105\,t^{24} + 70\,qt^{20} - 70\,t^{20} + 140\,q^2t^{16} - 105\,qt^{16} \\ & + 105\,qt^{15} - 35\,t^{16} + 735\,qt^{14} - 105\,t^{15} + 945\,q^2t^{12} - 735\,t^{14} - 4620\,qt^{12} + 735\,q^2t^{10} \\ & + 462\,qt^{11} + 3675\,t^{12} + 350\,q^3t^8 + 630\,q^2t^9 - 910\,qt^{10} - 462\,t^{11} + 70\,q^2t^8 \\ & - 4095\,qt^9 + 175\,t^{10} + 2100\,q^3t^6 + 5145\,q^2t^7 - 10290\,qt^8 + 3465\,t^9 + 301\,q^4t^4 \\ & + 1575\,q^3t^5 - 26985\,q^2t^6 - 41160\,qt^7 + 9870\,t^8 + q^6t + 63\,q^5t^2 + 1050\,q^4t^3 \\ & - 140\,q^3t^4 - 13020\,q^2t^5 + 98112\,qt^6 + 36015\,t^7 - 63\,q^5t - 3003\,q^4t^2 - 37380\,q^3t^3 \\ & - 100065\,q^2t^4 - 33285\,qt^5 - 73227\,t^6 + 1652\,q^4t + 55930\,q^3t^2 + 458745\,q^2t^3 \\ & + 799155\,qt^4 + 44730\,t^5 - 22435\,q^3t - 485800\,q^2t^2 - 2068815\,qt^3 - 699251\,t^4 \\ & + 159460\,q^2t + 1775039\,qt^2 + 1646400\,t^3 - 510524\,qt - 1342229\,t^2 + 371909\,t\end{aligned}$$

$$\begin{aligned}\overline{\chi}_{M_7^{\square}}(q,t) = {} & t^{128} + 28\,qt^{64} - 28\,t^{64} + 210\,qt^{48} - 210\,t^{48} + 280\,qt^{40} - 280\,t^{40} + 35\,qt^{35} \\ & - 35\,t^{35} + 266\,q^2t^{32} + 112\,qt^{32} - 378\,t^{32} + 840\,qt^{30} - 840\,t^{30} + 2940\,qt^{28} \\ & - 2940\,t^{28} + 728\,qt^{26} + 2940\,q^2t^{24} - 728\,t^{26} - 12600\,qt^{24} + 9660\,t^{24} + 5236\,qt^{22} \\ & + 4060\,q^2t^{20} + 756\,qt^{21} - 5236\,t^{22} - 2800\,qt^{20} - 756\,t^{21} + 5355\,q^2t^{18} - 1260\,t^{20} \\ & + 1050\,q^3t^{16} - 22680\,qt^{18} + 8610\,q^2t^{16} + 7140\,qt^{17} + 17325\,t^{18} + 2940\,q^2t^{15} \\ & - 63098\,qt^{16} - 7140\,t^{17} + 49980\,q^2t^{14} + 18900\,qt^{15} + 53438\,t^{16} + 11970\,q^3t^{12} \\ & + 2730\,q^2t^{13} - 393960\,qt^{14} - 21840\,t^{15} - 104370\,q^2t^{12} + 114660\,qt^{13} \\ & + 343980\,t^{14} + 17360\,q^3t^{10} + 72996\,q^2t^{11} + 135198\,qt^{12} - 117390\,t^{13} + 1701\,q^4t^8 \\ & + 18585\,q^3t^9 - 29960\,q^2t^{10} - 473704\,qt^{11} - 42798\,t^{12} + 36960\,q^3t^8 - 276885\,q^2t^9 \\ & - 1696240\,qt^{10} + 400708\,t^{11} + 15750\,q^4t^6 + 145530\,q^3t^7 - 728770\,q^2t^8 \\ & + 1687455\,qt^9 + 1708840\,t^{10} + 966\,q^5t^4 + 19950\,q^4t^5 - 367920\,q^3t^6 - 3185490\,q^2t^7 \\ & + 354760\,qt^8 - 1429155\,t^9 + q^7t + 127\,q^6t^2 + 5103\,q^5t^3 + 35574\,q^4t^4 - 332220\,q^3t^5 \\ & + 1810116\,q^2t^6 + 16611742\,qt^7 + 335349\,t^8 - 127\,q^6t - 13104\,q^5t^2 - 419643\,q^4t^3 \\ & - 4147920\,q^3t^4 - 8158920\,q^2t^5 - 389592\,qt^6 - 13571782\,t^7 + 7035\,q^5t \\ & + 562107\,q^4t^2 + 13243230\,q^3t^3 + 96374460\,q^2t^4 + 148295980\,qt^5 - 1068354\,t^6 \\ & - 215439\,q^4t - 12458460\,q^3t^2 - 191437890\,q^2t^3 - 734644638\,qt^4 - 139824790\,t^5 \\ & + 3831835\,q^3t + 143357809\,q^2t^2 + 1119690075\,qt^3 + 642381558\,t^4 - 37769977\,q^2t \\ & - 719052068\,qt^2 - 941080875\,t^3 + 169824305\,qt + 587603589\,t^2 - 135677633\,t\end{aligned}$$