Electronic Colloquium on Computational Complexity, Report No. 102 (2014)

Non-Malleable Codes Against Constant Split-State Tampering

Eshan Chattopadhyay*
Department of Computer Science,
University of Texas at Austin
eshanc@cs.utexas.edu

David Zuckerman †
Department of Computer Science,
University of Texas at Austin
diz@cs.utexas.edu

August 1, 2014

Abstract

Non-malleable codes were introduced by Dziembowski, Pietrzak and Wichs [DPW10] as an elegant generalization of the classical notions of error detection, where the corruption of a codeword is viewed as a tampering function acting on it. Informally, a non-malleable code with respect to a family of tampering functions $\mathcal{F}$ consists of a randomized encoding function Enc and a deterministic decoding function Dec such that for any m, $\text{Dec}(\text{Enc}(m)) = m$. Further, for any tampering function $f \in \mathcal{F}$ and any message m, $\text{Dec}(f(\text{Enc}(m)))$ is either m or is ϵ-close to a distribution D_f independent of m, where ϵ is called the error.

Of particular importance are non-malleable codes in the C-split-state model. In this model, the codeword is partitioned into C equal sized blocks and the tampering function family consists of functions $(f_1, \ldots, f_C)$ such that f_i acts on the i^{th} block. For $C = 1$ there cannot exist non-malleable codes. For $C = 2$, the best known explicit construction is by Aggarwal, Dodis and Lovett [ADL14] who achieve rate $= \Omega(n^{-6/7})$ and error $= 2^{-\Omega(n^{-1/7})}$, where n is the block length of the code.

In our main result, we construct efficient non-malleable codes in the C-split-state model for $C = 10$ that achieve constant rate and error $= 2^{-\Omega(n)}$. These are the first explicit codes of constant rate in the C-split-state model for any $C = o(n)$, that do not rely on any unproven assumptions. We also improve the error in the explicit non-malleable codes constructed in the bit tampering model by Cheraghchi and Guruswami [CG14b].

Our constructions use an elegant connection found between seedless non-malleable extractors and non-malleable codes by Cheraghchi and Guruswami [CG14b]. We explicitly construct such seedless non-malleable extractors for 10 independent sources and deduce our results on non-malleable codes based on this connection. Our constructions of extractors use encodings and a new variant of the sum-product theorem.

*Research supported in part by NSF Grant CCF-1218723.
†Research supported in part by NSF Grant CCF-1218723.

ISSN 1433-8092

1 Introduction

1.1 Non-malleable codes

Error-correcting codes encode a message m into a longer codeword c enabling recovery of m even after part of c is corrupted. We can view this corruption as a tampering function f acting on the codeword, where f is from some small allowable family $\mathcal{F}$ of tampering functions. The strict requirement of retrieving the encoded message m imposes restrictions on the kind of tampering functions that can be handled. Unique decoding is limited by the minimum distance of the codeword, and various bounds are known in the case of list decoding. Hence, many natural classes of tampering functions cannot be handled in this framework.

One might hope to achieve a weaker goal of only detecting errors, possibly with high probability. Cramer et al. [CDF+08] constructed one such class of error-detecting codes, known as Algebraic Manipulation Detection codes (AMD codes), where the allowable tampering functions consist of all functions of the form $f_a(x) = a + x$. However error detection is impossible with respect to the family of constant functions. This follows since one cannot hope to detect errors against a function that always outputs some fixed codeword.

Dziembowski, Pietrzak and Wichs [DPW10] introduced non-malleable codes as a natural generalization of error-detecting codes. Informally, a non-malleable code with respect to a tampering function family $\mathcal{F}$ is equipped with a randomized encoder Enc and a deterministic decoder Dec such that $\text{Dec}(\text{Enc}(m)) = m$ and for any tampering function $f \in \mathcal{F}$ the following holds: for any message m, $\text{Dec}(f(\text{Enc}(m)))$ is either the message m or is ϵ-close (in statistical distance) to a distribution D_f independent of m. The parameter ϵ is called the error.

Let $\mathcal{F}_n$ be the set of all functions on $\{0,1\}^n$. Note that there cannot exist a code with block length n which is non-malleable with respect to $\mathcal{F}_n$. This follows since the tampering function could then use the function Dec to decode the message m, get a message m' by flipping all the bits in m, and use the encoding function to pick any codeword in $\text{Enc}(m')$.

Therefore, it is natural to restrict the size of the family of tampering functions. It follows from the works in [DPW10], [CG14a] that there exists non-malleable codes with respect to any tampering function family of size bounded by $2^{2^{\delta n}}$ with rate close to $1-\delta$ and error $2^{-\Omega(n)}$, for any constant $\delta > 0$. The bounds obtained in these works are existential, and some progress has been made since then in giving explicit constructions against useful classes of tampering functions.

Non-malleable codes in the C-split-state model One of the most important families of tampering functions, both from an application point of view and from theoretical interest, is the family of tampering functions in the C-split-state model. In this model, each tampering function f is of the form $(f_1, \ldots, f_C)$ where $f_i \in \mathcal{F}_{n/C}$, and for any codeword $x = (x_1, \ldots, x_C) \in (\{0,1\}^{n/C})^C$ we define $(f_1, \ldots, f_C)(x_1, \ldots, x_C) = (f_1(x_1), \ldots, f_C(x_C))$. Thus each f_i independently tampers a fixed partition of the codeword. The relevance of this model comes from a practical point of view when a codeword is partitioned and stored in C different locations and different tampering functions acts independently on each part. Another motivation to study this model comes from the scenario where a codeword is sent through a channel that corrupts different parts independently. This suggests that even the case $C = n$ is interesting, but the case when C is independent of n is particularly important, especially when C is in fact a small integer.

There has been a lot of recent work on constructing explicit and efficient non malleable codes

in the C-split-state model. Since $C = 1$ includes all of $\mathcal{F}_n$, the best one can hope for is $C = 2$. A Monte-Carlo construction of non-malleable codes in this model was given in the original paper on non-malleable codes [DPW10] for $C = 2$ and then improved in [CG14a]. However, both of these constructions are inefficient. For $C = 2$, these Monte-Carlo constructions imply existence of codes of rate close to $\frac{1}{2}$ and corresponds to the hardest case. On the other extreme, when $C = n$, it corresponds to the case of bit tampering where each function f_i acts independently on a particular bit of the codeword.

The best known explicit construction of non-malleable codes in the C-split-state model for the case when $C = 2$ is due to the elegant work of Aggarwal, Dodis and Lovett [ADL14], who construct a code with rate $= \Omega(n^{-6/7})$ and error $= 2^{-\Omega(n^{-1/7})}$. Their proof of non-malleability uses sophisticated methods from additive combinatorics. The drawback of this construction is the polynomially small rate of the code.

Our main result on non-malleable codes is for the model of C-split-state adversaries when $C = 10$. We give explicit constructions of non-malleable codes in this model with rate $= \Omega(1)$ and error $= 2^{-\Omega(n)}$. In particular, we have the following result.

Theorem 1. *For all $n > 0$ there exists an explicit construction of efficient non-malleable codes on $\{0,1\}^n$ in the 10-split-state model with constant rate and error $= 2^{-\Omega(n)}$.*

We note that the best known non-malleable code in the $O(1)$-split-state prior to this work was the non-malleable code in the 2-split-state model from [ADL14], which as mentioned above, has rate $\Omega(n^{-6/7})$ and error is $2^{-\Omega(n^{-1/7})}$. Thus we give the first explicit construction of constant rate non-malleable codes in the split-state model for a fixed integer C that do not rely on any unproven assumptions; in fact, this is the first for $C = o(n)$. We further obtain optimal error.

For the case of bit tampering ($C = n$), the best known explicit constructions of non-malleable codes were given in the work of [CG14b] with rate $= (1 - o(1))$ and error $= 2^{-\Omega(n^{-1/7})}$. We improve upon the error and obtain the following result.

Theorem 2. *For all $n > 0$ there exists an explicit construction of efficient non-malleable codes on $\{0,1\}^n$ in the bit tampering model with rate $= (1 - o(1))$ and error $= 2^{-\Omega(n)}$.*

We obtain Theorem 2 from the following observation. The construction against bit tampering in [CG14b] uses a possibly sub-optimal rate non-malleable code against bit-tampering in its construction and shows a way to improve the rate to $(1 - o(1))$ while maintaining the error bound. The sub-optimal rate non-malleable code used was the code from [ADL14] which resulted in the sub-optimal error bound of $2^{-\Omega(n^{-1/7})}$. By plugging in our non-malleable code construction from Theorem 1 as the sub-optimal non-malleable code in the construction of [CG14b], we deduce Theorem 2.

Previous Work: Apart from the previous work stated above, there has been other work in constructing non-malleable codes. However they did not improve the parameters achieved in [ADL14] in the C-split model for $C = o(n)$. Before the work of [ADL14], the only unconditional efficient non-malleable code in the C-split-state model, for $C = o(n)$, was by Dziembowski, Kazana, and Obremski [DKO13]. However, they could encode only 1 bit messages.

There were also some conditional results. Liu and Lysyanskaya [LL12] constructed efficient constant rate non-malleable codes in the split-state model against computationally bounded adver-

saries. Their proof of non-malleability relies on the existence of robust public-key cryptosystems and existence of robust non-interactive zero-knowledge proof systems for some language in NP. They also use the common reference string (CRS) assumption which roughly states that one has access to an untampered random string. The recent work of Faust et al. [FMVW13] constructed almost optimal non-malleable codes against the class of polynomial sized circuits in the CRS framework. [CCP12], [CCFP11], [CKM11], and [FMNV14] considered non-malleable codes in other models.

Independent Work: Independently, Aggarwal, Dodis, Kazana and and Obremski [ADKO14] constructed non-malleable codes in the 2-split model with rate $\Omega(n^{-1/2})$. Furthermore, they gave a general reduction from 2 parts to a constant number of parts, incurring only a constant overhead in the rate, as long as the non-malleable extractor is strong, as ours is. As a result, after seeing a preliminary version of our work, they applied their reduction to our result to construct constant-rate non-malleable codes in the 2-split model.

1.2 Non-malleable extractors

We prove Theorem 1 by constructing an object called seedless non-malleable extractor, which is interesting in its own right. To motivate this, recall that the area of randomness extraction addresses the problem of efficiently generating nearly uniformly random bits from weak sources. The most widely used model of a weak source X measures the randomness in X in terms of its min-entropy $H_\infty(X)$. We say that X has min-entropy k if the maximum probability that X places on any point in its support is 2^{-k}. Unfortunately it is not possible to extract even a single bit from sources with min-entropy $n-1$. To overcome this, the notion of seeded extractors was considered in [NZ93] where one is allowed to extract from source X using a short uniformly random string Y. We now define strong seeded extractors, using $\circ$ to denote concatenation and $|D_1 - D_2|$ to denote the statistical distance between distributions D_1 and D_2 (see Section 2).

Definition 1.1. *A function* $\text{SExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* (k, ϵ)*-strong seeded extractor if the following holds : If* X *is a source on* $\{0,1\}^n$ *such that* $H_\infty(X) \geq k$ *and* Y *is a uniformly random string on* $\{0,1\}^d$ *independent of* X*, then*

$$|\text{SExt}(X, Y) \circ Y - U_m \circ Y| < \epsilon$$

From a series of works ending with [LRVW03],[GUV09],[DKSS09], we now have explicit constructions of strong seeded extractors for k as small as $O(\log n)$, which is optimal up to a constant factor.

A generalization of strong seeded extractors called seeded non-malleable extractors was introduced in the context of privacy amplification by Dodis and Wichs in [DW09]. Dodis and Wichs showed the existence of such extractors, and subsequently explicit constructions of seeded non-malleable extractors were given in the recent works of [DLWZ11], [CRS12], [Li12a] and [Li12b]. Recently Li [Li13] found applications of non-malleable extractors in constructing extractors for independent sources. To define non-malleable extractors, we need the following definition.

Definition 1.2. *For any function* $f : S \to S$*,* f *has a fixed point at* $s \in S$ *if* $f(s) = s$*. We say* f *has no fixed points in* $T \subseteq S$*, if* $f(t) \neq t$ *for all* $t \in T$*.* f *has no fixed points if* $f(s) \neq s$ *for all* $s \in S$*.*

We will need non-malleable extractors even if the seed is weak (not uniformly random), as in the following definition.

Definition 1.3. *A function* snmExt : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* (k_1, k_2, ϵ)*-seeded non-malleable extractor if the following holds : If* X *and* Y *are independent sources on* $\{0,1\}^n$ *and* $\{0,1\}^d$ *respectively such that* $H_\infty(X) \geq k_1$ *and* $H_\infty(Y) \geq k_2$ *and* $f : \{0,1\}^n \to \{0,1\}^n$ *has no fixed points, then*

$$|\text{snmExt}(X, Y) \circ \text{snmExt}(X, f(Y)) \circ Y - U_m \circ \text{snmExt}(X, f(Y)) \circ Y| < \epsilon$$

In the above definition f is called a tampering function.

In a recent work, Cheraghchi and Guruswami [CG14b] raised the natural question of constructing non-malleable extractors when we allow both X and Y to be tampered independently. They asked, roughly :

Construct a polytime function nmExt : $(\{0,1\}^n)^2 \to \{0,1\}^m$ such that the following holds : If X, Y are independent sources on $\{0,1\}^n$ such that $H_\infty(X), H_\infty(Y) \geq k$ and f, g are arbitrary tampering functions on $\{0,1\}^n$ such that at least one of f, g has no fixed points, then

$$|\text{nmExt}(X, Y) \circ \text{nmExt}(f(X), g(Y)) - U_m \circ \text{nmExt}(f(X), g(Y))| < \epsilon$$

Note that if both f and g are the identity function, then obviously there cannot be any such function nmExt. To avoid such technicalities, we have the restriction that at least one of f or g has no fixed points. It turns out that such functions, called seedless non-malleable extractors, exist for k as low as $O(\log n)$ and $\epsilon = 2^{-\Omega(k)}$ with $m = \Omega(k)$. This was shown in [CG14b] using clever techniques from the probabilistic method. However giving explicit constructions of such extractors turns out to be a very hard problem, even for $k = n$, and there are still no known constructions.

It appears nontrivial to extend existing constructions of seeded non-malleable extractors when both sources are tampered. For example for sources on $\mathbb{F}_p$, the function $\chi(x + y)$, where χ is the quadratic character[1], was shown to be a seeded non-malleable extractor [DLWZ11]. However it fails to work against tampering functions $f(x) = x + 1$ and $g(y) = y - 1$, even for full entropy.

In this paper we make progress on a relaxed version of this problem where we use a constant number of independent sources, each with min-entropy k, instead of just 2 sources. We note that prior to this work, there were no known results in this setting even for $k = n$.

We now give an informal definition of seedless non-malleable extractors for independent sources. We refer the reader to Section 3 for formal definitions.

Definition 1.4 (informal)**.** *A function* snmExt : $(\{0,1\}^n)^C \to \{0,1\}^m$ *is a* (k, ϵ)*-seedless non-malleable extractor for* C *independent sources if the following holds: If* $X_1, \ldots, X_C$ *are independent sources on* $\{0,1\}^n$ *such that* $H_\infty(X_i) \geq k$ *for all* $i = 1, \ldots, C$ *and* $f_1, \ldots, f_C$ *are arbitrary tampering functions such that there exists an* f_i *with no fixed points, then*

$$|\text{nmExt}(X_1, \ldots, X_C) \circ \text{nmExt}(f_1(X_1), \ldots, f_C(X_C)) - U_m \circ \text{nmExt}(f_1(X_1), \ldots, f_C(X_C))| < \epsilon$$

Our main result on non-malleable extractors is the following theorem.

[1]for any prime field $\mathbb{F}_p$, the quadratic character is given by $\chi(x) = x^{\frac{p-1}{2}}$

Theorem 3. *For some $\delta > 0$ there exists a polynomial time construction of a (k, ϵ)-seedless non-malleable extractor for 10 independent sources* $\mathrm{nmExt} : (\{0,1\}^n)^{10} \to \{0,1\}^m$ *with $k = (1-\delta)n$, $\epsilon = 2^{-\Omega(n)}$ and $m = \Omega(k)$.*

Theorem 1 now follows from an elegant reduction discovered in [CG14b], which shows how to use explicit constructions of seedless non-malleable extractors to construct non-malleable codes with an efficient decoder. This reduction however does not guarantee an efficient encoder for the constructed codes. Developing an efficient encoder for the non-malleable codes, which follow from the extractor construction in Theorem 3, requires some additional work. We build an efficient encoder using algorithms for almost uniformly sampling from algebraic varieties combined with the method of rejection sampling. The proof of correctness of the encoding algorithm relies on estimates on the number of rational points on algebraic varieties.

1.3 Organization

We discuss preliminaries in Section 2, and formally define non-malleable codes and seedless non-malleable extractors in Section 3. We recall the connection between non-malleable codes and seedless non-malleable extractors from [CG14b] and deduce Theorem 1 assuming Theorem 3 and Theorem 8.7 in Section 4. Our main technical contribution is the proof of Theorem 3. We use Section 5 to sketch the main ideas in proving Theorem 3. Section 5 can be skipped without any loss of continuation. We present the formal proof of Theorem 3 in Section 6. We require a sum-product estimate over $\mathbb{F}_p^4$ for proving Theorem 3. We prove this in Section 7. The proof of this estimate closely follows the arguments of a sum-product theorem over $\mathbb{F}_p^2$ by Bourgain [Bou05a]. We develop an efficient encoder for the constructed non-malleable codes in the 10-split-state model in Section 8. In Appendix B, we prove an additional property of the constructed seedless non-malleable extractor which might be useful in other explicit constructions.

2 Preliminaries

2.1 Notations

We use capital letters to denote distributions and their support. We use corresponding small letters to denote a sample from the source.
We use $[l]$ to denote the set $\{1, 2, \ldots, l\}$.
We use U_m to denote the uniform distribution over $\{0,1\}^m$.
For any set S, we use $|S|$ to denote its size.
$\mathbb{F}_p$ denotes the prime finite field with p elements.
For a vector $v \in \mathbb{F}_p^n$, we use $\Pi_S(v)$ to denote the projection of v to the coordinates indexed by the elements in $S \subset [n]$. We extend the action of Π_S to sets in the obvious manner. We use Π_i for $\Pi_{\{i\}}$.
For any set S, we use $s \sim S$ to denote a uniform draw from S.

2.2 Min entropy and flat distributions

Definition 2.1. *For a source X we define min-entropy of X as :*

$$H_\infty(X) = \min_{s \in \text{support}(X)} \left\{ \frac{1}{\log(\Pr[X = s])} \right\}$$

Definition 2.2. *We call a distribution (source) D to be flat if it is uniform over a set S.*

Definition 2.3. *A (n, k)-source is a distribution on $\{0,1\}^n$ with min-entropy k.*

It is a well known fact that any (n, k)-source is a convex combination of flat sources supported on sets of size 2^k.

2.3 Statistical distance, convex combination of distributions and probability lemmas

Definition 2.4. *Let D_1 and D_2 be two distributions on a set S. We define the statistical distance between D_1 and D_2 as :*

$$|D_1 - D_2| = \frac{1}{2} \sum_{s \in S} |\Pr[D_1 = u] - \Pr[D_2 = u]|$$

We say that a distribution D_1 is ϵ-close to another distribution D_2 if $|D_1 - D_2| \leq \epsilon$.

Definition 2.5. *The collision probability of a distribution D is defined as :* $\text{cp}(D) = \Pr[D = D']$, *where D' is independent and identicaly distributed as D.*

For the sake of convenience, we make the following definition.

Definition 2.6. *For a set A, define* $\text{cp}(A)$ *to be the collision probability of the uniform distribution on A.*

The following lemma was proved in [BIW06].

Lemma 2.7. *Let D be a distribution with* $\text{cp}(D) = \frac{1}{KL}$. *Then D is $L^{-1/2}$-close to a distribution with min-entropy at least* $\log K$.

Definition 2.8. *We say that a distribution D on a set S is a convex combination of distributions $D_1, \ldots, D_l$ on S if there exists non-negative constants (called weights) $w_1, \ldots, w_l$ with $\sum_{i=1}^{l} w_i = 1$ such that $\Pr[D = s] = \sum_{i=1}^{l} w_i \cdot \Pr[D_i = s]$ for all $s \in S$. We use the notation $D = \sum_{i=1}^{l} w_i \cdot D_i$ to denote the fact that D is a convex combination of the distributions $D_1, \ldots, D_l$ with weights $w_1, \ldots, w_l$.*

Definition 2.9. *For random variables X and Y, we use $X|Y$ to denote a random variable with distribution :* $\Pr[(X|Y) = x] = \sum_{y \in \text{support}(Y)} \Pr[Y = y] \cdot \Pr[X = x | Y = y]$.

We note the following lemma which follows from the above definitions.

Lemma 2.10. *Let X and Y be distributions on a set S such that $X = \sum_{i=1}^{l} w_i \cdot X_i$ and $Y = \sum_{i=1}^{l} w_i \cdot Y_i$. Then $|X - Y| \leq \sum_i w_i \cdot |X_i - Y_i|$.*

The following result follows from a lemma proved in [MW97].

Corollary 2.11. *Let X, Y be random variables with supports $S, T \subseteq V$ such that (X, Y) is ϵ-close to a distribution with min-entropy k. Further suppose that the random variable Y can take at most l values. Then*

$$\Pr_{y \sim Y}\left[(X|Y=y) \text{ is } 2\epsilon^{1/2}\text{-close to a source with min-entropy } k - \log l - \log\left(\frac{1}{\epsilon}\right)\right] \geq 1 - 2\epsilon^{1/2}$$

2.4 Some results from additive combinatorics

We recall some well known results from additive combinatorics. We refer the reader to the excellent book by Tao and Vu [TV06] for more details.

Definition 2.12. *For vectors $v, w \in \mathbb{F}_p^n$, where $v = (v_1, \ldots, v_n)$ and $w = (w_1, \ldots, w_n)$, we define*

$$v \odot w = (v_1 w_1, \ldots, v_n w_n)$$

Definition 2.13. *For subsets $A, B \subseteq \mathbb{F}_p^n$, define the sets :*

$$A + B = \{a + b : a \in A, b \in B\}$$

$$A \odot B = \{a \odot b : a \in A, b \in B\}$$

Observation 2.14. *$(\mathbb{F}_p^*)^n$ is a group under the operation $\odot$.*

Lemma 2.15 (Plünnecke-Ruzsa)**.** *Let A, B be finite subsets in an additive group G. Then*

$$|A + A| \leq \frac{|A + B|^4}{|A||B|^2}$$

Lemma 2.16 (Plünnecke-Ruzsa)**.** *Let A be a finite subset of any additive group G. Then*

$$|A - A| \leq \left(\frac{|A + A|}{|A|}\right)^3 |A|$$

Lemma 2.17 (Balog-Szemerédi-Gowers lemma [BS94],[Gow98])**.** *Let A, B be finite subsets of an additive group G and let $|A|^{1-\rho_1} \leq |B| \leq |A|^{1+\rho_1}$. If $\mathrm{cp}(A + B) \geq |A|^{-(1+\rho_2-\rho_1)}$, then there exists subsets $A' \subseteq A$, $B' \subseteq B$ such that $|A'| \geq |A|^{1-10\rho_2}$, $|B'| \geq |B|^{1-10\rho_2}$, and $|A' + B'| \leq |A|^{1+\rho_1+10\rho_2}$.*

2.5 Some known extractor constructions

We recall some known results on multi-source extractors and non-malleable extractors.

The following result on extracting from 2 independent sources is well known and a proof can be found in [Rao07].

Theorem 2.18. *For all $n > 0$ and any constant δ there exists an explicit function* $2\text{SExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, *$m = \Omega(\delta n)$, such that if X, Y are independent sources with min-entropy k_1, k_2 respectively satisfying $k_1 + k_2 \geq (1+\delta)n$, then*

$$|2\text{SExt}(X,Y) \circ X - U_m \circ X| \leq 2^{-\Omega(n)},$$
$$|2\text{SExt}(X,Y) \circ Y - U_m \circ Y| \leq 2^{-\Omega(n)}$$

We recall a 3-source extractor constructed in [Rao06].

Theorem 2.19 ([Rao06])**.** *For every n and and constant $\delta > 0$ there exists an explicit function* $3\text{ext} : \{0,1\}^n \to \{0,1\}^m$, *$m = \Omega(n)$, such that if X_1, X_2, X_3 are independent $(n, \delta n)$ sources then*

$$|3\text{Ext}(X_1, X_2, X_3) - U_m| < 2^{-\Omega(n)}$$

Explicit constructions of seeded non-malleable extractors follow from works of [DLWZ11] and [Li12b]. The output length in [DLWZ11] relies on an unproven but widely believed conjecture on primes while the output length in [Li12b] is unconditional. Further, either of the non-malleable extractors from [DLWZ11] or [Li12b] is also a strong 2-source extractor.

Theorem 2.20 ([DLWZ11],[Li12b])**.** *Let $\delta > 0$ be a constant. For all n, there exists an explicit function* $\text{snmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, *$m = \Omega(n)$, satisfying: Suppose X, Y are independent sources on $\{0,1\}^n$ with min-entropy k_1, k_2 respectively.*

1. *If $(k_1 + k_2) \geq (1+\delta)n$, then*

$$|\text{snmExt}(X,Y) \circ X - U_m \circ X| < 2^{-\Omega(n)},$$
$$|\text{snmExt}(X,Y) \circ Y - U_m \circ Y| < 2^{-\Omega(n)}$$

2. *If $k_1, k_2 > (1-\delta)n$ and f is any tampering function with no fixed points, then*

$$|\text{snmExt}(X,Y) \circ \text{snmExt}(X, f(Y)) - U_m \circ \text{snmExt}(X, f(Y))| < 2^{-\Omega(n)}$$

3 Non-malleable codes and Seedless non-malleable extractors

3.1 Non-malleable codes

We follow the presentation in [DPW10] and define non-malleable codes.

Definition 3.1 (Coding schemes)**.** *Let* $\text{Enc} : \{0,1\}^k \to \{0,1\}^n$ *and* $\text{Dec} : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$ *be functions such that* Enc *is a randomized function (i.e. it has access to a private randomness) and* Dec *is a deterministic function. We say that* (Enc, Dec) *is a coding scheme with block length n and message length k if for all $s \in \{0,1\}^k$,* $\Pr[\text{Dec}(Enc(s)) = s] = 1$ *(the probability is over the randomness in* Enc*).*

Definition 3.2 (Tampering functions)**.** *For any $n > 0$, let $\mathcal{F}_n$ denote the set of all functions $f : \{0,1\}^n \to \{0,1\}^n$. We call any subset $\mathcal{F} \subseteq \mathcal{F}_n$ to be a family of tampering functions.*

We do not specify the domain of tampering functions when it is implied from the context.

Definition 3.3. *For any function $f : S \to S$ and $T \subseteq S$, the maximum pre-image size of f in T is given by* $\max_{t \in T} |f^{-1}(t)|$. *The maximum pre-image size of f is* $\max_{s \in S} |f^{-1}(s)|$.

We now define non-malleable codes with respect to a family of tampering functions. We need to define the following function.

$$\mathrm{copy}(x, y) = \begin{cases} x & \text{if } x \neq same^\star \\ y & \text{if } x = same^\star \end{cases}$$

Definition 3.4 (Non-malleable codes)**.** *A coding scheme* $(\mathrm{Enc}, \mathrm{Dec})$ *with block length n and message length k is a non-malleable code with respect to a family of tampering functions $\mathcal{F} \subset \mathcal{F}_n$ and error ϵ if for every $f \in \mathcal{F}$ there exists a random variable D_f on $\{0,1\}^k \cup \{same^\star\}$ which is independent of the randomness in* Enc *such that for all messages $s \in \{0,1\}^k$, it holds that*

$$|\mathrm{Dec}(f(\mathrm{Enc}(s))) - \mathrm{copy}(D_f, s)| \leq \epsilon$$

The rate of a non-malleable code $\mathcal{C}$ is given by $\frac{k}{n}$.

As an easy example, suppose the tampering function family at hand is $\mathcal{F}_{constant}$, consisting of all constant functions, $f_c(x) = c$ for all x. We can use any coding scheme and for any tampering function $f_c \in \mathcal{F}_{constant}$, we may take D_{f_c} to be $\mathrm{Dec}(c)$ with probability 1.

3.1.1 Non-malleable codes in the C-split-state model

We formally define non-malleable codes in the C-split sate model.

Definition 3.5. *Let $\mathcal{F}_{n,C} = \{(f_1, \ldots, f_C) : f_i \in \mathcal{F}_{n/C}$ for all $i \in [C]\}$, where for any $x = (x_1, \ldots, x_C) \in (\{0,1\}^{n/C})^C$ we define $(f_1, \ldots, f_C)(x) = (f_1(x_1), \ldots, f_C(x_C))$. Non-malleable codes in the C-split-state model with block length n are non-malleable codes with respect to $\mathcal{F}_{n,C}$.*

We call $\mathcal{F}_{n,C}$ to be the family of tampering functions in the C-split-state model.

When $C = n$, note that this corresponds to the case of bit tampering. Also $C \geq 2$, since as discussed before, $C = 1$ is impossible.

3.2 Seedless non-malleable extractors

Seedless non-malleable extractors were first introduced by Cheraghchi and Guruswami in [CG14b]. We present a modified definition here.

Definition 3.6 (Seedless non-malleable extractors)**.** *A function* $\mathrm{nmExt} : \{0,1\}^n \to \{0,1\}^m$ *is said to be a seedless non-malleable extractor with respect to a class of sources $\mathcal{X}$ and a family of tampering functions $\mathcal{F}$ with error ϵ if for every distribution $X \in \mathcal{X}$ and every tampering function $f \in \mathcal{F}$, there exists a random variable $D_{X,f}$ on $\{0,1\}^m \cup \{same^\star\}$ which is independent of the source X such that*

$$|\mathrm{nmExt}(X) \circ \mathrm{nmExt}(f(X)) - U_m \circ \mathrm{copy}(D_{X,f}, U_m)| \leq \epsilon$$

where both U_m's refer to the same uniform m-bit string.

For example suppose nmExt is a deterministic extractor for the class of sources $\mathcal{X}$. If $f(x) = x$ is the identity function, then we may take $D_{X,f} = same^\star$ with probability 1.

We now define a special case of the above definition which is of particular interest to us.

Definition 3.7 (Seedless non-malleable multi-source extractors)**.** *For any constant C, we say that* $\text{nmExt} : (\{0,1\}^n)^C \to \{0,1\}^m$ *is a seedless non-malleable multi-source extractor for C independent sources with min-entropy k and error ϵ if whenever $X_1, X_2, \ldots, X_C$ are independent (n,k)-sources and $f_1, f_2 \ldots, f_C$ are arbitrary tampering functions in $\mathcal{F}_n$, there exists random variable D_f on $\{0,1\}^m \cup \{same^\star\}$ which is independent of the sources $X_1, \ldots, X_C$ such that*

$$|\text{nmExt}(X_1, \ldots, X_C) \circ \text{nmExt}(f_1(X_1), \ldots f_l(X_C)) - U_m \circ \text{copy}(D_f, U_m)| < \epsilon$$

where both U_m's refer to the same uniform m-bit string.

4 Non-malleable codes via Seedless non-malleable extractors

In this section we prove Theorem 1 assuming Theorem 3 and Theorem 8.7. The work by Cheraghchi and Guruswami [CG14b] shows a way to construct non-malleable codes with an efficient decoder from explicit constructions of seedless non-malleable extractors. We use this connection to construct non-malleable codes. An efficient encoder for the resulting non-malleable codes is constructed in Section 8.

The following theorem follows from the work in [CG14b]. We include a proof for the sake of completeness.

Theorem 4.1. *For any constant C, let* $\text{nmExt} : (\{0,1\}^n)^C \to \{0,1\}^m$, *$m = \Omega(n)$ be a polynomial time computable seedless non-malleable extractor for C-independent sources for min-entropy n with error $\epsilon = 2^{-\Omega(n)}$. Then there exists an explicit non-malleable code with an efficient decoder in the C-split-state model with block length $= Cn$, rate $= \Omega(1)$ and error $= 2^{-\Omega(n)}$.*

Proof. Let $\epsilon = 2^{-2\delta n}$ for some $\delta > 0$. If $m > \delta n$, we can modify nmExt such that its output is of length δn (without increasing the error). Thus, without loss of generality we can assume $m \leq \delta n$.

We now define the non-malleable code in the following way: For any message $s \in \{0,1\}^m$, the encoder $\text{Enc}(s)$ outputs a uniformly random string from the set $\text{nmExt}^{-1}(s) \subset \{0,1\}^{Cn}$. For any codeword $c \in \{0,1\}^{Cn}$, the decoder Dec outputs $\text{nmExt}(c)$. Thus, for any message s, $\text{Dec}(\text{Enc}(s)) = s$. We now prove that the code is non-malleable.

Let $X_1, \ldots, X_C$ be independent, uniformly random sources on $\{0,1\}^n$. Let $f_1, \ldots, f_C$ be arbitrary tampering functions on $\{0,1\}^n$. From the definition of non-malleable extractors, we know that there exists an independent random variable D_f such that

$$|\text{nmExt}(X_1, \ldots, X_C) \circ \text{nmExt}(f_1(X_1), \ldots, f_C(X_C)) - U_m \circ \text{copy}(D_f, U_m)| < \epsilon \tag{1}$$

where both U_m's refer to the same string.

Let S be a distribution which is uniformly random on $\{0,1\}^m$. For $c \in \{0,1\}^{Cn}$, let $f(c)$ denote the tampered string in $\{0,1\}^{Cn}$ which is obtained by partitioning c into blocks of length n and the i^{th} block of n bits is tampered by f_i.

We note that,

$$|\mathrm{Enc}(S) - U_{Cn}| = |\mathrm{nmExt}^{-1}(S) - U_{Cn}| < |(\mathrm{nmExt}^{-1}) \circ (\mathrm{nmExt}(U_{Cn})) - U_{Cn}| + \epsilon = \epsilon$$

Using this in (1), we have

$$|\mathrm{Dec}(\mathrm{Enc}(S)) \circ \mathrm{Dec}(\mathrm{Enc}(f(S)) - U_m \circ \mathrm{copy}(D_f, U_m)| < 2\epsilon$$

Therefore, for any $s \in \{0,1\}^m$,

$$|\mathrm{Dec}(\mathrm{Enc}(s)) \circ \mathrm{Dec}(\mathrm{Enc}(f(s)) - s \circ \mathrm{copy}(D_f, s)| < \frac{2\epsilon}{\Pr[S = s]} = 2^{m+1}\epsilon$$

Hence,

$$|\mathrm{Dec}(\mathrm{Enc}(f(s)) - \mathrm{copy}(D_f, s)| < 2^{-\delta n/2}$$

This shows that the constructed code is indeed non-malleable. The efficiency of the decoder follows from the fact that nmExt is efficiently computable. □

Thus composing Theorem 3 with Theorem 4.1 gives us an explicit construction of non-malleable codes in the 10 split-state model with an efficient decoder. An efficient encoder for this non-malleable code follows from Theorem 8.7. This proves Theorem 1.

5 Proof outline of Theorem 3

In this section we sketch the main ideas involved in proving Theorem 3. This section can be skipped without any loss of continuation. The formal proof of Theorem 3 is presented in Section 6.

Definition 5.1. *We call a set A satisfying the conclusion of Theorem* 6.2 *to be sum-product friendly. We call a flat distribution sum-product friendly if its support is sum-product friendly.*

Let $X_1, \ldots, X_8$ be independent $(n, (1-\delta)n)$-sources and X_9 be an independent $(2n, 2(1-\delta)n)$-source. We view each X_i, $i \in [8]$, as a source on $\mathbb{F}_p$ for some prime p, $2^n < p < 2^{n+1}$.

5.1 A first attempt

For simplicity, assume that we are dealing with tampering functions with no fixed points. Consider the sources $(X_i, f_i(X_i))$ on $\mathbb{F}_p^2$ with min-entropy $(1-\delta)\log p$. Following ideas of constructing multi-source extractors from the sum-product theorem (Theorem 6.1) in [BIW06], suppose we have that the source $(X_1 \cdot X_2 + X_3, f_1(X_1) \cdot f_2(X_2) + f_3(X_3))$ expands (in a statistical sense) and is $p^{-\Omega(1)}$-close to a source with min-entropy $(1+\delta)\log p$.

Since the maximum min-entropy in the source $f_1(X_1) \cdot f_2(X_2) + f_3(X_3)$ is $\log p$, we are in good shape. In particular by Corollary 2.11, $(X_1 \cdot X_2 + X_3)|(f_1(X_1) \cdot f_2(X_2) + f_3(X_3))$ is $p^{-\Omega(1)}$-close to a source with min-entropy $\Omega(\delta \log p)$ with probability $1 - p^{-\Omega(1)}$. Following this, we can thus group the sources in blocks of 3 and output

$$3\mathrm{Ext}(X_1 \cdot X_2 + X_3, X_4 \cdot X_5 + X_6, X_7 \cdot X_8 + X_9)$$

where 3Ext is the extractor from Theorem 2.19.

5.2 A simple counterexample to the approach above

It turns out that the source $(X_1 \cdot X_2 + X_3, f_1(X_1) \cdot f_2(X_2) + f_3(X_3))$ need not cross the $\log p$ min-entropy barrier. As an easy counter example consider the tampering functions $f_1(x) = 2x$, $f_2(x) = 2x$ and $f_3(x) = 4x$ (where we view the tampering functions as functions from $\mathbb{F}_p$ to $\mathbb{F}_p$). We see that

$$(X_1 \cdot X_2 + X_3, f_1(X_1) \cdot f_2(X_2) + f_3(X_3)) = (Y, 4Y)$$

for some distribution Y on $\mathbb{F}_p$. Thus the min-entropy expansion step in our attempted construction fails.

5.3 The actual construction

The high level idea is to make the approach of our first attempt work by characterizing all counterexamples to expansion and then using suitable encodings of the sources to avoid such counterexamples. We can ensure expansion from encodings under certain assumptions on the maximum pre-image size and number of fixed points of the tampering functions. We combine this with other extractor ideas to build seedless non-malleable multi-source extractors. We note that the idea of encoding sources was also used by Bourgain [Bou05b] for constructing extractors for 2 independent sources.

We now present the main steps involved in our construction. We assume $n \geq n_0$ for some constant n_0 (if $n < n_0$, we can do a constant time brute-force search for optimal extractors).

- We view each $(n, (1-\delta)n)$-source X_i, $i \in [8]$, as a source on $\mathbb{F}_p$, $2^n < p < 2^{n+1}$. We encode each x_i as $\mathrm{enc}(x_i) = (x_i, q(x_i))$ for some suitable $q()$ to be fixed later. Define the source

  $$X_{f,i,j} = (\mathrm{enc}(X_i) + \mathrm{enc}(X_j), \mathrm{enc}(f_i(X_i)) + \mathrm{enc}(f_j(X_j)))$$

  Note that $X_{f,i,j}$ is a source on $\mathbb{F}_p^4$.

  We find a suitable encoding such that the following claim holds.

  Claim 5.2 (informal)**.** *$X_{f,1,2} \odot X_{f,3,4} + X_{f,5,6} \odot X_{f,7,8}$ is $p^{-\Omega(1)}$-close to a source with min-entropy $(2 + 20\delta)\log p$ under the assumption that at least one of the f_i's has no fixed points and the maximum pre-image size of each of the f_i's is bounded.*

- To find a good encoding enc, we first derive a sum-product estimate over $\mathbb{F}_p^4$ in Theorem 6.2 which characterizes sets that do not expand. We roughly show that for a set $A \subset \mathbb{F}_p^4$ of size $p^{2-\delta}$ such that $\Pi_{\{1,2\}}(A), \Pi_{\{3,4\}}(A) > p^{1+\delta'}$ for $\delta' >> \delta$ and $|A \cap (\mathbb{F}_p^*)^4| > \frac{1}{2}|A|$, we have $|A + A| + |A \odot A| > p^{2+10\delta}$ unless A has a large intersection with a 2-dimensional plane of a certain form in $\mathbb{F}_p^4$.

  The sum-product estimate is proved in Section 7. It is obtained by closely following the proof of a sum product estimate over $\mathbb{F}_p^2$ obtained by Bourgain in [Bou05a] and extending the arguments to $\mathbb{F}_p^4$.

- The idea to prove Claim 5.2 is to adapt the machinery developed in [BIW06] for proving such expansion statements about min-entropy to a more general setting. We point out the key differences from [BIW06] and our contribution in making the proof work.

1. The sources $X_{f,i,j}$ are not flat sources. We show that each such $X_{f,i,j}$ is close to a convex combination of a constant number of flat sources. Since not all sets in $\mathbb{F}_p^4$ are sum-product friendly, we keep track of the supports of these flat sources.
2. Our key contribution here is to show that for the choice of $\mathrm{enc}(x) = (x, x^4 + x^2 + x)$, the flat sources corresponding to $X_{f,i,j}$ are sum-product friendly if at least one of f_i or f_j has no fixed points and the maximum pre-image size of f_i and f_j is bounded.
3. Thus we are dealing with convex combinations of distributions of the form $A \odot B + C \odot D$ where A, B, C, D are flat sources on $\mathbb{F}_p^4$ with the guarantee that at least one of the flat sources is sum-product friendly. We show that the proof technique of [BIW06] goes through even with this weaker guarantee.

- Define the following function.

  $$\mathrm{ext}_1(x_1, \ldots, x_8) = (\mathrm{enc}(x_1)+\mathrm{enc}(x_2)) \odot (\mathrm{enc}(x_3)+\mathrm{enc}(x_4)) + (\mathrm{enc}(x_5)+\mathrm{enc}(x_6)) \odot (\mathrm{enc}(x_7)+\mathrm{enc}(x_8))$$

  We use Claim 5.2 and Corollary 2.11 to conclude the following.

  Claim 5.3 (informal)**.** $\mathrm{ext}_1(X_1, \ldots, X_8) | \mathrm{ext}_1(f_1(X_1), \ldots, f_8(X_8))$ *is* $p^{-\Omega(1)}$*-close to a source with min-entropy* $10\delta \log p$ *with probability* $1 - p^{-\Omega(1)}$ *assuming that none of the* f_i*'s have large maximum pre-image size and at least one of the* f_i*'s have no fixed points.*

- We next prove that the requirement on pre-image size of the tampering functions in Claim 5.3 can be removed.

  Claim 5.4 (informal)**.** $\mathrm{ext}_1(X_1, \ldots, X_8) | \mathrm{ext}_1(f_1(X_1), \ldots, f_8(X_8))$ *is* $p^{-\Omega(1)}$*-close to a source with min-entropy* $10\delta \log p$ *with probability* $1 - p^{-\Omega(1)}$ *assuming that at least one of the* f_i*'s have no fixed points.*

  (We note that Claim 5.2 may not hold without the restriction on maximum pre-image size of the f_i's and hence we use some new observations for proving Claim 5.4)

- To motivate our final construction, we describe an extractor ext_2 in this step which we don't actually use in our construction. The formal proofs of the claims made in this step are not included in this paper.

  Let SExt be the strong 2-source extractor from Theorem 2.18.

  Let $\mathrm{ext}_2 : (\{0,1\}^n)^8 \times \{0,1\}^{2n} \to \{0,1\}^m$, $m = \Omega(n)$, be defined as:

  $$\mathrm{ext}_2(x_1, \ldots, x_9) = \mathrm{SExt}(\mathrm{ext}_1(x_1, \ldots, x_8), x_9)$$

  The following result follows from Claim 5.4.

  Claim 5.5 (informal)**.** *Let* $X_1, \ldots, X_8$ *be independent* $(n, (1-\delta)n)$*-sources and* X_9 *be an independent* $(2n, 2(1-\delta)n)$*-source. Then* $\mathrm{ext}_2(X_1, \ldots, X_9) | \mathrm{ext}_2(f_1(X_1), \ldots, f_9(X_9))$ *is* $p^{-\Omega(1)}$*-close to* U_m *with probability* $1 - p^{-\Omega(1)}$ *if there exists some* $i \in [8]$ *such that* f_i *has no fixed points.*

The proof of the above claim follows from the following observations. Define the random variable $W = \text{ext}_1(X_1, \ldots, X_8)$ and $V = \text{ext}_1(f_1(X_1), \ldots, f_8(X_8))$. We know by Claim 5.4 that for most fixings of $V = v$, W is $p^{-\Omega(1)}$-close to a source with min-entropy $10\delta \log p = 5\delta(2n)$. Since SExt is an extractor that for 2 independent sources on $\{0,1\}^{2n}$ with min-entropy k_1, k_2 satisfying $k_1 + k_2 \geq (2+\delta)n$, by Theorem 2.18 we have

$$|\text{SExt}(W, X_9) \circ V \circ X_9 - U_m \circ V \circ X_9| < 2^{-\Omega(n)}$$

The proof of Claim 5.5 now follows.

- However, ext_2 cannot be the required non-malleable extractor in Theorem 3. In particular when for all $i \in [8]$, f_i is the identity function and f_9 is any arbitrary tampering function with no fixed points, ext_2 does not work. Instead we replace SExt with snmExt and present our final construction.

  Let $\text{nmExt} : (\{0,1\}^n)^8 \times \{0,1\}^{2n} \to \{0,1\}^m$, $m = \Omega(n)$, be defined as:

  $$\text{nmExt}(x_1, \ldots, x_9) = \text{snmExt}(\text{ext}_1(x_1, \ldots, x_8), x_9)$$

  where snmExt is the seeded non-malleable extractor from Theorem 2.20. We prove the following claim.

  Claim 5.6 (informal)**.** *Let $X_1, \ldots, X_8$ be independent $(n, (1-\delta)n)$-sources and X_9 be an independent $(2n, 2(1-\delta)n)$-source. Then* $\text{nmExt}(X_1, \ldots, X_9) | \text{nmExt}(f_1(X_1), \ldots, f_9(X_9))$ *is $p^{-\Omega(1)}$-close to U_m with probability $1 - p^{-\Omega(1)}$ when at least one of the f_i's have no fixed points.*

  For an easier presentation of the main ideas, we outline the proof of the above claim in a simpler setting where each f_i is either the identity function or has no fixed points and at least one of the f_i's is not the identity function.

  The following cases arise depending on the f_i's.

  1. Suppose there is some $j \in [8]$ such that f_j has no fixed points. The conclusion in this case follows from Claim 5.5.
  2. Now suppose for all $j \in [8]$, f_j is the identity function. Thus f_9 has no fixed points.
     Set W to be the random variable $\text{ext}_1(X_1, \ldots, X_8)$.
     We show that W is $p^{-\Omega(1)}$-close to a source Z with min-entropy $2(1-2\delta)n$. Note that Z and $X_{9I(9)}$ are independent sources on $\{0,1\}^{2n}$, each with min-entropy rate $> (1-2\delta)$ and f_9^I has no fixed points. Thus by Theorem 2.20, we have

     $$|\text{snmExt}(Z, X_9) \circ \text{snmExt}(Z, f_9(X_9)) - U_m \circ \text{snmExt}(Z, f_9(X_9))| < 2^{-\Omega(n)}$$

     This concludes the proof of Claim 5.6.

We now summarize the proof of Theorem 3.

- Let S_i denote the support of each (flat) independent source X_i. We partition each S_i into S_{i0} and S_{i1} such that $f_i(x_{i0}) = x_{i0}$ for all $x_{i0} \in S_{i0}$ and f_i has no fixed points in S_{i1}. Let X_{i0} and X_{i1} be flat distributions supported on S_{i0} and S_{i1} respectively.

  For any 0-1 vector I, let $I(i)$ denote the i'th entry in I. Let $w_I = \prod_{i=1}^{9} \frac{|S_{iI(i)}|}{|S_i|}$ for $I \in \{0,1\}^9$.

  We use Claim 5.6 to prove the following.

Claim 5.7 (informal)**.** *Let $I \in \{0,1\}^9 \setminus \{\vec{0}\}$. Then*

$$w_I \cdot |\mathrm{nmExt}(X_{1I(1)}, \ldots, X_{9I(9)}) \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)})) - U_m \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)}))| < 2^{-\Omega(n)}$$

We also prove that nmExt is a multi-source extractor.

Claim 5.8 (informal)**.** *Let $Y_1, \ldots, Y_8$ be independent $(n, (1-2\delta)n)$-sources and Y_9 an independent $(2n, 2(1-2\delta)n)$-source . Then*

$$|\mathrm{nmExt}(Y_1, \ldots, Y_9) - U_m| < 2^{-\Omega(n)}$$

- Define the random variable D_f as:

$$D_f = w_{\vec{0}} \cdot \{same^\star\} + \sum_{I \in \{0,1\}^9 \setminus \{\vec{0}\}} w_I \cdot \mathrm{nmExt}(f_1(X'_{1I(1)}), \ldots, f_9(X'_{9I(9)}))$$

where for each $i \in [9]$ and $I \in \{0,1\}^9$, $X'_{iI(i)}$ is identically distributed as $X_{iI(i)}$ and independent of $X_1, \ldots, X_9$.

Recall that we need to prove :

$$|\mathrm{nmExt}(X_1, \ldots, X_9) \circ \mathrm{nmExt}(f_1(X_1), \ldots, f_9(X_9)) - U_m \circ \mathrm{copy}(D_f, U_m)| < 2^{-\Omega(n)} \quad (2)$$

We have,

$$\begin{aligned} &|\mathrm{nmExt}(X_1, \ldots, X_9) \circ \mathrm{nmExt}(f_1(X_1), \ldots, f_9(X_9)) - U_m \circ \mathrm{copy}(D_f, U_m)| \\ \leq (*) + &\sum_{I \in \{0,1\}^9 \setminus \{\vec{0}\}} w_I \cdot |\mathrm{nmExt}(X_{1I(1)}, \ldots, X_{9I(9)}) \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)})) \\ &- U_m \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)}))| \quad (3) \end{aligned}$$

where,

$$(*) = w_{\vec{0}} \cdot |\mathrm{nmExt}(X_{10}, \ldots, X_{90}) \circ \mathrm{nmExt}(f_1(X_{10}), \ldots, f_9(X_{90})) - \mathrm{copy}(same^\star, U_m) \circ U_m| \quad (4)$$

For $i \in [9]$, the support of X_{i0} is S_{i0} and by definition $f_i(x_{i0}) = x_{i0}$ for all $x_{i0} \in S_{i0}$. Using the definition of the function copy, we have

$$\begin{aligned} (*) &= w_{\vec{0}} \cdot |\mathrm{nmExt}(X_{10}, \ldots, X_{90}) \circ \mathrm{nmExt}(X_{10}, \ldots, X_{90}) - U_m \circ U_m| \\ &= w_{\vec{0}} \cdot |\mathrm{nmExt}(X_{10}, \ldots, X_{90}) - U_m| \quad (5) \\ &\leq 2^{-\Omega(n)} \end{aligned}$$

where the inequality in the last step is derived from Claim 5.8. Each term of the summation in the RHS of (3) can be bounded by $2^{-\Omega(n)}$ using Claim 5.7. Since there are $2^9 - 1$ terms in the summation, we have the required bound in (2). This concludes the proof outline of Theorem 3.

6 Proof of Theorem 3

6.1 A sum-product estimate

The following sum-product theorem over prime fields follows from [BKT04], [BGK06], and [Kon03].

Theorem 6.1 (Sum-product over prime fields)**.** *Let $\mathbb{F}_p$ be any prime field and let $A \subset \mathbb{F}_p$ be any non-empty subset such that $|A| < p^{1-\delta}$ for some constant $\delta > 0$. Then there exists a constant $\tau = \tau(\delta) > 0$, such that*

$$|A + A| + |A \cdot A| \geq |A|^{1+\tau}$$

An analogue of Theorem 6.1 over $\mathbb{F}_p \times \mathbb{F}_p$ was proved by Bourgain in [Bou05a]. We extend this to sets over $\mathbb{F}_p^4$ in the following theorem and use it in our proof of Theorem 3. It is stated in a convenient way.

Theorem 6.2. *There exists $\tau_0 > \tau_1 > 0$ such that the following holds : Let A be a subset of $\mathbb{F}_p^4$ satisfying $|A \cap (\mathbb{F}_p^*)^4| \geq \frac{|A|}{2}$. Suppose that for any subset $A_1 \subseteq A$ satisfying $|A_1| \geq p^{-\tau_1}|A|$, the following conditions holds.*

1. *$\Pi_{\{1,2\}}(A_1) \geq p^{1+\tau_0}$ and $\Pi_{\{3,4\}}(A_1) \geq p^{1+\tau_0}$.*
2. *$A_1 \nsubseteq P$, where P is a 2-dimensional linear subspace of $\mathbb{F}_p^4$ of the form*
   - *(a) $\{(x_1, x_2, c_1x_1, c_2x_2) : x_1 \in \mathbb{F}_p, x_2 \in \mathbb{F}_p\}$ or*
   - *(b) $\{(x_1, x_2, c_2x_2, c_1x_1) : x_1 \in \mathbb{F}_p, x_2 \in \mathbb{F}_p\}$.*

Then there exists a constant $\tau > 0$ (depending on τ_0, τ_1) such that if $|A| < p^{7/3-\tau_1}$, then

$$|A + A| + |A \odot A| > p^{\tau}|A|$$

We present the proof of Theorem 6.2 in Section 7. The proof of Theorem 6.2 closely follows and extends the arguments in the sum-product estimate over $\mathbb{F}_p^2$ proved by Bourgain.

Definition 6.3. *We call a set A satisfying the conclusion of Theorem* 6.2 *to be sum-product friendly. We call a flat distribution sum-product friendly if its support is sum-product friendly.*

6.2 A sum-product friendly encoding

Let τ, τ_0, τ_1 be the constants from Theorem 6.2. Let p be any prime satisfying : $p^{\tau_0} > 16$.

Define $\mathrm{enc} : \mathbb{F}_p \to \mathbb{F}_p^2$ in the following way.

$$\mathrm{enc}(x) = (x, x^4 + x^2 + x)$$

Lemma 6.4. *Let $S_1, S_2 \subset \mathbb{F}_p$ be subsets of size $p^{1-\delta}$, $p > 3$. Define the distribution*

$$X_{f,1,2} = (\mathrm{enc}(x_1) + \mathrm{enc}(x_2), \mathrm{enc}(f_1(x_1)) + \mathrm{enc}(f_2(x_2))) : x_1 \sim S_1, x_2 \sim S_2$$

where f_1, f_2 are arbitrary functions.

Then $X_{f,1,2}$ is $O(p^{-\delta})$-close to a convex combination of at most 4 flat distributions supported on sets of the form

$$T_i = \{(\mathrm{enc}(x_1) + \mathrm{enc}(x_2), \mathrm{enc}(f_1(x_1) + \mathrm{enc}(f_2(x_2))) : (x_1, x_2) \in G_i\},$$

where $G_i \subset \mathbb{F}_p^2$ and $|G_i| = |T_i| \geq p^{2-3\delta}$.

Proof. Let $T \subset \mathbb{F}_p^4$ denote the support of $X_{f,1,2}$. We partition T into at most 4 parts in the following way.

For any $t \in T$, let $s(t) \subset \mathbb{F}_p^2$ be the set of all $(x_1, x_2) \in S_1 \times S_2$ such that $(\mathrm{enc}(x_1) + \mathrm{enc}(x_2)$, $\mathrm{enc}(f_1(x_1)) + \mathrm{enc}(f_2(x_2))) = t$. Let $r(x)$ denote the cardinality of the set $s(x)$.

We claim that for any $t \in T$, $1 \leq r(x) \leq 4$. The upper bound follows from the following calculation. Let $t = (t_1, t_2, t_3, t_4) \in F_p^4$. Thus for any $(x_1, x_2) \in s(t)$, we have

$$\begin{aligned} x_1 + x_2 &= t_1 \\ x_1^4 + x_1^2 + x_1 + x_2^4 + x_2^2 + x_2 &= t_2 \end{aligned}$$

Substituting for x_2, we have

$$x_1^4 + (t_2 - x_1)^4 + q(x_1, t_1, t_2) = 0$$

where $q(x_1, t_1, t_2)$ has degree at most 2 in x_1. Thus x_1 must satisfy a polynomial of degree exactly 4. For each fixing of x_1, notice that x_2 also gets fixed. Thus $r(t) \leq 4$ for all $t \in T$.

For $i \in [4]$, we define the sets

$$T_i = \{t \in T : r(t) = i\}$$

Thus the T_i's form a partition of T.

Define sets $G_i \subset \mathbb{F}_p^2$, $i \in [4]$, such that for all $t \in T_i$, $|G_i \cap s(t)| = 1$. In other words G_i is constructed by picking exactly one element from $s(t)$ for each $t \in T_i$. Thus $|G_i| = |T_i|$.

We note that for any $t \in T_i$, $\Pr[X_{f,1,2} = t] = \frac{i}{|S_1||S_2|}$ and hence

$$\Pr[X_{f,1,2} \in T_i] = \frac{i|G_i|}{|S_1||S_2|}$$

Thus we have

$$X_{f,1,2} = \sum_i^4 w_i \cdot \big((\mathrm{enc}(x_1) + \mathrm{enc}(x_2), \mathrm{enc}(f_1(x_1)) + \mathrm{enc}(f_2(x_2))) : (x_1, x_2) \sim G_i\big)$$

where $w_i = \frac{i|G_i|}{|S_1||S_2|}$.

For some i, if $|G_i| < p^{2-3\delta}$ then $w_i \leq i \cdot p^{-\delta}$. Thus $X_{f,i,j}$ is $9 \cdot p^{-\delta}$-close to a distribution $X'_{f,i,j}$ defined as

$$X'_{f,1,2} = \sum_i^4 w'_i \cdot \big((\mathrm{enc}(x_1) + \mathrm{enc}(x_2), \mathrm{enc}(f_1(x_1)) + \mathrm{enc}(f_2(x_2))) : (x_1, x_2) \sim G_i\big)$$

where we set w'_i's as follows. Set $w'_i = 0$ for all i such that $w_i < i \cdot p^{-\delta}$. Pick a j such that $w_j \geq j \cdot p^{-\delta}$ and set $w'_j = w_j + \sum_{i:w_i < i \cdot p^{-\delta}} w_i$. For the remaining unset w'_i's, set it equal to w_i. □

Lemma 6.5. *Choose a small $\delta_1 > \tau_0$. Let f_1, f_2 be functions with maximum pre-image size bounded by p^{δ_1}. Further assume f_1 has no fixed points. Define the set $A = \{\text{enc}(x_1) + \text{enc}(x_2), \text{enc}(f_1(x_1)) + \text{enc}(f_2(x_2)) : (x_1, x_2) \in G\}$ where $G \subset \mathbb{F}_p^2$ is a subset of size at least $p^{1+10\delta_1}$. Then the set $A \subset \mathbb{F}_p^4$ is sum-product friendly.*

Proof. We begin by noting that $p^{1+9\delta_1} < |A| << p^{7/3}$.

We need the following claim.

Claim 6.6. *Define the set $B = \{(\text{enc}(y_1)+\text{enc}(y_2), \text{enc}(g_1(y_1))+\text{enc}(g_2(y_2))) : (y_1, y_2) \in H\}$ where $H \subset \mathbb{F}_p^2$ is a subset of size at least $p^{1+10\delta_1}$ and g_1, g_2 are tampering functions with pre-image size bounded by p^{δ_1}. Then following inequalities hold :*

- $|B \cap (\{0\} \times \mathbb{F}_p^3)| \leq p$
- $|B \cap (\mathbb{F}_p \times \{0\} \times \mathbb{F}_p^2)| \leq 4.p$
- $|B \cap (\mathbb{F}_p^2 \times \{0\} \times \mathbb{F}_p)| \leq p^{1+\delta_1}$
- $|B \cap (\mathbb{F}_p^3 \times \{0\})| \leq 4p^{1+\delta_1}$

Proof. We have,

$B = \{(y_1 + y_2, y_1^4 + y_1^2 + y_1 + y_2^4 + y_2^2 + y_2, g_1(y_1) + g_2(y_2), g_1(y_1)^4 + g_1(y_1)^2 + g_1(y_1) + g_2(y_2)^4 + g_2(y_2)^2 + g_2(y_2)) : (y_1, y_2) \in H\}$.

We prove the inequality:

$$|B \cap (\mathbb{F}_p^3 \times \{0\})| \leq 4p^{1+\delta_1}$$

The other inequalities follow using similar arguments.

Fix y_1 to some value in $\mathbb{F}_p$. We note that $g_2(y_2)$ is the root of a monic degree 4 polynomial and hence has at most 4 choices. Thus y_2 can take at most $4p^{\delta_1}$ values by using the bound on the pre-image size of g_2. The inequality now follows by observing that y_1 can take at most p values. □

Using Claim 6.6, we have $|A \cap (\mathbb{F}_p^*)^4| \geq (1 - p^{-7\delta_1})|A| > \frac{1}{2}|A|$.

Consider any subset $A_1 \subseteq A$ such that $|A_1| \geq p^{-\tau_1}|A|$. It follows that there exists $G_1 \subseteq G$ such that

$$A_1 = \{(\text{enc}(x_1) + \text{enc}(x_2), \text{enc}(f_1(x_1)) + \text{enc}(f_2(x_2))) : (x_1, x_2) \in G_1\}$$

Thus $|A_1| > p^{1+8\delta_1}$. We also note that $|G_1| \geq |A_1| > p^{1+8\delta_1}$.

We note that $|\Pi_{1,2}(A_1)| = |A_1| > p^{1+\tau_0}$. Further $|\Pi_{3,4}(A_1)| > |A_1|p^{-2\delta_1} > p^{1+6\delta_1} > p^{1+\tau_0}$.

The final part of the proof is to bound the intersection of A_1 with any 2-dimensional linear space P of the forms specified in Theorem 6.2.

Suppose $A_1 \subset P = \{(y_1, y_2, c_1y_1, c_2y_2) : y_1, y_2 \in \mathbb{F}_p\}$. Thus we have for all $(x_1, x_2) \in G_1$:

$$f_1(x_1) + f_2(x_2) = c_1(x_1 + x_2)$$
$$f_1(x_1)^4 + f_1(x_1)^2 + f_1(x_1) + f_2(x_2)^4 + f_2(x_2)^2 + f_2(x_2) = c_2(x_1^4 + x_1^2 + x_1 + x_2^4 + x_2^2 + x_2)$$

Fix $x_2 = \alpha$ such that $(x_1, \alpha) \in G_1$ for all $x_1 \in S_1 \subset \mathbb{F}_p$, $|S_1| \geq \frac{|G_1|}{p} \geq p^{8\delta_1}$. Let $f_2(\alpha) = \beta$. We thus have for all $x_1 \in S_1$,

$$f_1(x_1) = c_1 x_1 + c_1\alpha - \beta \tag{6}$$

$$f_1(x_1)^4 + f_1(x_1)^2 + f_1(x_1) + \beta^4 + \beta^2 + \beta = c_2(x_1^4 + x_1^2 + x_1 + \alpha^4 + \alpha^2 + \alpha) \tag{7}$$

$$\tag{8}$$

Thus for all $x \in S_1$, the following holds:

$$(c_1x_1 + c_1\alpha - \beta)^4 + (c_1x_1 + c_1\alpha - \beta)^2 + (c_1x_1 + c_1\alpha - \beta) + \beta^4 + \beta^2 + \beta \\ -c_2(x_1^4 + x_1^2 + x_1 + \alpha^4 + \alpha^2 + \alpha) = 0 \tag{9}$$

To derive a contradiction, we split it into the following cases.

- $c_1 \neq 0$, $c_1\alpha - \beta \neq 0$

  In this case notice that the LHS of (9) is of degree at least 3 and at most 4 in x_1 and hence can have at most 4 roots, which is a contradiction since $|S_1| \geq p^{8\delta_1} > 4$.

- $c_1 = 0$

  In this case we see that from (6), f_1 is constant on S_1 which contradicts the assumption that f_1 has pre-image size at most p^{δ_1}.

- $c_1\alpha - \beta = 0$, $c_1 \neq 0$

  Thus (9) simplifies to

  $$c_1^4x_1^4 + c_1^2x_1^2 + c_1x_1 + \beta^4 + \beta^2 + \beta - c_2(x_1^4 + x_1^2 + x_1 + \alpha^4 + \alpha^2 + \alpha) = 0 \tag{10}$$

  We see that this is at least a linear equation and at most a degree 4 equation in x_1 (and thus a contradiction, as argued above) unless $c_1^4 = c_1^2 = c_1 = c_2$. Thus $c_1 = 1$ (since $c_1 \neq 0$). But by (6), we then have $f_1(x_1) = x_1$ for all $x_1 \in S_1$. This contradicts the fact that f_1 has no fixed points.

This contradicts our assumption that $A_1 \subseteq \{(y_1, y_2, c_1y_1, c_2y_2) : y_1, y_2 \in \mathbb{F}_p\}$.

Now suppose $A_1 \subseteq P = \{(y_1, y_2, c_2y_2, c_1y_1) : y_1, y_2 \in \mathbb{F}_p\}$. We arrive at a contradiction using similar arguments as above. We have for all $(x_1, x_2) \in G_1$

$$f_1(x_1) + f_2(x_2) = c_2(x_1^4 + x_1^2 + x_1 + x_2^4 + x_2^2 + x_2)$$

$$f_1(x_1)^4 + f_1(x_1)^2 + f_1(x_1) + f_2(x_2)^4 + f_2(x_2)^2 + f_2(x_2) = c_1(x_1 + x_2)$$

Fix $x_2 = \alpha$ such that $(x_1, \alpha) \in G_1$ for all $x_1 \in S_1 \subset \mathbb{F}_p$, $|S_1| \geq \frac{|G_1|}{|p|} \geq p^{8\delta_1}$. Let $f_2(\alpha) = \beta$. We thus have for all $x \in S_1$,

$$f_1(x_1) = c_2(x_1^4 + x_1^2 + x_1 + \alpha^4 + \alpha^2 + \alpha) - \beta \tag{11}$$

$$f_1(x_1)^4 + f_1(x_1)^2 + f_1(x_1) + \beta^4 + \beta^2 + \beta = c_1(x_1 + x_2) \tag{12}$$

$$\tag{13}$$

It follows that for all $x \in S_1$,

$$\left(c_2(x_1^4 + x_1^2 + x_1 + \alpha^4 + \alpha^2 + \alpha) - \beta\right)^4 + \left(c_2(x_1^4 + x_1^2 + x_1 + \alpha^4 + \alpha^2 + \alpha) - \beta\right)^2 + \left(c_2(x_1^4 + x_1^2 + x_1 + \alpha^4 + \alpha^2 + \alpha) - \beta\right) + \beta^4 + \beta^2 + \beta - c_1(x_1 + \alpha) = 0 \tag{14}$$

We note that (14) is a degree 16 equation in x_1 (and hence a contradiction since $p^{8\delta_1} > 16$) unless $c_2 = 0$. But if $c_2 = 0$ then from (11) we have f_1 is constant on S_1 which contradicts our assumption that f_1 has pre-image size at most p^{δ_1}. This completes our proof that A is sum-product friendly.

□

In the following lemmas, we shall abuse notation and for any set A, we will also use A to denote the flat distribution with support A.

Choose δ_1 small enough such that for a sum-product friendly set A of size $p^{2-5\cdot 10^3\delta_1}$ we have $|A + A| + |A \odot A| > |A|p^{5\cdot 10^4\delta_1}$. This can be ensured by choosing $\delta_1 = 10^{-5} \cdot \tau$, where τ is the constant from Theorem 6.2.

Lemma 6.7. *Let $G_1, G_2, G_3 \subset \mathbb{F}_p^2$ be subsets of size at least $p^{2-\delta_1}$. Let $f_1, \ldots, f_6$ be functions with pre-image size at most $p^{10\delta_1}$. Further assume f_1 has no fixed points. For $i \in [3]$ define the sets $A_i = \{(\mathrm{enc}(x_{2i-1}) + \mathrm{enc}(x_{2i}), \mathrm{enc}(f_{2i-1}(x_{2i-1})) + \mathrm{enc}(f_{2i}(x_{2i}))) : (x_{2i-1}, x_{2i}) \in G_i\}$. Then $A_1 \odot A_2 + A_3$ is $O(p^{-\delta_1})$-close to a distribution with min-entropy $(2 + 10\delta_1)\log p$.*

To prove the above lemma, we borrow ideas from [BIW06] and use the proof technique developed in their work.

We begin by proving the following lemmas.

Lemma 6.8. *Let $A \subset (\mathbb{F}_p{}^*)^4$, $p^{2-300\delta_1} \leq |A| < p^2$ be such that any subset $A' \subseteq A$ of size greater than $p^{2-5\cdot 10^3\delta_1}$ is sum-product friendly. Supose that for some $B \subset (\mathbb{F}_p^*)^4$, we have $|A \odot B| \leq p^{2+300\delta_1}$, $p^{2-300\delta_1} \leq |B| < p^2$. Then for any $C \subset (\mathbb{F}_p^*)^4$ such that $p^{2-\delta_1} \leq |C| < p^2$, we have $\mathrm{cp}(A + C) \leq p^{-(2+12\delta_1)}$.*

Proof. Since $|A \odot B| \leq p^{2+300\delta_1}$, using Lemma 2.15 we have $|A \odot A| \leq |A|p^{2400\delta_1}$. Suppose there is some set C such that $|C| > p^{2-\delta_1}$ and $\mathrm{cp}(A + C) > p^{-(2+12\delta_1)}$. Using Lemma 2.17 with $\rho_1 = 200\delta_1$ and $\rho_2 = 220\delta_1$, it follows that there exists sets $A' \subseteq A$, $C' \subseteq C$, $|A' + C'| \leq p^{2+5\cdot 10^3\delta_1}$ and $|A'|, |C'| > p^{2-5\cdot 10^3\delta_1}$. Using Lemma 2.15, we get that $|A' + A'| \leq |A'|p^{4\cdot 10^4\delta_1}$. We also have $|A' \odot A'| \leq |A \odot A| \leq |A'|p^{10^4\delta_1}$. By our choice of δ_1, this contradicts A' being sum-product friendly. □

Switching the roles of addition and multiplication gives the following.

Lemma 6.9. *Let $A \subset (\mathbb{F}_p{}^*)^4$, $p^{2-300\delta_1} \leq |A| < p^2$ be such that any subset $A' \subseteq A$ of size at least $p^{2-5\cdot 10^3\delta_1}$ is sum-product friendly. Let $B \subset (\mathbb{F}_p^*)^4$ be a set such that $|A + B| \leq p^{2+300\delta_1}$, $p^{2-300\delta_1} \leq |B| < p^2$. Then for any $C \subset (\mathbb{F}_p^*)^4$ such that $p^{2-\delta_1} \leq |C| < p^2$, we have $\mathrm{cp}(A \odot C) \leq p^{-(2+12\delta_1)}$.*

We say that a set is plus-friendly if it satisfies the conclusion of Lemma 6.8. Similarly we say that a set is times-friendly if it satisfies the conclusion of Lemma 6.9.

Lemma 6.10. *Let $A_1 \subset \mathbb{F}_p^4$ be the set defined in Lemma 6.7. Then $A_1 = A_+ \cup A_\times \cup A_{11}$ such that the following hold:*

1. *A_+ is empty or plus-friendly*
2. *$A_\times$ is empty or times-friendly*
3. *$|A_{11}| \leq |A_1| p^{-\delta_1}$*

Proof of Lemma 6.10. We start out by replacing A_1 by $A_1 \cap (\mathbb{F}_p^*)^4$. We can do this without loss of generality since as observed in the proof of Lemma 6.5, $|A_1 \cap (\mathbb{F}_p^*)^4| > (1 - p^{-\delta_1})|A_1|$ and hence we add the set $A_1 \setminus (\mathbb{F}_p^*)^4$ to A_{11}.

Note that by Lemma 6.5, any subset of A_1 of size at least $p^{2-5\cdot 10^3 \delta_1}$ is sum-product friendly. Let $A_\times = A_1$ and $A_+ = \emptyset$. We maintain the invariance that A_+ is either plus-friendly or empty. If $A_\times$ is times-friendly then we are done. Else there exists some B of size at least $p^{2-\delta_1}$ such that $\mathrm{cp}(A_\times \odot B) > p^{-(2+12\delta_1)}$. Using Lemma 2.17 with $\rho_1 = 2\delta_1$ and $\rho_2 = 14\delta_1$, we have that there exists sets $A' \subseteq A_\times$, $B' \subseteq B$, $|A' \odot B'| \leq p^{2+284\delta_1}$ and $|A'|, |B'| \geq p^{2-282\delta_1}$. Thus, by Lemma 6.8, A' is plus-friendly. We remove A' from $A_\times$ and add it to A_+. Further it can be proved that unions of disjoint plus-friendly sets are also plus-friendly. We iterate as above till $A_\times$ is times-friendly or $|A_\times| \leq |A_1| p^{-\delta_1}$. □

Proof of Lemma 6.7. By Lemma 6.10 we have $A_1 = A_+ \cup A_\times \cup A'$. Using Claim 6.6, we have $|A_2 \cap (\mathbb{F}_p^*)^4| > (1 - p^{-\delta_1})|A_2|$ and $|A_3 \cap (\mathbb{F}_p^*)^4| > (1 - p^{-\delta_1})|A_3|$. Thus $A_1 \odot A_2 + A_3$ is $O(p^{-\delta_1})$-close to a convex combination of distributions of the form:

1. $A_+ \odot a_2 + A_3$, $a_2 \in A_2 \cap (\mathbb{F}_p^*)^4$
2. $A_\times \odot A_2 + a_3$, $a_3 \in A_3 \cap (\mathbb{F}_p^*)^4$

By Lemma 6.8 and Lemma 6.9, we thus have that $A_1 \odot A_2 + A_3$ is $O(p^{-\delta_1})$-close to a distribution with collision probability at most $p^{-(2+12\delta_1)}$. Thus by using Lemma 2.7, we have that $A_1 \odot A_2 + A_3$ is $O(p^{-\delta_1})$-close to a distribution with min-entropy $(2 + 10\delta_1) \log p$. □

Theorem 6.11. *Let $X_1, \ldots, X_8$ be independent sources on $\mathbb{F}_p$ with min-entropy $(1-\delta)\log p$. Let $f_1, f_2, \ldots, f_8$ be arbitrary functions such that at least one of the f_i's has no fixed points. Further suppose that the pre-image of each f_i is bounded by $p^{10\delta}$. Define the source*

$$X_{f,i,j} = \mathrm{enc}(X_i) + \mathrm{enc}(X_j), \mathrm{enc}(f_i(X_i)) + \mathrm{enc}(f_j(X_j))$$

Then $X_{f,1,2} \odot X_{f,3,4} + X_{f,5,6} \odot X_{f,7,8}$ is $O(p^{-\delta})$-close to a distribution with min-entropy $(2+10\delta)\log p$.

Proof. Without loss of generality suppose f_1 has no fixed points. For all $i \in [3]$, using Lemma 6.4 we have that $X_{f,2i-1,2i}$ is $O(p^{-\delta})$-close to a convex combination of at most 4 flat distributions A_{ij} of the form $(\mathrm{enc}(x_{2i-1}) + \mathrm{enc}(x_{2i}), \mathrm{enc}(f_1(x_{2i-1}) + \mathrm{enc}(f_2(x_{2i}))) : (x_{2i-1}, x_{2i}) \sim G_{ij}$ where $G_{ij} \subset \mathbb{F}_p^2$, $|G_{ij}| \geq p^{2-3\delta}$.

With probability $1 - O(p^{-\delta})$ over fixing of the sources X_7, X_8, we have $X_{f,1,2} \odot X_{f,3,4} + X_{f,5,6} \odot x_{f,7,8}$ is $O(p^{-\delta})$-close to a convex combination of at most 4^3 distributions of the form $A_{1j_1} \cdot A_{2j_2} +$

$\alpha \cdot A_{3j_3}$, $\alpha \in (\mathbb{F}_p^*)^4$. Since f_1 has no fixed points, by Lemma 6.7 with $\delta_1 = 3\delta$, we have that $A_{1j_1} \odot A_{2j_2} + \alpha \odot A_{3j_3}$ is $O(p^{-\delta})$-close to a distribution with min-entropy $(2 + 10\delta)\log p$. Hence, $X_{f,1,2} \odot X_{f,3,4} + X_{f,5,6} \odot X_{f,7,8}$ is $O(p^{-\delta})$-close to a distribution with min-entropy $(2 + 10\delta)\log p$.

□

6.3 Non-malleable extractors for functions with no fixed points

In this section we prove a special case of Theorem 3 where we have a restriction on the fixed points of the tampering functions. We use this result in the proof of Theorem 3.

Theorem 6.12. *There exists a constant $\delta > 0$ such that for every n there exists an explicit function* $\mathrm{nmExt} : (\{0,1\}^n)^8 \times \{0,1\}^{2n} \to \{0,1\}^m$, *such that if $X_1, X_2, \ldots, X_8$ are independent $(n, (1-\delta)n)$-sources, X_9 an independent $(2n, 2(1-\delta)n)$-source and $f_1, f_2, \ldots, f_9$ are arbitrary tampering functions such that there exists $j \in [8]$ such that f_j has no fixed points, then*

$$|\mathrm{nmExt}(X_1, \ldots, X_9) \circ \mathrm{nmExt}(f_1(X_1), \ldots, f_9(X_9)) - U_m \circ \mathrm{nmExt}(f_1(X_1), \ldots, f_9(X_9))| < 2^{-\Omega(n)}$$

Proof. We view each X_i, $i \in [8]$, as a source on $\mathbb{F}_p$ for a prime p satisfying $2^n < p < 2^{n+1}$. If $p^{\tau_0} \leq 16$, we do a brute-force search for nmExt (in constant time). Thus assume $p^{\tau_0} > 16$.

Let $\mathrm{snmExt} : \{0,1\}^{2n} \times \{0,1\}^{2n} \to \{0,1\}^m$, $m = \Omega(n)$, be the seeded non malleable extractor from Theorem 2.20. Define the functions

$$\mathrm{ext}_1(x_1, x_2, \ldots, x_8) = \sum_{i=0}^{1} \big(\mathrm{enc}(x_{4i+1}) + \mathrm{enc}(x_{4i+2})\big) \odot \big(\mathrm{enc}(x_{4i+3}) + \mathrm{enc}(x_{4i+4})\big)$$

$$\mathrm{nmExt}(x_1, \ldots, x_9) = \mathrm{snmExt}(\mathrm{ext}_1(x_1, \ldots, x_8), x_9)$$

We show that nmExt satisfies the conclusion of Theorem 6.12.

Let $S_i \subset \mathbb{F}_p$ be the support of the flat source X_i for all $i \in [8]$. Also let $S_9 \subset \{0,1\}^{2n}$ be the support of X_9. We partition each S_i into S_{i0} and S_{i1} based on the pre-image of f_i as follows.

$$S_{i0} = \{s \in S_i : |f_i^{-1}(s) \cap S_i| \leq p^{20\delta}\}, \;\; S_{i1} = S_i \setminus S_{i1}.$$

Let X_{ij} be the flat source on S_{ij} for $j = 0, 1$.

We thus have

$$\begin{aligned} |\mathrm{nmExt}(X_1, \ldots, X_9) \circ \mathrm{nmExt}(f_1(X_1), \ldots, f_9(X_9)) - U_m \circ \mathrm{nmExt}(f_1(X_1), \ldots, f_9(X_9))| \quad (15) \\ \leq \sum_{I \in \{0,1\}^9} w_I \cdot |\mathrm{nmExt}(X_{1I(1)}, \ldots, X_{9I(9)}) \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)})) - \\ U_m \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)}))| \quad (16) \end{aligned}$$

where $w_I = \prod_{i=1}^{9} \left(\frac{|S_{iI(i)}|}{|S_i|}\right)$.

We bound each term in (16). In particular we show that

$$\begin{aligned} w_I \cdot |\mathrm{nmExt}(X_{1I(1)}, \ldots, X_{9I(9)}) \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)})) - \\ U_m \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)}))| < 2^{-\Omega(n)} \quad (17) \end{aligned}$$

for each $I \in \{0,1\}^9$. Since there are $2^9 (= \textit{constant})$ such terms in (16), we get the required bound on (15).

We now prove (17). Fix any $I \in \{0,1\}^9$. The following two cases can occur.

1. Suppose for some $j \in [9]$, $|S_{jI(j)}| \leq p^{-\delta}|S_j|$. Then $w_I < p^{-\delta}$ and hence the bound in (17) follows.

2. Thus suppose $|S_{iI(i)}| \geq p^{-\delta}|S_j|$ for all $i \in [9]$.

   Define the random variables :

   $$W^I = \text{ext}_1(X_{1I(1)}, \ldots, X_{8I(8)}), \quad V^I = \text{ext}_1(f_1(X_{1I(1)}), \ldots, f_8(X_{8I(8)}))$$

   We prove that the following holds.

   $$\Pr_{v \sim V^I}[(W^I | V^I = v) \text{ is } O(p^{-\delta})\text{-close to a distribution with min-entropy at least } 10\delta \log p] \geq 1 - p^{-\delta} \tag{18}$$

   The following two cases arise depending on I.

   (a) Suppose $I(j) = 0$ for all $j \in [8]$. It follows from Theorem 6.11 that (W^I, V^I) is $p^{-\delta}$-close to a source with min-entropy $(2 + 20\delta)\log p$. Using Corollary 2.11 with $\epsilon = p^{-2\delta}$, we have that

   $$\Pr_{v \sim V_i^I}[(W_i^I | V_i^I = v_i) \text{ is } O(p^{-\delta})\text{-close to a distribution with min-entropy at least } 10\delta \log p] \geq 1 - p^{-\delta}$$

   (b) Suppose there exists some $j \in [8]$ such that $I(j) = 1$. Consider fixing $f_j(X_{jI(j)})$ and all $X_{iI(i)}$, $i \in [8] \setminus \{j\}$. Without loss of generality suppose $j = 1$.
   Under this fixing W^I has min-entropy at least $20\delta \log p$ unless sources $X_{3I(3)}, X_{4I(4)}$ are fixed such that $\text{enc}(x_{3I(3)}) + \text{enc}(x_{4I(4)}) \notin (F_p^*)^2$. But it follows from Claim 6.6 that $\Pr[\text{enc}(X_3) + \text{enc}(X_4) \notin (\mathbb{F}_p^*)^2] < p^{-\delta}$. Thus,

   $$\Pr_{v \sim V^I}[(W^I | V^I = v) \text{ is } O(p^{-\delta})\text{-close to a distribution with min-entropy at least } 20\delta \log p] = 1$$

   This completes the proof of (18).

   We continue with the proof of (17). For each $i \in [C']$, define the set

   $$\textit{Good}^I = \{v \in \text{support}(V^I) : (W^I | V^I = v) \text{ is } O(p^{-\delta})\text{-close to a distribution with min-entropy at least } 10\delta \log p\}$$

   It follows from (18) that $\Pr_{v \sim V^I}[v \in \textit{Good}^I] > 1 - p^{-\delta}$.

   It follows from Theorem 2.20 that snmExt is a strong 2-source extractor for independent sources on $2n$ bits with entropies k_1, k_2 respectively satisfying $k_1 + k_2 \geq (2 + \delta)n$.

Thus we have,

$$\begin{aligned}&|\text{snmExt}(W^I, X_{9I(9)}) \circ V^I \circ X_{9I(9)} - U_m \circ V^I \circ X_{9I(9)}|\\ &\leq (\Pr[V^I \notin \mathit{Good}^I]) + 2^{-\Omega(n)} + p^{-\delta} \leq 2p^{-\delta} + 2^{-\Omega(n)} = 2^{-\Omega(n)}\end{aligned}$$

Since $\text{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9,I(9)}))$ is a deterministic function of the random variables V^I and $X_{9I(9)}$, the bound in (17) is now immediate.

□

6.4 Non-malleable extractor for arbitrary functions

We now prove a slightly stronger version of Theorem 3.

Theorem 6.13 (Theorem 3 restated, stronger version)**.** *There exists a constant $\delta > 0$ such that for every n there exists an explicit function* $\text{nmExt} : (\{0,1\}^n)^8 \times \{0,1\}^{2n} \to \{0,1\}^m$*, $m = \Omega(n)$, such that if $X_1, X_2, \ldots, X_8$ are independent $(n, (1-\delta)n)$-sources, X_9 an independent $(2n, 2(1-\delta)n)$-source and $f_1, f_2, \ldots, f_9$ are arbitrary tampering functions then there exists random variable D_f on $\{0,1\}^m \cup \{\mathit{same}^\star\}$ which is independent of the sources $X_1, \ldots, X_9$ such that*

$$|\text{nmExt}(X_1, \ldots, X_9) \circ \text{nmExt}(f_1(X_1), \ldots, f_9(X_9)) - U_m \circ \text{copy}(D_f, U_m)| \leq 2^{-\Omega(n)}$$

where both U_m's refer to the same uniform m-bit string.

Proof. We view each X_i, $i \in [8]$, as a source on $\mathbb{F}_p$ for a prime p satisfying $2^n < p < 2^{n+1}$. We assume $p^{\tau_0} > 16$ (else we do a constant time brute-force search for nmExt).

Let $\text{snmExt} : \{0,1\}^{2n} \times \{0,1\}^{2n} \to \{0,1\}^m$, $m = \Omega(n)$, be the seeded non-malleable extractor from Theorem 2.20.

Define the functions

$$\text{ext}_1(x_1, x_2, \ldots, x_8) = \sum_{i=0}^{1} \big(\text{enc}(x_{4i+1}) + \text{enc}(x_{4i+2})\big) \odot \big(\text{enc}(x_{4i+3}) + \text{enc}(x_{4i+4})\big)$$

$$\text{nmExt}(x_1, \ldots, x_9) = \text{snmExt}(\text{ext}_1(x_1, \ldots, x_8), x_9)$$

We need the following claims.

Claim 6.14. *Let $Y_1, \ldots, Y_8$ be sources on $\mathbb{F}_p$ with min-entropy $(1-2\delta)\cdot\log p$. Then $\text{ext}_1(Y_1, \ldots, Y_8)$ is $2^{-\Omega(n)}$-close to a source with min-entropy $(1-2\delta)\cdot 2\log p$.*

Proof. We claim that $\text{enc}(Y_1) + \text{enc}(Y_2)$ is a source with min-entropy $2(1-2\delta)\log p - 2$. This follows from the fact that $(y_1 + y_2, y_1^4 + y_1^2 + y_1 + y_2 + y_2^2 + y_2) = (a, b)$ has at most 4 solutions in (y_1, y_2). Also it follows from Claim 6.6 that $\Pr[\text{enc}(Y_3) + \text{enc}(Y_4) \notin (\mathbb{F}_p^*)^2] < p^{-\delta}$. Thus $\text{ext}_1(Y_1, \ldots, Y_8)$ is $p^{-\delta}$-close to a source with min-entropy $2(1-2\delta)\log p - 2$. □

Claim 6.15. *Let $Y_1, \ldots, Y_8$ be independent $(n, (1-2\delta)n)$-sources and Y_9 an independent $(2n, 2(1-2\delta)n)$-source . Then*

$$|\text{nmExt}(Y_1, \ldots, Y_9) - U_m| < 2^{-\Omega(n)}$$

Proof. Follows directly from Claim 6.14 and Theorem 2.20. □

For each $i \in [8]$, let $S_i \subset \mathbb{F}_p$ be the support of the (flat) source X_i. Let $S_9 \subset \{0,1\}^{2n}$ be the support of X_9. We partition each S_i into S_{i0} and S_{i1} such that f_i has no fixed points in S_{i1}. Thus

$$S_{i0} = \{s \in S_i : f_i(s) = s\}, S_{i1} = S_i \setminus S_{i0}$$

Let X_{ij} be the flat source that is supported on S_{ij}, $i = 1, .., 9$, $j = 0, 1$. Let f_i^I denote f_i with its domain restricted to the set $S_{iI(i)}$. Thus f_i^I is a function from $S_{iI(i)}$ to $\mathbb{F}_p$.

For any 0-1 vector I, let $I(i)$ denote the i'th co-ordinate in I. Let $w_I = \prod_{i=1}^{9} \frac{|S_{iI(i)}|}{|S_i|}$ for $I \in \{0, 1\}^9$.

Define the random variable D_f as:

$$D_f = w_{\vec{0}} \cdot \{same^\star\} + \sum_{I \in \{0,1\}^9 \setminus \{\vec{0}\}} w_I \cdot \text{nmExt}(f_1(X'_{1I(1)}), \ldots, f_9(X'_{9I(9)}))$$

where for each $i \in [9]$ and $I \in \{0,1\}^9$, $X'_{iI(i)}$ is identically distributed as $X_{iI(i)}$ and independent of $X_1, \ldots, X_9$.

We show that nmExt and D_f satisfies the conclusion of Theorem 6.13.

Recall that to prove Theorem 6.13, we need to show the following bound.

$$|\text{nmExt}(X_1, \ldots, X_9) \circ \text{nmExt}(f_1(X_1), \ldots, f_9(X_9)) - U_m \circ \text{copy}(D_f, U_m)| < 2^{-\Omega(n)} \tag{19}$$

We have

$$(19) \leq (*) + \sum_{I \in \{0,1\}^9 \setminus \{\vec{0}\}} w_I \cdot |\text{nmExt}(X_{1I(1)}, \ldots, X_{9I(9)}) \circ \text{nmExt}(f_1^I(X_{1I(1)}), \ldots, f_9^I(X_{9I(9)})) - U_m \circ \text{nmExt}(f_1^I(X_{1I(1)}), \ldots, f_9^I(X_{9I(9)}))| \tag{20}$$

where

$$(*) = w_{\vec{0}} \cdot |\text{nmExt}(X_{10}, \ldots, X_{90}) \circ \text{nmExt}(f_1^{\vec{0}}(X_{10}), \ldots, f_9^{\vec{0}}(X_{90})) - \text{copy}(same^\star, U_m) \circ U_m|$$

For $i \in [9]$, the support of X_{i0} is S_{i0} and by definition $f_i^{\vec{0}}(x_{i0}) = x_{i0}$ for all $x_{i0} \in S_{i0}$. Using the definition of the function copy, we have

$$\begin{aligned}(*) &= w_{\vec{0}} \cdot |\text{nmExt}(X_{10}, \ldots, X_{90}) \circ \text{nmExt}(X_{10}, \ldots, X_{90}) - U_m \circ U_m| \\ &= w_{\vec{0}} \cdot |\text{nmExt}(X_{10}, \ldots, X_{90}) - U_m|\end{aligned} \tag{21}$$

We prove the following claims.

Claim 6.16.

$$(*) = w_{\vec{0}} \cdot |\text{nmExt}(X_{10}, \ldots, X_{90}) - U_m| < 2^{-\Omega(n)} \tag{22}$$

Claim 6.17. *For every $I \in \{0,1\}^9 \setminus \{\vec{0}\}$ the following holds:*

$$w_I \cdot |\mathrm{nmExt}(X_{1I(1)}, \ldots, X_{9I(9)}) \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)})) - U_m \circ \mathrm{nmExt}(f_1^I(X_{1I(1)}), \ldots, f_9^I(X_{9I(9)}))| < 2^{-\Omega(n)} \quad (23)$$

We use the above claims to conclude (19).

Proof of (19) using Claim 6.17 *and Claim* 6.16. Note that there are 2^9 terms in RHS of (20). Each term corresponding to a non-zero I is bounded by $2^{-\Omega(n)}$ using Claim 6.17 and the term corresponding to $I = \vec{0}$ is bounded by $2^{-\Omega(n)}$ using Claim 6.16. We can thus bound LHS of (19) by $2^{-\Omega(n)}$. □

Proof of Claim 6.16. We note that if for any $i \in [9]$, $\frac{|S_{i0}|}{|S_i|} < p^{-\delta}$, we have $w_{\vec{0}} < p^{-\delta}$. Thus suppose $|S_{i0}| > p^{-\delta}|S_i|$ for all $i \in [9]$. The bound now follows from Claim 6.15. □

Proof of Claim 6.17. Fix some $I \in \{0,1\}^9 \setminus \{\vec{0}\}$.

We split the proof into the following cases.

1. If for some $i \in [9]$, $|S_{iI(i)}| < p^{-\delta}|S_i|$, then $w_I < p^{-\delta}$ and hence the bound in (23) follows.
2. Thus suppose $|S_{iI(i)}| \geq p^{-\delta}|S_i|$ for all $i \in [9]$. We consider the following cases.
   (a) Suppose there exists some $j \in [8]$ such that $I(j) = 1$. In this case we use Theorem 6.12 to conclude the bound in (23).
   (b) Suppose for all $i \in [8]$, $I(i) = 0$. We note that $I(9) = 1$ since $I \neq \vec{0}$. Thus all f_i^I, $i \in [8]$, are the identity functions over their respective domains and f_9^I has no fixed points.
   Using Claim 6.14, we have $\mathrm{ext}_1(X_{1I(1)}, \ldots, X_{8I(8)})$ is $2^{-\Omega(n)}$-close to a source Z with min-entropy $(1-2\delta) \cdot 2n$.
   Define the random variable: $W^I = \mathrm{ext}_1(X_{1I(1)}, \ldots, X_{8I(8)})$.
   Thus we have

$$\begin{aligned} &|\mathrm{nmExt}(X_{1I(1)}, \ldots, X_{9I(9)}) \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)})) \\ &\quad - U_m \circ \mathrm{nmExt}(f_1(X_{1I(1)}), \ldots, f_9(X_{9I(9)}))| \\ &= |\mathrm{snmExt}(W^I, X_{9I(9)}) \circ \mathrm{snmExt}(W^I, f_9^I(X_{9I(9)})) - U_m \circ \mathrm{snmExt}(W^I, f_9^I(X_{9I(9)}))| \\ &\leq |\mathrm{snmExt}(Z, X_{9I(9)}) \circ \mathrm{snmExt}(Z, f_9^I(X_{9I(9)})) - U_m \circ \mathrm{snmExt}(Z, f_9^I(X_{9I(9)}))| + 2^{-\Omega(n)} \end{aligned}$$

   Note that Z and $X_{9I(9)}$ are independent sources on $\{0,1\}^{2n}$, each with min-entropy rate $> (1-2\delta)$ and f_9^I has no fixed points. Thus by Theorem 2.20, we have

$$|\mathrm{snmExt}(Z, X_{9I(9)}) \circ \mathrm{snmExt}(Z, f_9^I(X_{9I(9)})) - U_m \circ \mathrm{snmExt}(Z, f_9^I(X_{9I(9)}))| \leq 2^{-\Omega(n)}$$

   Thus, the bound in (23) follows.

This completes the proof of Claim 6.17. □

□

7 Proof of the sum-product estimate over $\mathbb{F}_p^4$

We closely follow the proof of the sum-product estimate by Bourgain in [Bou05a] and prove Theorem 6.2, which we restate.

Theorem 6.2. *Let $\tau_0 > \tau_1 > 0$ be any positive constants. Let A be a subset of $\mathbb{F}_p^4$ satisfying $|A \cap (\mathbb{F}_p^*)^4| \geq \frac{|A|}{2}$. Suppose that for any subset $A_1 \subseteq A$ satisfying $|A_1| \geq p^{-\tau_1}|A|$, the following conditions holds.*

1. *$\Pi_{\{1,2\}}(A_1) \geq p^{1+\tau_0}$ and $\Pi_{\{3,4\}}(A_1) \geq p^{1+\tau_0}$.*
2. *$A_1 \nsubseteq P$, where P is a 2-dimensional linear subspace of $\mathbb{F}_p^4$ of form*
   (a) *$\{(x_1, x_2, c_1x_1, c_2x_2) : x_1 \in \mathbb{F}_p, x_2 \in \mathbb{F}_p\}$ or*
   (b) *$\{(x_1, x_2, c_2x_2, c_1x_1) : x_1 \in \mathbb{F}_p, x_2 \in \mathbb{F}_p\}$.*

Then there exists some constant $\tau > 0$ (depending on τ_0, τ_1) such that if $|A| < p^{7/3-\tau_1}$, then

$$|A + A| + |A \odot A| > p^{\tau}|A|$$

We introduce some notations.

Definition 7.1. *Let $S \subseteq \mathbb{F}_p^n$ be any set of vectors. Define $S^{\odot 2} = S \odot S$ and $S^{\odot(k+1)} = S^{\odot k} \odot S$ for $k \geq 2$.*

We prove Theorem 6.2 using the following lemmas.

Lemma 7.2. *Let B be any subset of $\mathbb{F}_p^4$ such that $|\Pi_{\{1,2\}}(B)| \geq p^{1+\tau_0}$ and $|\Pi_{\{3,4\}}(B)| \geq p^{1+\tau_0}$. Then one of the following holds.*

1. *There exists constant $k = k(\tau_0)$ such that $|kB^{\odot k}| \geq p^{7/3}$ or*
2. *$B \subseteq P$ where P is a 2-dimensional linear subspace of $\mathbb{F}_p^4$ of the form*
   (a) *$\{(x_1, x_2, c_1x_1, c_2x_2) : x_1 \in \mathbb{F}_p, x_2 \in \mathbb{F}_p\}$ or*
   (b) *$\{(x_1, x_2, c_2x_2, c_1x_1) : x_1 \in \mathbb{F}_p, x_2 \in \mathbb{F}_p\}$.*

Lemma 7.3. *Let $B \subset (\mathbb{F}_p^*)^4$ such that $|B| \geq p^{1+\tau_0}$ and $|B + B| + |B \odot B| \leq p^{\tau}|B|$. Fix any $k > 0$. Then, there is a subset B_1 of B such that*

1. *$|B_1| \geq p^{-\tau_1}|B|$ and*
2. *$|kB_1^{\odot k}| \leq p^{\tau_1}|B_1|$*

where $\tau_1 = p^{3k^2}\tau$.

Proof of Theorem 6.2. We replace A with its intersection with $(\mathbb{F}_p^*)^4$. Choose τ small enough such for $k = k(\tau_0)$ (where $k(\tau_0)$ is the constant from Lemma 7.2), it holds that: $p^{3k^2}\tau < \tau_1$. Suppose that $|A + A| + |A \odot A| \leq p^{\tau}|A|$. Using Lemma 7.3, there exists a subset A_1, $|A_1| \geq p^{-\tau_1}|A|$, such

that $|kA_1^{\odot k}| \leq |A|p^{\tau_1}$. Further, we have that A_1 satisfies the hypothesis of Lemma 7.2. Suppose, conclusion (1) of Lemma 7.2 holds. This implies that $|A| \geq p^{7/3}$ which contradicts our assumption on the size of A. Further, from the assumptions on the structure of A_1, we see that conclusion (2) in Lemma 7.2 cannot hold. Thus, it must be that $|A+A| + |A \odot A| > p^{\tau}|A|$. □

Lemma 7.3 follows directly from Lemma 4 in [Bou05a] by noticing that their proof works over $(\mathbb{F}_p^*)^4$ as well. Hence we do not present the proof of Lemma 7.3.

Thus we focus on proving Lemma 7.2.

We require the following lemma which was proved by Bourgain [Bou05a].

Lemma 7.4. *For any $B \subseteq \mathbb{F}_p^2$ such that $|B| \geq p^{1+\tau_0}$ there exists a constant $k = k(\tau_0)$ such that $|kB^{\odot k}| = p^2$.*

We now proceed to prove Lemma 7.2.

Proof of Lemma 7.2. Let B_{ij} denote $\Pi_{\{i,j\}}(B)$. Using Lemma 7.4, there exists some k_0 such that $|kB_{12}^{\odot k}| = p^2$, $|kB_{34}^{\odot k}| = p^2$ for $k \geq k_0$. We split the proof into two cases.

1. Suppose there exists some $k \geq k_0$ such that $|kB^{\odot k}| > p^2$.

   Thus, it must be the case that the projection map $\Pi_{\{1,2\}}$ is not one-one on $kB^{\odot k}$. Thus there exists $b, b' \in kB^{\odot k}$ such that $\Pi_{\{1,2\}}(b) = \Pi_{\{1,2\}}(b')$ but $\Pi_{\{3,4\}}(b) \neq \Pi_{\{3,4\}}(b')$. Consider the set

   $$kB^{\odot k} - (b-b')kB^{\odot k} = \{\big(x_1, x_2, x_3 - (b_3 - b_3')y_3, x_4 - (b_4 - b_4')y_4\big) : \\ (x_1, x_2, x_3, x_4) \in kB^{\odot k}, (y_1, y_2, y_3, y_4) \in kB^{\odot k}\}$$

   Notice that (x_1, x_2) takes all values of $\mathbb{F}_p^2$ since $|\Pi_{\{1,2\}}(kB^{\odot k})| = p^2$. Similarly (y_3, y_4) takes all values of $\mathbb{F}_p \times \mathbb{F}_p$ since $|\Pi_{\{3,4\}}(kB^{\odot k})| = p^2$. Further, at least one of $(b_3 - b_3')$ or $(b_4 - b_4')$ is non zero. Without loss of generality, suppose $b_3 - b_3' \neq 0$. Then, for any fixing of $x \in kB^{\odot k})$, $\Pi_3(x - (b-b')kB^{\odot k}) = \mathbb{F}_p$ and hence $|kB^{\odot k} - (b-b')kB^{\odot k}| \geq p^3$.

   We observe that

   $$kB^{\odot k} - (b-b')kB^{\odot k} \subseteq kB^{\odot k} - (kB^{\odot k} - kB^{\odot k})kB^{\odot k} \subseteq k'B^{\odot k'} - k'B^{\odot k'}$$

   , where $k' = 3k^2$. Using Lemma 2.16 with $A = k'B^{\odot k'}$ and recalling that $|k'B^{\odot k'}| > p^2$, we have

   $$\begin{aligned} |k'B^{\odot k'} + k'B^{\odot k'}| &\geq \Big(|k'B^{\odot k'} - k'B^{\odot k'}||k'B^{\odot k'}|^2\Big)^{1/3} \\ &> (p^3p^4)^{1/3} = p^{7/3} \end{aligned}$$

   Setting a new $k = 2k'$, we have $|kB^{\odot k}| \geq p^{7/3}$.

2. Suppose $|kB^{\odot k}| = p^2$ for all $k \geq k_0$. Thus in particular we have

   $$|k_0B^{\odot k_0} + k_0B^{\odot k_0}| = |k_0B^{\odot k_0}|$$

and

$$|k_0B^{\odot k_0} \odot k_0B^{\odot k_0}| = |k_0B^{\odot k_0}|$$

Thus $k_0B^{\odot k_0}$ must be a 2-dimensional affine subspace of $\mathbb{F}_p^4$.

Let $k_0B^{\odot k_0} = \{z + \lambda v + \mu w : \lambda, \mu \in \mathbb{F}_p\}$, $z, v, w \in \mathbb{F}_p^4$. To complete the argument, we prove the following claims about the structure of z, v, w.

Claim 7.5. *We can assume* $v = (1, 0, \alpha_1, \alpha_2)$ *and* $w = (0, 1, \beta_1, \beta_2)$ *such that* $span\{(\alpha_1, \alpha_2), (\beta_1, \beta_2)\} = \mathbb{F}_p^2$

Proof. The proof follows from the observation that $\Pi_{\{1,2\}}(k_0B^{\odot k_0}) = \Pi_{\{3,4\}}(k_0B^{\odot k_0}) = \mathbb{F}_p^2$. □

Claim 7.6. *Let* $v = (1, 0, \alpha_1, \alpha_2)$ *and* $w = (0, 1, \beta_1, \beta_2)$. *Then* $\alpha_i\beta_i = 0$ *for* $i \in [2]$. *Further* $z = 0$.

We show how to complete the proof of Lemma 7.2, before proving the above claim.

Proof of Lemma 7.2 using Claim 7.5 and Claim 7.6. Since we have $\alpha_1\beta_1 = 0$, suppose $\alpha_1 = 0$. It follows from Claim 7.5 and Claim 7.6 that $\beta_1 \neq 0$, $\alpha_2 \neq 0$ and $\beta_2 = 0$.

Thus $k_0B^{\odot k_0} = \{z + \lambda v + \mu w : \lambda, \mu \in \mathbb{F}_p\} = \{(\lambda, \mu, \beta_1\mu, \alpha_2\lambda) : \lambda, \mu \in \mathbb{F}_p\}$.

Fix any $y = (y_1, y_2, y_3, y_4) \in k_0B^{\odot(k_0-1)} \cap (\mathbb{F}_p^*)^4$. Note that there exists such a y since $B \cap (F_p^*)^4 \neq \emptyset$ and $k_0B^{\odot k_0} \cap (F_p^*)^4 \neq \emptyset$.

For any $x = (x_1, x_2, x_3, x_4) \in B$, since $x \odot y \in k_0B^{\odot k_0} = \{(\lambda, \mu, \beta_1\mu, \alpha_2\lambda) : \lambda, \mu \in \mathbb{F}_p\}$, there exists λ, μ such that the following relations hold :

$$x_4 = y_4^{-1}\alpha_2 x_1 y_1, \quad x_3 = y_3^{-1}\beta_1 x_2 y_2$$

Thus

$$B \subseteq \{(x_1, x_2, c_2x_2, c_1x_1) : x_1, x_2 \in \mathbb{F}_p\}$$

where $c_1 = y_4^{-1}\alpha_2 y_1, c_2 = y_3^{-1}\beta_1 y_2$.

For the case when $\alpha_1 \neq 0$ (and hence $\beta_1 = 0$), we use an identical argument to derive that $B \subseteq \{(x_1, x_2, c_1x_1, c_2x_2 : x_1, x_2) \in \mathbb{F}_p\}$. □

We conclude by proving Claim 7.6.

Proof of Claim 7.6. Let $S = (k_0B^{\odot k_0}) \odot (k_0B^{\odot k_0})$. Recall that $k_0B^{\odot k_0} = \{z + \lambda v + \mu w : \lambda, \mu \in \mathbb{F}_p\}$ where $v = (1, 0, \alpha_1, \alpha_2), w = (0, 1, \beta_1, \beta_2)$ and $|S| = |k_0B^{\odot k_0}| = p^2$. Thus for each $i \in [4]$,

$$\Pi_i(S) = \{\pi_i(\lambda_1, \lambda_2, \mu_1, \mu_2) : \lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{F}_p\}$$

where

$$\begin{aligned}
\pi_1(\lambda_1,\lambda_2,\mu_1,\mu_2) &= \pi_1(\lambda_1,\lambda_2) = (z_1+\lambda_1)(z_1+\lambda_2)\\
&= \lambda_1\lambda_2 + (\lambda_1+\lambda_2)z_1 + z_1^2\\
\pi_2(\lambda_1,\lambda_2,\mu_1,\mu 2) &= \pi_2(\mu_1,\mu_2) = (z_2+\mu_1)(z_2+\mu_2)\\
&= \mu_1\mu_2 + (\mu_1+\mu_2)z_2 + z_2^2\\
\pi_3(\lambda_1,\lambda_2,\mu_1,\mu_2) &= (\lambda_1\alpha_1+\mu_1\beta_1+z_3)(\lambda_2\alpha_1+\mu_2\beta_1+z_3)\\
&= \lambda_1\lambda_2\alpha_1^2 + \mu_1\mu_2\beta_1^2 + \alpha_1\beta_1(\lambda_1\mu_2+\lambda_2\mu_1)+\\
&\qquad (\lambda_1+\lambda_2)\alpha_1 z_3 + (\mu_1+\mu_2)\beta_1 z_3 + z_3^2\\
\pi_4(\lambda_1,\lambda_2,\mu_1,\mu_2) &= (\lambda_1\alpha_2+\mu_1\beta_2+z_4)(\lambda_2\alpha_2+\mu_2\beta 2+z_4)\\
&= \lambda_1\lambda_2\alpha_2^2 + \mu_1\mu_2\beta_2^2 + \alpha_2\beta_2(\lambda_1\mu_2+\lambda_2\mu_1)+\\
&\qquad (\lambda_1+\lambda_2)\alpha_2 z_4 + (\mu_1+\mu_2)\beta_2 z_4 + z_4^2
\end{aligned}$$

- We prove $\alpha_i\beta_i = 0$, for $i = 1, 2$. Suppose not. Let $\alpha_1\beta_1 \neq 0$.
  Fix $\lambda_2 = a_2 \neq -z_1$ and let $\lambda_1 = a_1 \neq \lambda_2$ and let $\pi_1(a_1, a_2) = a$. Note that $\pi_1(a_1, a_2) = \pi(b_1, a_2)$ iff $a_1 = b_1$. Thus $|\{\pi_1(x, a_2) : x \in \mathbb{F}_p \setminus \{a_2\}\}| = p - 1$.
  We claim that for any such fixing of $\lambda_1 = a_1, \lambda_2 = a_2$, there exists μ_1, μ_2 such that $\pi_2(\mu_1, \mu_2) = b$ and $\pi_3(a_1, a_2, \mu_1, \mu_2) = c$ for at least $O(p^2)$ pairs $(b, c) \in \mathbb{F}_p^2$. Suppose

$$\begin{aligned}
\pi_2(\mu_1,\mu_2) &= \mu_1\mu_2 + (\mu_1+\mu_2)z_2 + z_2^2 = b\\
\pi_3(a_1,a_2,\mu_1,\mu_2) &= \beta_1^2\mu_1\mu_2 + \gamma_1\mu_1 + \gamma_2\mu_2 + \gamma_3 = c
\end{aligned}$$

  where $\gamma_1, \gamma_2, \gamma_3 \in \mathbb{F}_p$ are constants (does not depend on μ_1, μ_2) . By our choice of λ_1, λ_2, we have that $\gamma_1 \neq \gamma_2$ and hence the above system of equations has at most two pairs of values of (μ_1, μ_2) which satisfy it. Since (μ_1, μ_2) takes p^2 values, there at least $p^2/2$ distinct pairs (b, c) such that there $(\pi_2(\mu_1, \mu_2), \pi_3(\lambda_1, \lambda_2, \mu_1, \mu_2)) = (b, c)$.
  Thus we have shown that there exists $\lambda_1, \lambda_2, \mu_1, \mu_2$ such that $(\pi_1(\lambda_1, \lambda_2), \pi_2(\mu_1, \mu_2), \pi_3(\lambda_1, \lambda_2, \mu_1, \mu_2)) = (a, b, c)$ for at least $\frac{1}{2}(p-1)p^2$ distinct tuples $(a, b, c) \in \mathbb{F}_p^3$, which is a contradiction since $|S| = p^2$. Thus $\alpha_1\beta_1 = 0$. A similar argument implies that $\alpha_2\beta_2 = 0$.
- We now prove $z = 0$. Suppose $\alpha_1 = 0$. Thus $\beta_1 \neq 0$, $\alpha_2 \neq 0$ and $\beta_2 = 0$. We again fix $\lambda_2 = a_2 \neq -z_1$ and let $\lambda_1 = a_1 \neq \lambda_2$. Let $(b, c) \in \mathbb{F}_p^2$. We bound the number of (μ_1, μ_2) such that $(\pi_2(\mu 1, \mu_2), \pi_3(a_1, a_2, \mu_1, \mu_2) = (b, c)$. We have the following equations.

$$\begin{aligned}
\mu_1\mu_2 + (\mu_1+\mu_2)z_2 + z_2^2 &= b\\
\beta_1^2\mu_1\mu_2 + \beta_1 z_3(\mu_1+\mu_2) + \gamma_0 &= c
\end{aligned}$$

  We see that the number of solutions of the above pair of equations is bounded by 2 unless $z_3 = \beta_1 z_1$. It follows that if $z_3 \neq \beta_1 z_2$, there exists $(\lambda_1, \lambda_2, \mu_1, \mu_2)$ such that $(\pi_1(\lambda_1, \lambda_2), \pi_2(\mu_1, \mu_2), \pi_3(\lambda_1, \lambda_2, \mu_1, \mu_2)) = (a, b, c)$ for at least $\frac{1}{2}(p-1)p^2$ distinct tuples $(a, b, c) \in \mathbb{F}_p^3$, which is a contradiction. Thus suppose $z_3 = \beta_1 z_2$.
  Using an identical argument (but now fixing μ_1, μ_2 appropriately in π_2 and arguing about the range of π_1 and π_4 upon varying λ_1, λ_2), we get that $z_4 = \alpha_2 z_1$. Thus $z = (z_1, z_2, \beta_1 z_2, \alpha_2 z_1) = z_1 \cdot (1, 0, 0, \alpha_2) + z_2 \cdot (0, 1, \beta_1, 0) = z_1 \cdot v + z_2 \cdot w \in span\{v, w\}$. Hence we can take $z = 0$. □

□

8 Efficient algorithms for non-malleable codes in the 10-split-state model

In this section we prove efficiency of the non-malleable codes constructed in Theorem 1. Let nmExt be the function from Theorem 3. Recall that for any message s, its encoding is a uniform element from $\text{nmExt}^{-1}(s)$ and for any codeword c, the decoded message is $\text{nmExt}(c)$. Thus the efficiency of the decoder follows from the fact that nmExt is a polynomial time function.

We construct an efficient algorithm which takes as input a message $s \in \{0,1\}^n$ and samples from a distribution that is $2^{-\Omega(n)}$-close to uniform on $\text{nmExt}^{-1}(s)$ and use this as our encoder. This is indeed sufficient, since we only add an exponentially small error when we use this algorithm instead of sampling uniformly from $\text{nmExt}^{-1}(s)$.

Our sampling algorithm is based on the following observations.

- The uniform distribution on the set $\text{nmExt}^{-1}(s)$ is a convex combination of uniform distributions on algebraic varieties of low degree.
- Sampling almost uniformly from such algebraic sets can be done efficiently [CS09].
- Further, obtaining the weights in the convex combination reduces to approximately counting the size of such algebraic sets for which there are efficient algorithms [HW98]. However, the number of distributions in the convex combination can be exponentially large. To get around this difficulty, we use the method of rejection sampling. The proof of correctness of the algorithm relies on estimates on the number of rational points on algebraic varieties.

8.1 Tools from algebraic geometry

Let $g \in \mathbb{F}_p[x_1, \ldots, x_c]$ and let $\mathcal{H} \subseteq \mathbb{F}_p^c$ be its set of zeroes. We call $\mathcal{H}$ the algebraic hypersurface defined by g.

The following version of the Lang-Weil bound for hypersurfaces in $\mathbb{F}_p^c$ was proved in [CM06].

Theorem 8.1 (Lang-Weil bound)**.** *Let c, d be constant integers and let p be a large prime. Let $\mathcal{H} \subset \mathbb{F}_p^c$ be a hypersurface defined by a degree d polynomial. Then there exists an integer s, $0 \leq s \leq d$, such that*

$$||\mathcal{H}| - sp^{c-1}| \leq O(\text{sign}(s) \cdot p^{c-\frac{3}{2}} + p^{c-2})$$

where $\text{sign}(s) = 1$ *if* $s > 0$ *and* $\text{sign}(0) = 0$.

Lemma 8.2 (Schwartz-Zippel Lemma [Sch80], [Zip79])**.** *Let $g(x_1, \ldots, x_c)$ be a non-zero multivariate polynomial of degree d with coefficients in $\mathbb{F}_p$. Then the hypersurface $\mathcal{H} \subset \mathbb{F}_p^c$ defined by g is of size at most dp^{c-1}.*

We need some previous work on efficient sampling and approximate counting of algebraic varieties.

Theorem 8.3 ([CS09])**.** *Let c, k, d be constant integers such that $c > k$ and let p be a prime. There exists an efficient randomized algorithm $\mathcal{A}_1$ such that the following holds:*

Let $g_1, \ldots, g_k \in \mathbb{F}_p[x_1, \ldots, x_c]$ be arbitrary polynomials of degree at most d and let $S \subseteq \mathbb{F}_p^c$ be the set of common zeroes of $g_1, \ldots, g_k$. $\mathcal{A}_1$ takes as input the description of $g_1, \ldots, g_k$ and a parameter δ and outputs a sample from a distribution which is $O(1/p^{1-\delta})$-close to the uniform distribution on S. The worst-case running time of $\mathcal{A}_1$ is bounded by $\mathrm{poly}(\log p)$.

Theorem 8.4 ([HW98])**.** *Let $c, k, d > 0$ be constant integers and let p be a prime. There exists an efficient randomized algorithm $\mathcal{A}_2$ such that the following holds:*

Let $g_1, \ldots, g_k \in \mathbb{F}_p[x_1, \ldots, x_c]$ be arbitrary polynomials of degree at most d and let $S \subseteq \mathbb{F}_p^c$ be the set of common zeroes of $g_1, \ldots, g_k$. $\mathcal{A}_2$ takes as input the description of $g_1, \ldots, g_k$ and outputs an integer v such that

$$\frac{1}{|S|} \cdot |v - |S|| < O(p^{-1/2})$$

The worst-case running time of $\mathcal{A}_2$ is bounded by $\mathrm{poly}(\log p)$.

8.2 A new extractor

In the construction of the seedless non-malleable extractor nmExt in Theorem 3, we needed a seeded non-malleable extractor snmExt (with some additional properties, see Theorem 2.20). We carefully choose snmExt such that it is easy to sample almost uniformly from $\mathrm{nmExt}^{-1}(s)$. The main idea is to pick snmExt such that $\mathrm{nmExt}^{-1}(s)$ is a convex combination of algebraic varieties of low degree over a field with large characteristic. Thus, the constructions in [Li12b] look to be a good choice for the seeded non-malleable extractor. However, for this choice, we face the following difficulty:

Let $\sigma_M : \mathbb{F}_p \to \mathbb{Z}_M$ be defined as $\sigma_M(x) = x \pmod M$. nmExt is of the form $\sigma_M \circ \mathrm{ext}_2 \circ \mathrm{ext}_1$, where $\mathrm{ext}_1 : \mathbb{F}_p^{10} \to \mathbb{F}_p^4$, $\mathrm{ext}_2 : \mathbb{F}_q^2 \to \mathbb{F}_q$, and p, q are primes satisfying $p^2 \leq q \leq 2p^2$ (and interpreting the output of ext_1 as an element in $\mathbb{F}_q^2$). Changing the characteristic of the field destroys the low degree properties of the function $\mathrm{ext}_2 \circ \mathrm{ext}_1$.

To fix this, we construct a new extractor for ext_2 (satisfying the conditions of Theorem 2.20) which allows us to work over the same field as ext_1. The extractor is a variation of a construction by Bourgain [Bou05b]. The proof uses ideas from [Li12b], but requires more work.

Theorem 8.5. *Let p be a prime. Define the functions* $\mathrm{ext}_2 : (\mathbb{F}_p^2) \times (\mathbb{F}_p^2) \to \mathbb{F}_p$ *and* $\mathrm{snmExt} : (\mathbb{F}_p^2) \times (\mathbb{F}_p^2) \to \mathbb{Z}_M$ *in the following way:*

$$\mathrm{ext}_2((x_1, x_2), (y_1, y_2)) = \sum_{j=1}^{2} (x_j y_j + x_j^2 y_j^2), \qquad \mathrm{snmExt}(x, y) = \sigma_M(\mathrm{ext}_2(x, y))$$

where $\sigma_M(x) = x \ (mod\ M)$. Suppose X, Y are independent sources on $\mathbb{F}_p^2$ with min-entropies k_1, k_2 respectively.

1. *If $(k_1 + k_2) \geq (2 + \delta) \log p$, then*

$$|\mathrm{snmExt}(X, Y) \circ X - U_M \circ X| < p^{-\Omega(1)}, \ |\mathrm{snmExt}(X, Y) \circ Y - U_M \circ Y| < p^{-\Omega(1)}$$

2. *If $k_1, k_2 > (2-\delta)\log p$ and f is any tampering function with no fixed points, then*

$$|\text{snmExt}(X,Y) \circ \text{snmExt}(X, f(Y)) - U_M \circ \text{snmExt}(X, f(Y))| < p^{-\Omega(1)}.$$

The proof of Theorem 8.5 is presented in Appendix A.

8.3 A generic sampling algorithm

We construct an algorithm for almost uniformly sampling from certain structured sets.

Theorem 8.6. *Let S_1, S_2, and S_3 be finite sets. For arbitrary functions $g : S_2 \to S_3$, $h : S_1 \to S_2$, there exists a sampling algorithm $\mathcal{B}$ which takes as input $z \in S_3$ and a parameter $\epsilon \geq \epsilon_0$, runs in time* $\text{poly}(\log(|S_1| \cdot |S_2|), \log(\frac{1}{\epsilon}))$*, and outputs a sample from a distribution that is $O(\epsilon)$-close to uniform on the set $(g \circ h)^{-1}(z)$, if the following conditions hold:*

1. *There exists an algorithm $\mathcal{B}_1$, which takes as input $z \in S_3$, runs in time* $\text{poly}(\log(|S_2|))$*, and outputs a sample from a distribution that is uniform on the set $g^{-1}(z)$.*
2. *There exists an algorithm $\mathcal{B}_2$, which takes as input $y \in S_2$ and ϵ, runs in time* $\text{poly}(\log(|S_1|), \log(\frac{1}{\epsilon}))$*, and outputs a sample from a distribution that is ϵ-close to uniform on the set $h^{-1}(y)$.*
3. *There exists an algorithm $\mathcal{B}_3$, which takes as input $y \in S_2$ and ϵ, runs in time* $\text{poly}(\log(|S_1|), \log(\frac{1}{\epsilon}))$*, and outputs an approximation A_y for $|h^{-1}(y)|$ with a multiplicative error of at most ϵ, i.e., $1 - \epsilon \leq \frac{A_y}{|h^{-1}(y)|} \leq 1 + \epsilon$.*
4. *There exist constants $\beta > 0$ and $\lambda \geq 1$, and an efficiently computable value* max *such that for all $\epsilon \geq \epsilon_0$ the following holds: There exists a subset $S_2' \subseteq S_2$ such that for all $y \in S_2'$, $\frac{\max}{\lambda} \leq |h^{-1}(y)| \leq \max$. Further, $\frac{1}{|(g\circ h)^{-1}(z)|} \sum_{y \in S_2 \setminus S_2'} |h^{-1}(y)| \leq \epsilon$ and $\frac{|S_2'|}{|S_2|} > \beta$.*

Proof. The idea is to use the method of rejection sampling.

Algorithm $\mathcal{B}$ (given input $z \in S_3$ and error parameter ϵ):

1. Use $\mathcal{B}_1$ to sample y from $g^{-1}(z)$. Compute an approximation A_y for $|h^{-1}(y)|$ with error ϵ using algorithm $\mathcal{B}_3$. If $A_y < \max \cdot (\frac{1}{\lambda} - \epsilon)$, reject y. Else accept y with probability $wt(y) = \frac{A_y}{\max}$. Iterate this step till some y is accepted. If no sample is accepted after $O(\log \frac{1}{\epsilon})$ iterations, accept the next sample.
2. Once y is accepted, sample from $h^{-1}(y)$ using $\mathcal{B}_2$ (with error ϵ).

Proof of correctness of Algorithm $\mathcal{B}$: Consider any subset $T \subseteq (g \circ h)^{-1}(z)$. Let $p_{T,1}$ be the probability that some element from T is picked by $\mathcal{B}$ in one iteration.

Then:

$$p_{T,1} = \sum_{y \in g^{-1}(z)} \frac{1}{|g^{-1}(z)|} \cdot \left(\frac{|h^{-1}(y)|}{\max} \pm \epsilon \right) \cdot \left(\frac{|T \cap h^{-1}(y)|}{|h^{-1}(y)|} \pm \epsilon \right)$$

The above expression is derived in the following way: Consider any $y \in g^{-1}(z)$. Let A_y be the approximation of $|h^{-1}(y)|$ computed by algorithm $\mathcal{B}_3$. The probability of y being picked by $\mathcal{B}_1$

is $\frac{1}{|g^{-1}(z)|}$. The probability that this y is accepted is given by $\frac{A_y}{\max} = \frac{|h^{-1}(y)|}{\max} \pm \epsilon$. Further, if y is accepted, $\frac{|T \cap h^{-1}(y)|}{|h^{-1}(y)|} \pm \epsilon$ is the probability that some element from the set T is picked by algorithm $\mathcal{B}_2$ (since $\mathcal{B}_2$ samples from a distribution ϵ-close to uniform on $h^{-1}(y)$).

It follows that,

$$|p_{T,1} - \frac{|T|}{\max \cdot |g^{-1}(z)|}| = O(\epsilon)$$

Let $N = |(g \circ h)^{-1}(z)|$. The probability that an iteration of Step (1) fails to accept a sample is:

$$p_{reject} = \left(1 - \frac{N}{\max \cdot |g^{-1}(z)|}\right) \pm O(\epsilon)$$

Let $k = O(\log \frac{1}{\epsilon})$. The probability p_T that some element from T is picked by $\mathcal{B}$ in at most k iterations is given by:

$$\begin{aligned} p_T &= p_{T,1} \sum_{i=0}^{k-1} (p_{reject})^i \\ &= \left(\frac{|T|}{\max \cdot |g^{-1}(z)|} \pm O(\epsilon)\right) \cdot \sum_{i=0}^{k-1} \left(1 - \frac{N}{\max \cdot |g^{-1}(z)|} \pm O(\epsilon)\right)^i \end{aligned}$$

Thus,

$$\begin{aligned} \left| p_T - \frac{|T|}{N} \right| &\leq \left(1 - \frac{N}{\max \cdot |g^{-1}(z)|}\right)^k + O(\epsilon) \\ &\leq e^{-\frac{Nk}{\max \cdot |g^{-1}(z)|}} + O(\epsilon) = O(\epsilon) \end{aligned}$$

where the equality in the last step follows from the fact that $\frac{N}{\max \cdot |g^{-1}(z)|} = O(1)$ (by Condition (4) in the hypothesis).

The probability that no sample is accepted by $\mathcal{B}$ in k iterations is bounded by:

$$\left(1 - \frac{N}{\max \cdot |g^{-1}(z)|}\right)^k + O(\epsilon) = O(\epsilon)$$

Let $\mathcal{B}(z, \epsilon)$ denote the output distribution of algorithm $\mathcal{B}$. Thus,

$$\begin{aligned} |\mathcal{B}(z, \epsilon) - U_{(g \circ h)^{-1}(z)}| &= \max_{T \subseteq (g \circ h)^{-1}(z)} \left| \Pr[\mathcal{B}(z, \epsilon) \in T] - \frac{|T|}{N} \right| \\ &\leq \left| p_T + O(\epsilon) - \frac{|T|}{N} \right| = O(\epsilon) \end{aligned}$$

□

8.4 An efficient encoder

We recall the seedless non-malleable extractor constructed in Theorem 3.

Let $\text{enc} : \mathbb{F}_p \to \mathbb{F}_p^2$ be defined as $\text{enc}(x) = (x, x^4 + x^2 + x)$.

Then $\text{nmExt} : \mathbb{F}_p^{10} \to \mathbb{Z}_M$ is defined to be:

$$\text{nmExt}(x_1, \ldots, x_{10}) = \text{ext}_3(\text{ext}_2(\text{ext}_1(x_1, \ldots, x_{10})))$$

where, $\text{ext}_1 : \mathbb{F}_p^{10} \to \mathbb{F}_p^4$, $\text{ext}_2 : \mathbb{F}_p^4 \to \mathbb{F}_p$, and $\text{ext}_3 : \mathbb{F}_p \to \mathbb{Z}_M$ are defined in the following way:

$$\text{ext}_1(x_1, \ldots, x_{10}) = \big(\sum_{i=0}^{1}(\text{enc}(x_{4i+1}) + \text{enc}(x_{4i+2})) \odot (\text{enc}(x_{4i+3}) + \text{enc}(x_{4i+4})), x_9, x_{10}\big)$$

$$\text{ext}_2(y_1, y_2, z_1, z_2) = \sum_{j=1}^{2}(y_j z_j + y_j^2 z_j^2), \;\; \text{ext}_3(w) = \sigma_M(w) = w \pmod M$$

We set $M = p^\delta$ such that the error in the extractor nmExt is $\epsilon = p^{-2\delta}$. Note that, as discussed before, we use the extractor from Subsection 8.2 for ext_2 in nmExt instead of the constructions in [DLWZ11], [Li12b].

An efficient encoder for the constructed non-malleable codes in the 10-split-state model follows from the following theorem.

Theorem 8.7. *There exists a randomized algorithm which takes as input $z \in \mathbb{Z}_M$ and a parameter $\epsilon > O(p^{-1/2})$ and samples from a distribution $O(\epsilon)$-close to uniform on the set $(\text{nmExt})^{-1}(z)$. The worst case running time of the algorithm is bounded by $\text{poly}(\log p, \log(\frac{1}{\epsilon}))$.*

We prove Theorem 8.7 using the following lemma.

Lemma 8.8. *For $s \in \mathbb{Z}_M$, let $T_s = \text{ext}_3^{-1}(s) \subset \mathbb{F}_p$ and $S = \text{nmExt}^{-1}(s)$. For $a \in \mathbb{F}_p$, define $W_a = (\text{ext}_2 \circ \text{ext}_1)^{-1}(a) \subset \mathbb{F}_p^{10}$. Define $I_s = \{a \in T_s : \frac{|W_a|}{p^9} \leq 0.9\}$ and $W = \bigcup_{a \in I_s} W_a$.*

Then

$$\frac{|W|}{|S|} < p^{-(1-\delta)}, \quad \frac{|I_s|}{|T_s|} < \frac{18}{19}$$

Proof of Theorem 8.7 assuming Lemma 8.8. We show that for $g = \text{ext}_3$ and $h = \text{ext}_2 \circ \text{ext}_1$, all the conditions of Theorem 8.6 are satisfied.

1. It is easy to uniformly sample from $g^{-1}(z)$.
2. An efficient algorithm for almost uniformly sampling from $h^{-1}(y)$ follows from Lemma 8.3.
3. An efficient algorithm for approximately counting $h^{-1}(y)$ follows from Lemma 8.4.
4. Using Lemma 8.8, we have that for at least $(1/19)^{th}$ fraction of the y's in $g^{-1}(z)$, $0.9p^9 < |h^{-1}(y)| \leq 18p^9$.

Define $I = \{y \in g^{-1}(z) : |h^{-1}(y)| \leq 0.9p^9\}$. It follows from Lemma 8.8 that:

$$\frac{1}{|(g \circ h)^{-1}(z)|} \sum_{y \in I} |h^{-1}(y)| < p^{-(1-\delta)}$$

Thus by Theorem 8.6, there exists an efficient algorithm to sample almost uniformly from the set $(\text{nmExt})^{-1}(z)$. □

Proof of Lemma 8.8. We begin by proving some claims.

Claim 8.9. *For any $s \in \mathbb{Z}_M$,*

$$p^{10-\delta}(1 - p^{-\delta}) < |\text{nmExt}^{-1}(s)| < (p^{10-\delta})(1 + p^{-\delta})$$

Proof. Let $X_1, \ldots, X_{10}$ be uniform on $\mathbb{F}_p$. Using the fact that nmExt is an extractor for independent sources with error at most $\epsilon = p^{-2\delta}$, we have $|\Pr[\text{nmExt}(X_1, \ldots, X_{10}) = s] - \frac{1}{M}| < \epsilon$. The bound on $|\text{nmExt}^{-1}(s)|$ now follows. □

Claim 8.10. *For any $a \in \mathbb{F}_p$, let $W_a = (\text{ext}_2 \circ \text{ext}_1)^{-1}(a) \subset \mathbb{F}_p^{10}$. Then there exists a polynomial $g \in \mathbb{F}_p[x_1, \ldots, x_{10}]$ of degree at most 18 with coefficients in $\mathbb{F}_p$ such that W_a is the set of zeroes of g.*

Proof. Define $g(x_1, \ldots, x_{10}) = \text{ext}_2 \circ \text{ext}_1(x_1, \ldots, x_{10}) - a$. □

For $a \in \mathbb{F}_p$, define $N_a = |W_a|$. Note that $|T_s| = p^{1-\delta}$.

Using Claim 8.9, we have

$$p^{10-\delta} - p^{10-2\delta} \leq \sum_{a \in T_s} N_a \leq p^{10-\delta} + p^{10-2\delta}$$

It follows from Lemma 8.2 and Claim 8.10 that for any $a \in \mathbb{F}_p$, $N_a \leq 18p^9$. Further, Theorem 8.1 and Claim 8.10 imply that if $N_a < 0.9p^9$ for some $a \in \mathbb{F}_p$, then $N_a < Cp^8$ for some constant C.

Thus,

$$p^{10-\delta} - p^{10-2\delta} \leq |I_s| \cdot Cp^8 + (|T_s| - |I_s|) \cdot 18p^9 \tag{24}$$

for some constant C.

Since $|I_s| \leq |T_s| = p^{1-\delta}$, $|I_s| \cdot Cp^8 \leq Cp^{9-\delta}$. It follows that,

$$p^{1-\delta} - o(1) \leq 18(|T_S| - |I_s|) \tag{25}$$

Rearranging, we have

$$\frac{|I_s|}{|T_s|} \leq \frac{17}{18} + o(1) < \frac{18}{19}$$

Further,

$$\frac{|W|}{|S|} < \frac{1}{|S|} \cdot |I_s| \cdot Cp^8 \leq C \cdot \frac{p^{9-\delta}}{p^{10-\delta} - p^{10-2\delta}} < p^{-(1-\delta)}.$$

□

Acknowledgments

We are grateful to Divesh Aggarwal and Yevgeniy Dodis for very useful comments. We also thank the anonymous referees for helpful comments.

References

[ADKO14] D. Aggarwal, Y. Dodis, T. Kazana, and M. Obremski. Non-malleable reductions and applications. Unpublished manuscript, 2014.

[ADL14] Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *STOC*, 2014.

[BGK06] J. Bourgain, A. A. Glibichuk, and S. V. Konyagin. Estimates for the number of sums and products and for exponential sums in fields of prime order. *Journal of the London Mathematical Society*, 73:380–398, 4 2006.

[BIW06] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. *SIAM J. Comput.*, 36(4):1095–1118, December 2006.

[BKT04] Jean Bourgain, Nets Katz, and Terence Tao. A sum-product estimate in finite fields, and applications. *Geometric and Functional Analysis GAFA*, 14(1):27–57, 2004.

[Bou05a] J. Bourgain. Mordell's exponential sum estimate revisited. *Journal of the American Mathematical Society*, 18, No. 2 Apr.:477–499, 2005.

[Bou05b] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 01(01):1–32, 2005.

[BS94] Antal Balog and Endre Szemerdi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

[CCFP11] Hervé Chabanne, Gérard D. Cohen, Jean-Pierre Flori, and Alain Patey. Non-malleable codes from the wire-tap channel. *CoRR*, abs/1105.3879, 2011.

[CCP12] Hervé Chabanne, Gérard D. Cohen, and Alain Patey. Secure network coding and non-malleable codes: Protection against linear tampering. In *ISIT*, pages 2546–2550, 2012.

[CDF+08] Ronald Cramer, Yevgeniy Dodis, Serge Fehr, Carles Padró, and Daniel Wichs. Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors. In *EUROCRYPT*, pages 471–488, 2008.

[CG14a] Mahdi Cheraghchi and Venkatesan Guruswami. Capacity of non-malleable codes. In *ITCS*, pages 155–168, 2014.

[CG14b] Mahdi Cheraghchi and Venkatesan Guruswami. Non-malleable coding against bit-wise and split-state tampering. In *TCC*, pages 440–464, 2014.

[CKM11] SeungGeol Choi, Aggelos Kiayias, and Tal Malkin. Bitr: Built-in tamper resilience. In DongHoon Lee and Xiaoyun Wang, editors, *Advances in Cryptology ASIACRYPT 2011*, volume 7073 of *Lecture Notes in Computer Science*, pages 740–758. 2011.

[CM06] Antonio Cafure and Guillermo Matera. Improved explicit estimates on the number of solutions of equations over a finite field. *Finite Fields Appl.*, 12(2):155–185, April 2006.

[CRS12] Gil Cohen, Ran Raz, and Gil Segev. Non-malleable extractors with short seeds and applications to privacy amplification. In *IEEE Conference on Computational Complexity*, pages 298–308, 2012.

[CS09] Mahdi Cheraghchi and Amin Shokrollahi. Almost-uniform sampling of points on high-dimensional algebraic varieties. In *STACS*, pages 277–288, 2009.

[DKO13] Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. Cryptology ePrint Archive, Report 2013/498, 2013.

[DKSS09] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. Extensions to the method of multiplicities, with applications to kakeya sets and mergers. In *FOCS*, pages 181–190, 2009.

[DLWZ11] Yevgeniy Dodis, Xin Li, Trevor D. Wooley, and David Zuckerman. Privacy amplification and non-malleable extractors via character sums. In *FOCS*, pages 668–677, 2011.

[DPW10] Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. In *ICS*, pages 434–452, 2010.

[DW09] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *STOC*, pages 601–610, 2009.

[FMNV14] Sebastian Faust, Pratyay Mukherjee, Jesper Buus Nielsen, and Daniele Venturi. Continuous non-malleable codes. In *TCC*, pages 465–488, 2014.

[FMVW13] Sebastian Faust, Pratyay Mukherjee, Daniele Venturi, and Daniel Wichs. Efficient non-malleable codes and key-derivation for poly-size tampering circuits. *IACR Cryptology ePrint Archive*, 2013:702, 2013.

[Gow98] W. T. Gowers. A new proof of szemeredi's theorem for arithmetic progressions of length four. *Geometric and Functional Analysis GAFA*, 8(3):529–551, 1998.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *J. ACM*, 56(4), 2009.

[HW98] Ming-Deh A. Huang and Yiu-Chung Wong. An algorithm for approximate counting of points on algebraic sets over finite fields. In *ANTS*, pages 514–527, 1998.

[Kon03] Sergei Konyagin. A sum-product estimate in fields of prime order. arXiv:math/0304217, 2003.

[Li12a] Xin Li. Design extractors, non-malleable condensers and privacy amplification. In *STOC*, pages 837–854, 2012.

[Li12b] Xin Li. Non-malleable extractors, two-source extractors and privacy amplification. In *FOCS*, pages 688–697, 2012.

[Li13] Xin Li. New independent source extractors with exponential improvement. In *STOC*, pages 783–792, 2013.

[LL12] Feng-Hao Liu and Anna Lysyanskaya. Tamper and leakage resilience in the split-state model. In *CRYPTO*, pages 517–532, 2012.

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Extractors: optimal up to constant factors. In *STOC*, pages 602–611, 2003.

[MW97] Ueli M. Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In *CRYPTO*, pages 307–321, 1997.

[NZ93] Noam Nisan and David Zuckerman. More deterministic simulation in logspace. In *STOC*, pages 235–244, 1993.

[Rao06] Anup Rao. Extractors for a constant number of polynomially small min-entropy independent sources. In *STOC*, pages 497–506, 2006.

[Rao07] Anup Rao. An exposition of Bourgain's 2-source extractor. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(034), 2007.

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, October 1980.

[TV06] Terence Tao and Van H. Vu. Cambridge University Press, 2006.

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposiumon on Symbolic and Algebraic Computation*, EUROSAM '79, pages 216–226, London, UK, UK, 1979. Springer-Verlag.

A Proof of Theorem 8.5

We restate Theorem 8.5 for the sake of convenience.

Theorem 8.5 (restated)**.** *Let p be a prime. Define the functions* $\mathrm{ext}_2 : (\mathbb{F}_p^2) \times (\mathbb{F}_p^2) \to \mathbb{F}_p$ *and* $\mathrm{snmExt} : (\mathbb{F}_p^2) \times (\mathbb{F}_p^2) \to \mathbb{Z}_M$ *in the following way:*

$$\mathrm{ext}_2((x_1, x_2), (y_1, y_2)) = \sum_{j=1}^{2} (x_j y_j + x_j^2 y_j^2), \qquad \mathrm{snmExt}(x, y) = \sigma_M(\mathrm{ext}_2(x, y))$$

where $\sigma_M(x) = x \ (mod\ M)$. Suppose X, Y are independent sources on $\mathbb{F}_p^2$ with min-entropies k_1, k_2 respectively.

1. *If* $(k_1 + k_2) \geq (2 + \delta) \log p$, *then*

$$|\text{snmExt}(X, Y) \circ X - U_M \circ X| < p^{-\Omega(1)}, \ |\text{snmExt}(X, Y) \circ Y - U_M \circ Y| < p^{-\Omega(1)}$$

2. *If* $k_1, k_2 > (2 - \delta) \log p$ *and* f *is any tampering function with no fixed points, then*

$$|\text{snmExt}(X, Y) \circ \text{snmExt}(X, f(Y)) - U_M \circ \text{snmExt}(X, f(Y))| < p^{-\Omega(1)}.$$

We recall some results in order to prove Theorem 8.5.

Notation For any distribution D and non-negative integers c_1, c_2, let $c_1 D - c_2 D$ be the distribution obtained by drawing independent samples $d_1, \ldots, d_{c_1+c_2}$ from D and outputting $(d_1 + \ldots + d_{c_1}) - (d_{c_1+1} + \ldots + d_{c_1+c_2})$.

Define $e_N : \mathbb{Z}_N \to \mathbb{C}$ as $e_N(x) = e^{\frac{2\pi i x}{N}}$. Recall that any nontrivial character ϕ of the additive group $\mathbb{Z}_N$ is of the form $\phi(x) = e_N(\alpha x)$ for some $\alpha \in \mathbb{Z}_N \setminus \{0\}$. For an introduction to Fourier analysis on Abelian groups, we refer the reader to [TV06].

The following XOR lemma was proved in [Rao07].

Lemma A.1 (XOR lemma)**.** *Let* D *be a distribution over* $\mathbb{Z}_N$ *such that for every nontrivial character* ψ *of* $\mathbb{Z}_N$, *we have* $|\mathbb{E}[\psi(D)]| \leq \epsilon$. *Then, for any* $M < N$, *we have*

$$|\sigma_M(D) - U_M| = O(\epsilon \log N \sqrt{M}) + O(M/N)$$

We need a slightly modified version of an XOR Lemma proved in [DLWZ11].

Lemma A.2. *Let* D_1, D_2 *be distributions over* $\mathbb{Z}_N$ *such that for arbitrary characters* ψ, ϕ *of* $\mathbb{Z}_N$, *we have* $|\mathbb{E}[\psi(D_1)\phi(D_2)]| \leq \epsilon$, *whenever* ψ *is nontrivial. Then, for any* $M < N$, *we have*

$$|(\sigma_M(D_1), \sigma_M(D_2)) - (U_M, \sigma_M(D_2))| = O(\epsilon (\log N)^2 M) + O(M/N)$$

For vectors $v, w \in \mathbb{F}_p^C$, let $\langle v, w \rangle$ denote the standard inner product over $\mathbb{F}_p$. The following results are well known and can be found in [Rao07].

Lemma A.3. *Let* $\delta > 0$ *be a constant. Let* X, Y *be independent sources on* $\mathbb{F}_p^C$ *with min-entropies* k_1, k_2 *respectively, such that* $k_1 + k_2 \geq (C + \delta) \log p$. *Then for any nontrivial character* ϕ *of* $\mathbb{F}_p$, $|\mathbb{E}[\phi(\langle X, Y \rangle)]| < p^{-\Omega(1)}$.

Lemma A.4. *Let* $\delta > 0$ *be a constant. Let* X, Y *be independent sources on* $\mathbb{F}_p^C$ *with the following property: There exist constants* c_1, c_2 *such that the sources* $c_1 X, c_2 Y$ *with min-entropy* k_1, k_2 *respectively, satisfying* $k_1 + k_2 \geq (C + \delta) \log p$. *Then for any nontrivial character* ϕ *of* $\mathbb{F}_p$, $|\mathbb{E}[\phi(\langle X, Y \rangle)]| < p^{-\Omega(1)}$.

Lemma A.5. *Let* $\text{ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, $m = \Omega(n)$, *be a 2-source extractor with error* $2^{-\Omega(n)}$, *for independent sources with min-entropies* k_1, k_2 *respectively, satisfying* $k_1 + k_2 > (1 + \delta)n$. *Then* ext *is a strong 2-source extractor with error* $2^{-\Omega(n)}$ *for independent sources with min-entropies* k_1, k_2 *respectively, satisfying* $k_1 + k_2 > (1 + 2\delta)n$.

Proof of Theorem 8.5.

- Suppose $k_1 + k_2 > (2+\delta)\log p$.

  We prove the following claim.

  Claim A.6. *For any nontrivial character ϕ of $\mathbb{F}_p$, we have $|\mathbb{E}[\phi(\mathrm{ext}_2(X,Y)]| < p^{-\Omega(1)}$.*

  Proof. We note that,

  $$\mathrm{ext}_2(X,Y) = \langle (x_1, x_2, x_1^2, x_2^2), (y_1, y_2, y_1^2, y_2^2) \rangle$$

  Define the distributions X', Y' (over $\mathbb{F}_p^4$) in the following way:

  $$X' = (x_1, x_2, x_1^2, x_2^2) : (x_1, x_2) \sim X, \quad Y' = (y_1, y_2, y_1^2, y_2^2) : (y_1, y_2) \sim Y$$

  Thus

  $$\mathrm{ext}_2(X,Y) = \langle X', Y' \rangle$$

  We claim that $2X'$ has min-entropy at least $2k_1 - 2$. To prove this, consider arbitrary $(a, b, c, d) \in \mathbb{F}_p^4$. $x_1 + \bar{x}_1 = a$, $x_1^2 + \bar{x}_1^2 = c$ has at most 2 solutions in $(x_1, \bar{x}_1)$. In a similar way, $x_2 + \bar{x}_2 = b$, $x_2^2 + \bar{x}_2^2 = d$ has at most 2 solutions in $(x_2, \bar{x}_2)$. Thus,

  $$\Pr[2X' = (a,b,c,d)] \leq 4 \cdot 2^{-2k_1}$$

  Hence $2X'$ has min-entropy at least $2k_1 - 2$. Similarly, the distribution $2Y'$ has entropy at least $2k_2 - 2$.

  Thus,

  $$H_\infty(2X') + H_\infty(2Y') \geq 2(k_1 + k_2) - 4 > (4+\delta)\log p$$

  The proof now follows using Claim A.4. □

  By combining Claim A.6 with Lemma A.1 and Lemma A.5, we have

  $$|\mathrm{snmExt}(X,Y) \circ X - U_M \circ X| < p^{-\Omega(1)}, \ |\mathrm{snmExt}(X,Y) \circ Y - U_M \circ Y| < p^{-\Omega(1)}$$

- Now suppose $k_1, k_2 > (2-\delta)\log p$ and f is any tampering function with no fixed points.

  Let f_1, f_2 be functions such that for any $(y_1, y_2) \in \mathbb{F}_p^2$,

  $$f(y_1, y_2) = (f_1(y_1, y_2), f_2(y_1, y_2))$$

  We show that for arbitrary characters ϕ, ψ of $\mathbb{F}_p$, such that ϕ is nontrivial,

  $$|\mathbb{E}[\phi(\mathrm{ext}_2(X,Y))\psi(\mathrm{ext}_2(X, f(Y)))]| < p^{-\Omega(1)} \tag{26}$$

  Let $\phi(x) = e_p(\alpha x)$, $\psi(x) = e_p(\alpha\beta x)$ for some $\alpha \in \mathbb{F}_p^*$ and $\beta \in \mathbb{F}_p$. To prove (26), it is enough to prove the following claim.

  Claim A.7. *For arbitrary $\alpha \in \mathbb{F}_p^*$ and $\beta \in \mathbb{F}_p$,*

  $$|\mathbb{E}[e_p(\alpha \cdot \mathrm{ext}_2(X,Y) + \alpha\beta \cdot \mathrm{ext}_2(X, f(Y))]| < p^{-\Omega(1)}$$

Proof. If $\beta = 0$, the proof follows by Claim A.6. Thus suppose $\beta \in \mathbb{F}_p^*$.

Define the distributions X' and $Y'_{\beta,f}$ (over $\mathbb{F}_p^4$) in the following way:

$$X' = (x_1, x_2, x_1^2, x_2^2) : (x_1, x_2) \sim X$$

$$Y'_{\beta,f} = (y_1 + \beta f_1(y_1, y_2), y_2 + \beta f_2(y_1, y_2), y_1^2 + \beta f_1(y_1, y_2)^2, y_2^2 + \beta f_2(y_1, y_2)^2) : (y_1, y_2) \sim Y$$

We note that,

$$\mathbb{E}[e_p(\alpha \cdot \mathrm{ext}_2(X, Y) + \alpha\beta \cdot \mathrm{ext}_2(X, f(Y))]| = \mathbb{E}[\phi(\langle X', Y'_{\beta,f}\rangle)]$$

where $\phi(x) = e_p(\alpha x)$ is a nontrivial character of $\mathbb{F}_p$.

It follows from the proof of Claim A.6 that the source $2X'$ has min-entropy at least $2k_1 - 2$.

We now claim that the distribution $Y'_{\beta,f}$ has min-entropy at least $k_2 - \log p$. To prove this consider arbitrary $(a, b, c, d) \in \mathbb{F}_p^4$. To get an upper bound on $\Pr[Y'_{\beta,f} = (a, b, c, d)]$, we bound the number of pairs (y_1, y_2) satisfying the following equations:

$$y_1 + \beta f_1(y_1, y_2) = a \tag{27}$$

$$y_1^2 + \beta f_1(y_1, y_2)^2 = c \tag{28}$$

$$y_2 + \beta f_2(y_1, y_2) = b \tag{29}$$

$$y_2^2 + \beta f_2(y_1, y_2)^2 = d \tag{30}$$

Eliminating $f_i(y_1, y_2)$ from equations (28) and (30), we have

$$(1 + \beta)y_1^2 - 2ay_1 + (a^2 - \beta c) = 0 \tag{31}$$

$$(1 + \beta)y_2^2 - 2by_2 + (b^2 - \beta b) = 0 \tag{32}$$

1. If $\beta \neq -1$, clearly there are at most 4 possible values of (y_1, y_2) satisfying the equations (31) and (32).
2. Now suppose $\beta = -1$.
   (a) If $ab \neq 0$, then (y_1, y_2) can take exactly 1 value.
   (b) Consider the case where $a = 0$. This forces $c = 0$ for (31) to hold. Hence by (27), $f_1(y_1, y_2) = y_1$.
      i. If $b \neq 0$, then y_2 can take at most 1 value and hence there are at most p values of (y_1, y_2) satisfying the equations (31) and (32).
      ii. If $b = 0$, then d is forced to be 0 for (32) to hold. By (29), $f_2(y_1, y_2) = y_2$ and hence $f(y_1, y_2) = (y_1, y_2)$. Since f has no fixed points, there can be no solutions in this case.

We thus have,

$$\Pr[Y'_{\beta,f} = (a, b, c, d)] \leq p \cdot 2^{-k_2}$$

and hence $Y'_{\beta,f}$ has min-entropy at least $k_2 - \log p$.

Thus,

$$H_\infty(2X') + H_\infty(Y'_{\beta,f}) \geq 2k_1 + k_2 - \log p - 2 \geq 5 \log p - 3\delta \log p - 2 >> (4 + \delta) \log p$$

The proof now follows using Claim A.4. □

By combining (26) with Lemma A.2, we have

$$|\text{snmExt}(X,Y) \circ \text{snmExt}(X, f(Y)) - U_M \circ \text{snmExt}(X, f(Y))| < p^{-\Omega(1)}.$$

This concludes the proof of Theorem 8.5.

□

B An additional property of the constructed seedless non-malleable extractor

We include an additional property of the seedless non-malleable extractor from Theorem 6.13, which might find application in other explicit constructions. Independently, Aggarwal, Dodis, Kazanaand and Obremski [ADKO14] gave a general reduction from 2 parts to a constant number of parts, incurring only a constant overhead in the rate, as long as the non-malleable extractor satisfies this property. As a result, after seeing a preliminary version of our work, they applied their reduction to our result to construct constant-rate non-malleable codes in the 2-split model.

Theorem B.1. *Let $X_1, \ldots, X_8$ be independent (n,n)-sources and let X_9 be an independent $(2n,2n)$-source. Let* $\text{nmExt} : (\{0,1\}^n)^8 \times \{0,1\}^{2n} \to \{0,1\}^m$*,$m = \Omega(n)$, be the seedless non-malleable extractor with error $\epsilon = 2^{-\Omega(n)}$ from Theorem* 6.13*. Then:*

$$|\text{nmExt}(X_1, \ldots, X_9) \circ X_{i_1} \circ \ldots X_{i_8} - U_m \circ X_{i_1} \circ \ldots X_{i_8}| < 2^{-\Omega(n)}$$

for arbitrary $1 \leq i_1 < \ldots < i_8 \leq 9$.

Proof. Recall that $\text{nmExt}(X_1 \ldots, X_9) = \text{snmExt}(\text{ext}_1(X_1, \ldots, X_8), X_9)$, where the functions snmExt and ext_1 are defined in the proof of Theorem 6.13. We split the proof into two cases.

- Suppose $i_8 \neq 9$. Consider a fixing of the sources $X_1, \ldots, X_7$ to arbitrary values $x_1, \ldots, x_7$ such that $(\text{enc}(x_5) + \text{enc}(x_6)) \in (\mathbb{F}_p^*)^2$. It follows from Claim 6.6 that the probability over $X_1 \ldots, X_7$ of a fixing satisfying this condition is at least $1 - 2^{-\Omega(n)}$. We prove that:

$$|\text{nmExt}(x_1, \ldots x_7, X_8, X_9) \circ X_8 - U_m \circ X_8| < 2^{-\Omega(n)} \tag{33}$$

  It follows from the structure of ext_1 that $\text{ext}_1(x_1, \ldots, x_7, X_8)$ is a (2n,n)-source. By applying Theorem 2.20 on the sources $\text{ext}_1(x_1, \ldots, x_7, X_8)$ and X_9, we have

  $$|\text{snmExt}(\text{ext}_1(x_1, \ldots, x_7, X_8), X_9) \circ \text{ext}_1(x_1, \ldots, x_7, X_8) - U_m \circ \text{ext}_1(x_1, \ldots, x_7, X_8)| < 2^{-\Omega(n)}.$$

  Now (33) follows by observing that there is a deterministic one-one map between the random variables X_8 and $\text{ext}_1(x_1, \ldots, x_7, X_8)$.

  Using Lemma 2.10, it follows that

  $$|\text{nmExt}(X_1, \ldots, X_9) \circ X_1 \circ \ldots X_8 - U_m \circ X_1 \circ \ldots X_8| < 2^{-\Omega(n)}$$

- Now suppose $i_8 = 9$. Let $1 \leq j \leq 8$ be such that $j \neq i_l$ for any $l \in [8]$. Consider any fixing of the sources $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_8$ such that $\mathrm{ext}_1(x_1, \ldots, x_{j-1}, X_j, x_{j+1}, \ldots, x_8)$ is a $(2n, n)$-source. As argued in the previous case, the probability (over $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_8$) of a fixing satisfying this condition is at least $1 - 2^{-\Omega(n)}$. Further recall that X_9 is a $(2n, 2n)$-source. Using Theorem 2.20, we have

$$|\mathrm{snmExt}(\mathrm{ext}_1(x_1, \ldots, x_{j-1}, X_j, x_{j+1}, \ldots, x_8), X_9) \circ X_9 - U_m \circ X_9| < 2^{-\Omega(n)}$$

  Using Lemma 2.10, it follows that

$$|\mathrm{nmExt}(X_1, \ldots, X_9) \circ X_{i_1} \circ \ldots X_{i_8} - U_m \circ X_{i_1} \circ \ldots X_{i_8}| < 2^{-\Omega(n)}.$$

□

ECCC ISSN 1433-8092

http://eccc.hpi-web.de