2014 IEEE Annual Symposium on Foundations of Computer Science

# Non-Malleable Codes Against Constant Split-State Tampering

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

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

***Abstract*—Non-malleable codes were introduced by Dziembowski, Pietrzak and Wichs [1] 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 $\epsilon$-close to a distribution $D_f$ independent of $m$, where $\epsilon$ 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 [2] 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 [3].**

**Our constructions use an elegant connection found between seedless non-malleable extractors and non-malleable codes by Cheraghchi and Guruswami [3]. 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.**

***Index Terms*—non-malleable codes; non-malleable extractors; coding theory; randomness extractors.**

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

## I. Introduction

### A. 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. [4] 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 [1] 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 $\epsilon$-close (in statistical distance) to a distribution $D_f$ independent of $m$. The parameter $\epsilon$ is called the error.

Let $\mathcal{F}_n$ be the set of all functions on $\{0,1\}^n$. Note

0272-5428 2014
DOI 10.1109/FOCS.2014.40

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 [1], [5] 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 [1] for $C = 2$ and then improved in [5]. 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 [2], 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 [2], 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 [3] 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 [3] 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 [2] 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 [3], 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 [2] in the $C$-split model for $C = o(n)$. Before the work of [2], the only unconditional efficient non-malleable code in the $C$-split-state model, for $C = o(n)$, was by Dziembowski, Kazana, and Obremski [6]. However, they could encode only 1 bit messages.

There were also some conditional results. Liu and Lysyanskaya [7] constructed efficient constant rate non-malleable codes in the split-state model against computationally bounded adversaries. 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. [8] constructed almost optimal non-malleable codes against the class of polynomial sized circuits in the CRS framework. [9], [10], [11], and [12] considered non-malleable codes in other models.

**Independent Work:** Independently, Aggarwal, Dodis, Kazana and and Obremski [13] 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.

### *B. 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 [14] 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 II).

**Definition I.1.** *A function* $\text{SExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* $(k,\epsilon)$*-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 [15],[16],[17], 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 [18]. 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 [19], [20], [21] and [22]. Recently Li [23] found applications of non-malleable extractors in constructing extractors for independent sources. To define non-malleable extractors, we need the following definition.

**Definition I.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 I.3.** *A function* $\text{snmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* $(k_1,k_2,\epsilon)$*-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, $f$ is called a tampering function.

In a recent work, Cheraghchi and Guruswami [3] 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 $\text{nmExt} : (\{0,1\}^n)^2 \to \{0,1\}^m$ 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 [3] 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 $\chi$ is the quadratic character, was shown to be a seeded non-malleable extractor [19]. 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 III for formal definitions.

**Definition I.4** (informal)**.** *A function* snmExt : $(\{0,1\}^n)^C \to \{0,1\}^m$ *is a* $(k,\epsilon)$*-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.

**Theorem 3.** *For some* $\delta > 0$ *there exists a polynomial time construction of a* $(k,\epsilon)$*-seedless non-malleable extractor for* 10 *independent sources* 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 [3], 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.

### C. Organization

We discuss preliminaries in Section $II$, and formally define non-malleable codes and seedless non-malleable extractors in Section $III$. We recall the connection between non-malleable codes and seedless non-malleable extractors from [3] and deduce Theorem 1 assuming Theorem 3 in Section $IV$. Our main technical contribution is the proof of Theorem 3. We use Section $V$ to sketch the main ideas in proving Theorem 3. We require a sum-product estimate over $\mathbb{F}_p^4$ for proving Theorem 3. We state this theorem in Section VI. The proof of this estimate closely follows the arguments of a sum-product theorem over $\mathbb{F}_p^2$ by Bourgain [24]. We give high level ideas of constructing an efficient encoder for the constructed non-malleable codes in the 10-split-state model in Section VII. In Section VIII, we state an additional property of the constructed seedless non-malleable extractor which might be useful in other explicit constructions.

## II. Preliminaries

### A. Notations

Let $[l]$ denote the set $\{1,2,\ldots,l\}$. Let $U_m$ denote the uniform distribution over $\{0,1\}^m$. 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 $\Pi_S$ to sets in the obvious manner. We use $\Pi_i$ for $\Pi_{\{i\}}$.

### B. Min entropy and flat distributions

For a source $X$ we define min-entropy of $X$ as: $H_\infty(X) = \min_{s \in \text{support}(X)} \left\{ \frac{1}{\log(\Pr[X=s])} \right\}$. A $(n,k)$-source is a distribution on $\{0,1\}^n$ with min-entropy $k$. We call a distribution (source) $D$ to be flat if it is uniform

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

### C. Statistical distance, convex combination of distributions and probability lemmas

**Definition II.1.** *Let $D_1$ and $D_2$ be two distributions on a set $S$. The statistical distance between $D_1$ and $D_2$ is*

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

**Definition II.2.** *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 that $D$ is a convex combination of the distributions $D_1,\ldots,D_l$ with weights $w_1,\ldots,w_l$.*

**Definition II.3.** *For random variables $X$ and $Y$, let $X|Y$ 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 record the following simple lemma.

**Lemma II.4.** *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 $|X - Y| \le \sum_i w_i \cdot |X_i - Y_i|$.*

The following result follows from a lemma proved in [25].

**Corollary II.5.** *Let $X, Y$ be random variables with supports $S, T \subseteq V$ such that $(X,Y)$ is $\epsilon$-close to a distribution with min-entropy $k$. Further suppose that the random variable $Y$ can take at most $l$ values. Then*

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

### D. 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 [26].

**Theorem II.6.** *For all $n > 0$ and any constant $\delta$ 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 $k_1 + k_2 \ge (1+\delta)n$, then*

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

Explicit constructions of seeded non-malleable extractors follow from works of [19] and [22]. The output length in [19] relies on an unproven but widely believed conjecture on primes while the output length in [22] is unconditional. Further, either of the non-malleable extractors from [19] or [22] is also a strong 2-source extractor.

**Theorem II.7** ([19],[22])**.** *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 $(k_1 + k_2) \ge (1+\delta)n$, 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*

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

## III. Non-malleable codes and Seedless non-malleable extractors

### A. Non-malleable codes

We follow the presentation in [1] and define non-malleable codes.

**Definition III.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* $(\text{Enc}, \text{Dec})$ *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 III.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 need to define the following function.

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

**Definition III.3** (Non-malleable codes)**.** *A coding scheme* $(\text{Enc}, \text{Dec})$ *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* $\epsilon$ *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*

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

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 $\text{Dec}(c)$ with probability 1.

*1) Non-malleable codes in the $C$-split-state model:* We formally define non-malleable codes in the $C$-split sate model.

**Definition III.4.** *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.

### *B. Seedless non-malleable extractors*

Seedless non-malleable extractors were first introduced by Cheraghchi and Guruswami in [3]. We define a special case of such non-malleable which is of particular interest to us.

**Definition III.5** (Seedless non-malleable multi-source extractors)**.** *For any constant $C$, we say that* 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 $\epsilon$ 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.*

## IV. Non-malleable codes via Seedless non-malleable extractors

In this section we prove Theorem 1 assuming Theorem 3. The work by Cheraghchi and Guruswami [3] 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 VII.

The following theorem follows from the work in [3].

**Theorem IV.1.** *For any constant $C$, let* 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 $= 2^{-\Omega(n)}$.*

Thus composing Theorem 3 with Theorem $IV.1$ 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 is constructed in Section VII. This proves Theorem 1.

## V. Proof outline of Theorem 3

In this section we sketch the main ideas involved in proving Theorem 3. The formal proof of Theorem 3 is deferred to the full version.

**Definition V.1.** *We call a set $A$ satisfying the conclusion of Theorem $VI.1$ to be sum-product friendly. We call a flat distribution sum-product friendly if its support is sum-product friendly.*

**Definition V.2.** *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 $\max_{s \in S} |f^{-1}(s)|$.*

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

### A. 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 over prime fields ([27], [28], [29]) in [30], 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 II.5, $(X_1 \cdot X_2 + X_3) | (f_1(X_1) \cdot f_2(X_2) + f_3(X_3))$ 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

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

where 3Ext is an extractor for 3 independent sources.

### B. 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.

### C. The actual construction

The high level idea is to make the previous approach 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 [31] 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).

- 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 $\text{enc}(x_i) = (x_i, q(x_i))$ for some suitable $q()$ to be fixed later. Define the source

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

  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 V.3** (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 VI.1 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 statement of sum-product estimate is presented in Section $VI$. It is obtained by closely following the proof of a sum product estimate over $\mathbb{F}_p^2$ obtained by Bourgain in [24] and extending the arguments to $\mathbb{F}_p^4$. We defer its proof to the full version.
- The idea to prove Claim $V.3$ is to adapt the machinery developed in [30] for proving such expansion statements about min-entropy to a more general setting. We point out the key differences from [30] 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 $\text{enc}(x) = (x, x^4 + x^2 + x)$, 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 [30] goes through even with this weaker guarantee.

- Define the following function.

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

We use Claim $V.3$ and Corollary $II.5$ to conclude the following.

**Claim V.4** (informal)**.** *Let $X_1,\ldots,X_8$ be independent $(n,(1-\delta)n)$-sources. Then $\text{ext}_1(X_1,\ldots,X_8)|\text{ext}_1(f_1(X_1),\ldots,f_8(X_8))$ 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 $V.4$ can be removed.

**Claim V.5** (informal)**.** *Let $X_1,\ldots,X_8$ be independent $(n,(1-\delta)n)$-sources. Then $\text{ext}_1(X_1,\ldots,X_8)|\text{ext}_1(f_1(X_1),\ldots,f_8(X_8))$ 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 $V.3$ 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 $V.5$)

- To motivate our final construction, we describe an extractor $\text{ext}_2$ in this step which we don't actually use in our construction.

Let SExt be the strong 2-source extractor from Theorem $II.6$. Let $\text{ext}_2 : (\{0,1\}^n)^8 \times \{0,1\}^{2n} \to \{0,1\}^m$, $m=\Omega(n)$, be defined as:

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

The following result follows from Claim $V.5$.

**Claim V.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 $\text{ext}_2(X_1,\ldots,X_9)|\text{ext}_2(f_1(X_1),\ldots,f_9(X_9))$ 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 $V.5$ 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 $II.6$ we have

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

The proof of Claim $V.6$ now follows.

- However, if $f_i$ is the identity function for all $i\in[8]$, and $f_9$ is any arbitrary tampering function with no fixed points, $\text{ext}_2$ does not work. 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 $II.7$. We prove the following claim.

**Claim V.7** (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.*

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 $V.6$.
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 $II.7$,

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

   This concludes the proof of Claim $V.7$.

Theorem 3 follows from Claim V.7 with some additional work.

## VI. The sum-product estimate over $\mathbb{F}_p^4$

We closely follow the proof of the sum-product estimate by Bourgain in [24] and obtain the following theorem.

**Theorem VI.1.** *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 $\tau_0, \tau_1$) such that if $|A| < p^{7/3-\tau_1}$, then*

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

We defer the proof of Theorem $VI.1$ to the full version.

## VII. Efficient algorithms for non-malleable codes in the 10-split-state model

In this section we give the high level ideas of the proof of efficiency of the non-malleable codes constructed in Theorem 1. The detailed proof is deferred to the full version. 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 nmExt being polynomial time function.

We construct an efficient algorithm which 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 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 [32].
- Further, obtaining the weights in the convex combination reduces to approximately counting the size of such algebraic sets for which there are efficient algorithms [33]. 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.

### A. 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 II.7). We carefully choose snmExt such that it is easy to sample almost uniformly from $\text{nmExt}^{-1}(s)$. The main idea is to pick snmExt such that $\text{nmExt}^{-1}(s)$ is a convex combination of algebraic varieties of low degree over a field with large characteristic. Thus, the constructions in [22] 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 $\sigma_M \circ \text{ext}_2 \circ \text{ext}_1$, where $\text{ext}_1 : \mathbb{F}_p^{10} \to \mathbb{F}_p^4$, $\text{ext}_2 : \mathbb{F}_q^2 \to \mathbb{F}_q$, and $p, q$ are primes satisfying $p^2 \leq q \leq 2p^2$ (and interpreting the output of $\text{ext}_1$ as an element in $\mathbb{F}_q^2$). Changing the characteristic of the field destroys the low degree properties of the function $\text{ext}_2 \circ \text{ext}_1$.

To fix this, we construct a new extractor for $\text{ext}_2$ (satisfying the conditions of Theorem II.7) which allows us to work over the same field as $\text{ext}_1$. The extractor is a variation of a construction by Bourgain [31]. The proof uses ideas from [22], but requires more work.

**Theorem VII.1.** *Let $p$ be a prime. Define the functions $\text{ext}_2 : (\mathbb{F}_p^2) \times (\mathbb{F}_p^2) \to \mathbb{F}_p$ and $\text{snmExt} : (\mathbb{F}_p^2) \times (\mathbb{F}_p^2) \to \mathbb{Z}_M$ in the following way:*

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

*where $\sigma_M(x) = x \pmod M$. 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*

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

The proof of Theorem VII.1 is deferred to the full version.

## VIII. An additional property of the constructed seedless non-malleable extractor

We include an additional property of the seedless non-malleable extractor from Theorem 3, which might find application in other explicit constructions. We do not use Theorem VIII.1 in this paper.

**Theorem VIII.1.** *Let $X_1, \ldots, X_8$ be independent $(n,n)$-sources and let $X_9$ be an independent $(2n,2n)$-source. Let* 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 3.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 $1 \leq i_1 < \ldots < i_8 \leq 9$.*

## Acknowledgments

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

## References

[1] S. Dziembowski, K. Pietrzak, and D. Wichs, "Non-malleable codes," in *ICS*, pp. 434–452, 2010.

[2] D. Aggarwal, Y. Dodis, and S. Lovett, "Non-malleable codes from additive combinatorics," in *STOC*, 2014.

[3] M. Cheraghchi and V. Guruswami, "Non-malleable coding against bit-wise and split-state tampering," in *TCC*, pp. 440–464, 2014.

[4] R. Cramer, Y. Dodis, S. Fehr, C. Padró, and D. Wichs, "Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors," in *EUROCRYPT*, pp. 471–488, 2008.

[5] M. Cheraghchi and V. Guruswami, "Capacity of non-malleable codes," in *ITCS*, pp. 155–168, 2014.

[6] S. Dziembowski, T. Kazana, and M. Obremski, "Non-malleable codes from two-source extractors." Cryptology ePrint Archive, Report 2013/498, 2013.

[7] F.-H. Liu and A. Lysyanskaya, "Tamper and leakage resilience in the split-state model," in *CRYPTO*, pp. 517–532, 2012.

[8] S. Faust, P. Mukherjee, D. Venturi, and D. Wichs, "Efficient non-malleable codes and key-derivation for poly-size tampering circuits," *IACR Cryptology ePrint Archive*, vol. 2013, p. 702, 2013.

[9] H. Chabanne, G. D. Cohen, and A. Patey, "Secure network coding and non-malleable codes: Protection against linear tampering," in *ISIT*, pp. 2546–2550, 2012.

[10] H. Chabanne, G. D. Cohen, J.-P. Flori, and A. Patey, "Non-malleable codes from the wire-tap channel," *CoRR*, vol. abs/1105.3879, 2011.

[11] S. Choi, A. Kiayias, and T. Malkin, "Bitr: Built-in tamper resilience," in *Advances in Cryptology ASIACRYPT 2011* (D. Lee and X. Wang, eds.), vol. 7073 of *Lecture Notes in Computer Science*, pp. 740–758, 2011.

[12] S. Faust, P. Mukherjee, J. B. Nielsen, and D. Venturi, "Continuous non-malleable codes," in *TCC*, pp. 465–488, 2014.

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

[14] N. Nisan and D. Zuckerman, "More deterministic simulation in logspace," in *STOC*, pp. 235–244, 1993.

[15] C.-J. Lu, O. Reingold, S. P. Vadhan, and A. Wigderson, "Extractors: optimal up to constant factors," in *STOC*, pp. 602–611, 2003.

[16] V. Guruswami, C. Umans, and S. P. Vadhan, "Unbalanced expanders and randomness extractors from parvaresh–vardy codes," *J. ACM*, vol. 56, no. 4, 2009.

[17] Z. Dvir, S. Kopparty, S. Saraf, and M. Sudan, "Extensions to the method of multiplicities, with applications to kakeya sets and mergers," in *FOCS*, pp. 181–190, 2009.

[18] Y. Dodis and D. Wichs, "Non-malleable extractors and symmetric key cryptography from weak secrets," in *STOC*, pp. 601–610, 2009.

[19] Y. Dodis, X. Li, T. D. Wooley, and D. Zuckerman, "Privacy amplification and non-malleable extractors via character sums," in *FOCS*, pp. 668–677, 2011.

[20] G. Cohen, R. Raz, and G. Segev, "Non-malleable extractors with short seeds and applications to privacy amplification," in *IEEE Conference on Computational Complexity*, pp. 298–308, 2012.

[21] X. Li, "Design extractors, non-malleable condensers and privacy amplification," in *STOC*, pp. 837–854, 2012.

[22] X. Li, "Non-malleable extractors, two-source extractors and privacy amplification," in *FOCS*, pp. 688–697, 2012.

[23] X. Li, "New independent source extractors with exponential improvement," in *STOC*, pp. 783–792, 2013.

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

[25] U. M. Maurer and S. Wolf, "Privacy amplification secure against active adversaries," in *CRYPTO*, pp. 307–321, 1997.

[26] A. Rao, "An exposition of Bourgain's 2-source extractor," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 14, no. 034, 2007.

[27] J. Bourgain, N. Katz, and T. Tao, "A sum-product estimate in finite fields, and applications," *Geometric and Functional Analysis GAFA*, vol. 14, no. 1, pp. 27–57, 2004.

[28] 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*, vol. 73, pp. 380–398, 4 2006.

[29] S. Konyagin, "A sum-product estimate in fields of prime order," vol. arXiv:math/0304217, 2003.

[30] B. Barak, R. Impagliazzo, and A. Wigderson, "Extracting randomness using few independent sources," *SIAM J. Comput.*, vol. 36, pp. 1095–1118, Dec. 2006.

[31] J. Bourgain, "More on the sum-product phenomenon in prime fields and its applications," *International Journal of Number Theory*, vol. 01, no. 01, pp. 1–32, 2005.

[32] M. Cheraghchi and A. Shokrollahi, "Almost-uniform sampling of points on high-dimensional algebraic varieties," in *STACS*, pp. 277–288, 2009.

[33] M.-D. A. Huang and Y.-C. Wong, "An algorithm for approximate counting of points on algebraic sets over finite fields," in *ANTS*, pp. 514–527, 1998.