Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 70 (2018)

Non-Malleable Extractors and Codes for Composition of Tampering, Interleaved Tampering and More

Eshan Chattopadhyay*
Cornell University
eshanc@cornell.edu

Xin Li†
Department of Computer Science,
John Hopkins University
lixints@cs.jhu.edu

November 2, 2018

Abstract

Non-malleable codes were introduced by Dziembowski, Pietrzak, and Wichs (JACM 2018) as a generalization of standard error correcting codes to handle severe forms of tampering on codewords. This notion has attracted a lot of recent research, resulting in various explicit constructions, which have found applications in tamper-resilient cryptography and connections to other pseudorandom objects in theoretical computer science.

We continue the line of investigation on explicit constructions of non-malleable codes in the information theoretic setting, and give explicit constructions for several new classes of tampering functions. These classes strictly generalize several previously studied classes of tampering functions, and in particular extend the well studied split-state model which is a "compartmentalized" model in the sense that the codeword is partitioned *a prior* into disjoint intervals for tampering. Specifically, we give explicit non-malleable codes for the following classes of tampering functions.

- Interleaved split-state tampering: Here the codeword is partitioned in an unknown way by an adversary, and then tampered with by a split-state tampering function.
- Linear function composed with split-state tampering: In this model, the codeword is first tampered with by a split-state adversary, and then the whole tampered codeword is further tampered with by a linear function. In fact our results are stronger, and we can handle linear function composed with interleaved split-state tampering.
- Bounded communication split-state tampering: In this model, the two split-state tampering adversaries are allowed to participate in a communication protocol with a bounded communication budget.

Our results are the first explicit constructions of non-malleable codes in any of these tampering models. We derive all these results from explicit constructions of seedless non-malleable extractors, which we believe are of independent interest.

Using our techniques, we also give an improved seedless extractor for an unknown interleaving of two independent sources.

*Part of this work was done when the author was a postdoctoral researcher at the Institute for Advanced Study, Princeton, partially supported by NSF Grant CCF-1412958 and the Simons Foundation.

†Partially supported by NSF Grant CCF-1617713.

ISSN 1433-8092

1 Introduction

1.1 Non-malleable Codes

Non-malleable codes were introduced by Dziembowski, Pietrzak, and Wichs [DPW18] as an elegant relaxation and generalization of standard error correcting codes, where the motivation is to handle much larger classes of tampering functions on the codeword. Traditionally, error correcting codes only provide meaningful guarantees (e.g., unique decoding or list-decoding) when *part* of the codeword is modified (i.e., the modified codeword is close in Hamming distance to an actual codeword), whereas in practice an adversary can possibly use much more complicated functions to modify the entire codeword. In the latter case, it is easy to see that error correction or even error detection becomes generally impossible, for example an adversary can simply change all codewords into a fixed string. On the other hand, non-malleable codes can still provide useful guarantees here, and thus partially bridge this gap. Informally, a non-malleable code guarantees that after tampering, the decoding either correctly gives the original message or gives a message that is completely unrelated and independent of the original message. This captures the notion of non-malleability: that an adversary cannot modify the codeword in a way such that the tampered codeword decodes back to a related but different message.

The original intended application of non-malleable codes is in tamper-resilient cryptography [DPW18], where they can be used generally to prevent an adversary from learning secret information by observing the input/output behavior of modified ciphertexts. Subsequently, non-malleable codes have found applications in non-malleable commitments [GPR16], public-key encryptions [CMTV15], non-malleable secret sharing schemes [GK18a, GK18b], and privacy amplification protocols [CKOS18]. Furthermore, interesting connections were found to non-malleable extractors [CG14b], and very recently to spectral expanders [RS18]. Along the way, the constructions of non-malleable codes used various components and sophisticated ideas from additive combinatorics [ADL14, CZ14] and randomness extraction [CGL16], and some of these techniques have also found applications in constructing extractors for independent sources [Li17]. Until today, non-malleable codes have become fundamental objects at the intersection of coding theory and cryptography. They are well deserved to be studied in more depth in their own right, as well as to find more connections to other well studied objects in theoretical computer science.

We first introduce some notation before formally defining non-malleable codes.

Definition 1.1. *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 $s \in S$.*

Definition 1.2 (Tampering functions)**.** *For any $n > 0$, let $\mathcal{F}_n$ denote the set of all functions $f : \{0,1\}^n \to \{0,1\}^n$. Any subset of $\mathcal{F}_n$ is a family of tampering functions.*

We use the statistical distance to measure the distance between distributions.

Definition 1.3. *The statistical distance between two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ over some universal set Ω is defined as $|\mathcal{D}_1 - \mathcal{D}_2| = \frac{1}{2}\sum_{d\in\Omega} |\mathbf{Pr}[\mathcal{D}_1 = d] - \mathbf{Pr}[\mathcal{D}_2 = d]|$. We say $\mathcal{D}_1$ is ϵ-close to $\mathcal{D}_2$ if $|\mathcal{D}_1 - \mathcal{D}_2| \leq \epsilon$ and denote it by $\mathcal{D}_1 \approx_\epsilon \mathcal{D}_2$.*

To handle fixed points, we need to define the following function.

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

Following the treatment in [DPW18], we first define coding schemes.

Definition 1.4 (Coding schemes)**.** *Let* $\mathrm{Enc} : \{0,1\}^k \to \{0,1\}^n$ *and* $\mathrm{Dec} : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$ *be functions such that* Enc *is a randomized function (i.e., it has access to 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* $s \in \{0,1\}^k$, $\Pr[\mathrm{Dec}(Enc(s)) = s] = 1$, *where the probability is taken over the randomness in* Enc.

We can now define non-malleable codes.

Definition 1.5 (Non-malleable codes)**.** *A coding scheme* $\mathcal{C} = (\mathrm{Enc}, \mathrm{Dec})$ *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*

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

We say the code is explicit if both the encoding and decoding can be done in polynomial time. The rate of $\mathcal{C}$ *is given by* k/n.

Relevant prior work on non-malleable codes. There has been a lot of exciting research on non-malleable codes, and we do not even attempt to provide a comprehensive survey of them. Instead we focus on relevant explicit constructions in the information theoretic setting, which is also the focus of this paper. One of the most studied classes of tampering functions is the so called *split-state* tampering, where the codeword is divided into (at least two) disjoint intervals and the adversary can tamper with each interval arbitrarily but independently. This model arises naturally in situations where the codeword may be stored in different parts of memory or different devices. Following a very successful line of work [DKO13,ADL14,CG14b,CZ14,ADKO15,CGL16,Li17,Li18], we now have explicit constructions of non-malleable codes in the 2-split state model with constant rate and constant error, or rate $\Omega(\log\log n/\log n)$ with exponentially small error [Li18]. For larger number of states, recent work of Kanukurthi, Obbattu, and Sruthi [KOS17], and that of Gupta, Maji and Wang [GMW18] gave explicit constructions in the 4-split-state model and 3-split-state model respectively, with constant rate and negligible error.

The split state model is a "compartmentalized" model, where the codeword is partitioned *a prior* into disjoint intervals for tampering. Recently, there has been progress towards handling non-compartmentalized tampering functions. A work of Agrawal, Gupta, Maji, Pandey and Prabhakaran [AGM+15] gave explicit constructions of non-malleable codes with respect to tampering functions that permute or flip the bits of the codeword. Ball, Dachman-Soled, Kulkarni and Malkin [BDKM16] gave explicit constructions of non-malleable codes against t-local functions for $t \leq n^{1-\epsilon}$. However in all these models, each bit of the tampering function only depends on part of the codeword. A recent work of Chattopadhyay and Li [CL17] gave the first explicit constructions of non-malleable codes where each bit of the tampering function may depend on all bits of the codeword. Specifically, they gave constructions for the classes of linear functions and small-depth (unbounded fain-in) circuits. The rate of the non-malleable code with respect to small-depth circuits was exponentially improved by a subsequent work of Ball, Dachman-Soled, Guo, Malkin, and Tan [BDG+18].

Given all these exciting results, a major goal of the research on non-malleable codes remains to give explicit constructions for broader classes of tampering functions, as one can use the probabilistic method to show the existence of non-malleable codes with rate close to $1 - \delta$ for any class $\mathcal{F}$ of tampering functions with $|\mathcal{F}| \leq 2^{2^{\delta n}}$ [CG14a].

Our results. We continue the line of investigation on explicit constructions of non-malleable codes, and give explicit constructions for several new classes of non-compartmentalized tampering functions, where in some classes each bit of the tampering function can depend on all the bits of the codeword. The new classes strictly generalize several previous studied classes of tampering functions. In particular, we consider the following three classes.

1. *Interleaved* 2*-split-state tampering*, where the adversary can divide the codeword into two arbitrary disjoint intervals and tamper with each interval arbitrarily but independently. This model generalizes the split-state model and captures the situation where the codeword is partitioned into two halves in an unknown way by the adversary before applying a 2-split-state tampering function. Constructing non-malleable codes for this class of tampering functions was left as an open problem by Cheraghchi and Guruswami [CG14b].

2. *Composition of tampering*, where the adversary takes two tampering functions and compose them together to get a new tampering function. We note that function composition is a natural strategy for an adversary to achieve more powerful tampering, and it has been studied widely in other fields (e.g., computational complexity and communication complexity). Thus we believe that studying non-malleable codes for the composition of known classes of tampering functions is also a natural and important direction.

3. *Bounded communication* 2*-split-state tampering*, where the two tampering functions in a 2-split state model are allowed to have some bounded communication.

We now formally define these classes and some related classes below. We use the notation that for any permutation $\pi : [n] \to [n]$ and any string $x \in [r]^n$, $y = x_\pi$ denotes the length n string such that $y_{\pi(i)} = x_i$.

- The family of 2-split-state functions $2\text{SS} \subset \mathcal{F}_{2n}$: Any $f \in 2\text{SS}$ comprises of two functions $f_1 : \{0,1\}^n \to \{0,1\}^n$ and $f_2 : \{0,1\}^n \to \{0,1\}^n$, and for any $x, y \in \{0,1\}^n$, $f(x,y) = (f_1(x), f_2(x))$. This family of tampering functions has been extensively studied, with a long line of work achieving near optimal explicit constructions of non-malleable codes.

- The family of linear functions $\text{Lin} \subset \mathcal{F}_n$: Any $f \in \text{Lin}$ is a linear function from $\{0,1\}^n$ to $\{0,1\}^n$ (viewing $\{0,1\}^n$ as $\mathbb{F}_2^n$).

- The family of interleaved 2-split-state functions $2\text{ISS} \subset \mathcal{F}_{2n}$: Any $f \in 2\text{ISS}$ comprises of two functions $f_1 : \{0,1\}^n \to \{0,1\}^n$, $f_2 : \{0,1\}^n \to \{0,1\}^n$, and a permutation $\pi : [2n] \to [2n]$. For any $z = (x \circ y)_\pi \in \{0,1\}^{2n}$, where $x, y \in \{0,1\}^n$, let $f(z) = (f_1(x) \circ f_2(y))_\pi$ (where $\circ$ denotes the string concatenation operation).

- The family of bounded communication 2-split-state functions $(2,t) - \text{CSS}$: Consider the following natural extension of the 2-split-state model. Let $c = (x, y)$ be a codeword in $\{0,1\}^{2n}$, where x is the first n bits of c and y is the remaining n bits of c. Let Alice and Bob be two tampering adversaries, where Alice has access to x and Bob has access to y. Alice and Bob run a (deterministic) communication protocol based on x and y respectively, which can last for an arbitrary number of rounds but each party sends at most t bits in total. Finally, based on the transcript and x Alice outputs $x' \in \{0,1\}^n$, similarly based on the transcript and y Bob outputs $y' \in \{0,1\}^n$. The tampered codeword is $c' = (x', y')$.

- For any tampering function families $\mathcal{F}, \mathcal{G} \subset \mathcal{F}_n$, define the family $\mathcal{F} \circ \mathcal{G} \subset \mathcal{F}_n$ to be the set of all functions of the form $f \circ g$, where $f \in \mathcal{F}$, $g \in \mathcal{G}$ and $\circ$ denotes function composition.

We now formally state our results. Our main result is an explicit non-malleable code with respect to the tampering class of $\mathrm{Lin} \circ \mathrm{2ISS}$, i.e, linear function composed with interleaved 2-split-state tampering. Specifically, we have the following theorem.

Theorem 1. *There exists a constant $\delta > 0$ such that for all integers $n > 0$ there exists an explicit non-malleable code with respect to* $\mathrm{Lin} \circ \mathrm{2ISS}$ *with rate $1/n^{\delta}$ and error $2^{-n^{\delta}}$.*

This immediately gives the following corollaries, which give explicit non-malleable codes for interleaved 2-split-state tampering, and linear function composed with 2-split-state tampering.

Corollary 2. *There exists a constant $\delta > 0$ such that for all integers $n > 0$ there exists an explicit non-malleable code with respect to* $\mathrm{2ISS}$ *with rate $1/n^{\delta}$ and error $2^{-n^{\delta}}$.*

Corollary 3. *There exists a constant $\delta > 0$ such that for all integers $n > 0$ there exists an explicit non-malleable code with respect to* $\mathrm{Lin} \circ \mathrm{2SS}$ *with rate $1/n^{\delta}$ and error $2^{-n^{\delta}}$.*

Next we give an explicit non-malleable code with respect to bounded communication 2-split-state tampering.

Theorem 4. *There exists a constant $\delta > 0$ such that for all integers $n, t > 0$ with $t \leq \delta n$, there exists an explicit non-malleable code with respect to $(2, t) - \mathrm{CSS}$ with rate $\Omega(\log\log n/\log n)$ and error $2^{-\Omega(n \log\log n/\log n)}$.*

Prior to our work, no explicit non-malleable code of any rate was known for these tampering classes.

1.2 Seedless non-malleable extractors

Our results on non-malleable codes are based on new constructions of seedless non-malleable extractors, which we believe are of independent interest. Before defining seedless non-malleable extractors formally, we first recall some basic notation from the area of randomness extraction.

Randomness extraction is motivated by the problem of purifying imperfect (or defective) sources of randomness. The concern stems from the fact that natural random sources often have poor quality, while most applications require high quality (e.g., uniform) random bits. We use the standard notion of min-entropy to measure the amount of randomness in a distribution.

Definition 1.6. *The min-entropy $H_{\infty}(\mathbf{X})$ of a probability distribution $\mathbf{X}$ is defined to be $\min_x(-\log(\Pr[\mathbf{X} = x]))$. We say a probability distribution $\mathbf{X}$ on $\{0,1\}^n$ is an $(n, H_{\infty}(\mathbf{X}))$-source and the min-entropy rate is $H_{\infty}(\mathbf{X})/n$.*

It turns out that it is impossible to extract from a single general weak random source even for min-entropy $n - 1$. There are two possible ways to bypass this barrier. The first one is to relax the extractor to be a *seeded extractor*, which takes an additional independent short random seed to extract from a weak random source. The second one is to construct deterministic extractors for special classes of weak random sources.

Both kinds of extractors have been studied extensively. Recently, they have also been generalized to stronger notions where the inputs to the extractor can be tampered with by an adversary. Specifically, Dodis and Wichs [DW09] introduced the notion of *seeded non-malleable extractor* in the context of privacy amplification against an active adversary. Informally, such an extractor satisfies the stronger property that the output of the extractor is independent of the output of the extractor on a tampered seed. Similarly, and more relevant to this paper, a seedless variant

of non-malleable extractors was introduced by Cheraghchi and Guruswami [CG14b] as a way to construct non-malleable codes. Apart from their original applications, both kinds of non-malleable extractors are of independent interest. They are also related to each other and have applications in constructions of extractors for independent sources [Li17].

We now define seedless non-malleable extractors. For simplicity, the definition here assumes that the tampering function has no fixed points. See Section 3 for a more formal definition.

Definition 1.7 (Seedless non-malleable extractors)**.** *Let $\mathcal{F} \subset \mathcal{F}_n$ be a family of tampering functions such that no function in $\mathcal{F}$ has any fixed points. A function* $\mathrm{nmExt} : \{0,1\}^n \to \{0,1\}^m$ *is a seedless (n, m, ϵ)-non-malleable extractor with respect to $\mathcal{F}$ and a class of sources $\mathcal{X}$ if for every distribution $\mathbf{X} \in \mathcal{X}$ and every tampering function $f \in \mathcal{F}$,*

$$|\mathrm{nmExt}(\mathbf{X}), \mathrm{nmExt}(f(\mathbf{X})) - \mathbf{U}_m, \mathrm{nmExt}(f(\mathbf{X}))| \leq \epsilon.$$

Further, we say that nmExt *is ϵ'-invertible, if there exists a polynomial time sampling algorithm $\mathcal{A}$ that takes as input $y \in \{0,1\}^m$, and outputs a sample from a distribution that is ϵ'-close to the uniform distribution on the set* $\mathrm{nmExt}^{-1}(y)$.

In the above definition, when the class of sources $\mathcal{X}$ is the distribution $\mathbf{U}_n$, we simply say that nmExt is a seedless (n, m, ϵ)-non-malleable extractor with respect to $\mathcal{F}$.

Relevant prior work on seedless non-malleable extractors. The first construction of seedless non-malleable extractors was given by Chattopadhyay and Zuckerman [CZ14] with respect to the class of 10-split-state tampering. Subsequently, a series of works starting with the work of Chattopadhyay, Goyal and Li [CGL16] gave explicit seedless non-malleable extractors for 2-split-state tampering. The only known construction with respect to a class of tampering functions different from split state tampering is the work of Chattopadhyay and Li [CL17], which gave explicit seedless non-malleable extractors with respect to the tampering class Lin and small depth circuits. We note that constructing explicit seedless non-malleable extractors with respect to 2ISS was also posed as an open problem in [CG14b].

Our results. We give the first explicit constructions of seedless non-malleable extractors with respect to the tampering classes $\mathrm{Lin} \circ \mathrm{2ISS}$ and $(2, t) - \mathrm{CSS}$. Note that the first construction also directly implies non-malleable extractors with respect to the classes 2ISS and $\mathrm{Lin} \circ \mathrm{2SS}$. The non-malleable extractors with respect to $\mathrm{Lin} \circ \mathrm{2ISS}$ is a fundamentally new construction. The non-malleable extractor with respect to $(2, t) - \mathrm{CSS}$ is obtained by showing a reduction to seedless non-malleable extractors for 2SS, where excellent constructions are known (e.g., a recent construction of Li [Li18]).

We now formally state our main results.

Theorem 5. *For all $n > 0$ there exists an efficiently computable seedless $(n, n^{\Omega(1)}, 2^{-n^{\Omega(1)}})$-non-malleable extractor with respect to* $\mathrm{Lin} \circ \mathrm{2ISS}$*, that is $2^{-n^{\Omega(1)}}$-invertible.*

This immediately gives the following two corollaries.

Corollary 6. *For all $n > 0$ there exists an efficiently computable seedless $(n, n^{\Omega(1)}, 2^{-n^{\Omega(1)}})$-non-malleable extractor with respect to* 2ISS*, that is $2^{-n^{\Omega(1)}}$-invertible.*

Corollary 7. *For all $n > 0$ there exists an efficiently computable seedless $(n, n^{\Omega(1)}, 2^{-n^{\Omega(1)}})$-non-malleable extractor with respect to* $\mathrm{Lin} \circ \mathrm{2SS}$*, that is $2^{-n^{\Omega(1)}}$-invertible.*

Next we give the non-malleable extractor with respect to $(2,t)-\mathrm{CSS}$.

Theorem 8. *There exists a constant $\delta > 0$ such for all integers $n, t > 0$ with $t \leq \delta n$, there exists an efficiently computable seedless $(n, \Omega\left(\frac{n\cdot(\log\log n)}{\log n}\right), 2^{-\Omega(n\log\log n/\log n)})$-non-malleable extractor with respect to $(2,t)-\mathrm{CSS}$, that is $2^{-\Omega(n\log\log n/\log n)}$-invertible.*

We derive our results on non-malleable codes using the above explicit constructions of non-malleable extractors. In particular we use the following theorem proved by Cheraghchi and Guruswami [CG14b] that connects non-malleable extractors and codes.

Theorem 1.8 ([CG14b])**.** *Let* $\mathrm{nmExt} : \{0,1\}^n \to \{0,1\}^m$ *be an efficient seedless (n, m, ϵ)-non-malleable extractor with respect to a class of tampering functions $\mathcal{F}$ acting on $\{0,1\}^n$. Further suppose* nmExt *is ϵ'-invertible.*

Then there exists an efficient construction of a non-malleable code with respect to the tampering family $\mathcal{F}$ with block length $= n$, relative rate $\frac{m}{n}$ and error $2^m\epsilon + \epsilon'$.

1.3 Extractors for interleaved sources

Our techniques also yield improved explicit constructions of extractors for interleaved sources, which generalize extractors for independent sources in the following way: the inputs to the extractor are samples from a few independent sources mixed (interleaved) in an unknown (but fixed) way. Raz and Yehudayoff [RY11] showed that such extractors have applications in communication complexity and proving lower bounds for arithmetic circuits. In a subsequent work, Chattopadhyay and Zuckerman [CZ16b] showed that such extractors can also be used to construct extractors for certain samplable sources, extending a line of work initiated by Trevisan and Vadhan [TV00]. We now define interleaved sources formally.

Definition 1.9 (Interleaved Sources)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_r$ be arbitrary independent sources on $\{0,1\}^n$ and let $\pi : [rn] \to [rn]$ be any permutation. Then $Z = (\mathbf{X}_1 \circ \ldots \circ \mathbf{X}_r)_\pi$ is an r-interleaved source.*

Relevant prior work on interleaved extractors. Raz and Yehudayoff [RY11] gave explicit extractors for 2-interleaved sources when both the sources have min-entropy at least $(1-\delta)n$ for a tiny constant $\delta > 0$. Their construction is based on techniques from additive combinatorics and can output $\Omega(n)$ bits with exponentially small error. Subsequently, Chattopadhyay and Zuckerman [CZ16b] constructed extractors for 2-interleaved sources where one source has entropy $(1-\gamma)n$ for a small constant $\gamma > 0$ and the other source has entropy $\Omega(\log n)$. They achieve output length $O(\log n)$ bits with error $n^{-\Omega(1)}$.

A much better result (in terms of the min-entropy) is known if the extractor has access to an interleaving of more sources. For a large enough constant C, Chattopadhyay and Li [CL16] gave an explicit extractor for C-interleaved sources where each source has entropy $k \geq \mathrm{poly}(\log n)$. They achieve output length $k^{\Omega(1)}$ and error $n^{-\Omega(1)}$.

Our results. Our main result is an explicit extractor for 2-interleaved sources where each source has min-entropy at least $2n/3$. The extractor outputs $\Omega(n)$ bits with error $2^{-n^{\Omega(1)}}$.

Theorem 9. *For any constant $\delta > 0$ and all integers $n > 0$, there exists an efficiently computable function* $\mathrm{i}\ell\mathrm{Ext} : \{0,1\}^{2n} \to \{0,1\}^m$*, $m = \Omega(n)$, such that for any two independent sources $\mathbf{X}$ and $\mathbf{Y}$, each on n bits with min-entropy at least $(2/3+\delta)n$, and any permutation $\pi : [2n] \to [2n]$,*

$$|\mathrm{i}\ell\mathrm{Ext}((\mathbf{X} \circ \mathbf{Y})_\pi) - \mathbf{U}_m| \leq 2^{-n^{\Omega(1)}}.$$

1.4 Open questions

Non-malleable codes for composition of function classes We gave efficient constructions of non-malleable codes for the tampering class $\mathrm{Lin} \circ \mathrm{2SS}$ (and more generally $\mathrm{Lin} \circ \mathrm{ISS}$). Many natural questions remain to be answered. For instance, one open problem is to efficiently construct non-malleable codes for the tampering class $\mathrm{2SS} \circ \mathrm{Lin}$. It looks like one needs substantially new ideas to give such constructions. More generally, for what other interesting classes of functins $\mathcal{F}$ and $\mathcal{G}$ can we construct non-malleable codes for the composed class $\mathcal{F} \circ \mathcal{G}$? Is it possible to efficiently construct non-malleable codes for the tampering class $\mathcal{F} \circ \mathcal{G}$ if we have efficient non-malleable codes for the classes $\mathcal{F}$ and $\mathcal{G}$?

Other applications for seedless non-malleable extractors The explicit seedless non-malleable extractors that we construct satisfy strong pseudorandom properties. A natural question is to find more applications of these non-malleable extractors in explicit constructions of other interesting objects.

Improved seedless extractors We construct an extractor for 2-interleaved sources that works for min-entropy rate $2/3$. It is easy to verify that there exists extractors for sources with min-entropy as low as $C \log n$, and a natural question here is to come up with such explicit constructions. Given the success in constructing 2-source extractors for low min-entropy [CZ16a, Li18], we are hopeful that more progress can be made on this problem.

1.5 Organization

The rest of the paper is organized as follows. We use Section 2 to present an overview of our results and techniques. We use Section 3 to introduce some background and notation. We present our seedless non-malleable extractor construction with respect to $\mathrm{Lin} \circ \mathrm{2ISS}$ in Section 4. We use Section 5 to present our non-malleable extractor construction with respect to $(2, t) - \mathrm{CSS}$. We present efficient sampling algorithms for our seedless non-malleable extractor constructions in Section 6. We use Section 7 to present an explicit construction of an extractor for interleaved sources.

2 Overview of constructions and techniques

Our results on non-malleable codes are derived from explicit constructions of invertible seedless non-malleable extractors (see Theorem 1.8). In this section, we focus on explicit constructions of seedless non-malleable extractors with respect to the relevant classes of tampering functions, and explicit extractors for interleaved sources.

Seedless non-malleable extractors with respect to $\mathrm{Lin} \circ \mathrm{2ISS}$. We construct a seedless non-malleable extractor $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, $m = n^{\Omega(1)}$ such that the following hold: Let $\mathbf{X}$ and $\mathbf{Y}$ be two independent uniform sources, each on n bits. Let $h : \{0,1\}^{2n} \to \{0,1\}^{2n}$ be an arbitrary linear function, $f : \{0,1\}^n \to \{0,1\}^n$, $g : \{0,1\}^n \to \{0,1\}^n$ be two arbitrary functions, and $\pi : [2n] \to [2n]$ be an arbitrary permutation. Then,

$$\mathrm{nmExt}((\mathbf{X}, \mathbf{Y})_\pi), \mathrm{nmExt}(h((f(\mathbf{X}) \circ g(\mathbf{Y}))_\pi)) \approx_\epsilon \mathbf{U}_m, \mathrm{nmExt}(h((f(\mathbf{X}) \circ g(\mathbf{Y}))_\pi)),$$

where $\epsilon = 2^{-n^{\Omega(1)}}$. Notice that such an extractor is not possible to construct in general, for example when all f, g, h are the identify function. However, such an extractor exists when the composed function does not have fixed points. For simplicity, we ignore this issue related to fixed points in the proof sketch, and just mention that we have a reduction from the problem of constructing non-malleable codes to the problem of constructing a non-malleable extractor with no fixed points. The argument is similar to the argument in [CG14b] but more complicated since here we are dealing with more powerful adversaries. We refer the reader to Section 4 for more details.

Our first step is to reduce the problem of constructing non-malleable codes with respect to Lin $\circ$ 2ISS to constructing non-malleable extractors with the following guarantee. For strings x, $y \in \{0,1\}^n$, we use $x + y$ (or equivalently $x - y$) to denote the bit-wise xor of the two strings. Let $\mathbf{X}$ and $\mathbf{Y}$ to be two independent $(n, n - n^\delta)$-sources and $f_1, f_2, g_1, g_2 \in \mathcal{F}_n$ be four functions that satisfy the following condition:

- $\forall x \in \mathit{support}(\mathbf{X})$ and $y \in \mathit{support}(\mathbf{Y})$, $f_1(x) + g_1(y) \neq x$ or
- $\forall x \in \mathit{support}(\mathbf{X})$ and $y \in \mathit{support}(\mathbf{Y})$, $f_2(x) + g_2(y) \neq y$.

Then,

$$|\mathrm{nmExt}((\mathbf{X}, \mathbf{Y})_\pi), \mathrm{nmExt}((f_1(\mathbf{X}) + g_1(\mathbf{Y}), f_2(\mathbf{X}) + g_2(\mathbf{Y}))_\pi) - \mathbf{U}_m, \mathrm{nmExt}((f_1(\mathbf{X}) + g_1(\mathbf{Y}), f_2(\mathbf{X}) + g_2(\mathbf{Y}))_\pi)| \leq 2^{-n^{\Omega(1)}}.$$

The reduction can be seen in the following way: Define $\overline{x} = (x, 0^n)_\pi$ and $\overline{y} = (0^n, y)_\pi$. Similarly define $\overline{f(x)} = h((f(x), 0^n)_\pi)$ and $\overline{g(y)} = h((0^n, g(y))_\pi)$. Thus, $(x, y)_\pi = \overline{x} + \overline{y}$ and $h((f(x), g(y))_\pi) = \overline{f(x)} + \overline{g(y)}$. Define functions $h_1 : \{0,1\}^{2n} \to \{0,1\}^n$ and $h_2 : \{0,1\}^{2n} \to \{0,1\}^n$ such that $h((f(x), g(y))_\pi) = (h_1(x, y), h_2(x, y))_\pi$. Since $h((f(x), g(y))_\pi) = \overline{f(x)} + \overline{g(y)}$, it follows that there exists functions $f_1, g_1, f_2, g_2 \in \mathcal{F}_n$ such that for all $x, y \in \{0,1\}^n$, the following hold:

- $h_1(x, y) = f_1(x) + g_1(y)$, and
- $h_2(x, y) = f_2(x) + g_2(y)$.

Thus, $h((f(x), g(y))_\pi) = ((f_1(x) + g_1(y)), (f_2(x) + g_2(y)))_\pi$. The loss of entropy in $\mathbf{X}$ and $\mathbf{Y}$ in the reduction (from n to $n - n^\delta$) is due to the fact that we have to handle issues related to fixed points of the tampering functions, and we ignore it for the proof sketch here.

The idea now is to use the framework of advice generators and correlation breakers with advice to construct the non-malleable extractor [Coh15, CGL16]. We informally define these objects below as we describe our explicit constructions.

We start with the construction of the advice generator advGen : $\{0,1\}^{2n} \to \{0,1\}^a$. Informally, advGen is a weaker object than a non-malleable extractor, and we only need that advGen$((\mathbf{X}, \mathbf{Y})_\pi) \neq$ advGen$((f_1(\mathbf{X}) + g_1(\mathbf{Y}), f_2(\mathbf{X}) + g_2(\mathbf{Y}))_\pi)$ (with high probability). Further, it is crucial that $a \ll n$, and in particular think of $a = n^\gamma$ for a small constant $\gamma > 0$. Without loss of generality, suppose that $\forall x \in \mathit{support}(\mathbf{X})$ and $y \in \mathit{support}(\mathbf{Y})$, $f_1(x) + g_1(y) \neq x$.

Let $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})_\pi$. Let $n_0 = n^{\delta'}$ for some small constant $\delta' > 0$. We take two slices from $\mathbf{Z}$, say $\mathbf{Z}_1$ and $\mathbf{Z}_2$ of lengths $n_1 = n_0^{c_0}$ and $n_2 = 10n_0$, for some constant $c_0 > 1$. Next, we use a good linear error correcting code (let the encoder of this code be E) to encode $\mathbf{Z}$ and sample n^γ coordinates (let $\mathbf{S}$ denote this set) from this encoding using $\mathbf{Z}_1$ (the sampler is based on seeded extractors [Zuc97]). Let $\mathbf{W}_1 = E(\mathbf{Z})_\mathbf{S}$. Next, using $\mathbf{Z}_2$, we sample a random set of indices $\mathbf{T} \subset [2n]$, and let $\mathbf{Z}_3 = \mathbf{Z}_\mathbf{T}$.

We now use an extractor for interleaved sources, i.e., an extractor that takes as input an unknown interleaving of two independent sources and outputs uniform bits (see Section 1.3). Let iℓExt be this extractor (say from Theorem 9), and we apply it to $\mathbf{Z}_3$ to get $\mathbf{R} = \text{i}\ell\text{Ext}(\mathbf{Z}_3)$. Finally, let $\mathbf{W}_2$ be the output of a linear seeded extractor[1] LExt on $\mathbf{Z}$ with $\mathbf{R}$ as the seed. The output of the advice generator is $\mathbf{Z}_1 \circ \mathbf{Z}_2 \circ \mathbf{Z}_3 \circ \mathbf{W}_1 \circ \mathbf{W}_2$.

The intuition that this works is as follows. We use the notation that if $\mathbf{W} = h((\mathbf{X}, \mathbf{Y})_\pi)$ (for some function h), then $\mathbf{W}'$ or $(\mathbf{W})'$ stands for the corresponding random variable after tampering, i.e., $h(((f_1(\mathbf{X}) + g_1(\mathbf{Y})), (f_2(\mathbf{X}) + g_2(\mathbf{Y})))_\pi)$. Further, let $\mathbf{X}_i$ be the bits of $\mathbf{X}$ in $\mathbf{Z}_i$ for $i = 1, 2, 3$ and $\mathbf{X}_4$ be the remaining bits of $\mathbf{X}$. Similarly define $\mathbf{Y}_i$'s, $i = 1, 2, 3, 4$. Without loss of generality suppose that $|\mathbf{X}_1| \geq |\mathbf{Y}_1|$, (where $|\alpha|$ denotes the length of the string α).

The correctness of advGen is direct if $\mathbf{Z}_i \neq \mathbf{Z}'_i$ for some $i \in \{1, 2, 3\}$. Thus, assume $\mathbf{Z}_i = \mathbf{Z}'_i$ for $i = 1, 2, 3$. It follows that hence $\mathbf{S} = \mathbf{S}', \mathbf{T} = \mathbf{T}'$ and $\mathbf{R} = \mathbf{R}'$. Recall that $(\mathbf{X}, \mathbf{Y})_\pi = \overline{\mathbf{X}} + \overline{\mathbf{Y}}$ and $h((f(\mathbf{X}), g(\mathbf{Y}))_\pi) = \overline{f(\mathbf{X})} + \overline{g(\mathbf{Y})}$. Since E is a linear code and LExt is a linear seeded extractor, the following hold:

$$\mathbf{W}_1 - \mathbf{W}'_1 = (E(\overline{\mathbf{X}} + \overline{\mathbf{Y}} - \overline{f(\mathbf{X})} - \overline{g(\mathbf{Y})}))_\mathbf{S},$$
$$\mathbf{W}_2 - \mathbf{W}'_2 = \text{LExt}(\overline{\mathbf{X}} + \overline{\mathbf{Y}} - \overline{f(\mathbf{X})} - \overline{g(\mathbf{Y})}, \mathbf{R}).$$

Now the idea is the following: Either (i) we can fix $\overline{\mathbf{X}} - \overline{f(\mathbf{X})}$ and claim that $\mathbf{X}_1$ still has enough min-entropy, or (ii) we can claim that $\overline{\mathbf{X}} - \overline{f(\mathbf{X})}$ has enough min-entropy conditioned on the fixing of $(\mathbf{X}_2, \mathbf{X}_3)$. Let us first discuss why this is enough. Suppose we are in the first case. Then, we can fix $\overline{\mathbf{X}} - \overline{f(\mathbf{X})}$ and $\mathbf{Y}$ and argue that $\mathbf{Z}_1$ is a deterministic function of $\mathbf{X}$ and contains enough entropy. Note that $\overline{\mathbf{X}} + \overline{\mathbf{Y}} - \overline{f(\mathbf{X})} - \overline{g(\mathbf{Y})}$ is now fixed, and in fact it is fixed to a non-zero string (using the assumption that $f_1(x) + g_1(y) \neq x$). Thus, $E(\overline{\mathbf{X}} + \overline{\mathbf{Y}} - \overline{f(\mathbf{X})} - \overline{g(\mathbf{Y})})$ is a string with a constant fraction of the coordinates set to 1 (since E is an encoder of a linear error correcting code with constant relative distance), and it follows that with high probability $(E(\overline{\mathbf{X}} + \overline{\mathbf{Y}} - \overline{f(\mathbf{X})} - \overline{g(\mathbf{Y})}))_\mathbf{S}$ contains a non-zero entry (using the fact that $\mathbf{S}$ is sampled using $\mathbf{Z}_1$, which has enough entropy). This finishes the proof in this case since it implies $\mathbf{W}_1 \neq \mathbf{W}'_1$ with high probability.

Now suppose we are in case (ii). We use the fact that $\mathbf{Z}_2$ contains entropy to conclude that the sampled bits $\mathbf{Z}_3$ contain almost equal number of bits from $\mathbf{X}$ and $\mathbf{Y}$ (with high probability over $\mathbf{Z}_2$). Now we can fix $\mathbf{Z}_2$ without loosing too much entropy from $\mathbf{Z}_3$ (by making the size of $\mathbf{Z}_3$ to be significantly larger than $\mathbf{Z}_2$). Next, we observe that $\mathbf{Z}_3$ is an interleaved source, and hence $\mathbf{R}$ is close to uniform. We now fix $\mathbf{X}_3$, and argue that $\mathbf{R}$ continues to be uniform. This follows roughly from the fact that any 2-source extractor is strong [Rao07], which easily extends to extractors for 2 interleaved sources. Thus, $\mathbf{R}$ now becomes a deterministic function of $\mathbf{Y}$ while at the same time, $\overline{\mathbf{X}} - \overline{f(\mathbf{X})}$ still has enough min-entropy. Hence, $\text{LExt}(\overline{\mathbf{X}} - \overline{f(\mathbf{X})}, \mathbf{R})$ is close to uniform even conditioned on $\mathbf{R}$. We can now fix $\mathbf{R}$ and $\text{LExt}(\overline{\mathbf{Y}} - \overline{g(\mathbf{Y})}, \mathbf{R})$ without affecting the distribution $\text{LExt}(\overline{\mathbf{X}} - \overline{f(\mathbf{X})}, \mathbf{R})$, since $\text{LExt}(\overline{\mathbf{Y}} - \overline{g(\mathbf{Y})}, \mathbf{R})$ is a deterministic function of $\mathbf{Y}$ while $\text{LExt}(\overline{\mathbf{X}} - \overline{f(\mathbf{X})}, \mathbf{R})$ is a deterministic function of $\mathbf{X}$ conditioned on the previous fixing of $\mathbf{R}$. It follows that after these fixings, $\mathbf{W}_2 - \mathbf{W}'_2$ is close to a uniform string and hence $\mathbf{W}_2 - \mathbf{W}'_2 \neq 0$ with probability $1 - 2^{-n^{\Omega(1)}}$, which completes the proof.

The fact that we can only consider case (i) and case (ii) relies on a careful convex combination argument, which is in turn based on the pre-image size of the function $\tau : \{0,1\}^n \to \{0,1\}^{2n}$ defined as $\tau(x) = (x, 0^n)_\pi - h((f(x), 0^n)_\pi) = \overline{x} - \overline{f(x)}$. The intuition is as follows. If conditioned on the

[1] A linear seeded extractor is a seeded extractor where for any fixing of the seed, the output is a linear function of the source.

fixing of $\tau(\mathbf{X}) = \overline{\mathbf{X}} - \overline{f(\mathbf{X})}$ we have that $\mathbf{X}$ still has very high min-entropy, then we can take the slice $\mathbf{Z}_1$ such that $\mathbf{X}_1$ still has enough entropy conditioned on the fixing of $\tau(\mathbf{X})$. On the other hand, if conditioned on the fixing of $\tau(\mathbf{X})$ we have that $\mathbf{X}$ does not have high min-entropy, then $\tau(\mathbf{X})$ itself must have a large support size (or relatively high entropy). Therefore we can take the slice $\mathbf{Z}_2$ and sample a short string $\mathbf{Z}_3$ such that conditioned on the fixing of $(\mathbf{X}_2, \mathbf{X}_3)$, $\tau(\mathbf{X})$ still has enough min-entropy. To make the whole argument work, we need to carefully choose the sizes of the three slices $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$. In particular, we need to ensure that the size of $(\mathbf{Z}_2, \mathbf{Z}_3)$ is much smaller than that of $\mathbf{Z}_1$.

We now discuss the other crucial component in the construction, the advice correlation breaker $\mathrm{ACB} : \{0,1\}^{2n} \times \{0,1\}^a \to \{0,1\}^m$. Informally, ACB takes 2 inputs, a source $\mathbf{Z}$ (that contains some min-entropy) and an advice string $s \in \{0,1\}^a$, and outputs a distribution on $\{0,1\}^m$ with the following guarantee. If $\mathbf{Z}'$ is the distribution of $\mathbf{Z}$ after tampering, and $s' \in \{0,1\}^a$ is another advice such that $s \neq s'$, then $\mathrm{ACB}(\mathbf{Z}, s), \mathrm{ACB}(\mathbf{Z}', s') \approx \mathbf{U}_m, \mathrm{ACB}(\mathbf{Z}', s')$. Typically, we also assume some structures in $\mathbf{Z}$ (e.g., it consists of two independent sources or an interleaving of two independent sources). Our main result is an advice correlation breaker that satisfies

$$\mathrm{ACB}(\overline{\mathbf{X}} + \overline{\mathbf{Y}}, s), \mathrm{ACB}(\overline{f(\mathbf{X})} + \overline{g(\mathbf{Y})}, s') \approx_\epsilon \mathbf{U}_m, \mathrm{ACB}(\overline{f(\mathbf{X})} + \overline{g(\mathbf{Y})}, s'),$$

for any fixed strings $s, s' \in \{0,1\}^a$ where $s \neq s'$. We note that correlation breakers have found important applications in explicit constructions of seedless extractors [Coh15, Li16, Li18], thus we believe the above correlation breaker can be of independent interest and potentially find other applications. By composing the advice generator advGen and the correlation breaker ACB in the natural way, we get the non-malleable extractor. Here we only briefly mention that the above advice correlation breaker crucially exploits the "sum-structure" of the source and the tampering function, the fact that extractors are samplers [Zuc97], and previous constructions of correlation breakers using linear seeded extractors [CL17]. We refer the reader to Section 4.2 for more details.

Finally, it is far from obvious how to efficiently invert the extractor, or in other words, sample from the pre-image of this non-malleable extractor. This is important since the encoder of the corresponding non-malleable code is doing exactly the sampling and thus we need it to be efficient. We use Section 6 to suitably modify our extractor to support efficient sampling. Here we briefly sketch some high level ideas involved. Recall $\mathbf{Z} = (\mathbf{X} \circ \mathbf{Y})_\pi$. The first modification is that in all applications of seeded extractors in our construction, we specifically use linear seeded extractors. This allows us to argue that the pre-image we are trying to sample from is in fact a convex combination of distributions supported on subspaces. The next crucial observation is the fact that we can use smaller disjoint slices of $\mathbf{Z}$ to carry out various steps outlined in the construction. This is to ensure that the dimensions of the subspaces that we need to sample from, do not depend on the values of random variables that we fix. For the steps where we use the entire source $\mathbf{Z}$ (in the construction of the advice correlation breaker), we replace $\mathbf{Z}$ by a large enough slice of $\mathbf{Z}$. However this is problematic if we choose the slice deterministically, since in an arbitrary interleaving of two sources, a slice of length less than n might have bits only from one source. We get around this by pseudorandomly sampling enough coordinates from $\mathbf{Z}$ (by first taking a small slice of $\mathbf{Z}$ and using a sampler that works for weak sources [Zuc97]).

We now use an elegant trick introduced by Li [Li17] where the output of the non-malleable extractor described above (with the modifications that we have specified) is now used as a seed to a linear seeded extractor applied on an even larger pseudorandom slice of $\mathbf{Z}$. The linear seeded extractor that we use has the property that for any fixing of the seed, the rank of the linear map corresponding to the extractor is the same, and furthermore one can efficiently sample from the pre-image of any output of the extractor. The final modification needed is a careful choice of the

error correcting code used in the advice generator. For this we use a dual BCH code, which allows us to argue that we can discard some output bits of the advice generator without affecting its correctness (based on the dual distance of the code). This is crucial in order to argue that the rank of the linear restriction imposed on the free variables of $\mathbf{Z}$ does not depend on the values of the bits fixed so far. We refer the reader to Section 6 for more details.

Non-malleable extractors for $(2,t)-$CSS. We show that any 2-source non-malleable extractor that works for min-entropy $n-2\delta n$ can be used as non-malleable extractor with respect to $(2, t)-$CSS for $t \leq \delta n$. The tampering function h that is based on the communication protocol can be rephrased in terms of functions in the following way. Suppose the protocol lasts for ℓ rounds, there exist deterministic functions f_i and g_i for $i = 1, \ldots, \ell$, and $f : \{0,1\}^n \times \{0,1\}^{2t} \to \{0,1\}^n$ and $g : \{0,1\}^n \times \{0,1\}^{2t} \to \{0,1\}^n$ such that the communication protocol between Alice and Bob corresponds to computing the following random variables: $\mathbf{S}_1 = f_1(\mathbf{X}), \mathbf{R}_1 = g_1(\mathbf{Y}, \mathbf{S}_1), \mathbf{S}_2 = f_2(\mathbf{X}, \mathbf{S}_1, \mathbf{R}_1), \ldots, \mathbf{S}_i = f_i(\mathbf{X}, \mathbf{S}_1, \ldots, \mathbf{S}_{i-1}, \mathbf{R}_1, \ldots, \mathbf{R}_{i-1}), \mathbf{R}_i = g_i(\mathbf{Y}, \mathbf{S}_1, \ldots, \mathbf{S}_i, \mathbf{R}_{i,}, \ldots, \mathbf{R}_{i-1}), \ldots, \mathbf{R}_\ell = g_\ell(\mathbf{Y}, \mathbf{S}_1, \ldots, \mathbf{S}_\ell, \mathbf{R}_1, \ldots, \mathbf{R}_{\ell-1})$.

Finally, $\mathbf{X}' = f(\mathbf{X}, \mathbf{R}_1, \ldots, \mathbf{R}_\ell, \mathbf{S}_1, \ldots, \mathbf{S}_\ell)$ and $\mathbf{Y}' = g(\mathbf{Y}, \mathbf{R}_1, \ldots, \mathbf{R}_\ell, \mathbf{S}_1, \ldots, \mathbf{S}_\ell)$ correspond to the output of Alice and the output of Bob respectively. Thus, $h(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}', \mathbf{Y}')$.

Similar to the way we argue about alternating extraction protocols, we fix random variables in the following order: Fix $\mathbf{S}_1$, and it follows that $\mathbf{R}_1$ is now a deterministic function of $\mathbf{Y}$. We fix $\mathbf{R}_1$, and thus $\mathbf{S}_2$ is now a deterministic function of $\mathbf{X}$. Thus, continuing in this way, we can fix all the random variables $\mathbf{S}_1, \ldots, \mathbf{S}_\ell$ and $\mathbf{R}_1, \ldots, \mathbf{R}_\ell$ while maintaining that $\mathbf{X}$ and $\mathbf{Y}$ are independent. Further, invoking Lemma 3.1, with probability at least $1-2^{-\Omega(n)}$, both $\mathbf{X}$ and $\mathbf{Y}$ have min-entropy at least $n-t-\delta n \geq n-2\delta n$ since both parties send at most t bits.

Note that now, $\mathbf{X}'$ is a deterministic function of X and $\mathbf{Y}'$ is a deterministic function of Y. Thus, any invertible 2-source non-malleable extractor for min-entropy $n-2\delta n$ can be used. Our result follows by using such a construction from a recent work of Li [Li18].

Extractors for interleaved sources. We construct an explicit extractor $\mathrm{i}\ell\mathrm{Ext} : \{0,1\}^{2n} \to \{0,1\}^m$, $m = \Omega(n)$ that satisfies the following: Let $\mathbf{X}$ and $\mathbf{Y}$ be independent (n,k)-sources with $k \geq (2/3+\delta)n$, for any constant $\delta > 0$. Let $\pi : [2n] \to [2n]$ be any permutation. Then,

$$|\mathrm{i}\ell\mathrm{Ext}((\mathbf{X} \circ \mathbf{Y})_\pi) - \mathbf{U}_m| \leq \epsilon.$$

We present our construction and also explain the proof along the way, as this gives more intuition to the different steps of the construction. Let $\mathbf{Z} = (\mathbf{X} \circ \mathbf{Y})_\pi$. We start by taking a large enough slice $\mathbf{Z}_1$ from $\mathbf{Z}$ (say, of length $(2/3+\delta/2)n$). Let $\mathbf{X}$ have more bits in this slice than $\mathbf{Y}$. Let $\mathbf{X}_1$ be the bits of $\mathbf{X}$ in $\mathbf{Z}_1$ and $\mathbf{X}_2$ be the remaining bits of $\mathbf{X}$. Similarly define $\mathbf{Y}_1$ and $\mathbf{Y}_2$. Notice that $\mathbf{X}_1$ has linear entropy and also that $\mathbf{X}_2$ has linear entropy conditioned on $\mathbf{X}_1$. We fix $\mathbf{Y}_1$ and use a condenser (from works of Barak et al. [BRSW12] and Zuckerman [Zuc07]) to condense $\mathbf{Z}_1$ into a matrix with a constant number such that at least one of the row has entropy rate at least 0.9. Notice that this matrix is a deterministic function of $\mathbf{X}$. The next step is to use $\mathbf{Z}$ and each row of the matrix as a seed to a linear seeded extractor get longer rows. This requires some care for the choice of the linear seeded extractor since the seed has some deficiency in entropy. After this step, we use the advice correlation breaker from [CL16] on $\mathbf{Z}$ and each row of the somewhere random source with the row number as the advice (similar to as done before in the construction of seedless non-malleable extractors for 2ISS), and compute the bit-wise XOR of the different outputs that we produce. Let $\mathbf{V}$ denote this random variable. Finally, to output $\Omega(n)$ bits we use a linear seeded

extractor on $\mathbf{Z}$ with $\mathbf{V}$ as the seed. The correctness of various steps in the proof exploit the fact that $\mathbf{Z}$ can be written as the bit-wise sum of two independent sources, and the fact that we use linear seeded extractors. We refer the reader to Section 7 for more details.

3 Background and notation

We use $\mathbf{U}_m$ to denote the uniform distribution on $\{0,1\}^m$.
For any integer $t > 0$, $[t]$ denotes the set $\{1, \ldots, t\}$.
For a string y of length n, and any subset $S \subseteq [n]$, we use y_S to denote the projection of y to the coordinates indexed by S.
We use bold capital letters for random variables and samples as the corresponding small letter, e.g., $\mathbf{X}$ is a random variable, with x being a sample of $\mathbf{X}$.
For strings $x, y \in \{0,1\}^n$, we use $x + y$ (or equivalently $x - y$) to denote the bit-wise xor of the two strings.

3.1 A probability lemma

The following result on min-entropy was proved by Maurer and Wolf [MW97].

Lemma 3.1. *Let $\mathbf{X}, \mathbf{Y}$ be random variables such that the random variable $\mathbf{Y}$ takes at ℓ values. Then*

$$\mathbf{Pr}_{y \sim \mathbf{Y}}[H_\infty(\mathbf{X}|\mathbf{Y} = y) \geq H_\infty(\mathbf{X}) - \log \ell - \log(1/\epsilon)] > 1 - \epsilon.$$

3.2 Conditional min-entropy

Definition 3.2. *The average conditional min-entropy of a source $\mathbf{X}$ given a random variable $\mathbf{W}$ is defined as*

$$\widetilde{H}_\infty(\mathbf{X}|\mathbf{W}) = -\log\left(\mathbf{E}_{w \sim W}\left[\max_x \Pr[\mathbf{X} = x|\mathbf{W} = w]\right]\right) = -\log\left(\mathbf{E}\left[2^{-H_\infty(\mathbf{X}|\mathbf{W}=w)}\right]\right).$$

We recall some results on conditional min-entropy from the work of Dodis et al. [DORS08].

Lemma 3.3 ([DORS08])**.** *For any $\epsilon > 0$,*

$$\mathbf{Pr}_{w \sim \mathbf{W}}\left[H_\infty(\mathbf{X}|\mathbf{W} = w) \geq \widetilde{H}_\infty(\mathbf{X}|\mathbf{W}) - \log(1/\epsilon)\right] \geq 1 - \epsilon.$$

Lemma 3.4 ([DORS08])**.** *If a random variable $\mathbf{Y}$ has support of size 2^ℓ, then $\widetilde{H}_\infty(\mathbf{X}|\mathbf{Y}) \geq H_\infty(\mathbf{X}) - \ell$.*

Definition 3.5. *A function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k, ϵ)-seeded extractor if for any source $\mathbf{X}$ of min-entropy k, $|\mathrm{Ext}(\mathbf{X}, \mathbf{U}_d) - \mathbf{U}_m| \leq \epsilon$. Ext is called a strong seeded extractor if $|(\mathrm{Ext}(\mathbf{X}, \mathbf{U}_d), \mathbf{U}_d) - (\mathbf{U}_m, \mathbf{U}_d)| \leq \epsilon$, where $\mathbf{U}_m$ and $\mathbf{U}_d$ are independent.*

Further, if for each $s \in \mathbf{U}_d$, $\mathrm{Ext}(\cdot, s) : \{0,1\}^n \to \{0,1\}^m$ is a linear function, then Ext is called a linear seeded extractor.

We require extractors that can extract uniform bits when the source only has sufficient conditional min-entropy.

Definition 3.6. *A (k, ϵ)-seeded average case seeded extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *for min-entropy k and error ϵ satisfies the following property: For any source $\mathbf{X}$ and any arbitrary random variable $\mathbf{Z}$ with $\tilde{H}_\infty(\mathbf{X}|\mathbf{Z}) \geq k$,*

$$\mathrm{Ext}(\mathbf{X}, \mathbf{U}_d), \mathbf{Z} \approx_\epsilon \mathbf{U}_m, \mathbf{Z}.$$

It was shown in [DORS08] that any seeded extractor is also an average case extractor.

Lemma 3.7 ([DORS08])**.** *For any $\delta > 0$, if* Ext *is a (k, ϵ)-seeded extractor, then it is also a $(k + \log(1/\delta), \epsilon + \delta)$-seeded average case extractor.*

3.3 Samplers and extractors

Zuckerman [Zuc97] showed that seeded extractors can be used as samplers given access to weak sources. This connection is best presented by a graph theoretic representation of seeded extractors. A seeded extractor $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ can be viewed as an unbalanced bipartite graph G_{Ext} with 2^n left vertices (each of degree 2^d) and 2^m right vertices. Let $\mathcal{N}(x)$ denote the set of neighbors of x in G_{Ext}.

Theorem 3.8 ([Zuc97])**.** *Let* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a seeded extractor for min-entropy k and error ϵ. Let $D = 2^d$. Then for any set $R \subseteq \{0,1\}^m$,*

$$|\{x \in \{0,1\}^n : ||\mathcal{N}(x) \cap R| - \mu_R D| > \epsilon D\}| < 2^k,$$

where $\mu_R = |R|/2^m$.

Theorem 3.9 ([Zuc97])**.** *Let* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a seeded extractor for min-entropy k and error ϵ. Let $\{0,1\}^d = \{r_1, \ldots, r_D\}$, $D = 2^d$. Define* $\mathrm{Samp}(x) = \{\mathrm{Ext}(x, r_1), \ldots, \mathrm{Ext}(x, r_D)\}$*. Let $\mathbf{X}$ be an $(n, 2k)$-source. Then for any set $R \subseteq \{0,1\}^m$,*

$$\mathbf{Pr}_{\mathbf{x} \sim \mathbf{X}}[||\mathrm{Samp}(\mathbf{x}) \cap R| - \mu_R D| > \epsilon D] < 2^{-k},$$

where $\mu_R = |R|/2^m$.

3.4 Explicit extractors from prior work

We recall an optimal construction of strong-seeded extractors.

Theorem 3.10 ([GUV09])**.** *For any constant $\alpha > 0$, and all integers $n, k > 0$ there exists a polynomial time computable strong-seeded extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *with $d = O(\log n + \log(1/\epsilon))$ and $m = (1 - \alpha)k$.*

The following are explicit constructions of linear seeded extractors.

Theorem 3.11 ([Tre01, RRV02])**.** *For every $n, k, m \in \mathbb{N}$ and $\epsilon > 0$, with $m \leq k \leq n$, there exists an explicit strong linear seeded extractor* $\mathrm{LExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *for min-entropy k and error ϵ, where $d = O\left(\log^2(n/\epsilon)/\log(k/m)\right)$.*

A drawback of the above construction is that the seeded length is $\omega(\log n)$ for sub-linear min-entropy. A construction of Li [Li15] achieves $O(\log n)$ seed length for even polylogarithmic min-entropy.

Theorem 3.12 ([Li15]). *There exists a constant $c > 1$ such that for every $n, k \in \mathbb{N}$ with $c \log^8 n \leq k \leq n$ and any $\epsilon \geq 1/n^2$, there exists a polynomial time computable linear seeded extractor* $\mathrm{LExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *for min-entropy k and error ϵ, where $d = O(\log n)$ and $m \leq \sqrt{k}$.*

A different construction achieves seed length $O(\log(n/\epsilon))$ for high entropy sources.

Theorem 3.13 ([CGL16, Li17]). *For all $\delta > 0$ there exist $\alpha, \gamma > 0$ such that for all integers $n > 0$, $\epsilon \geq 2^{-\gamma n}$, there exists an efficiently computable linear strong seeded extractor* $\mathrm{LExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{\alpha d}$, *$d = O(\log(n/\epsilon))$ for min-entropy δn. Further, for any $y \in \{0,1\}^d$, the linear map* $\mathrm{LExt}(\cdot, y)$ *has rank αd.*

The above theorem is stated in [Li17] for $\delta = 0.9$, but it is straightforward to see that the proof extends for any constant $\delta > 0$.

We use a property of linear seeded extractors proved by Rao [Rao09].

Lemma 3.14 ([Rao09]). *Let* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a linear seeded extractor for min-entropy k with error $\epsilon < \frac{1}{2}$. Let X be an affine (n,k)-source. Then*

$$\Pr_{u \sim U_d}[|\mathrm{Ext}(X, u) - U_m| > 0] \leq 2\epsilon.$$

We recall a two-source extractor construction for high entropy sources based on the inner product function.

Theorem 3.15 ([CG88]). *For all $m, r > 0$, with $q = 2^m, n = rm$, let $\mathbf{X}, \mathbf{Y}$ be independent sources on $\mathbb{F}_q^r$ with min-entropy k_1, k_2 respectively. Let* IP *be the inner product function over the field $\mathbb{F}_q$. Then, we have:*

$$|\mathrm{IP}(\mathbf{X}, \mathbf{Y}), \mathbf{X} - \mathbf{U}_m, \mathbf{X}| \leq \epsilon, \quad |\mathrm{IP}(\mathbf{X}, \mathbf{Y}), \mathbf{Y} - \mathbf{U}_m, \mathbf{Y}| \leq \epsilon$$

where $\epsilon = 2^{-(k_1+k_2-n-m)/2}$.

3.5 Advice correlation breakers

We use a primitive called 'correlation breaker' in our construction. Consider a situation where we have arbitrarily correlated random variables $\mathbf{Y}^1, \ldots, \mathbf{Y}^r$, where each $\mathbf{Y}^i$ is on ℓ bits. Further suppose $\mathbf{Y}^1$ is a 'good' random variable (typically, we assume $\mathbf{Y}^1$ is uniform or has almost full min-entropy). A correlation breaker CB is an explicit function that takes some additional resource $\mathbf{X}$, where $\mathbf{X}$ is typically additional randomness (an (n,k)-source) that is independent of $\{\mathbf{Y}^1, \ldots, \mathbf{Y}^r\}$. Thus using $\mathbf{X}$, the task is to break the correlation between $\mathbf{Y}^1$ and the random variables $\mathbf{Y}^2, \ldots, \mathbf{Y}^r$, i.e., $\mathrm{CB}(\mathbf{Y}^1, \mathbf{X})$ is independent of $\{\mathrm{CB}(\mathbf{Y}^2, \mathbf{X}), \ldots, \mathrm{CB}(\mathbf{Y}^r, \mathbf{X})\}$. A weaker notion is that of an advice correlation breaker that takes in some advice for each of the $\mathbf{Y}^i$'s as an additional resource in breaking the correlations. This primitive was implicitly constructed in [CGL16] and used in explicit constructions of non-malleable extractors, and has subsequently found many applications in explicit constructions of extractors for independent sources and non-malleable extractors.

We recall an explicit advice correlation breaker constructed in [CL16]. This correlation breaker works even with the weaker guarantee that the 'helper source' $\mathbf{X}$ is now allowed to be correlated to the sources random variables $\mathbf{Y}^1, \ldots, \mathbf{Y}^r$ in a structured way. Concretely, we assume the source to be of the form $\mathbf{X} + \mathbf{Z}$, where $\mathbf{X}$ is assumed to be an (n,k)-source that is uncorrelated with $\mathbf{Y}^1, \ldots, \mathbf{Y}^r, \mathbf{Z}$. We now state the result more precisely.

Theorem 3.16 ([CL16]). *For all integers $n, n_1, n_2, k, k_1, k_2, t, d, h, \lambda$ and any $\epsilon > 0$, such that $d = O(\log^2(n/\epsilon))$, $k_1 \geq 2d + 8tdh + \log(1/\epsilon)$, $n_1 \geq 2d + 10tdh + (4ht+1)n_2^2 + \log(1/\epsilon)$, and $n_2 \geq 2d + 3td + \log(1/\epsilon)$, let*

- *$\mathbf{X}$ be an (n, k_1)-source, $\mathbf{X}'$ a r.v on n bits, $\mathbf{Y}^1$ be an $(n_1, n_1 - \lambda)$-source, $\mathbf{Z}, \mathbf{Z}'$ are r.v's on n bits, and $\mathbf{Y}^2, \ldots, \mathbf{Y}^t$ be r.v's on n_1 bits each, such that $\{\mathbf{X}, \mathbf{X}'\}$ is independent of $\{\mathbf{Z}, \mathbf{Z}', \mathbf{Y}^1, \ldots, \mathbf{Y}^t\}$,*
- *$id^1, \ldots, id^t$ be bit-strings of length h such that for each $i \in \{2, t\}$, $id^1 \neq id^i$.*

Then there exists an efficient algorithm $\mathrm{ACB} : \{0,1\}^{n_1} \times \{0,1\}^n \times \{0,1\}^h \to \{0,1\}^{n_2}$ which satisfies the following: let

- *$\mathbf{Y}_h^1 = \mathrm{ACB}(\mathbf{Y}^1, \mathbf{X} + \mathbf{Z}, id^1)$,*
- *$\mathbf{Y}_h^i = \mathrm{ACB}(\mathbf{Y}^i, \mathbf{X}' + \mathbf{Z}', id^i)$, $i \in [2, t]$*

Then,

$$\mathbf{Y}_h^1, \mathbf{Y}_h^2, \ldots, \mathbf{Y}_h^t, \mathbf{X}, \mathbf{X}' \approx_{O((h+2^\lambda)\epsilon)} \mathbf{U}_{n_2}, \mathbf{Y}_h^2, \ldots, \mathbf{Y}_h^t, \mathbf{X}, \mathbf{X}'.$$

4 NM extractors for linear composed with interleaved split-state adversaries

The main result of this section is an explicit non-malleable extractor against the tampering family $\mathrm{Lin} \circ \mathrm{2ISS} \subset \mathcal{F}_{2n}$.

Theorem 4.1. *For all integers $n > 0$ there exists an explicit function $\mathrm{nmExt} : \{0,1\}^{2n} \to \{0,1\}^m$, $m = n^{\Omega(1)}$, such that the following holds: For any linear function $h : \{0,1\}^{2n} \to \{0,1\}^{2n}$, arbitrary tampering functions $f, g \in \mathcal{F}_n$, any permutation $\pi : [2n] \to [2n]$ and independent uniform sources $\mathbf{X}$ and $\mathbf{Y}$ each on n bits, there exists a distribution $\mathcal{D}_{h,f,g,\pi}$ on $\{0,1\}^m \cup \{same^\star\}$, such that*

$$|\mathrm{nmExt}((\mathbf{X} \circ \mathbf{Y})_\pi), \mathrm{nmExt}(h((f(\mathbf{X}) \circ g(\mathbf{Y}))_\pi)) - \mathbf{U}_m, \mathrm{copy}(\mathcal{D}_{h,f,g,\pi}, \mathbf{U}_m)| \leq 2^{-n^{\Omega(1)}}.$$

Our first step is to show that in order to prove Theorem 4.1 it is enough to construct a non-malleable extractor satisfying Theorem 4.2.

Theorem 4.2. *There exists a $\delta > 0$ such that for all integers $n, k > 0$ with $n \geq k \geq n - n^\delta$, there exists an explicit function $\mathrm{nmExt} : \{0,1\}^{2n} \to \{0,1\}^m$, $m = n^{\Omega(1)}$, such that the following holds: Let $\mathbf{X}$ and $\mathbf{Y}$ be independent $(n, n - n^\delta)$-sources, $\pi : [2n] \to [2n]$ any arbitrary permutation and arbitrary tampering functions $f_1, f_2, g_1, g_2 \in \mathcal{F}_n$ that satisfy the following condition:*

- *$\forall x \in support(\mathbf{X})$ and $y \in support(\mathbf{Y})$, $f_1(x) + g_1(y) \neq x$ or*
- *$\forall x \in support(\mathbf{X})$ and $y \in support(\mathbf{Y})$, $f_2(x) + g_2(y) \neq y$.*

Then,

$$|\mathrm{nmExt}((\mathbf{X} \circ \mathbf{Y})_\pi), \mathrm{nmExt}(((f_1(\mathbf{X}) + g_1(\mathbf{Y})) \circ (f_2(\mathbf{X}) + g_2(\mathbf{Y})))_\pi) - \mathbf{U}_m, \mathrm{nmExt}(((f_1(\mathbf{X}) + g_1(\mathbf{Y})) \circ (f_2(\mathbf{X}) + g_2(\mathbf{Y})))_\pi)| \leq 2^{-n^{\Omega(1)}}.$$

Proof of Theorem 4.1 assuming Theorem 4.2. Define $\overline{f(x)} = h((f(x) \circ 0^n)_\pi)$ and $\overline{g(y)} = h((0^n \circ y)_\pi)$. Thus, $h((f(x) \circ g(y))_\pi) = \overline{f(x)} + \overline{g(y)}$. Define functions $h_1 : \{0,1\}^{2n} \to \{0,1\}^n$ and $h_2 : \{0,1\}^{2n} \to \{0,1\}^n$ such that $h((f(x) \circ g(y))_\pi) = (h_1(x,y) \circ h_2(x,y))_\pi$. Since $h(f(x), g(y)) = \overline{f(x)} + \overline{g(y)}$, it follows that there exists functions $f_1, g_1, f_2, g_2 \in \mathcal{F}_n$ such that for all $x, y \in \{0,1\}^n$, the following hold:

- $h_1(x,y) = f_1(x) + g_1(y)$, and
- $h_2(x,y) = f_2(x) + g_2(y)$.

Thus, $h((f(x) \circ g(y))_\pi) = ((f_1(x) + g_1(y)) \circ (f_2(x) + g_2(y)))_\pi$.

Now, the idea is to show that $((\mathbf{X} \circ \mathbf{Y})_\pi, ((f_1(\mathbf{X}) + g_1(\mathbf{Y})) \circ (f_2(\mathbf{X}) + g_2(\mathbf{Y})))_\pi)$ is $2^{-n^{\Omega(1)}}$-close to a convex combination of $((\mathbf{X} \circ \mathbf{Y})_\pi, (\mathbf{X} \circ \mathbf{Y})_\pi)$ and distributions of the form $((\mathbf{X}' \circ \mathbf{Y}')_\pi$, $((\eta_1(\mathbf{X}) + \nu_1(\mathbf{Y})) \circ (\eta_2(\mathbf{X}) + \nu_2(\mathbf{Y})))_\pi)$, where $\mathbf{X}'$ and $\mathbf{Y}'$ are independent $(n, n - n^\delta)$-sources and $\eta_1, \eta_2, \nu_1, \nu_2$ are deterministic functions in $\mathcal{F}_n$ satisfying the conditions that:

- $\forall x \in \mathit{support}(\mathbf{X}')$ and $y \in \mathit{support}(\mathbf{Y}')$, $\eta_1(x) + \nu_1(y) \neq x$ or
- $\forall x \in \mathit{support}(\mathbf{X}')$ and $y \in \mathit{support}(\mathbf{Y}')$, $\eta_2(x) + \nu_2(y) \neq y$.

Theorem 4.1 is then direct from from Theorem 4.2.

Let $n_0 = n^\delta$. For any $y \in \{0,1\}^n$ and any function $\eta : \{0,1\}^n \to \{0,1\}^n$, let $\eta^{-1}(y)$ denote the set $\{z \in \{0,1\}^n : \eta(z) = y\}$. We partition $\{0,1\}^n$ into the following two sets:

$$\Gamma_1 = \{y \in \{0,1\}^n : |g_1^{-1}(g_1(y))| \geq 2^{n-n_0}\}, \qquad \Gamma_2 = \{0,1\}^n \setminus \Gamma_1.$$

Let $\mathbf{Y}_1$ be uniform on Γ_1 and $\mathbf{Y}_2$ be uniform on Γ_2. Clearly, $\mathbf{Y}$ is a convex combination of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ with weights $w_i = |\Gamma_1|/2^n$, $i = 1, 2$. If $w_i \leq 2^{-n_0/2}$, we ignore the corresponding source and add an error of $2^{-n_0/2}$ to the extractor. Thus, suppose $w_i \geq 2^{-n_0/2}$ for $i = 1, 2$. Thus, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ each have min-entropy at least $n - n_0/2$.

We claim that $g_1(\mathbf{Y}_2)$ has min-entropy at least $n_0/2$. This can be seen in the following way. For any $y \in \Gamma_2$, $|g_1^{-1}(g_1(y))| \leq 2^{n-n_0}$, and hence it follows $g_1(\mathbf{Y}_2)$ has min-entropy at least $(n - n_0/2) - (n - n_0) = n_0/2$. Thus, clearly for any $x \in \{0,1\}^n$, $x + g_1(\mathbf{Y}_2) \neq x$ with probability at least $1 - 2^{-n_0/2}$. We add a term of $2^{-n^{\Omega(1)}}$ to the error and assume that $\mathbf{X} + g_1(\mathbf{Y}_2) \neq \mathbf{X}$. Thus, $(\mathbf{X} \circ \mathbf{Y}_2)_\pi, ((f_1(\mathbf{X}) + g_1(\mathbf{Y}_2)) \circ (f_1(\mathbf{X}) + g_1(\mathbf{Y}_2)))_\pi$ is indeed $2^{-n^{\Omega(1)}}$close to a convex combination of distributions of the required form.

Next, we claim that for any fixing of $g_1(\mathbf{Y}_1)$, the random variable $\mathbf{Y}_1$ has min-entropy at least $n - n_0$. This is direct from the fact that for any $y \in \Gamma_2$, $|g_1^{-1}(g_1(y))| > 2^{n-n_0}$. We fix $g_1(\mathbf{Y}_1) = g$, and let $f_{1,g}(x) = f_1(x) + g$. Thus, $f_{1,g}(\mathbf{X}) = f_1(\mathbf{X}) + g_1(\mathbf{Y}_1)$. We now partition $\{0,1\}^n$ according to the fixed points of $f_{1,g}$. Let

$$\Delta_1 = \{x : f_1'(x) = x\}, \qquad \Delta_2 = \{0,1\}^n \setminus \Delta_1.$$

Let $\mathbf{X}_1$ be a flat distribution on Δ_1 and $\mathbf{X}_2$ be a flat distribution on Δ_2. If $|\Delta_1| < 2^{n-n_0/2}$, we ignore the distribution $\mathbf{X}_1$ and add an error of $2^{n-n_0/2}$ to the analyis of the non-malleable extractor. Further, it is direct from definition that $f_1(\mathbf{X}_2) + g \neq \mathbf{X}_2$. We now handle to case when $\Delta_1 > 2^{n-n_0/2}$. Note that in this case, $H_1(\mathbf{X}_1) \geq n - n_0/2$. The idea is now to partition Δ_1 into two sets based on the pre-image size of f_2 similar to the way we partioned the support of $\mathbf{Y}$ based on the pre-image size of g_1. Define the sets

$$\Delta_{11} = \{x \in \Delta_1 : |f_2^{-1}(f_2(x)) \cap \Delta_1| \geq 2^{n-n_0}\}, \qquad \Delta_{12} = \Delta_1 \setminus \Delta_{11}.$$

Let $\mathbf{X}_{11}$ be flat on Δ_{11} and $\mathbf{X}_{12}$ be flat on Δ_{12}. Clearly, $\mathbf{X}_1$ is a convex combination of the sources $\mathbf{X}_{11}$ and $\mathbf{X}_{12}$. If Δ_{11} or Δ_{12} is smaller than $2^{n-3n_0/4}$, we ignore the corresponding distribution and add an error of $2^{-n_0/4}$ to the error analysis of the non-malleable extractor. Thus suppose $\Delta_{1i} \geq 2^{n-3n_0/4}$ for $i = 1, 2$. Thus, $\mathbf{X}_{11}$ and $\mathbf{X}_{12}$ both have min-entropy at least $n - 3n_0/4$.

We claim that $f_2(\mathbf{X}_{12})$ has min-entropy at least $n_0/4$. This can be seen in the following way. For any $x \in \Delta_{12}$, $|f_2^{-1}(f_2(x)) \cap \Delta_1| \leq 2^{n-n_0}$, and hence it follows $f_2(\mathbf{X}_{12})$ has min-entropy at least $(n - 3n_0/4) - (n - n_0) = n_0/4$. Thus, clearly $f_2(\mathbf{X}_{12}) + g_2(\mathbf{Y}_1) \neq \mathbf{Y}_1$ with probability at least $1 - 2^{-n_0/4}$. As before, we add an error of $2^{-n^{\Omega(1)}}$ to the error, and assume that $f_2(\mathbf{X}_{12}) + g_2(\mathbf{Y}_1) \neq \mathbf{Y}_1$. Thus, $(\mathbf{X}_{12} \circ \mathbf{Y}_1)_\pi, ((f_1(\mathbf{X}_{12}) + g_1(\mathbf{Y}_2)) \circ (f_1(\mathbf{X}_{12}) + g_1(\mathbf{Y}_2)))_\pi$ is indeed $2^{-n^{\Omega(1)}}$-close to a convex combination of distributions of the required form.

Next, we claim that for any fixing of $f_2(\mathbf{X}_{11})$, the random variable $\mathbf{X}_{11}$ has min-entropy at least $n - n_0$. This is direct from the fact that for any $x \in \Delta_1$, $|f_2^{-1}(f_1(x)) \cap \Delta_1| > 2^{n-n_0}$. We fix $f_2(\mathbf{X}_{11}) = \lambda$, and let $g_{2,\lambda}(y) = \lambda + g_2(y)$. Thus, $g_{2,\lambda}(\mathbf{Y}) = f_1(\mathbf{X}) + g_1(\mathbf{Y}_1)$. We now partition Γ_1 according to the fixed points of $f_{1,g}$. Let

$$\Gamma_{11} = \{y : g_{2,\lambda}(y) = y\}, \qquad \Gamma_{12} = \{0,1\}^n \setminus \Gamma_{11}.$$

Let $\mathbf{Y}_{11}$ be a flat distribution on Γ_{11} and $\mathbf{Y}_{12}$ be a flat distribution on Γ_{12}. It follows from definition that $(f_1(\mathbf{X}_{11}) + g_1(\mathbf{Y}_{11}), f_2(\mathbf{X}_{11}) + g_2(\mathbf{Y}_{11})) = (\mathbf{X}_{11}, \mathbf{Y}_{11})$. Further, $f_2(\mathbf{X}_{11}) + g_2(\mathbf{Y}_{12}) \neq \mathbf{Y}_{12}$, and hence $(\mathbf{X}_{11} \circ \mathbf{Y}_{12})_\pi, ((f_1(\mathbf{X}_{11}) + g_1(\mathbf{Y}_{12})) \circ (f_1(\mathbf{X}_{11}) + g_1(\mathbf{Y}_{12})))_\pi$ is $2^{-n^{\Omega(1)}}$-close to a convex combination of distributions of the required form. This completes the proof. □

In the rest of the section, we prove Theorem 4.2. We assume the setup given from Theorem 4.2. Thus, $\mathbf{X}$ and $\mathbf{Y}$ are independent $(n, n - n^\delta)$-sources, $\pi : [2n] \to [2n]$ is an arbitrary permutation and $f_1, f_2, g_1, g_2 \in \mathcal{F}_n$ satisfy the following conditions:

- $\forall x \in \mathit{support}(\mathbf{X})$ and $y \in \mathit{support}(\mathbf{Y})$, $f_1(x) + g_1(y) \neq x$ or
- $\forall x \in \mathit{support}(\mathbf{X})$ and $y \in \mathit{support}(\mathbf{Y})$, $f_2(x) + g_2(y) \neq y$.

We use the following notation: if $\mathbf{W} = h((\mathbf{X} \circ \mathbf{Y})_\pi)$ (for some function h), then we use to $\mathbf{W}'$ or $(\mathbf{W})'$ to denote the random variable $h(((f_1(\mathbf{X}) + g_1(\mathbf{Y})) \circ (f_2(\mathbf{X}) + g_2(\mathbf{Y})))_\pi)$. Further, define $\overline{\mathbf{X}} = (\mathbf{X} \circ 0^n)_\pi$, $\overline{\mathbf{Y}} = (0^n \circ \mathbf{Y})_\pi$, $\overline{f_1(\mathbf{X})} = (f_1(\mathbf{X}) \circ 0^n)_\pi$, $\overline{f_2(\mathbf{X})} = (0^n \circ f_2(\mathbf{X}))_\pi$, $\overline{g_1(\mathbf{Y})} = (g_1(\mathbf{Y}) \circ 0^n)_\pi$ and $\overline{g_2(\mathbf{Y})} = (0^n \circ g_2(\mathbf{Y}))_\pi$. It follows that $\mathbf{Z} = \overline{\mathbf{X}} + \overline{\mathbf{Y}}$ and $\mathbf{Z}' = \overline{f_1(\mathbf{X})} + \overline{g_1(\mathbf{Y})} + \overline{f_2(\mathbf{X})} + \overline{g_2(\mathbf{Y})}$.

We use Section 4.1 to construct an advice generator and Section 4.2 to construct an advice correlation breaker. Finally, we present the non-malleable extractor construction in Section 4.3.

4.1 An advice generator

Lemma 4.3. *There exists an efficiently computable function* $\mathrm{advGen} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{n_4}$, $n_4 = n^\delta$, *such that with probability at least* $1 - 2^{-n^{\Omega(1)}}$ *over the fixing of the random variables* $\{\mathrm{advGen}((\mathbf{X} \circ \mathbf{Y})_\pi), \mathrm{advGen}(((f_1(\mathbf{X}) + g_1(\mathbf{Y})) \circ (f_2(\mathbf{X}) + g_2(\mathbf{Y})))_\pi)\}$, *the following hold:*

- $\{\mathrm{advGen}((\mathbf{X} \circ \mathbf{Y})_\pi) \neq \mathrm{advGen}(((f_1(\mathbf{X}) + g_1(\mathbf{Y})) \circ (f_2(\mathbf{X}) + g_2(\mathbf{Y})))_\pi)\}$,
- $\mathbf{X}$ *and* $\mathbf{Y}$ *are independent,*
- $H_\infty(\mathbf{X}) \geq k - 2n^\delta$, $H_\infty(\mathbf{Y}) \geq k - 2n^\delta$.

We prove the above lemma in the rest of this subsection. We claim that the function advGen computed by Algorithm 1 satisfies the above lemma. We first set up some parameters and ingredients.

- Let C be a large enough constant and $\delta' = \delta/C$.
- Let $n_0 = n^{\delta'}, n_1 = n_0^{c_0}, n_2 = 10n_0$, for some constant c_0 that we set below.
- Let $E : \{0,1\}^{2n} \to \{0,1\}^{n_3}$ be the encoding function of a linear error correcting code $\mathcal{C}$ with constant rate α and constant distance β.
- Let $\mathrm{Ext}_1 : \{0,1\}^{n_1} \times \{0,1\}^{d_1} \to \{0,1\}^{\log(n_3)}$ be a $(n_1/20, \beta/10)$-seeded extractor instantiated using Theorem 3.10. Thus $d_1 = c_1 \log n_1$, for some constant c_1. Let $D_1 = 2^{d_1} = n_1^{c_1}$.
- Let $\mathrm{Samp}_1 : \{0,1\}^{n_1} \to [n_3]^{D_1}$ be the sampler obtained from Theorem 3.9 using Ext_1.
- Let $\mathrm{Ext}_2 : \{0,1\}^{n_2} \times \{0,1\}^{d_2} \to \{0,1\}^{\log(2n)}$ be a $(n_2/20, 1/n_0)$-seeded extractor instantiated using Theorem 3.10. Thus $d_2 = c_2 \log n_2$, for some constant c_2. Let $D_2 = 2_2^d$. Thus $D_2 = 2^{d_2} = n_2^{c_2}$.
- Let $\mathrm{Samp}_2 : \{0,1\}^{n_2} \to [2n]^{D_2}$ be the sampler obtained from Theorem 3.9 using Ext_2.
- Set $c_0 = 2c_2$.
- Let $\mathrm{i}\ell\mathrm{Ext} : \{0,1\}^{D_2} \to \{0,1\}^{n_0}$ be the extractor from Theorem 7.1.
- Let $\mathrm{LExt} : \{0,1\}^{2n} \times \{0,1\}^{n_0} \to \{0,1\}^{n_0}$ be a linear seeded extractor instantiated from Theorem 3.15 set to extract from min-entropy $n_1/100$ and error $2^{-\Omega(\sqrt{n_0})}$.

Algorithm 1: advGen(z)

Input: Bit-string $z = (x \circ y)_\pi$ of length $2n$, where x and y are each n bit-strings and $\pi : [2n] \to [2n]$ is a permutation.

Output: Bit string v of length n_4.

1 Let $z_1 = \mathrm{Slice}(z, n_1), z_2 = \mathrm{Slice}(z, n_2)$.
2 Let $S = \mathrm{Samp}_1(z_1)$.
3 Let $T = \mathrm{Samp}_2(z_2)$ and $z_3 = z_T$.
4 Let $r = \mathrm{i}\ell\mathrm{Ext}(z_3)$.
5 Let $w_1 = (E(z))_S$.
6 Let $w_2 = \mathrm{LExt}(z, r)$.
7 Output $v = z_1 \circ z_2 \circ z_3 \circ w_1 \circ w_2$.

Lemma 4.4. *With probability at least* $1 - 2^{-n^{\Omega(1)}}$, $\mathbf{V} \neq \mathbf{V}'$.

Proof. We prove the lemma assuming $f_1(\mathbf{X}) + g_1(\mathbf{Y}) \neq \mathbf{X}$. The proof in the other case (i.e., $f_2(\mathbf{X}) + g_2(\mathbf{Y}) \neq \mathbf{Y}$) is similar and we skip it.

First observe that the lemma is direct if $\mathbf{Z}_1 \neq \mathbf{Z}'_1$ or $\mathbf{Z}_2 \neq \mathbf{Z}'_2$ or $\mathbf{Z}_3 \neq \mathbf{Z}'_3$. Thus, we can assume $\mathbf{Z}_i = \mathbf{Z}'_i$ for $i = 1, 2, 3$. It is easy to see that $\mathbf{S} = \mathbf{S}', \mathbf{T} = \mathbf{T}'$, and $\mathbf{Z}_4 = \mathbf{Z}'_4$.

Now observe that

$$\mathbf{Z} - \mathbf{Z}' = \overline{\mathbf{X}} + \overline{\mathbf{Y}} - \overline{f_1(\mathbf{X})} - \overline{g_1(\mathbf{Y})} - \overline{f_2(\mathbf{X})} - \overline{g_2(\mathbf{Y})}.$$

Note that $\mathbf{Z} - \mathbf{Z}' \neq 0$ which follows from our assumption that $f_1(\mathbf{X}) + g_1(\mathbf{Y}) \neq \mathbf{X}$.

Now define the function $h_1 : \{0,1\}^{2n} \to \{0,1\}^{2n}$ as $h_1(z) = z - f_1(z) - f_2(z)$ and $h_2 : \{0,1\}^{2n} \to \{0,1\}^{2n}$ as $h_2(z) = z - g_1(z) - g_2(z)$.

Thus,

$$\mathbf{Z} - \mathbf{Z}' = h_1(\overline{\mathbf{X}}) + h_2(\overline{\mathbf{Y}}).$$

Let $\mathbf{X}_i$ be the bits of $\mathbf{X}$ in $\mathbf{Z}_i$ for $i = 1, 2, 3$ and $\mathbf{X}_4$ be the remaining bit of $\mathbf{X}$. Similarly define $\mathbf{Y}_i$'s, $i = 1, 2, 3, 4$. Without loss of generality suppose that $|\mathbf{X}_1| \geq |\mathbf{Y}_1|$, (where $|\alpha|$ denotes the length of the string α).

Let $\Gamma \subset \{0,1\}^{2n}$ denote the support of the source $\overline{\mathbf{X}}$. We partition Γ into two sets Γ_a and Γ_b according to the pre-image size of the function h_1 in the following way. For any $z \in \{0,1\}^{2n}$, let $h_1^{-1}(z)$ denote the set $\{y \in \{0,1\}^{2n} : h_1(y) = z\}$.

Let $n_p = n_1/50$. Define

$$\Gamma_a = \{z \in \Gamma : |h_1^{-1}(h_1(z)) \cap \Gamma| \geq 2^{n-n_p}\}, \quad \Gamma_b = \Gamma \setminus \Gamma_1.$$

Let $p_a = \Pr[\overline{\mathbf{X}} \in \Gamma_a]$ and $p_b = \Pr[\overline{\mathbf{X}} \in \Gamma_b]$. Let $\overline{\mathbf{X}}_a$ be the source supported on Γ_a with the probability law $\Pr[\overline{\mathbf{X}}_a = z] = \frac{1}{p_a} \cdot \Pr[\overline{\mathbf{X}} = z]$. Also define $\overline{\mathbf{X}}_b$ supported on Γ_b with the probability law $\Pr[\overline{\mathbf{X}}_a = z] = \frac{1}{p_b} \cdot \Pr[\overline{\mathbf{X}} = z]$.

Clearly $\overline{\mathbf{X}}$ is a convex combination of the distributions $\overline{\mathbf{X}}_a$ and $\overline{\mathbf{X}}_b$, with weights p_a and p_b respectively. If any of p_a or p_b is less that 2^{-n_0}, we ignore the corresponding source and add it to the error. Thus suppose both p_a and p_b are at least 2^{-n_0}. This implies that both $\overline{\mathbf{X}}_a$ and $\overline{\mathbf{X}}_b$ have min-entropy at least $n - 2n_0$. We record the following two bounds that are direct from the above definitions.

- For any fixing of $h_1(\overline{\mathbf{X}}_a) = x_a$, $\overline{\mathbf{X}}_a$ has min-entropy at least $n - n_p$.
- The distribution $h_1(\mathbf{X}_b)$ has min-entropy at least $n_p - 2n_0$.

We introduce some notation. For any random variable $\nu = \eta(\overline{\mathbf{X}}, \overline{\mathbf{Y}})$ (where η is an arbitrary deterministic function), we add an extra a or b to the subscript and use ν_a to denote the random variable $\eta(\overline{\mathbf{X}}_a, \overline{\mathbf{Y}})$ and ν_b to denote the random variables $\eta(\overline{\mathbf{X}}_b, \overline{\mathbf{Y}})$ respectively. For example, $\mathbf{Z}'_{1,a} = \overline{f_1(\mathbf{X}_a)} + \overline{g_1(\mathbf{Y})} + \overline{f_2(\mathbf{X}_a)} + \overline{g_2(\mathbf{Y})}$. Further we use $\mathbf{X}_a$ to denote the distribution on n bits such that $\overline{\mathbf{X}}_a = (\mathbf{X}_a \circ 0^n)_\pi$. We similarly define the distribution $\mathbf{X}_b$.

We prove the following two statements:

1. $\mathbf{W}_{1,a} - \mathbf{W}'_{1,a} \neq 0$ with probability $1 - 2^{-n^{\Omega(1)}}$.
2. $\mathbf{W}_{2,b} - \mathbf{W}'_{2,b} \neq 0$ with probability $1 - 2^{-n^{\Omega(1)}}$.

It is direct that the lemma follows from the above two inequalities.

We begin with the proof of (1). Since E is a linear code, we have

$$\begin{aligned}\mathbf{W}_{1,a} - \mathbf{W}'_{1,a} &= (E(\mathbf{Z}_a - \mathbf{Z}'_a))_{\mathbf{S}_a}. \\ &= (E(h_1(\overline{\mathbf{X}_a}) + h_2(\overline{\mathbf{Y}})))_{\mathbf{S}_a}.\end{aligned}$$

Now fix the random $h_1(\mathbf{X}_a)$, and it follows that $\mathbf{X}_a$ has min-entropy at least $n - n_p$. Recall that we assumed $|\mathbf{X}_1| \geq |\mathbf{Y}_1|$. Thus, $\mathbf{X}_{1,a}$ has min-entropy at least $n_1/2 - n_p - n_0 > n_1/10$ with probability

at least $1-2^{-n_0}$. Further fix $\mathbf{Y}$, and note that this does not affect the distribution of $\mathbf{X}_{1,a}$. This fixes $E(\mathbf{Z}_a-\mathbf{Z}'_a)$. Further $\mathbf{Z}_a \neq \mathbf{Z}'_a$, the $E(\mathbf{Z}_a-\mathbf{Z}'_a)$ contains 1's at least β fraction of its coordinates. Recalling that $\mathbf{S}_a = \mathrm{Samp}_1(\mathbf{Z}_{1,a})$, it now follows from Theorem 3.9 that with probability at least $1-2^{-n^{\Omega(1)}}$, $(E(\mathbf{Z}_a-\mathbf{Z}'_a))_{\mathbf{S}}$ is a non-zero string (and hence $\mathbf{W}_{1,a}-\mathbf{W}'_{1,a} \neq 0$). This completes the proof of this case.

We now proceed to prove (2). Using the fact that LExt is a linear seeded extractor, it follows that

$$\begin{aligned}\mathbf{W}_{2,b}-\mathbf{W}'_{2,b} &= \mathrm{LExt}(\mathbf{Z}_b-\mathbf{Z}'_b,\mathbf{R}_b)\\ &= \mathrm{LExt}(h_1(\mathbf{X}_b),\mathbf{R}_b)+\mathrm{LExt}(h_2(\mathbf{Y}_b),\mathbf{R}_b).\end{aligned}$$

Without loss of generality, suppose $\mathbf{X}$ has more bits in $\mathbf{Z}_2$ (the argument is identical in the other case). Since $\mathbf{X}_{2,b}$ has min-entropy at least $n-2n_0$, it follows that $\mathbf{X}_{2,b}$ has min-entropy at least $\frac{n_2}{2}-3n_0 > \frac{n_2}{10}$ with probability at least $1-2^{-n_0}$. Fix the bits of $\mathbf{Y}$ in $\mathbf{Z}_2$, and thus $\mathbf{Z}_{2,b}$ is a deterministic function of $\mathbf{X}_{2,b}$. Recall that $\mathbf{T}_b = \mathrm{Samp}_2(\mathbf{Z}_2)$. It is now straightforward to see that with probability $1-2^{-n^{\Omega(1)}}$ over the fixing of $\mathbf{X}_{2,b}$, $|\mathbf{T}_b|\cdot(1/2-o(1)) \leq |\mathbf{T}_b \cap \pi([n])| \leq |\mathbf{T}_b|\cdot(1/2+o(1)$. Recall $|T_b| = D_2$. We fix $\mathbf{X}_{2,b}$ such that $(1/2-o(1))D_2 \leq |\mathbf{T}_b \cap \pi([n])| \leq (1/2+o(1))D_2$. Thus, $\mathbf{Z}_3$ contains at least $(1/2-o(1))D_2$ bits from both $\mathbf{X}_b$ and $\mathbf{Y}$. It follows that both $\mathbf{X}_{3,b}$ and $\mathbf{Y}$ both have min-entropy at least $(1/2-o(1))D_2-2n_0-n_2 = (1/2-o(1))D_2$ (even with the conditionings so far), and hence $\mathbf{R}_b$ is $2^{-n^{\Omega(1)}}$-close to uniform. We argue this this hold even conditioned on $\mathbf{X}_{3,b}$. This follows roughly from the fact that any 2-source extractor is strong [Rao07] which easily extends to interleaved extractors. We fix $\mathbf{X}_{3,b}$, and thus $\mathbf{R}_b$ is now a deterministic function of $\mathbf{Y}$.

Next, we note that $h_1(\mathbf{X}_b)$ has min-entropy at least $(n-2n_0)-(n-n_p)-n_2-D_2-n_0 = n_p-3n_0-D_2-n_2 > n_p/2$ (with probability $1-2^{-n^{\Omega(1)}}$). Thus, $\mathrm{LExt}(h_1(\mathbf{X}_b),\mathbf{R}_b)$ is $2^{-n^{\Omega(1)}}$-close to uniform. We fix $\mathbf{R}_b$ and $\mathrm{LExt}(h_1(\mathbf{X}_b),\mathbf{R}_b)$ continues to be close to uniform using the fact that LExt is a strong-seeded extractor. Further, $\mathrm{LExt}(h_1(\mathbf{X}_b),\mathbf{R}_b)$ is now a deterministic function of $\mathbf{X}_b$ and we can fix $\mathrm{LExt}(h_2(\mathbf{Y}_b),\mathbf{R}_b)$ which is a deterministic function of $\mathbf{Y}$. It thus follows that $\mathbf{W}_{2,b}-\mathbf{W}'_{2,b} \neq 0$ with probability $1-2^{-n^{\Omega(1)}}$ using the fact that $\mathrm{LExt}(h_1(\mathbf{X}_b),\mathbf{R}_b)$ is close to uniform. This completes the proof of (2). The fact that $\mathbf{V}$ and $\mathbf{V}'$ can be fixed such that $\mathbf{X}$ and $\mathbf{Y}$ remain independent with min-entropy at least $k-2n^{\delta}$ (with probability $1-2^{-n^{\Omega(1)}}$) is easy to verify from the construction. This completes the proof of Lemma 4.4. □

4.2 An Advice Correlation Breaker

We recall the setup of Theorem 4.2. $\mathbf{X}$ and $\mathbf{Y}$ are independent (n,k)-sources, $k \geq n-n^{\delta}$, $\pi : [2n] \to [2n]$ is an arbitrary permutation and $f_1, f_2, g_1, g_2 \in \mathcal{F}_n$ satisfy the following conditions:

- $\forall x \in \mathit{support}(\mathbf{X})$ and $y \in \mathit{support}(\mathbf{Y})$, $f_1(x)+g_1(y) \neq x$ or
- $\forall x \in \mathit{support}(\mathbf{X})$ and $y \in \mathit{support}(\mathbf{Y})$, $f_2(x)+g_2(y) \neq y$.

Further, we defined the following: $\overline{\mathbf{X}} = (\mathbf{X}\circ 0^n)_\pi$, $\overline{\mathbf{Y}} = (0^n \circ \mathbf{Y})_\pi$, $\overline{f_1(\mathbf{X})} = (f_1(\mathbf{X})\circ 0^n)_\pi$, $\overline{f_2(\mathbf{X})} = (0^n \circ f_2(\mathbf{X}))_\pi$, $\overline{g_1(\mathbf{Y})} = (g_1(\mathbf{Y})\circ 0^n)_\pi$ and $\overline{g_2(\mathbf{Y})} = (0^n \circ g_2(\mathbf{Y}))_\pi$. It follows that $\mathbf{Z} = \overline{\mathbf{X}}+\overline{\mathbf{Y}}$ and $\mathbf{Z}' = \overline{f_1(\mathbf{X})}+\overline{g_1(\mathbf{Y})}+\overline{f_2(\mathbf{X})}+\overline{g_2(\mathbf{Y})}$. Thus, for some functions $f,g \in \mathcal{F}_{2n}$, $\mathbf{Z}' = f(\overline{\mathbf{X}})+g(\overline{\mathbf{Y}})$. Let $\overline{\mathbf{X}'} = f(\overline{\mathbf{X}})$ and $\overline{\mathbf{Y}'} = g(\overline{\mathbf{Y}})$.

The following is the main result of this section. Assume that we have some random variables such that $\mathbf{X}$ and $\mathbf{Y}$ continue to be independent, and $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq k-2n^{\delta}$.

Lemma 4.5. *There exists an efficiently computable function* $\mathrm{ACB} : \{0,1\}^{2n} \times \{0,1\}^{n_1} \to \{0,1\}^m$, $n_1 = n^\delta$ *and* $m = n^{\Omega(1)}$, *such that*

$$\mathrm{ACB}(\overline{\mathbf{X}} + \overline{\mathbf{Y}}, w), \mathrm{ACB}(\overline{f(\mathbf{X})} + \overline{g(\mathbf{Y})}, w') \approx_\epsilon \mathbf{U}_m, \mathrm{ACB}(\overline{f(\mathbf{X})} + \overline{g(\mathbf{Y})}, w'),$$

for any fixed strings $w, w' \in \{0,1\}^{n_1}$ *with* $w \neq w'$.

We use the rest of the section to prove the above lemma. In particular, we prove that the function ACB computed by Algorithm 2 satisfies the conclusion of Lemma 4.5.

We start by setting up some ingredients and parameters.

- Let $\delta > 0$ be a small enough constant.
- Let $n_2 = n^{\delta_1}$, where $\delta_1 = 2\delta$.
- Let $\mathrm{LExt}_1 : \{0,1\}^{n_2} \times \{0,1\}^d \to \{0,1\}^{d_1}$, $d_1 = \sqrt{n_2}$, be a linear-seeded extractor instantiated from Theorem 3.11 set to extract from entropy $k_1 = n_2/10$ with error $\epsilon_1 = 1/10$. Thus $d = C_1 \log n_2$, for some constant C_1. Let $D = 2^d = n^{\delta_2}$, $\delta_2 = 2C_1\delta$.
- Set $\delta' = 20C_1\delta$.
- Let $\mathrm{LExt}_2 : \{0,1\}^{2n} \times \{0,1\}^{d_1} \to \{0,1\}^{n_4}$, $n_4 = n^{8\delta_3}$ be a linear-seeded extractor instantiated from Theorem 3.11 set to extract from entropy $k_2 = 0.9k$ with error $\epsilon_2 = 2^{-\Omega(\sqrt{d_1})} = 2^{-n^{\Omega(1)}}$, such that the seed length of the extractor LExt_2 (by Theorem 3.11) is d_1.
- Let $\mathrm{ACB}' : \{0,1\}^{n_{1,acb'}} \times \{0,1\}^{n_{acb'}} \times \{0,1\}^{h_{acb'}} \to \{0,1\}^{n_{2,acb'}}$, be the advice correlation breaker from Theorem 3.16 set with the following parameters: $n_{acb'} = 2n, n_{1,acb'} = n_4$, $n_{2,acb'} = m = O(n^{2\delta_2}), t_{acb'} = 2D, h_{acb'} = n_1 + d, \epsilon_{acb'} = 2^{-n^\delta}$, $d_{acb'} = O(\log^2(n/\epsilon_{acb'}))$, $\lambda_{acb'} = 0$. It can be checked that by our choice of parameters, the conditions required for Theorem 3.16 indeed hold for $k_{1,acb'} \geq n^{2\delta_2}$.

Algorithm 2: $\mathrm{ACB}(z)$

Input: Bit-strings $z = (x \circ y)_\pi$ of length $2n$ and bit string w of length n_1, where x and y are each n bit-strings and $\pi : [2n] \to [2n]$ is a permutation.
Output: Bit string of length m.

Let $z_1 = \mathrm{Slice}(z, n_2)$.
Let v be a $D \times n_3$ matrix, with its i'th row $v_i = \mathrm{LExt}_1(z_1, i)$.
Let r be a $D \times n_4$ matrix, with its i'th row $r_i = \mathrm{LExt}_2(z, v_i)$.
Let s be a $D \times m$ matrix, with its i'th row $s_i = \mathrm{ACB}'(r_i, z, w \circ i)$.
Output $\oplus_{i=1}^D s_i$.

Let $\mathbf{X}_1$ be the bits of $\mathbf{X}$ in $\mathbf{Z}_1$ and $\mathbf{X}_2$ be the remaining bit of $\mathbf{X}$. Define $\mathbf{Y}_1$ and $\mathbf{Y}_2$ similarly. Without loss of generality suppose that $|\mathbf{X}_1| \geq |\mathbf{Y}_1|$. Let $\overline{\mathbf{X}}_1 = \mathrm{Slice}(\overline{\mathbf{X}}, n_2)$ and $\overline{\mathbf{Y}}_1 = \mathrm{Slice}(\overline{\mathbf{Y}}, n_2)$. Define $\overline{\mathbf{X}}'_1 = \mathrm{Slice}(f(\overline{\mathbf{X}}), n_2)$ and $\overline{\mathbf{Y}_1}' = \mathrm{Slice}(g(\overline{\mathbf{Y}}), n_2)$. It follows that $\mathbf{Z}_1 = \overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1$ and $\mathbf{Z}'_1 = \overline{\mathbf{X}}'_1 + \overline{\mathbf{Y}}'_1$.

Claim 4.6. *Conditioned on the random variables* $\mathbf{Y}_1, \overline{\mathbf{Y}}'_1$, $\{\mathrm{LExt}_2(\overline{\mathbf{X}}, \mathrm{LExt}_1(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, i))\}_{i=1}^D$, $\{\mathrm{LExt}_2(\overline{\mathbf{X}}', \mathrm{LExt}_1(\overline{\mathbf{X}}'_1 + \overline{\mathbf{Y}}'_1, i))\}_{i\in[D]}$, $\mathbf{X}_1$ *and* $\overline{\mathbf{X}}'_1$, *the following hold:*

- *the matrix* $\mathbf{R}$ *is* $2^{-n^{\Omega(1)}}$*-close to a somewhere random source,*
- $\mathbf{R}$ *and* $\mathbf{R}'$ *are deterministic functions of* $\mathbf{Y}$*,*
- $H_\infty(\mathbf{X}) \geq n - n^{\delta'}$*,* $H_\infty(\mathbf{Y}) \geq n - n^{\delta'}$*.*

Proof. By construction, we have that for any $j \in [D]$,

$$\begin{aligned}\mathbf{R}_j &= \mathrm{LExt}_2(\mathbf{Z}, \mathrm{LExt}_1(\mathbf{Z}_1, j)) \\ &= \mathrm{LExt}_2(\overline{\mathbf{X}} + \overline{\mathbf{Y}}, \mathrm{LExt}_1(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, j)) \\ &= \mathrm{LExt}_2(\overline{\mathbf{X}}, \mathrm{LExt}_1(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, j)) + \mathrm{LExt}_2(\overline{\mathbf{Y}}, \mathrm{LExt}_1(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, j))\end{aligned}$$

Similarly,

$$\mathbf{R}'_j = \mathrm{LExt}_2(\overline{\mathbf{X}}', \mathrm{LExt}_1(\overline{\mathbf{X}}'_1 + \overline{\mathbf{Y}}'_1, j)) + \mathrm{LExt}_2(\overline{\mathbf{Y}}', \mathrm{LExt}_1(\overline{\mathbf{X}}'_1 + \overline{\mathbf{Y}}'_1, j)).$$

Fix the random variables $\mathbf{Y}_1, \overline{\mathbf{Y}}'_1$. Note that after these fixings, $\overline{\mathbf{Y}}$ has min-entropy at least $k - 2n_1 - n_2 > 0.9k$. Now, since LExt_2 is a strong seeded extractor for entropy $0.9k$, it follows that there exists a set $T \subset \{0,1\}^{d_1}$, $|T| \geq (1 - \sqrt{\epsilon_2})2^{d_1}$, such that for any $j \in [T]$, $|\mathrm{LExt}_2(\overline{\mathbf{Y}}, j) - \mathbf{U}_{n_4}| \leq \sqrt{\epsilon_2}$.

Now viewing LExt_1 as a sampler (see Section 3.3) using the weak source $\overline{\mathbf{X}}_{1,y_1} = \overline{\mathbf{X}}_1 + \overline{y_1}$, it follows by Theorem 3.9 that

$$\Pr[|\{\mathrm{LExt}_1(\overline{\mathbf{X}}_{1,y_1}, i) : i \in \{0,1\}^d\} \cap T| > (1 - \sqrt{\epsilon_2} - \epsilon_1)D] \geq 1 - 2^{0.2n_2} = 1 - 2^{-n^{\Omega(1)}}.$$

We fix $\overline{\mathbf{X}}_1$, and it follows that with probability at least $1 - 2^{-n^{\Omega(1)}}$, $\{\mathrm{LExt}_1(\overline{\mathbf{X}}_{1,y_1}, i) : i \in \{0,1\}^d\} \cap T \neq \emptyset$, and thus there exists a $j \in [D]$ such that $\mathrm{LExt}_2(\overline{\mathbf{Y}}, \mathrm{LExt}_1(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, j))$ is $2^{-n^{\Omega(1)}}$-close to $\mathbf{U}_{n_2}$ and is a deterministic function of $\mathbf{Y}$.

We now fix the random variables $\overline{\mathbf{X}}'_1$, $\{\mathrm{LExt}_2(\overline{\mathbf{X}}, \mathrm{LExt}_1(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, i))\}_{i=1}^D$, $\{\mathrm{LExt}_2(\overline{\mathbf{X}}', \mathrm{LExt}_1(\overline{\mathbf{X}_1}' + \overline{\mathbf{Y}_1}', i))\}_{i=1}^D$, and note that $\mathrm{LExt}_2(\overline{\mathbf{Y}}, \mathrm{LExt}_1(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, j))$ continues to be $2^{-n^{\Omega(1)}}$-close to $\mathbf{U}_{n_2}$. It follows that $\mathbf{R}_j$ is $2^{-n^{\Omega(1)}}$-close to $\mathbf{U}_{n_2}$. Further, for any $i \in [D]$, the random variables $\mathbf{R}_i$ and $\mathbf{R}'_i$ are deterministic functions of $\mathbf{Y}$. Finally, note that $\mathbf{X}$ and $\mathbf{Y}$ remain independent after these conditionings, and $H_\infty(\mathbf{X}) \geq n - 3n_1 - 2n_2 - 2Dn_4 \geq n - n^{10\delta_2}$ and $H_\infty(\mathbf{Y}) \geq n - 3n_1 - n_2 > n - n^{\delta_2}$. □

Lemma 4.5 is now direct from the next claim.

Claim 4.7. *There exists* $j \in [D]$ *such that*

$$\mathbf{S}_j, \{\mathbf{S}_i\}_{i \in [D] \setminus j} \approx_{2^{-n^{\Omega(1)}}} \mathbf{U}_m, \{\mathbf{S}_i\}_{i \in [D] \setminus j}.$$

Proof. Fix the random variables: $\mathbf{W}, \mathbf{W}', \mathbf{Y}_1, \overline{\mathbf{Y}}'_1$, $\{\mathrm{LExt}_2(\overline{\mathbf{X}}, \mathrm{LExt}_1(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, i))\}_{i=1}^D$, $\{\mathrm{LExt}_2(\overline{\mathbf{X}}', \mathrm{LExt}_1(\overline{\mathbf{X}}'_1 + \overline{\mathbf{Y}}'_1, i))\}_{i \in [D]}$, $\mathbf{X}_1$ and $\overline{\mathbf{X}}'_1$. By Lemma 4.3, we have that with probability at least $1 - 2^{-n^{\Omega(1)}}$, $\mathbf{W} \neq \mathbf{W}'$. Further, by Claim 4.6 we have that $\mathbf{R}$ and $\mathbf{R}'$ are deterministic functions of $\mathbf{Y}$, and with probability at least $1 - 2^{-n^{\Omega(1)}}$, there exists $j \in [D]$ such that $\mathbf{R}_j$ is $2^{-n^{\Omega(1)}}$-close to uniform, and $H_\infty(\overline{\mathbf{X}}) \geq \frac{1}{2}n_{acb} - n^{\delta'} > n^{2\delta_2}$. Recall that $\mathbf{Z} = \overline{\mathbf{X}} + \overline{\mathbf{Y}}$ and $\mathbf{Z}' = \overline{\mathbf{X}}' + \overline{\mathbf{Y}}'$. It now follows by Theorem 3.16 that

$$\begin{aligned}\mathrm{ACB}'(\mathbf{R}_j, \mathbf{Z}, \mathbf{W} \circ j), \{\mathrm{ACB}'(\mathbf{R}_i, \overline{\mathbf{X}} + \overline{\mathbf{Y}}, \mathbf{W} \circ i)\}_{i \in [D] \setminus j}, \{\mathrm{ACB}'(\mathbf{R}'_i, \overline{\mathbf{X}}' + \overline{\mathbf{Y}}', \mathbf{W}' \circ i)\}_{i \in [D]} \approx_{2^{-n^{\Omega(1)}}} \\ \mathbf{U}_m, \{\mathrm{ACB}'(\mathbf{R}_i, \overline{\mathbf{X}} + \overline{\mathbf{Y}}, \mathbf{W} \circ i)\}_{i \in [D] \setminus j}, \{\mathrm{ACB}'(\mathbf{R}'_i, \overline{\mathbf{X}}' + \overline{\mathbf{Y}}', \mathbf{W}' \circ i)\}_{i \in [D]}\end{aligned}$$

This completes the proof of the claim. □

4.3 The non-malleable extractor

We are now ready to present the construction of iℓNM that satisfies the requirements of Theorem 4.2.

- Let $\delta > 0$ be a small enough constant, $n_1 = n^\delta$ and $m = n^{\Omega(1)}$.
- Let advGen : $\{0,1\}^{2n} \to \{0,1\}^{n_1}$, $n_1 = n^\delta$, be the advice generator from Lemma 4.3.
- Let ACB : $\{0,1\}^{2n} \times \{0,1\}^{n_1} \to \{0,1\}^m$ be the advice correlation breaker from Lemma 4.5.

Algorithm 3: iℓNM(z)

Input: Bit-string $z = (x \circ y)_\pi$ of length $2n$, where x and y are each n bit-strings, and $\pi : [2n] \to [2n]$ is a permutation.

Output: Bit string of length m.

1 Let $w = \text{advGen}(z)$.

2 Output $\text{ACB}(z, w)$

We prove that the function iℓNM computed by Algorithm 3 satisfies the conclusion of Theorem 4.2 as follows. Fix the random variables $\mathbf{W}, \mathbf{W}'$. By Lemma 4.3, it follows that $\mathbf{X}$ remains independent of $\mathbf{Y}$, and with probability at least $1 - 2^{-n^{\Omega(1)}}$, $H_\infty(\mathbf{X}) \geq k - 2n_1$ and $H_\infty(\mathbf{Y}) \geq k - 2n_1$ (recall $k \geq n - n^\delta$). Theorem 4.2 is now direct using Lemma 4.5.

5 Non-malleable extractors for split-state adversaries with bounded communication

Let $\mathcal{F}_{n,t} \subset \mathcal{F}_{2n}$ be the set of all functions that can be computed in the following way. Let $c = (x, y)$ be the input in $\{0,1\}^{2n}$, where x is the first n bits of c and y is the remaining n bits of c. Let Alice and Bob be two tampering adversaries, where Alice has access to x and Bob has access to y. Alice and Bob run a (deterministic) communication protocol based on x and y respectively, which can last for an arbitrary number of rounds but each party sends at most t bits. Finally, based on the transcript and x Alice outputs $x' \in \{0,1\}^n$, similarly based on the transcript and y Bob outputs $y' \in \{0,1\}^n$. The function outputs $c' = (x', y')$. The following is our main result.

Theorem 5.1. *There exists a constant $\delta > 0$ such that for all integers $n, t > 0$ with $t \leq \delta n$, there exists an efficiently computable function* nmExt : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, $m = \Omega(n)$, *such that the following holds: let $\mathbf{X}$ and $\mathbf{Y}$ be uniform independent sources each on n bits, and let h be an arbitrary tampering function in $\mathcal{F}_{n,t}$. Then, there exists a distribution $\mathcal{D}_h$ on $\{0,1\}^m \cup \{same^\star\}$ that is independent of $\mathbf{X}$ and $\mathbf{Y}$ such that*

$$|\text{nmExt}(\mathbf{X}, \mathbf{Y}), \text{nmExt}(h(\mathbf{X}, \mathbf{Y})) - \mathbf{U}_m, \text{copy}(\mathcal{D}_h, \mathbf{U}_m)| \leq 2^{-\Omega(n \log\log n / \log n)}.$$

Further, nmExt *is $2^{-\Omega(n \log\log n / \log n)}$-invertible.*

Proof. We show that any 2-source non-malleable extractor that works for min-entropy $n - 2\delta n$ can be used as the required non-malleable extractor in the above theorem. The tampering function h that is based on the communication protocol can be rephrased in terms of functions in the following way. Suppose the protocol lasts for ℓ rounds, there exist deterministic functions f_i and

g_i for $i = 1, \ldots, \ell$, and $f : \{0,1\}^n \times \{0,1\}^{2t} \to \{0,1\}^n$ and $g : \{0,1\}^n \times \{0,1\}^{2t} \to \{0,1\}^n$ such that the communication protocol between Alice and Bob corresponds to computing the following random variables: $\mathbf{S}_1 = f_1(\mathbf{X}), \mathbf{R}_1 = g_1(\mathbf{Y}, \mathbf{S}_1), \mathbf{S}_2 = f_2(\mathbf{X}, \mathbf{S}_1, \mathbf{R}_1), \ldots, \mathbf{S}_i = f_i(\mathbf{X}, \mathbf{S}_1, \ldots, \mathbf{S}_{i-1}, \mathbf{R}_1, \ldots, \mathbf{R}_{i-1}), \mathbf{R}_i = g_i(\mathbf{Y}, \mathbf{S}_1, \ldots, \mathbf{S}_i, \mathbf{R}_{i,}, \ldots, \mathbf{R}_{i-1}), \ldots, \mathbf{R}_\ell = g_\ell(\mathbf{Y}, \mathbf{S}_1, \ldots, \mathbf{S}_\ell, \mathbf{R}_1, \ldots, \mathbf{R}_{\ell-1})$.

Finally, $\mathbf{X}' = f(\mathbf{X}, \mathbf{R}_1, \ldots, \mathbf{R}_\ell, \mathbf{S}_1, \ldots, \mathbf{S}_\ell)$ and $\mathbf{Y}' = g(\mathbf{Y}, \mathbf{R}_1, \ldots, \mathbf{R}_\ell, \mathbf{S}_1, \ldots, \mathbf{S}_\ell)$ correspond to the output of Alice and the output of Bob respectively. Thus, $h(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}', \mathbf{Y}')$.

Similar to the way we argue about alternating extraction protocols, we fix random variables in the following order: Fix $\mathbf{S}_1$, and it follows that $\mathbf{R}_1$ is now a deterministic function of $\mathbf{Y}$. We fix $\mathbf{R}_1$, and thus $\mathbf{S}_2$ is now a deterministic function of $\mathbf{X}$. Thus, continuing in this way, we can fix all the random variables $\mathbf{S}_1, \ldots, \mathbf{S}_\ell$ and $\mathbf{R}_1, \ldots, \mathbf{R}_\ell$ while maintaining that $\mathbf{X}$ and $\mathbf{Y}$ are independent. Further, invoking Lemma 3.1, with probability at least $1 - 2^{-\Omega(n)}$, both $\mathbf{X}$ and $\mathbf{Y}$ have min-entropy at least $n - t - \delta n \geq n - 2\delta n$ since both parties send at most t bits.

Note that now, $\mathbf{X}' = \eta(\mathbf{X})$ for some deterministic function η and $\mathbf{Y}' = \nu(\mathbf{X})$ for some deterministic function ν. Thus, for any 2-source non-malleable extractor nmExt that works for min-entropy $n - 2\delta n$ with error ϵ, we have that there exists a distribution $\mathcal{D}_{\eta,\nu}$ over $\{0,1\}^m \cup \{same^\star\}$ that is independent of $\mathbf{X}$ and $\mathbf{Y}$ such that

$$|\mathrm{nmExt}(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}(\eta(\mathbf{X}), \nu(\mathbf{Y})) - \mathbf{U}_m, \mathrm{copy}(\mathcal{D}_{\eta,\nu}, \mathbf{U}_m)| \leq \epsilon + .2^{-\Omega(n)}$$

The theorem now follows by plugging in such a construction from a recent work of Li ([Li18], Theorem 1.12). We note the non-malleable extractor in [Li18] is indeed $2^{-\Omega(n \log\log n/\log n)}$-invertible. □

6 Efficient sampling algorithms

In this section, we provide efficient sampling algorithms for the seedless non-malleable extractor construction presented in Section 4. This is crucial to get efficient encoding algorithms for the corresponding non-malleable codes. We do not know how to invert the non-malleable extractor constructions in Theorem 4.1, but we show that the constructions can suitably modified in a way that admits efficient sampling from the pre-image of the extractor.

6.1 An invertible non-malleable extractor with respect to linear composed with interleaved adversaries

The main idea is to ensure that on fixing appropriate random variables that are generated in computing the non-malleable extractor, the source is now restricted onto a known subspace of fixed dimension (i.e., the dimension does not depend on value of the fixed random variables). Once we can ensure this, sampling from the pre-image can simply be done by first uniformly sampling the fixed random variables, and then sampling the other variables uniformly from the known subspace. To carry this out, we need an efficient construction of a linear seeded extractor that has the property that for any fixing of the seed the linear map corresponding linear seeded extractor has the same rank. Such a linear seeded extractor was constructed in prior works [CGL16, Li17] (see Theorem 3.13).

One additional care we need to take is the choice of the error correcting code we use in the advice generator construction. We ensure that the linear constraints imposed by fixing the advice string does not depend on the value of the advice string. This is subtle since the advice generator comprises of a sample from an error correction of the sources as well as the output of a linear

seeded extractor on the source. The basic idea is to remove a few sampled coordinates of the error corrected sources and show that this suffices to remove any linear dependencies.

We use the following notation: For any linear map $L : \{0,1\}^r \to \{0,1\}^s$ given by $L(\alpha) = M\alpha$ for some matrix M, we use con_L to denote a maximal set of linearly independent rows of M.

We now set up some parameters and ingredients for our construction of an invertible non-malleable extractor.

- Let $\delta > 0$ be a small enough constant and C a large constant.
- Let $\delta' = \delta/C$.
- Let $\mathcal{C}$ be a BCH code with parameters: $[n_b, n_b - t_b \log n_b, 2t_b]_2$, $t_b = \sqrt{n_b}/100$, where we fix n_b in the following way. Let dBCH be the dual code. From standard literature, it follows that dBCH is a $[n_b, t_b \log n_b, \frac{n_b}{2} - t_b\sqrt{n_b}]_2$-code. Set n_b such that $t_b \cdot \log n_b = \sqrt{n_b} \log n_b = 2n$. Let E be the encoder of dBCH. Note that by our choice of parameters, the relative minimum distance of dBCH is at least $1/3$.
- Let $n_0 = n^{\delta'}, n_1 = n_0^{c_0}, n_2 = 10n_0$, for some constant c_0 that we set below.
- Let $n_3 = n^{C\delta}, n_4 = n^{C^2\delta}/5, n_5 = n^{C^3\delta}, n_6 = n - \sum_{i=1}^{5} n_i$. We ensure that $n_6 = n(2 - o(1))$.
- Let $\mathrm{Ext}_1 : \{0,1\}^{n_1} \times \{0,1\}^{d_1} \to \{0,1\}^{\log(n_b)}$ be a $(n_1/20, 1/10)$-seeded extractor instantiated using Theorem 3.10. Thus $d_1 = c_1 \log n_1$, for some constant c_1. Let $D_1 = 2^{d_1} = n_1^{c_1}$.
- Let $\mathrm{Samp}_1 : \{0,1\}^{n_1} \to [n_b]^{D_1}$ be the sampler obtained from Theorem 3.9 using Ext_1.
- Let $\mathrm{Ext}_2 : \{0,1\}^{n_2} \times \{0,1\}^{d_2} \to \{0,1\}^{\log(n_6)}$ be a $(n_2/20, 1/n_0)$-seeded extractor instantiated using Theorem 3.10. Thus $d_2 = c_2 \log n_2$, for some constant c_2. Let $D_2 = 2_2^d$. Thus $D_2 = 2^{d_2} = n_2^{c_2}$.
- Let $\mathrm{Samp}_2 : \{0,1\}^{n_2} \to [n_6]^{D_2}$ be the sampler obtained from Theorem 3.9 using Ext_2.
- Set $c_0 = 2c_2$.
- Let $\mathrm{i}\ell\mathrm{Ext} : \{0,1\}^{D_2} \to \{0,1\}^{n_0}$ be the extractor from Theorem 7.1.
- Let $\mathrm{LExt}_0 : \{0,1\}^{2n} \times \{0,1\}^{n_0} \to \{0,1\}^{\sqrt{n_0}}$ be a linear seeded extractor instantiated from Theorem 3.15 set to extract from min-entropy $n_1/100$ and error $2^{-\Omega(\sqrt{n_0})}$.
- Let $\mathrm{Ext}_3 : \{0,1\}^{n_3} \times \{0,1\}^{d_3} \to \{0,1\}^{\log(n_6 - D_2)}$ be a $(n_3/8, 1/100)$-seeded extractor instantiated using Theorem 3.10. Thus $d_3 = C_1 \log n_3$, for some constant C_1.
- Let $\mathrm{Samp}_3 : \{0,1\}^{n_3} \to [n_6 - D_2]^{n_7}$ be the sampler obtained from Theorem 3.9 using Ext_3. Thus $n_7 = 2^{d_3} = n_3^{C_1}$.
- Let $\mathrm{Ext}_4 : \{0,1\}^{n_4} \times \{0,1\}^{d_4} \to \{0,1\}^{n_6 - n_7 - D_2}$ be a $(n_4/8, 1/100)$-seeded extractor instantiated using Theorem 3.10. Thus $d_3 = C_1 \log n_4$.
- Let $\mathrm{Samp}_4 : \{0,1\}^{n_4} \to [n_5 - n_7 - D_2]^{n_8}$ be the sampler obtained from Theorem 3.9 using Ext_4. Thus $n_8 = 2^{d_3} = n_4^{C_1}$.
- Let $\mathrm{LExt}_1 : \{0,1\}^{n_5} \times \{0,1\}^{d} \to \{0,1\}^{d_5}$, $d_5 = \sqrt{n_5}$, be a linear-seeded extractor instantiated from Theorem 3.11 set to extract from entropy $k_1 = n_2/10$ with error $\epsilon_1 = 1/10$. Thus $d = C_2 \log n_5$, for some constant C_2. Let $D = 2^d$.

- Let $\mathrm{LExt}_2 : \{0,1\}^{n_7} \times \{0,1\}^{d_5} \to \{0,1\}^{m_1}$, $m_1 = \sqrt{n_7}$ be a linear-seeded extractor instantiated from Theorem 3.11 set to extract from entropy $k_2 = n_7/100$ with error $\epsilon_2 = 2^{-\Omega(\sqrt{d_4})} = 2^{-n^{\Omega(1)}}$, such that the seed length of the extractor LExt_2 (by Theorem 3.11) is d_5.
- Let $\mathrm{ACB} : \{0,1\}^{n_{1,acb}} \times \{0,1\}^{n_{acb}} \times \{0,1\}^{h_{acb}} \to \{0,1\}^{n_{2,acb}}$, be the advice correlation breaker from Theorem 3.16 set with the following parameters: $n_{acb} = n_7, n_{1,acb} = m_1, n_{2,acb} = n_9 = D^2, t_{acb} = 2D, h_{acb} = n^\delta + d, \epsilon_{acb} = 2^{-n^{\delta'}}$, $d_{acb} = O(\log^2(n/\epsilon_{acb}))$, $\lambda_{acb} = 0$. It can be checked that by our choice of parameters, the conditions required for Theorem 3.16 indeed hold for $k_{1,acb} \geq n^{C\delta}$.
- Let $\mathrm{LExt}_3 : \{0,1\}^{n_8} \times \{0,1\}^{n_9} \to \{0,1\}^m$ be the linear seeded extractor from Theorem 3.13 set to extract from min-entropy rate $1/10$ and error $\epsilon = 2^{-n^{\Omega(1)}}$ (such that the seed-length is indeed n_9). Thus, $m = \alpha n_9$, for some small contant α that arises out of Theorem 3.13.

Algorithm 4: $\mathrm{i}\ell\mathrm{NM}(z)$

Input: Bit-string $z = (x \circ y)_\pi$ of length $2n$, where x and y are each n bit-strings, and $\pi : [2n] \to [2n]$ is a permutation.

Output: Bit string of length m.

Let $z_i = z_1 \circ z_2 \circ z_3 \circ z_4 \circ z_5 \circ z_6$, where z_i is of length n_i.
Let $T_i = \mathrm{Samp}_i(z_i)$, $i = 1, 2, 3, 4$.
Let $\overline{z}_2 = (z_6)_{T_2}$.
Let $z_2' = \mathrm{i}\ell\mathrm{Ext}(\overline{z}_2)$.
Let $z_2'' = \mathrm{LExt}_0(z, z_2')$.
For any set $Q \subseteq [2n]$, define the linear function $E : \{0,1\}^{2n} \to \{0,1\}^{|Q|}$ as $E_Q(x) = (E(x))_Q$.
Pick a subset $\overline{T_1} \subset T_1$ of size $D_1 - \sqrt{n_0}$ such that $con_{E_{\overline{T_1}}}$ is linearly independent of $con_{LExt_0(\cdot, z_2')}$. If there is no such set $\overline{T_1}$, then output 0^m.
Let $w = z_1 \circ z_2 \circ \overline{z}_2 \circ (E(z))_{\overline{T_1}} \circ z_2''$.
Let v be a $D \times d_4$ matrix, with its i'th row $v_i = \mathrm{LExt}_1(z_5, i)$.
Let z_6' be the bits in z_6 outside T_2. Let $\overline{z_6} = (z_6')_{T_3}$.
Let r be a $D \times n_4$ matrix, with its i'th row $r_i = \mathrm{LExt}_2(\overline{z_6}, v_i)$.
Let s be a $D \times m$ matrix, with its i'th row $s_i = \mathrm{ACB}(r_i, \overline{z_6}, w \circ i)$.
Let $\tilde{s} = \oplus_{i=1}^{D} s_i$.
Let z_7 be the bits in z_6 outside the coordinates $T_2 \cup T_3$.
Let $\overline{z_7} = (z_7)_{T_4}$. Let z_8 be the bits in z_6 outside the coordinates $T_2 \cup T_3 \cup T_4$.
Output $g = \mathrm{LExt}_3(\overline{z_7}, \tilde{s})$.

Theorem 6.1. *For all integers $n > 0$ there exists an explicit function $\mathrm{nmExt} : \{0,1\}^{2n} \to \{0,1\}^m$, $m = n^{\Omega(1)}$, such that the following holds: For any linear function $h : \{0,1\}^{2n} \to \{0,1\}^{2n}$, arbitrary tampering functions $f, g \in \mathcal{F}_n$, any permutation $\pi : [2n] \to [2n]$ and independent uniform sources $\mathbf{X}$ and $\mathbf{Y}$ each on n bits, there exists a distribution $\mathcal{D}_{h,f,g,\pi}$ on $\{0,1\}^m \cup \{same^\star\}$, such that*

$$|\mathrm{nmExt}((\mathbf{X} \circ \mathbf{Y})_\pi), \mathrm{nmExt}(h((f(\mathbf{X}) \circ g(\mathbf{Y}))_\pi)) - \mathbf{U}_m, \mathrm{copy}(\mathcal{D}_{h,f,g,\pi}, \mathbf{U}_m)| \leq 2^{-n^{\Omega(1)}}.$$

The proof that $\mathrm{i}\ell\mathrm{NM}$ computed by Algorithm 4 satisfies Theorem 4.1 is very similar, and we omit the details. We include a discussion of the key differences and subtleties that arise from the modifications done in the above construction as compared to Algorithm 2.

The first key difference is Step 7, where we discard some bits from the advice generator's output. The existence of the subset $\overline{T}_1$ is guaranteed by the fact that E has dual distance $t_b = \Omega(n/\log n)$. Thus, for any T, it must be that $\mathrm{Con}_{E_{T_1}}$ is a set of size $|T_1| = D_1$. Further, $con_{LExt_0(\cdot,z_2')}$ is a set with cardinality at most $\sqrt{n_0}$. Thus, indeed there exists such a set $\overline{T}_1$. An important detail to notice is that $|T_1 \setminus \overline{T_1}| = o(D_1)$ and the distance of the code computed by E is $\Omega(1)$. Thus, the fact that we discard the bits indexed by the set $T_1 \setminus \overline{T_1}$ from the string $E(\mathbf{Z})_{T_1}$ (and thus from the output of the advice generator) does not affect the correctness of the advice generator.

Another difference is that in the steps where we transform the somewhere random matrix v into a matrix with longer rows, and the subsequent step where the advice correlation breaker is applied is now done using a pseudorandomly sampled subset of coordinates from $\mathbf{Z}$ (as opposed to the entire $\mathbf{Z}$ which we did before). It is not hard to prove that this does not make a difference as long as we sample enough bits. Finally, another difference is the final step where we use a linear seeded extractor, with $\overline{\mathbf{Z}_6}$ as the seed. As done many times in the paper, we use the sum structure of $\overline{\mathbf{Z}_7}$ (into a source that depends on $\mathbf{X}$ and a source that depends on $\mathbf{Y}$) along with the fact that LExt_3 is linear seeded to show that the output is close to uniform.

We now focus on the problem of efficiently sampling from the pre-image of this extractor. The following lemma almost immediately implies a simple sampling algorithm.

Lemma 6.2. *With probability $1-2^{-n^{\Omega(1)}}$ over the fixing of the variables $z_1, z_2, \overline{z_2}, z_2'', z_3, z_4, z_5, \overline{z_6}, w$, and any $g \in \{0,1\}^m$, the set $\mathrm{i}\ell\mathrm{NM}^{-1}(g)$ is a linear subspace of fixed dimension.*

Proof. Consider any fixing of z_1, z_2, z_3, z_4. Clearly, these fix the sets T_i, $i = 1,2,3,4$. Next, note that given $\overline{z}_2$, we have the value of z_2'. We note that by Lemma 3.14 that with probability $1-2^{-n^{\Omega(1)}}$, the linear map $\mathrm{LExt}_0(, z_2')$ has full rank. Using Algorithm 2, determine the set $\overline{T_1}$ (if it exists). Fix $E(z)_{\overline{T}_1}$ and z_2'', noting that the value of w is now determined. Now given $z_5, \overline{z_6}$, we can compute $r, s, \tilde{s}$. Next observe that given g and $\tilde{s}$, Theorem 3.13 implies the value of $\overline{z_7}$ belongs to a subspace whose dimension does not depend on the values of g and $\tilde{s}$. Finally, we are left to see how to compute z_8. Note that the constraints on z_8 are imposed by the fixings of z_2'' and $E(C)_{\overline{T_1}}$. However, by construction (Step 7 of our algorithm), the number of independent linear constraints on z_8 is exactly equal to D_1 as long as $\mathrm{LExt}_0(, z_2')$ has full rank (which as noted before occurs with probability at least $1-2^{-n^{\Omega(1)}}$). This completes the proof. □

Given Lemma 6.2, the sampling algorithm is now straightforward:

Input $g \in \{0,1\}^m$; Output z that is uniform on the set $\mathrm{i}\ell\mathrm{NM}^{-1}(g)$.

1. Sample z_i, $i = 1,2,3,4,5$ uniformly at random. Compute T_1, T_2, T_3, T_4 following Algorithm 2.
2. Sample $\overline{z_2}$ uniformly, and compute z_2'. Further, sample z_2'' uniformly.
3. Compute $\overline{T_1}$, and sample $(E(z))_{\overline{T_1}}$ uniformly at random.
4. Compute $w, v, r, s, \tilde{s}$ using Algorithm 2.
5. Sample $\overline{z_7}$ from $(\mathrm{LExt}_3(\cdot, \tilde{s}))^{-1}(g)$ efficiently using Theorem 3.13.
6. Sample z_8 as described in Lemma 6.2. Compute the string z_6.
7. Output $z = z_1 \circ z_2 \circ z_3 \circ z_4 \circ z_5 \circ z_6$.

7 Extractors for interleaved sources

Our techniques yield improved explicit constructions of extractors for interleaved sources. Our extractor works when both sources have entropy at least $2n/3$, and outputs $\Omega(n)$ bits that are $2^{-n^{\Omega(1)}}$-close to uniform.

The following is our main result.

Theorem 7.1. *For any constant $\delta > 0$ and all integers $n > 0$, there exists an efficiently computable function* $\mathrm{i}\ell\mathrm{Ext} : \{0,1\}^{2n} \to \{0,1\}^m$, *$m = \Omega(n)$, such that for any two independent sources $\mathbf{X}$ and $\mathbf{Y}$, each on n bits with min-entropy at least $(2/3+\delta)n$, and any permutation $\pi : [2n] \to [2n]$, we have*

$$|\mathrm{i}\ell\mathrm{Ext}((\mathbf{X} \circ \mathbf{Y})_\pi) - \mathbf{U}_m| \leq 2^{-n^{\Omega(1)}}.$$

We use the rest of the section to prove Theorem 7.1. An important ingredient in our construction is an explicit somewhere condenser for high-entropy sources constructed in the works of Barak et al. [BRSW12] and Zuckerman [Zuc07].

Theorem 7.2. *For all constants β, δ and all integers $n > 0$, there exists an efficiently computable function* $\mathrm{Con} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^\ell$, *$d = 0(1)$ and $\ell = \Omega(n)$ such that the following holds: for any $(n, \delta n)$-source $\mathbf{X}$ there exists a $y \in \{0,1\}^d$ such that* $\mathrm{Con}(\mathbf{X}, y)$ *is $2^{-\Omega(n)}$-close to a source with min-entropy $(1-\beta)\ell$.*
We call such a function Con *to be a $(\delta, 1-\beta)$-condenser.*

We prove that Algorithm 5 computes the required extractor. We begin by setting up some ingredients and parameters.

- Let $\kappa > 0$ be a small enough constant.
- Let $n_1 = (2/3 + \delta/2)n$ and $n_2 = n^{5\kappa}$.
- Let β be a parameter which we fix later. Let $\mathrm{Con} : \{0,1\}^{n_1} \times \{0,1\}^d \to \{0,1\}^\ell$ be a $(\delta/4, 1-\beta)$-condenser instantiated from Theorem 7.2. Thus $\ell = n/C'$, for some constant C' that depends on δ, β. Let $D = 2^d$. Note that $D = O(1)$.
- Let $\mathrm{LExt}_1 : \{0,1\}^{2n} \times \{0,1\}^\ell \to \{0,1\}^{n_2}$ be the linear seeded extractor from Theorem 3.13 set to extract from min-entropy rate $1/12$ and error $\epsilon_1 = 2^{-2\beta\ell}$. The seed-length is at most $3C\beta\ell$, some constant C that arises out of Theorem 3.13. We choose $\beta = min\{1/3C, \gamma\}$, where γ is the constant in Theorem 3.13. Note that the seed-length of LExt_1 is indeed at most ℓ.
- Let $\mathrm{ACB} : \{0,1\}^{n_{1,acb}} \times \{0,1\}^{n_{acb}} \times \{0,1\}^{h_{acb}} \to \{0,1\}^{n_{2,acb}}$, be the advice correlation breaker from Theorem 3.16 set with the following parameters: $n_{acb} = 2n, n_{1,acb} = n_2, n_{2,acb} = n_3 = n^{2\kappa}, t_{acb} = D, h_{acb} = d, \epsilon_{acb} = 2^{-n^\kappa}, d_{acb} = O(\log^2(n/\epsilon_{acb})), \lambda_{acb} = 0$. It can be checked that by our choice of parameters, the conditions required for Theorem 3.16 indeed hold for $k_{1,acb} \geq n^{2\kappa}$.
- Let $\mathrm{LExt}_2 : \{0,1\}^{2n} \times \{0,1\}^{n_3} \to \{0,1\}^m$, $m = \Omega(n)$, be a linear-seeded extractor instantiated from Theorem 3.11 set to extract from entropy $k_1 = n/10$ with error $\epsilon_1 = 2^{-\alpha\sqrt{n_3}}$, for an appropriately picked small constant α.

Algorithm 5: $\mathrm{i}\ell\mathrm{Ext}(z)$

Input: Bit-string $z = (x \circ y)_\pi$ of length $2n$, where x and y are each n bit-strings, and $\pi : [2n] \to [2n]$ is a permutation.

Output: Bit string of length m.

Let $z_1 = \mathrm{Slice}(z, n_1)$.
Let v be a $D \times n_2$ matrix, with its i'th row $v_i = \mathrm{Con}(z_1, i)$.
Let r be a $D \times n_3$ matrix, with its i'th row $r_i = \mathrm{LExt}_1(z, v_i)$.
Let s be a $D \times m$ matrix, with its i'th row $s_i = \mathrm{ACB}(r_i, z, i)$.
Let $\tilde{s} = \oplus_{i=1}^{D} s_i$.
Output $\mathrm{LExt}_2(z, \tilde{s})$.

We use the following notation: Let $\mathbf{X}_1$ be the bits of $\mathbf{X}$ in $\mathbf{Z}_1$ and $\mathbf{X}_2$ be the remaining bit of $\mathbf{X}$. Let $\mathbf{Y}_1$ be the bits of $\mathbf{Y}$ in $\mathbf{Z}_1$ and $\mathbf{Y}_2$ be the remaining bits of $\mathbf{Y}$. Without loss of generality assume $|\mathbf{X}_1| \geq |\mathbf{Y}_1|$. Define $\overline{\mathbf{X}} = (\mathbf{X} \circ 0^n)_\pi$ and $\overline{\mathbf{Y}} = (\mathbf{Y} \circ 0^n)_\pi$. Further, let $\overline{\mathbf{X}}_1 = \mathrm{Slice}(\overline{\mathbf{X}}, n_1)$ and $\overline{\mathbf{Y}}_1 = \mathrm{Slice}(\overline{\mathbf{Y}}, n_1)$. It follows that $\mathbf{Z} = \overline{\mathbf{X}} + \overline{\mathbf{Y}}$, and $\mathbf{Z}_1 = \overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}$. Further, let $k_x = k_y = (2/3 + \delta)n$.

We begin by proving the following claim.

Claim 7.3. *Conditioned on the random variables* $\mathbf{X}_1, \mathbf{Y}_1, \{\mathrm{LExt}_1(\overline{\mathbf{X}}, \mathrm{Con}(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, i))\}_{i=1}^{D}$, *the following hold:*

- *the matrix* $\mathbf{R}$ *is* $2^{-\Omega(n)}$*-close to a somewhere random source,*
- $\mathbf{R}$ *is a deterministic functions of* $\mathbf{Y}$*,*
- $H_\infty(\mathbf{X}) \geq \delta n/4$*,* $H_\infty(\mathbf{Y}) \geq n/6$*.*

Proof. By construction, we have that for any $j \in [D]$,

$$\begin{aligned}
\mathbf{R}_j &= \mathrm{LExt}_1(\mathbf{Z}, \mathrm{Con}(\mathbf{Z}_1, j)) \\
&= \mathrm{LExt}_1(\overline{\mathbf{X}} + \overline{\mathbf{Y}}, \mathrm{Con}(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, j)) \\
&= \mathrm{LExt}_2(\overline{\mathbf{X}}, \mathrm{Con}(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, j)) + \mathrm{LExt}_2(\overline{\mathbf{Y}}, \mathrm{Con}(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, j))
\end{aligned}$$

Fix the random variables $\mathbf{Y}_1$, and $\overline{\mathbf{Y}}$ has min-entropy at least $k_y - n_1/2 \geq n/6 + 3\delta n/4$. Further, note that $\overline{\mathbf{X}_1}$ has min-entropy at least $n_1/2 - (n - k_x) \geq \delta n/4$. Now, by Theorem 7.2, we know that there exists a $j \in [D]$ such that $\mathrm{Con}(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, j)$ is $2^{-\Omega(n)}$-close to a source with min-entropy at least $(1 - \beta)\ell$. Further, note that $\mathbf{V}$ is a deterministic function of $\mathbf{X}$.

Now, since LExt_1 is a strong seeded extractor set to extract from min-entropy $n/6$, it follows that

$$|\mathrm{LExt}_1(\overline{\mathbf{Y}}, \mathrm{Con}(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, j)) - \mathbf{U}_{n_2}| \leq 2^{\beta\ell}\epsilon_1 + 2^{-\Omega(n)} \leq 2^{-\beta\ell+1}.$$

We now fix the random variables $\overline{\mathbf{X}}_1$ and note that $\mathrm{LExt}_1(\overline{\mathbf{Y}}, \mathrm{Con}(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, j))$ continues to be $2^{-\Omega(\ell)}$-close to $\mathbf{U}_{n_2}$. This follows from the fact that LExt_1 is a strong seeded extractor. Note that the random variables $\{\mathrm{Con}(\overline{\mathbf{X}}_1 + \overline{\mathbf{Y}}_1, i)) : i \in [D]\}$ are now fixed. Next, fix the random variables $\{\mathrm{LExt}_1(\overline{\mathbf{X}}, \mathrm{Con}(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, i))\}_{i=1}^{D}$ noting that they are deterministic functions of $\mathbf{X}$. Thus $\mathbf{R}_j$ is $2^{-\Omega(n)}$-close to $\mathbf{U}_{n_2}$ and for any $i \in [D]$, the random variables $\mathbf{R}_i$ are deterministic functions of $\mathbf{Y}$. Finally, note that $\mathbf{X}$ and $\mathbf{Y}$ remain independent after these conditionings, and $H_\infty(\mathbf{X}) \geq k_x - n_1 - Dn_2$ and $H_\infty(\mathbf{Y}) \geq k_y - n_1/2$. $\square$

The next claim almost gets us to Theorem 7.1.

Claim 7.4. *There exists $j \in [D]$ such that*

$$\mathbf{S}_j, \{\mathbf{S}_i\}_{i\in[D]\setminus j}, \mathbf{X} \approx_{2^{-n^{\Omega(1)}}} \mathbf{U}_{n_3}, \{\mathbf{S}_i\}_{i\in[D]\setminus j}, \mathbf{X}.$$

Proof. Fix the random variables: $\mathbf{X}_1, \mathbf{Y}_1, \{\mathrm{LExt}_1(\overline{\mathbf{X}}, \mathrm{Con}(\overline{\mathbf{X}_1} + \overline{\mathbf{Y}_1}, i))\}_{i=1}^{D}$. By Claim 7.3 we have that $\mathbf{R}$ is a deterministic function of $\mathbf{Y}$, and with probability at least $1-2^{-\Omega(n)}$, there exists $j \in [D]$ such that $\mathbf{R}_j$ is $2^{-n^{\Omega(1)}}$-close to uniform, and $H_\infty(\overline{\mathbf{X}}) \geq \delta n/4$. Recall that $\mathbf{Z} = \overline{\mathbf{X}} + \overline{\mathbf{Y}}$. It now follows by Theorem 3.16 that

$$\mathrm{ACB}(\mathbf{R}_j, \mathbf{Z}, \mathbf{W} \circ j), \{\mathrm{ACB}(\mathbf{R}_i, \overline{\mathbf{X}} + \overline{\mathbf{Y}}, \mathbf{W} \circ i)\}_{i\in[D]\setminus j}, \mathbf{X} \approx_{2^{-n^{\Omega(1)}}} \mathbf{U}_{n_3}, \{\mathrm{ACB}(\mathbf{R}_i, \overline{\mathbf{X}} + \overline{\mathbf{Y}}, \mathbf{W} \circ i)\}_{i\in[D]\setminus j}, \mathbf{X}.$$

□

It follows by Claim 7.4 that $\widetilde{\mathbf{S}}$ is $2^{-n^{\Omega(1)}}$-close to uniform even conditioned on $\mathbf{X}$. Thus, noting that $\mathrm{LExt}_2(\mathbf{Z}, \widetilde{\mathbf{S}}) = \mathrm{LExt}_2(\overline{\mathbf{X}}, \widetilde{\mathbf{S}}) + \mathrm{LExt}_2(\overline{\mathbf{Y}}, \widetilde{\mathbf{S}})$, it follows that we can fix $\widetilde{\mathbf{S}}$ and $\mathrm{LExt}_2(\overline{\mathbf{X}}, \widetilde{\mathbf{S}})$ remains $2^{-n^{\Omega(1)}}$-close to uniform and is a deterministic function of $\mathbf{X}$. Next, we fix $\mathrm{LExt}_2(\overline{\mathbf{Y}}, \widetilde{\mathbf{S}})$ without affecting the distribution of $\mathrm{LExt}_2(\overline{\mathbf{X}}, \widetilde{\mathbf{S}})$. It follows that $\mathrm{LExt}_2(\mathbf{Z}, \widetilde{\mathbf{S}})$ is $2^{-n^{\Omega(1)}}$-close to uniform. This completes the proof of Theorem 7.1.

References

[ADKO15] D. Aggarwal, Y. Dodis, T. Kazana, and M. Obremski. Non-malleable reductions and applications. To appear in STOC, 2015.

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

[AGM+15] Shashank Agrawal, Divya Gupta, Hemanta K. Maji, Omkant Pandey, and Manoj Prabhakaran. A rate-optimizing compiler for non-malleable codes against bit-wise tampering and permutations. In *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I*, pages 375–397, 2015.

[BDG+18] Marshall Ball, Dana Dachman-Soled, Siyao Guo, Tal Malkin, and Li-Yang Tan. Non-malleable codes for small-depth circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 2018.

[BDKM16] Marshall Ball, Dana Dachman-Soled, Mukul Kulkarni, and Tal Malkin. Non-malleable codes for bounded depth, bounded fan-in circuits. In *TCC*, 2016.

[BRSW12] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl-Wilson construction. *Annals of Mathematics*, 176(3):1483–1543, 2012. Preliminary version in STOC '06.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[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.

[CGL16] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *STOC*, 2016.

[CKOS18] Eshan Chattopadhyay, Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, and Sruthi Sekar. Privacy amplification from non-malleable codes. *IACR Cryptology ePrint Archive*, 2018:293, 2018.

[CL16] Eshan Chattopadhyay and Xin Li. Extractors for sumset sources. In *STOC*, 2016.

[CL17] Eshan Chattopadhyay and Xin Li. Non-malleable codes and extractors for small-depth circuits, and affine functions. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1171–1184. ACM, 2017.

[CMTV15] Sandro Coretti, Ueli Maurer, Björn Tackmann, and Daniele Venturi. From single-bit to multi-bit public-key encryption via non-malleable codes. In *Theory of Cryptography Conference*, pages 532–560. Springer, 2015.

[Coh15] Gil Cohen. Local correlation breakers and applications to three-source extractors and mergers. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, 2015.

[CZ14] Eshan Chattopadhyay and David Zuckerman. Non-malleable codes against constant split-state tampering. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science*, pages 306–315, 2014.

[CZ16a] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *STOC*, 2016.

[CZ16b] Eshan Chattopadhyay and David Zuckerman. New extractors for interleaved sources. In *CCC*, 2016.

[DKO13] Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In *CRYPTO (2)*, pages 239–257, 2013.

[DORS08] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 38:97–139, 2008.

[DPW18] Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. *J. ACM*, 65(4):20:1–20:32, April 2018.

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

[GK18a] Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 685–698. ACM, 2018.

[GK18b] Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing for general access structures. In *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part I*, pages 501–530, 2018.

[GMW18] Divya Gupta, Hemanta K Maji, and Mingyuan Wang. Constant-rate non-malleable codes in the split-state model. Technical report, Technical Report Report 2017/1048, Cryptology ePrint Archive, 2018.

[GPR16] Vipul Goyal, Omkant Pandey, and Silas Richelson. Textbook non-malleable commitments. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 1128–1141. ACM, 2016.

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

[KOS17] Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, and Sruthi Sekar. Four-state non-malleable codes with explicit constant rate. In *Theory of Cryptography Conference*, pages 344–375. Springer, 2017.

[Li15] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. Technical Report TR15-125, ECCC, 2015.

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 168–177. IEEE, 2016.

[Li17] Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 1144–1156, 2017.

[Li18] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. *Electronic Colloquium on Computational Complexity (ECCC)*, 2018.

[MW97] Ueli Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In *Advances in Cryptology — CRYPTO '97*, volume 1294, pages 307–321, August 1997.

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

[Rao09] Anup Rao. Extractors for low-weight affine sources. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity*, 2009.

[RRV02] Ran Raz, Omer Reingold, and Salil Vadhan. Extracting all the randomness and reducing the error in Trevisan's extractors. *JCSS*, 65(1):97–128, 2002.

[RS18] Peter M. R. Rasmussen and Amit Sahai. Expander graphs are non-malleable codes. *CoRR*, 2018.

[RY11] Ran Raz and Amir Yehudayoff. Multilinear formulas, maximal-partition discrepancy and mixed-sources extractors. *Journal of Computer and System Sciences*, 77:167–190, 2011.

[Tre01] Luca Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, pages 860–879, 2001.

[TV00] Luca Trevisan and Salil P. Vadhan. Extracting Randomness from Samplable Distributions. In *IEEE Symposium on Foundations of Computer Science*, pages 32–42, 2000.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11:345–367, 1997.

[Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, pages 103–128, 2007.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il