Computational and Information-Theoretic Two-Source (Non-Malleable) Extractors

Divesh Aggarwal[*] Maciej Obremski[†] João Ribeiro[‡] Mark Simkin[§]
Luisa Siniscalchi[¶]

Abstract

Two-source non-malleable extractors are pseudorandom objects which extract randomness even when an adversary is allowed to learn the behavior of the extractor on tamperings of the input weak sources, and they have found important applications in non-malleable coding and secret sharing. We begin by asking how hard it is to improve upon the best known constructions of such objects (Chattopadhyay, Goyal, Li, STOC 2016, and Li, STOC 2017). We show that even small improvements to these constructions lead to explicit low-error two-source extractors for very low linear min-entropy, a longstanding open problem in pseudorandomness.

Given the result above in the information-theoretic setting, we turn to studying two-source non-malleable extractors *in the computational setting*, namely in the CRS model first considered in (Garg, Kalai, Khurana, Eurocrypt 2020). We enforce that both the sampling process for the input sources and the tampering functions must be efficient, but we do not necessarily put such a constraint on the adversary distinguishing the output of the extractor from uniform. We obtain results about two-source non-malleable extractors in the CRS model under different types of hardness assumptions:

- Under standard assumptions, we show that small improvements upon state-of-the-art statistical two-source non-malleable extractors also yield explicit low-error two-source non-malleable extractors *in the CRS model* for low min-entropy against computationally unbounded distinguishers. Remarkably, all previous results on computational extractors require much stronger assumptions;
- Under a quasi-polynomial hardness assumption, we give explicit constructions of low-error two-source non-malleable extractors in the CRS model with much lower min-entropy requirements than their best statistical counterparts, against a computationally bounded distinguisher;
- Assuming the existence of nearly optimal collision-resistant hash functions, we give a simple explicit construction of a low-error two-source non-malleable extractors in the CRS model for very low min-entropy, against a computationally *unbounded* distinguisher.

1 Introduction

The problem of constructing explicit low-error two-source extractors for low min-entropy sources was an important focus of research in pseudorandomness over more than 25 years, with fundamental connections to combinatorics. A deep line of work, culminating in the groundbreaking work of

[*] National University of Singapore. dcsdiva@nus.edu.sg
[†] National University of Singapore. obremski.math@gmail.com
[‡] Imperial College London. j.lourenco-ribeiro17@imperial.ac.uk
[§] Aarhus University. simkin@cs.au.dk
[¶] Concordium Blockchain Research Center, Aarhus University. lsiniscalchi@cs.au.dk

Chattopadhyay and Zuckerman [CZ19], succeeded in constructing explicit 1-bit two-source extractors for polylogarithmic min-entropy with polynomially small error (this was quickly improved to larger output length [Li16] and near-logarithmic min-entropy [BDT17], which is nearly optimal). In particular, these results yield explicit constructions of bipartite Ramsey graphs with very good parameters, which is a fundamental problem in combinatorics.

Unfortunately, the results described above are not appropriate for cryptographic applications, which are some of the main motivations for constructing randomness extractors. Indeed, the running time of the proposed constructions is polynomial in $1/\varepsilon$, where ε denotes the error of the extractor. Therefore, in order to ensure the constructions are asymptotically efficient, the error must be non-negligible in the source length.

In the low-error (i.e., negligible in the source length) regime, much less is known. Chor and Goldreich [CG88] showed that the inner product function is a low-error two-source extractor for n-bit sources with min-entropy $(1/2+\gamma)n$, where $\gamma > 0$ is an arbitrarily small constant. This was improved by Bourgain [Bou05], who gave an explicit low-error two-source extractor for sources with min-entropy $(1/2-\gamma)n$, where $\gamma > 0$ is a small constant. An improved analysis by Lewko [Lew19] shows that Bourgain's extractor can handle sources with min-entropy $4n/9$. In an incomparable result, Raz [Raz05] gave an explicit low-error two-source extractor where one of the sources must have min-entropy $(1/2+\gamma)n$ for an arbitrarily small constant $\gamma > 0$, while the other source is allowed to have logarithmic min-entropy. A standard application of the probabilistic method shows that (inefficient) low-error two-source extractors exist for polylogarithmic min-entropy. However, given the state-of-the-art, it remains a major open problem to construct explicit low-error two-source extractors for min-entropy δn for a small constant $\delta > 0$. Motivated by this, several works have studied related problems, such as constructing low-error two-source condensers with small entropy gap for low min-entropy sources [Rao08, BCDT19], showing reductions from explicit two-source extractors to other pseudorandom objects with as yet unattained parameters [ZB11, BCD$^+$18], and constructing low-error two-source extractors for low min-entropy in the computational setting, under different hardness assumptions [TV00, KLR09, GKK19].

More recently, a strengthened version of two-source extractors, called two-source *non-malleable* extractors (also known as seedless non-malleable extractors, in contrast with seeded non-malleable extractors [DW09]), was introduced by Cheraghchi and Guruswami [CG17]. Roughly speaking, a function $\mathsf{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is said to be a non-malleable extractor if the output of the extractor remains close to uniform (in statistical distance), even conditioned on the output of the extractor at several inputs correlated with the original sources. In other words, we require that

$$\begin{aligned}\mathsf{nmExt}(X,Y), \mathsf{nmExt}(f_1(X), g_1(Y)), \dots, \mathsf{nmExt}(f_r(X), g_r(Y))&\\ \approx_\varepsilon U_m, \mathsf{nmExt}(f_1(X), g_1(Y)), \dots, \mathsf{nmExt}(f_r(X), g_r(Y)),&\end{aligned}$$

where X and Y are independent sources with enough min-entropy, $f_i, g_i : \{0,1\}^n \to \{0,1\}^n$ for $i = 1, \dots, r$ are arbitrary tampering functions such that (f_i, g_i) has no fixed points, U_m is uniform over $\{0,1\}^m$ and independent of the rest, and $\approx_\varepsilon$ means the two distributions are ε-close in statistical distance (for small ε). The motivation for studying explict two-source non-malleable extractors stems from the fact that they directly yield explicit split-state non-malleable codes and secret sharing schemes [DPW18, GK18] (provided the extractor also supports efficient preimage sampling).

The situation regarding explicit low-error two-source non-malleable extractors is, as expected, much direr than for regular two-source extractors. For n-bit sources, the state-of-the-art constructions by Chattopadhyay, Goyal, and Li [CGL16] and Li [Li17] require min-entropy $(1-\mathrm{poly}(1/r))n$ to handle r tamperings. In particular, if r is constant, then the existing explicit non-malleable extractors require min-entropy $(1-\gamma)n$ for a very small constant $\gamma > 0$.

Our contributions. Making a parallel with the approach towards regular low-error two-source extractors, it is natural to consider the following question:

How hard is it to improve upon the best known low-error two-source non-malleable extractors?

In the first part of our work, we show that small improvements to the parameters of [CGL16, Li17], *without even requiring efficient preimage sampling*, lead to explicit low-error two-source extractors for min-entropy δn with a very small constant $\delta > 0$. Put differently, the constructions of [CGL16, Li17] are almost the best we can hope for without solving a longstanding open problem in pseudorandomness along the way. From another perspective, this result also gives an additional compelling reason for studying two-source non-malleable extractors, besides their applications to non-malleable codes and secret sharing.

Given our first result above, we turn to constructing two-source non-malleable extractors for lower min-entropy *in the computational setting*. More precisely, motivated by recent work of Garg, Kalai, and Khurana [GKK19] on the Common Reference String (CRS) model, we study the following open-ended question:

Under which hardness assumptions can we construct two-source non-malleable extractors for low min-entropy in the CRS model, and with which parameters?

At a high-level, in the CRS model, a CRS is sampled once and for all, and we consider three adversaries, all with full access to the CRS: The first adversary (the sampler) samples independent randomness sources (with enough min-entropy), the second adversary (the tamperer) is allowed to tamper with the sources generated by the sampler, and the third adversary (the distinguisher) attempts to distinguish the output of the extractor from a uniform distribution given also access to the extractor's outputs on tampered versions generated by the tamperer. While we always constrain the sampler and tamperer to be computationally bounded, at times we will allow the distinguisher to be computationally unbounded. We present three constructions of two-source non-malleable extractors in the CRS model based on hardness assumptions of different flavors, remarkably including one result that can be instantiated with a large range of standard assumptions. More precisely, from weakest to strongest assumption:

1. Assuming the existence of *any* family of collision-resistant hash functions, small improvements on the constructions of [CGL16, Li17] also yield explicit low-error two-source *non-malleable* extractors in the CRS model for low min-entropy against a computationally unbounded distinguisher. Previous works on computational extractors require strong hardness assumptions [KLR09, DRV12, GKK19], or put severe constraints on the trade-off between security and running time of the extractor [TV00].

2. Assuming quasi-polynomial hardness of the DDH assumption[1], we construct a low-error two-source non-malleable extractor in the CRS model for much lower min-entropy and handling many more tamperings than its statistical counterparts in [CGL16, Li17], against a computationally bounded distinguisher.

3. Assuming the existence of nearly optimal collision-resistant hash functions, we give a simple low-error two-source non-malleable extractor in the CRS model for very low min-entropy against a computationally *unbounded* distinguisher;

[1] By quasi-polynomial hardness of the DDH assumption we mean no algorithm running in time $n^{\log n}$ solves the Decisional Diffie-Hellman problem with non-negligible (in n) advantage.

We put our results in context of previous work in Section 1.1, and provide a more technical discussion of our results and techniques in Section 1.2.

1.1 Comparison to previous work

Our connection between explicit two-source non-malleable extractors with slightly improved parameters and explicit low-error two-source extractors for low min-entropy fits within a set of results that show improved versions of different pseudorandom objects imply such two-source extractors. Zewi and Ben-Sasson [ZB11] show an implication of this type from explicit affine seeded extractors with good parameters, assuming the Polynomial Freiman-Ruzsa conjecture. More recently, Ben-Aroya et al. [BCD$^+$18] adapt the approach of [CZ19] to show explicit *seeded* non-malleable extractors with improved seed length lead to explicit low-error two-source extractors for low min-entropy. Although the main result of [BCD$^+$18] yields two-source extractors with *unbalanced* sources, it is possible to obtain two-source extractors for balanced sources by further strengthening the parameters of the initial seeded non-malleable extractor.

Some works have also focused on constructing extractors in computational settings. Early work by Trevisan and Vadhan [TV00] can be interpreted as giving explicit extractors for a single source with logarithmic min-entropy in the CRS model (a similar remark was already made in [DRV12]). Under strong hardness assumptions, they also construct explicit deterministic extractors for high min-entropy sources samplable by bounded-size circuits. However, they prove the strong negative result that, for both settings above, the running time of the extractor must be larger than the time needed to sample the source. In particular, if one wishes to extract randomness from all efficiently samplable sources in the CRS model, then the extractor in question cannot be efficient. Dodis, Ristenpart, and Vadhan [DRV12] implicitly show that this negative result can be avoided if one instead focuses on single-source *condensers* in the CRS model, assuming the existence of nearly optimal collision-resistant hash functions. In a different setting, Kalai, Li, and Rao [KLR09] studied two-source extractors for information-theoretic sources (without a CRS) against a *computationally bounded* distinguisher. They succeed in constructing such extractors for linear min-entropy sources, under the assumption that nearly optimal exponentially secure one-way permutations exist. To avoid the reliance on such strong assumptions, Garg, Kalai, and Khurana [GKK19] initiate the study of two-source extractors in the CRS model. They focus solely on the setting with efficiently samplable sources and computationally bounded distinguishers, and assume the subexponential hardness of the DDH assumption[2] (a weaker assumption relative to that required by [KLR09]). Under these conditions, they construct a special type of two-source extractor that lies between seeded and two-source non-malleable extractors, in the sense that neither source is required to be uniform, but only the second source is allowed to be tampered. They give such explicit extractors in the CRS model with balanced sources for min-entropy matching that of the best explicit statistical two-source extractors. Then, they exploit this extractor and results of [BCD$^+$18] to construct an extractor of the same type for unbalanced sources with lower min-entropy. We remark that the assumption in [GKK19] can be weakened to quasi-polynomial hardness of the DDH assumption if one is aiming to match the min-entropy requirements of the best explicit statistical two-source extractors, as is done in the first part of [GKK19]. To go below such min-entropy requirements, a subexponential hardness assumption appears to be necessary.

In this work, we consider the same setting as [GKK19], although for some of our constructions we allow the distinguisher to be computationally unbounded. We focus on constructing two-source non-malleable extractors, while, as mentioned above, [GKK19] considers only one-sided tampering.

[2] By subexponential hardness of the DDH assumption we mean that there exists a constant $c \in (0,1)$ such that no algorithm running in time 2^{n^c} solves the Decisional Diffie-Hellman problem with non-negligible (in n) advantage.

Furthermore, we study what kind of constructions and parameters are achievable under different types of hardness assumptions. Recall that our first result shows how to construct two-source non-malleable extractors in the CRS model for low min-entropy (against an unbounded distinguisher) from collision-resistant hash functions and statistical two-source non-malleable extractors for very high min-entropy. In comparison, as discussed above, previous work on low-error computational (even malleable) extractors for low min-entropy requires at least subexponentially secure hardness assumptions. For our second construction, we make use of a quasi-polynomial hardness assumption, and similarly to [GKK19] consider a computationally bounded distinguisher. We are able to match the min-entropy requirements of the best explicit statistical two-source extractors. Finally, our last construction of a two-source non-malleable extractor in the CRS model (against an unbounded distinguisher) is extremely simple, but requires the same strong hardness assumption as [DRV12] (nearly optimal collision-resistant hash functions).

1.2 Technical overview

In this section, we provide a more in-depth overview of our contributions.

1.2.1 Slightly better two-source non-malleable extractors imply great two-source extractors.

As our first contribution, we show that small improvements to the parameters of the best known explicit two-source non-malleable extractors [CGL16, Li17] yield explicit low-error (malleable) two-source extractors for sources with low linear min-entropy. This result can be seen from two perspectives: On the one hand, it suggests that improving significantly upon [CGL16, Li17] is very challenging, as it would entail solving a longstanding open problem. On the other hand, it shows we are tantalizingly close to making significant progress on constructing explicit low-error two-source extractors for low min-entropy, and provides yet one more application of two-source non-malleable extractors.

The starting points of our construction are an explicit two-source non-malleable extractor nmExt for high min-entropy sources handling enough tamperings, and two independent n-bit sources X and Y with min-entropy δn, for some small constant $\delta > 0$. In other words, X and Y have min-entropy rate δ. If we had access to a uniformly random seed (as is the case for seeded extractors), we could apply seeded condensers to transform X and Y into shorter sources X' and Y' which are (statistically close to) sources with high min-entropy rate. This would allow us to compute $\mathsf{nmExt}(X', Y')$ and conclude that its output is close to uniform, without even exploiting the non-maleability of nmExt. Unfortunately, the strategy above is impossible to realize without a uniform seed, as is our case.

Although deterministic condensers do not exist, there does exist a deterministic object with related properties, called a *somewhere-condenser.* Roughly speaking, a somewhere-condenser SCond receives as input a source X with min-entropy rate δ, and outputs $X' = \mathsf{SCond}(X)$ composed of ℓ blocks $(X'_1, X'_2, \dots, X'_\ell)$, with the property that for some random variable I it holds that X'_I is statistically close to a source with min-entropy rate $1 - \gamma$. Importantly, we can write the blocks X'_i for $i \neq I$ as randomized tamperings of the *good* block X'_I. Analogously, computing $Y' = \mathsf{SCond}(Y)$ leads to ℓ blocks $(Y'_1, Y'_2, \dots, Y'_\ell)$ and a random index J such that Y'_J is close to a source with high min-entropy rate, and Y'_j for $j \neq J$ can be written as randomized tamperings of Y'_J. Combined with the non-malleability properties of nmExt, these observations naturally lead to the candidate two-source extractor Ext given by

$$\mathsf{Ext}(X,Y) = \bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(X'_i\|p_i, Y'_j\|p_j), \tag{1}$$

where p_i and p_j are the binary representations of indices i and j, respectively. Intuitively, these suffixes are added to ensure that the tamperings induced by the somewhere-condenser SCond do not have fixed points. In order to prove that Ext indeed extracts from the low min-entropy sources X and Y, it is enough to show that $\mathsf{nmExt}(X'_I\|p_I, Y'_J\|p_J)$ is close to uniform given the side information $\mathsf{nmExt}(X'_i\|p_i, Y'_j\|p_j)$ for $(i,j) \neq (I,J)$. This is equivalent to requiring that nmExt resists $\ell^2 - 1$ tamperings. Explicit constructions of somewhere-condensers with good parameters are known [BKS+10, Raz05, Zuc06, Li11]. In particular, we can take the number of blocks ℓ to be a constant depending only on δ and γ, and the error to be exponentially small in the length of the output blocks. Therefore, our argument goes through provided we have an explicit two-source non-malleable extractor for min-entropy rate $1-\gamma$ handling ℓ^2-1 tamperings. Moreover, the resulting extractor Ext has low error if nmExt does so.

Overall, our reduction above trades the number of tamperings handled with lowering the original min-entropy requirement of the underlying two-source non-malleable extractor. We leave formal details of our general result for Section 3, and present here one important case.

Theorem 1 (Informal)**.** *For every constant $\gamma > 0$ there exists a constant C_δ such that if there exists an explicit low-error two-source non-malleable extractor nmExt for min-entropy rate $1-\gamma$ handling C_δ tamperings, then there exists an explicit low-error two-source extractor for min-entropy rate δ. In particular, if nmExt handles $r = \omega(1)$ tamperings, then for every constant $\delta > 0$ there exists an explicit low-error two-source extractor for min-entropy rate δ.*

Interestingly, by [Li17] (see Proposition 5) we have explicit constructions of low-error non-malleable extractors for constant min-entropy rate $1-\gamma$ (with γ a small constant) and a constant number of tamperings, and $r = \omega(1)$ tamperings for *any* min-entropy rate $1 - o(1)$. If this result is improved to handle *any* superconstant number of tamperings with *some* constant min-entropy rate, then Corollary 2 implies that we have explicit low error two-source extractors for *any* linear min-entropy rate. Even improving the number of tamperings handled to a large enough constant for some constant min-entropy rate would already yield significantly improved explicit low-error two-source extractors. We remark also that small improvements on the two-source non-malleable extractor from [CGL16] are enough to make our argument go through as well. We discuss this in detail in Section 3. Finally, note that the two-source non-malleable extractors we require for our reduction are very far from optimal. Indeed, it is known that, for any constant $\delta > 0$, with high probability a random function is a two-source non-malleable extractor for n-bit sources with min-entropy δn handling $r = n^{\Omega(1)}$ tamperings with error $2^{-\Omega(n)}$ [CGGL19].

1.2.2 Two-source non-malleable extractors in the CRS model

In the second part of our work, we focus on constructing two-source non-malleable extractors in the CRS model first explicitly considered by Garg, Kalai, and Khurana [GKK19], under hardness assumptions of different strength.

The CRS model. We begin by describing the CRS model for two-source non-malleable extractors in more detail than in Section 1. Formal definitions can be found in Section 2.4. In this model, we assume that a CRS (denoted CRS) is first efficiently sampled and set once and for all. Our goal is to extract either computationally or statistically perfect randomness from independent weak sources X and Y which are sampled from CRS by a *computationally bounded* sampler. As side information, we disclose the output of the extractor on tampered versions of X and Y. More precisely, for arbitrary computationally bounded functions g_1 and g_2, we reveal the output of the extractor on $\overline{X} = g_1(X, \mathsf{CRS})$ and $\overline{Y} = g_2(Y, \mathsf{CRS})$. We say a candidate function cnmExt is a two-source non-malleable extractor in the CRS model if it holds that

$$\mathsf{cnmExt}(X, Y, \mathsf{CRS}), \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{CRS} \approx U, \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{CRS},$$

where U is uniformly distributed and independent of the remaining random variables, and $\approx$ denotes either computational or statistical indistinguishability.

Although we do not discuss it in the following paragraphs, we allow more than one tampering of X and Y, and also allow the sampler to leak additional auxiliary information about X to help the distinguisher. Note that the CRS is quite different from an independent uniform seed, since both the sources and the tampering functions are allowed to depend adversarially on the CRS.

Finally, we remark that the well-known upper bound of $2n$ tamperings for statistical two-source non-malleable extractors also holds in the CRS model[3]. This is unlike *one-sided* tampering, in which case an unbounded (polynomial) number of tamperings is allowed in the computational setting.

Slightly better two-source non-malleable extractors imply great two-source non-malleable extractors in the CRS model, under a standard assumption. Given our approach in Section 1.2.1, it is natural to wonder whether the two-source extractor for low min-entropy we obtain there can be made non-malleable. Unfortunately, it is not clear how to achieve that in the information-theoretic setting. Indeed, one can tamper X into $\overline{X} \neq X$ such that $\mathsf{SCond}(X) = \mathsf{SCond}(\overline{X})$, and this is enough to break the (information-theoretic) non-malleability of the extractor Ext defined in (1). We move this problem to the CRS model, and ask instead whether Ext can be made non-malleable in the CRS model. We show that this can be done assuming *any* family of collision-resistant hash functions secure against polynomial-time adversaries with not too long output (namely, output length $o(n)$, where n is the length of X). These can be instantiated from a large range of standard assumptions.

Intuitively, the only way to break non-malleability of Ext is to proceed as in the paragraph above, by finding valid tamperings of X and Y that lead to collisions at the input to the underlying two-source non-malleable extractor nmExt. However, in the CRS model we need only to deal with efficiently samplable sources and efficient tampering functions. Therefore, we can get around this problem by sampling a collision-resistant hash function H, and including the hashes $H(X)$ and $H(Y)$ as input to nmExt. In other words, we use the intuition above to show that for $\mathsf{CRS} = H$, the function

$$\mathsf{cnmExt}(X, Y, H) = \bigoplus_{i,j \in [\ell]} \mathsf{nmExt}(X'_i \| p_i \| H(X), Y'_j \| p_j \| H(Y)),$$

where $X' = \mathsf{SCond}(X)$, $Y' = \mathsf{SCond}(Y)$, and p_i denotes the binary representation of index i, is a low-error two-source *non-malleable* extractor in the CRS model for low min-entropy, provided the underlying nmExt can handle $\ell^2 - 1$ tamperings, similarly to the result presented in Section 1.2.1. The reason for this is that the efficient tampering functions will find non-trivial collisions for $H(X)$ and $H(Y)$ only with negligible probability. Remarkably, since we only require that collisions are

[3] Since there exist pairs (a, b) and (a', b) such that $\mathsf{nmExt}(a, b) \neq \mathsf{nmExt}(a', b)$, we can learn one bit of X by applying efficient tampering functions g_1 such that $g_1(x) = a$ if $x_i = 0$ and $g_1(x) = a'$ otherwise, and g_2 such that $g_2(y) = b$ for all y. We can then perform analogous tamperings for Y in place of X.

hard to find at the sampling and tampering levels, the output of cnmExt is *statistically* close to uniform given the tamperings and CRS. We leave the formal statement of our result to Section 4, and instead present a special case very similar to Theorem 1.

Theorem 2 (Informal)**.** *Suppose there exists a family $\mathcal{H}$ of collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{m_h}$ with $m_h = o(n)$. Then, for every constant $\gamma > 0$ there exists a constant C_δ such that if there exists an explicit low-error two-source non-malleable extractor* nmExt *for min-entropy rate $1 - \gamma$ handling C_δ tamperings, then there exists an explicit low-error two-source non-malleable extractor in the CRS model for min-entropy rate δ against a computationally unbounded distinguisher. In particular, if* nmExt *handles $r = \omega(1)$ tamperings, then for every constant $\delta > 0$ there exists an explicit low-error two-source non-malleable extractor in the CRS model for min-entropy rate δ.*

As discussed in Section 1.2.1, the two-source non-malleable extractors from [CGL16, Li17] come very close to the desired parameters.

Two-source non-malleable extractors in the CRS model from quasi-polynomial hardness. Since our previous result is conditional on small improvements on the extractors from [CGL16, Li17], we turn to obtaining unconditional results about two-source non-malleable extractors in the CRS model under a stronger hardness assumption. To achieve this goal, we employ tools similar to those used by Garg, Kalai, and Khurana [GKK19], and earlier by Braverman, Hassidin, and Kalai [BHK11].

The basis for our extractor is a family $\mathcal{F}$ of *lossy functions*, first introduced and constructed by Peikert and Waters [PW11]. Roughly speaking, $\mathcal{F}$ is a family of functions $f : \{0,1\}^n \to \{0,1\}^n$ containing both injective and lossy functions, i.e., functions with small image size. The security of $\mathcal{F}$ ensures that for $f \in \mathcal{F}$ injective with probability 1/2 and lossy with probability 1/2 no computationally bounded adversary can guess whether f is injective or lossy with non-negligible advantage. Moreover, we also require a family of collision-resistant hash functions $\mathcal{H}$ with output length not too large (namely, polylogarithmic in the input size).

We show that the approach of [GKK19] for one-sided tamperings can be modified and extended to two-sided tampering to show that cnmExt is a two-source non-malleable extractor in the CRS model for much lower min-entropy than the statistical non-malleable extractors from [CGL16, Li17], under the quasi-polynomial hardness of the DDH assumption.

To set the CRS, first we sample a hash function $h \leftarrow \mathcal{H}$ with output length m. Then, we sample $b \leftarrow \{0,1\}^{2m}$, and sample f_{ij} from $\mathcal{F}$ for $i \in [2m]$ and $j \in \{0,1\}$ such that f_{ib_i} is injective and f_{i1-b_i} is lossy for every i. Given such CRS, we define our candidate two-source non-malleable extractor in the CRS model cnmExt as

$$\mathsf{cnmExt}(X, Y, \mathsf{CRS}) = \mathsf{Ext}(f_{h(X)\|h(Y)}(X), f_{h(X)\|h(Y)}(Y)),$$

where Ext is a statistical strong two-source extractor, and

$$f_a(x) = f_{1a_1}(f_{2a_2}(\cdots(f_{2ma_{2m}}(x))\cdots))$$

Let $\overline{X}$ and $\overline{Y}$ denote tamperings of X and Y, respectively. First, due to the security properties of the family of lossy functions $\mathcal{F}$ under the quasi-polynomial hardness of the DDH assumption, we show that we can assume that $h(X)\|h(Y) = b$ and $h(X)\|h(Y) \neq h(\overline{X})\|h(\overline{Y})$ hold simultaneously. Under these conditions, it follows that $f_{h(X)\|h(Y)}$ is an injective function and $f_{h(\overline{X})\|h(\overline{Y})}$ has small image size. Our final goal is to show that $\mathsf{cnmExt}(X, Y, \mathsf{CRS})$ is computationally close to uniform given $\mathsf{cnmExt}(\overline{X},\overline{Y},\mathsf{CRS})$. Since $f_{h(\overline{X})\|h(\overline{Y})}$ has small image size and h has small output length, it follows that X and Y do not lose much min-entropy when we reveal $f_{h(\overline{X})\|h(\overline{Y})}(\overline{X})$, $f_{h(\overline{X})\|h(\overline{Y})}(\overline{Y})$, and all the hashes. Revealing the information above is stronger than simply disclosing $\mathsf{cnmExt}(\overline{X},\overline{Y},\mathsf{CRS})$, and it ensures that X and Y stay independent and do not lose much min-entropy. This allows us to invoke the statistical properties of Ext to obtain the desired result.

Instantiating Ext with the best known statistical two-source extractors [Bou05, Lew19, CZ19] yields cnmExt with much lower min-entropy requirements than its best known statistical counterparts [CGL16, Li17] (recall that we have proved that improving upon the parameters of [CGL16, Li17] in the information-theoretic setting appears to be very challenging). Formal statements and more details can be found in Section 5.

Theorem 3 (Informal)**.** *Assuming the quasi-polynomial hardness of the DDH assumption, there exists an explicit two-source non-malleable extractor in the CRS model for min-entropy* $0.46n$ *and negligible error handling* $n^{\Omega(1)}$ *tamperings. Moreover, there also exists an explicit two-source non-malleable extractor in the CRS model for min-entropy* $n^{\Omega(1)}$ *and error* $1/n^{\Omega(1)}$ *handling* $n^{\Omega(1)}$ *tamperings.*

Simple two-source non-malleable extractors in the CRS model from nearly optimal collision-resistant hash functions, against an unbounded distinguisher. Since our previous result holds only for a computationally bounded distinguisher, we ask whether we can devise an explicit two-source non-malleable extractor in the CRS model secure against computationally unbounded distinguishers, potentially by strengthening the underlying hardness assumption. We show that this is possible with a simple construction, provided we assume the existence of nearly optimal collision-resistant hash functions. In practice, this is not a far-fetched assumption: For most widely deployed hash functions such as SHA-256, SHA-512, and SHA-3 there are currently not better attacks than brute-force.

Suppose for a moment that we have access to a seedless condenser Cond (which does not exist information-theoretically) and a statistical two-source non-malleable extractor nmExt. Then, similarly to what was discussed regarding our first result in the CRS model, the only reason why

$$\mathsf{nmExt}(\mathsf{Cond}(X),\mathsf{Cond}(Y))$$

is not non-malleable is that we can tamper X to $\overline{X}\neq X$ such that $\mathsf{Cond}(X)=\mathsf{Cond}(\overline{X})$. In order to make this attack unfeasible for efficient tamperings, we can replace Cond by $H\leftarrow\mathcal{H}$ for a family of collision-resistant hash functions $\mathcal{H}$. However, it is not clear that the outputs $H(X)$ and $H(Y)$ are (statistically close to) high-min entropy sources, more so when conditioned on H. Fortunately, Dodis, Ristenpart, and Vadhan [DRV12] showed that this is indeed the case (with very good parameters), provided $\mathcal{H}$ is a family of nearly optimal collision-resistant hash functions (in the sense that the birthday attack is essentially the best attack possible). In other words, they show that such hash functions are so-called *seed-dependent condensers*. As a result, it readily follows that, for $\mathsf{CRS}=H$, the simple function cnmExt defined as

$$\mathsf{cnmExt}(X,Y,H)=\mathsf{nmExt}(H(X),H(Y))$$

is a low-error two-source non-malleable extractor in the CRS model for low min-entropy against a computationally *unbounded* distinguisher, and handling the same number of tamperings as nmExt. Instantiating nmExt with the two-source non-malleable extractor from [Li17], we obtain the following informal result. Formal statements and more details can be found in Section 6.

Theorem 4 (Informal)**.** *If there exist nearly optimal collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^m$ for some $m = \Omega(\mathrm{polylog}(n))$, then there exists an explicit low-error two-source non-malleable extractor for n-bit sources with min-entropy m.*

1.3 Organization

The remainder of the paper is organized as follows: In Section 2, we introduce notation and preliminary concepts and results. Section 3 discusses the reduction from low-error two-source extractors for low min-entropy to low-error two-source non-malleable extractors for high min-entropy. In Section 4, we discuss the related reduction in the CRS model from standard assumptions. Section 5 focuses on non-malleable extractors in the CRS model obtained from the quasi-polynomial hardness of the DDH assumption. Finally, simple non-malleable extractors in the CRS model obtained from nearly optimal collision-resistant hash functions are analyzed in Section 6.

2 Preliminaries

2.1 Notation

Random variables are usually denoted by uppercase letters such as X, Y, and Z. Sets are usually denoted by uppercase calligraphic letters such as $\mathcal{S}$ and $\mathcal{T}$. Given two strings x and y, we denote their concatenation by $x\|y$. The base-2 logarithm is denoted by log. We say an algorithm is *size-t* if it can be computed by a (possibly randomized) circuit of size at most t. Moreover, we use $\mathrm{poly}(n)$ to denote an *arbitrary* polynomial in n.

2.2 Statistical distance and min-entropy

In this section, we introduce the basic concepts of statistical distance and min-entropy, along with useful lemmas.

Definition 1 (Statistical distance)**.** *Given two distributions X and Y over a set $\mathcal{X}$, the* statistical distance between X and Y*, denoted by $\Delta(X;Y)$, is defined as*

$$\Delta(X;Y) = \max_{\mathcal{S}\subseteq\mathcal{X}} |\Pr[X\in\mathcal{S}] - \Pr[Y\in\mathcal{S}]| = \frac{1}{2}\sum_{x\in\mathcal{X}} |\Pr[X=x] - \Pr[Y=x]|.$$

*We may write $\Delta(X;Y|Z)$ as shorthand for $\Delta(X,Z;Y,Z)$, and say that X and Y are ε-*close*, also written $X \approx_\varepsilon Y$, if $\Delta(X;Y) \le \varepsilon$. For a random variable $X \in \{0,1\}$, we informally call $\Delta(X;U_1) = |\Pr[X=1] - 1/2|$ the* bias *of X.*

Definition 2 (Min-entropy)**.** *Given a distribution X over $\mathcal{X}$, the* min-entropy of X*, denoted by $\mathbf{H}_\infty(X)$, is defined as*

$$\mathbf{H}_\infty(X) = -\log\left(\max_{x\in\mathcal{X}} \Pr[X=x]\right).$$

Definition 3 (Average min-entropy)**.** *Given distributions X and Z, the* average min-entropy of X given Z*, denoted by $\widetilde{\mathbf{H}}_\infty(X|Z)$, is defined as*

$$\widetilde{\mathbf{H}}_\infty(X|Z) = -\log\left(\mathbb{E}_{z\leftarrow Z}\left[\max_{x\in\mathcal{X}} \Pr[X=x|Z=z]\right]\right).$$

Lemma 1 ([DORS08]). *Given arbitrary distributions X and Z such that $|\mathsf{supp}(Z)| \leq 2^{\lambda}$, we have*

$$\widetilde{\mathbf{H}}_{\infty}(X|Z) \geq \mathbf{H}_{\infty}(X, Z) - \lambda \geq \mathbf{H}_{\infty}(X) - \lambda.$$

Lemma 2 ([MW97]). *For arbitrary distributions X and Z, it holds that*

$$\Pr_{z \leftarrow Z}[\mathbf{H}_{\infty}(X|Z = z) \geq \widetilde{\mathbf{H}}_{\infty}(X|Z) - s] \geq 1 - 2^{-s}.$$

Lemma 3. *Suppose X and Z are random variables such that $\widetilde{\mathbf{H}}_{\infty}(X|Z) \geq k$ and E is an event with $\Pr[E] \geq p$. Then, it holds that*

$$\widetilde{\mathbf{H}}_{\infty}(X|E, Z) := \widetilde{\mathbf{H}}_{\infty}((X|E)|Z) \geq k - \log(1/p).$$

Proof. We have

$$\begin{aligned}
\mathbb{E}_{z \leftarrow Z}\left[\max_x \Pr[X = x|E, Z = z]\right] &= \sum_z \Pr[Z = z|E] \cdot \max_x \frac{\Pr[X = x, E|Z = z]}{\Pr[E|Z = z]} \\
&\leq \sum_z \Pr[Z = z|E] \cdot \max_x \frac{\Pr[X = x|Z = z]}{\Pr[E|Z = z]} \\
&= \sum_z \frac{\Pr[Z = z]}{\Pr[E]} \cdot \max_x \Pr[X = x|Z = z] \\
&\leq \frac{1}{p} \cdot \mathbb{E}_{z \leftarrow Z}\left[\max_x \Pr[X = x|Z = z]\right] \\
&\leq \frac{2^{-k}}{p},
\end{aligned}$$

where the second inequality follows from $\Pr[E] \geq p$ and the last inequality follows from the fact that $\widetilde{\mathbf{H}}_{\infty}(X|Z) \geq k$. □ □

2.3 Extractors and condensers

We present some important objects from pseudorandomness.

Definition 4 ((n, k)-source). *A distribution $X \in \{0, 1\}^n$ is said to be an (n, k)*-source *if $\mathbf{H}_{\infty}(X) \geq k$. Moreover, X is said to be* flat *if it is uniformly distributed over a set of size at least 2^k.*

Definition 5 ((k, ε)-extractor). *A function $\mathsf{Ext} : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}^m$ is said to be a* (strong, average-case) (k_1, k_2, ε)-extractor *if for independent random variables X and (Y, W) such that X is an (n, k_1)-source and $\widetilde{\mathbf{H}}_{\infty}(Y|W) \geq k_2$ we have*

$$\mathsf{Ext}(X, Y), X, W \approx_{\varepsilon} U_m, X, W.$$

If $k_1 = k_2 = k$, we say Ext is a (strong, average-case) (k, ε)-extractor.

It is easy to see that every non-average-case (k, ε)-extractor Ext is also an average-case $(k + \log(1/\gamma), \varepsilon + \gamma)$-extractor for any $\gamma > 0$. We will need the following explicit two-source extractors.

Proposition 1 ([Bou05, Lew19]). *There exists an explicit strong average-case (k, ε)-extractor $\mathsf{Ext} : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ with $k = 0.45n$ and $\varepsilon = 2^{-\Omega(n)}$.*

Proposition 2 ([Raz05]). *For any constant $\gamma > 0$ there exists an explicit strong average-case (k_1, k_2, ε)-extractor* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *with $k_1 = O(\log n)$, $k_2 = (1/2 + \gamma)n$, and $\varepsilon = 2^{-\Omega(n)}$.*

Proposition 3 ([CZ19]). *There exists an explicit strong average-case (k, ε)-extractor* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *with $k = \mathrm{polylog}(n)$ and $\varepsilon = n^{-\Omega(1)}$.*

Definition 6 ((k, ε, r)-non-malleable extractor). *A function* $\mathsf{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is said to be a* (strong, average-case) (k, ε, r)-non-malleable extractor *if for every pair of independent distributions X and (Y, W) such that X is an (n, k)-source and $\widetilde{\mathbf{H}}_\infty(Y|W) \geq k$ and every family of tampering functions $g_{1i}, g_{2i} : \{0,1\}^n \to \{0,1\}^n$ where one of g_{1i} and g_{2i} has no fixed points for all $i = 1, \ldots, r$, we have*

$$\Delta(\mathsf{nmExt}(X, Y); U_m | X, W, \mathsf{nmExt}(g_{11}(X), g_{21}(Y)), \ldots, \mathsf{nmExt}(g_{1r}(X), g_{2r}(Y))) \leq \varepsilon.$$

Proposition 4 ([CGL16, GKP$^+$18]). *There exists an explicit strong average-case (k, ε, r)-non-malleable extractor* $\mathsf{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *where $k = n - n^{\Omega(1)}$, $\varepsilon = 2^{-n^{\Omega(1)}}$, $r = n^{\Omega(1)}$, and $m = n^{\Omega(1)}$.*

Proposition 5 ([Li17, GKP$^+$18]). *For every constant r there exists a small enough constant $\gamma > 0$ such that there exists an explicit strong average-case (k, ε, r)-non-malleable extractor* $\mathsf{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *where $k = (1 - \gamma)n$, $\varepsilon = 2^{-\Omega(n/\log n)}$, and $m = \Omega(n)$.*

Moreover, if $k = (1 - o(1))n$, then there is $r = \omega(1)$ such that there exists an explicit strong (k, ε, r)-non-malleable extractor with $\varepsilon = 2^{-n^{\Omega(1)}}$ and $m = n^{\Omega(1)}$.

Although Li [Li17] presents its non-malleable extractor for the case $r = 1$ only, it is straightforward to check that it can be extended to more than one tampering as above.

The following lemma states that non-malleable extractors are also resilient against tampering functions with independent shared randomness.

Lemma 4. *Let* $\mathsf{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *be a (k, ε, r)-non-malleable extractor, and let R be an arbitrary distribution over some set $\mathcal{R}$. Then, for any tuple of functions $(g_{1i}, g_{2i})_{i \in [r]}$ of the form $g_{1i}, g_{2i} : \{0,1\}^n \times \mathcal{R} \to \{0,1\}^n$ such that for every fixing $R = \mathsf{rand}$ and $i = 1, \ldots, r$ either $g_{1i}(\cdot, \mathsf{rand})$ or $g_{2i}(\cdot, \mathsf{rand})$ has no fixed points, it holds that*

$$\Delta(\mathsf{nmExt}(X, Y); U_m | \mathsf{nmExt}(g_{11}(X, R), g_{21}(Y, R)), \ldots, \mathsf{nmExt}(g_{1r}(X, R), g_{2r}(Y, R)), R) \quad \leq \quad \varepsilon$$

whenever X and Y are independent (n, k)-sources also independent of R.

Moreover, if nmExt *is strong, $(g_{1i}, g_{2i})_{i \in [r]}$ are as above and $F : \{0,1\}^n \times \mathcal{R} \to \{0,1\}^*$ is an arbitrary function, we have*

$$\Delta(\mathsf{nmExt}(X, Y); U_m | F(X, R), \mathsf{nmExt}(g_{11}(X, R), g_{21}(Y, R)), \ldots, \mathsf{nmExt}(g_{1r}(X, R), g_{2r}(Y, R)), R) \leq \varepsilon$$

Proof. The claim follows from the fact that the desired inequality holds for every fixing $R = \mathsf{rand}$ by the definition of non-malleable extractor (in the case of *strong* non-malleable extractors, also because $F(X, \mathsf{rand})$ is a function of X only). □ □

Definition 7 (Somewhere-k sources). *A distribution $Y = (Y_1, \ldots, Y_\ell) \in \{0,1\}^{m \cdot \ell}$ is said to be an* elementary somewhere-k source *if there is $i \in [\ell]$ such that $\mathbf{H}_\infty(Y_i) \geq k$. Then, a distribution $Y \in \{0,1\}^{m \cdot \ell}$ is said to be a* somewhere-k source *if Y is a convex combination of elementary somewhere-k sources.*

Definition 8 (Somewhere-condenser)**.** *A function* $\mathsf{SCond} : \{0,1\}^n \to \{0,1\}^{m\cdot\ell}$ *is said to be a* $(\delta \to \delta', \varepsilon)$-somewhere condenser *if for every* $(n, \delta n)$*-source* X *there exists a somewhere-*$(\delta' m)$ *source* $Y \in \{0,1\}^{m\cdot\ell}$ *such*

$$\mathsf{SCond}(X) \approx_\varepsilon Y.$$

We will need the following two somewhere condensers due to Zuckerman and Li [Zuc06, Li11]. The first one transforms an input source with potentially low min-entropy rate into a somewhere-k source with constant min-entropy rate. The second somewhere condenser transforms an input source with constant min-entropy rate into a somewhere-k source with potentially large min-entropy rate. We note that other somewhere-condensers have also been constructed in [BKS+10, Raz05].

We begin by stating a somewhere-condenser that condenses sources to min-entropy rate 3/4, due to Zuckerman [Zuc06].

Lemma 5. *For* δ *and* n *such that* $\delta n = \omega(1)$ *there is an explicit* $(\delta \to 3/4, \varepsilon)$*-somewhere condenser* $\mathsf{SCond} : \{0,1\}^n \to \{0,1\}^{m\cdot\ell}$ *with* $\ell = \mathrm{poly}(1/\delta)$*,* $m = n/\mathrm{poly}(1/\delta)$*, and* $\varepsilon = 2^{-\Omega(m)}$*.*

Improving upon the analysis of [Zuc06], Li [Li11] obtained the following somewhere-condenser that condenses sources to potentially very high min-entropy rate. A version of this somewhere-condenser also appears in [BDT16][4].

Lemma 6. *For every* $T = T(n) < n$ *there exists a* $(3/4 \to 1 - 1/T, \varepsilon)$*-somewhere-condenser* $\mathsf{SCond} : \{0,1\}^n \to \{0,1\}^{m\cdot\ell}$ *with* $\ell = T^{5/2}$*,* $m = n/T^{\frac{5\log(3/2)}{2}}$*, and* $\varepsilon = 2^{-n/T^c}$ *for some* $c > 1$*, provided* n *is large enough.*

Combining Lemmas 5 and 6 immediately leads to the following corollary.

Corollary 1. *For every constant* $\delta > 0$ *and every* $T = T(n) < n$ *there exists a* $(\delta \to 1 - 1/T, \varepsilon)$*-somewhere-condenser* $\mathsf{SCond} : \{0,1\}^n \to \{0,1\}^{m\cdot\ell}$ *with* $\ell = O_\delta(T^{5/2})$*,* $m = \Omega_\delta(n/T^{\frac{5\log(3/2)}{2}})$*, and* $\varepsilon = 2^{-\Omega_\delta(n/T^c)}$ *for some absolute constant* $c > 1$*, provided* n *is large enough.*

2.4 Computational extractors in the CRS model

In this section, we present the relevant definitions of computational pseudorandom objects in the CRS model. As usual, all parameters are functions of a single security parameter λ. For the sake of clarity, we do not write this dependence explicitly in the rest of the paper.

Definition 9 ((t,k)-samplable sources in the CRS model)**.** *A tuple*

$$(X, \mathsf{AUX}, Y) \in \{0,1\}^n \times \{0,1\}^a \times \{0,1\}^n$$

is said to be a tuple of (t,k)-samplable sources in the CRS model *if there exists* $\mathsf{CRS} \in \{0,1\}^c$ *such that the following hold:*

- *There exists a size-*t *circuit* $\mathcal{G}$ *such that* $(X, \mathsf{AUX}, Y) \leftarrow \mathcal{G}(\mathsf{CRS})$*.*
- (X, AUX) *and* Y *are conditionally independent given* CRS*.*
- $\mathbf{H}_\infty(X|\mathsf{CRS} = \mathsf{crs}) \geq k$ *and* $\mathbf{H}_\infty(Y|\mathsf{CRS} = \mathsf{crs}) \geq k$ *for every fixing* $\mathsf{CRS} = \mathsf{crs}$*.*

[4] The work [BDT16] has been retracted. However, the somewhere-condenser presented there is a restatement of the one of Li [Li11], and is correct.

When AUX *is the empty string, we say* (X, Y) *are* (t, k)-samplable sources without auxiliary information.

Moreover, we say (X, AUX, Y) *is a tuple of* (t, k_1, k_2)-samplable sources in the CRS model *if in the above we have* $\mathbf{H}_\infty(X|\mathsf{CRS} = \mathsf{crs}) \geq k_1$ *and* $\mathbf{H}_\infty(Y|\mathsf{CRS} = \mathsf{crs}) \geq k_2$.

Definition 10 ($(t, t', t'', k, \varepsilon, r)$-non-malleable extractor in the CRS model)**.** *A function* $\mathsf{cnmExt} : \{0, 1\}^n \times \{0, 1\}^n \times \{0, 1\}^c \to \{0, 1\}^m$ *is said to be a* $(t, t', t'', k, \varepsilon, r)$-non-malleable extractor in the CRS model *if there is* $\mathsf{CRS} \in \{0, 1\}^c$ *such that the following holds:*

For every tuple (X, Y, AUX) *of* (t, k)*-samplable sources from* CRS*, every tuple of deterministic size-*t' *circuits* $g_{11}, \dots, g_{1r} : \{0, 1\}^n \times \{0, 1\}^a \times \{0, 1\}^c \to \{0, 1\}^n$ *and* $g_{21}, \dots, g_{2r} : \{0, 1\}^n \times \{0, 1\}^c \to \{0, 1\}^n$ *such that for every* $i \in [r]$ *and every fixing* $\mathsf{CRS} = \mathsf{crs}$ *either* $g_{2i}(\cdot, \mathsf{crs})$ *has no fixed points or* $g_{1i}(\cdot, \mathsf{aux}, \mathsf{crs})$ *has no fixed points for every fixing* $\mathsf{AUX} = \mathsf{aux}$*, and every size-*$t''$ *adversary* $\mathcal{A}$ *we have*

$$\begin{aligned}|\Pr[\mathcal{A}(\mathsf{cnmExt}(X, Y, \mathsf{CRS}), L_1, \dots, L_r, \mathsf{AUX}, \mathsf{CRS}) = 1] \\ - \Pr[\mathcal{A}(U_m, L_1, \dots, L_r, \mathsf{AUX}, \mathsf{CRS}) = 1]| \leq \varepsilon(t),\end{aligned}$$

where $L_i = \mathsf{cnmExt}(g_{1i}(X, \mathsf{AUX}, \mathsf{CRS}), g_{2i}(Y, \mathsf{CRS}), \mathsf{CRS})$*. We set* $t'' = \infty$ *to denote that* $\mathcal{A}$ *is allowed to be computationally unbounded.*

We say cnmExt *is a* $(t, t', t'', k, \varepsilon, r)$-non-malleable extractor without auxiliary information *if the above holds for all* (t, k)*-samplable sources* (X, Y) *without auxiliary information.*

Moreover, we say cnmExt *is a* $(t, t', t'', k_1, k_2, \varepsilon, r)$-non-malleable extractor in the CRS model *if the above holds for* (t, k_1, k_2)*-samplable sources.*

Observe that every non-malleable extractor resilient to auxiliary information is, in particular, strong.

2.5 Other relevant computational objects

In this section, we present other computational objects that will prove useful throughout the paper.

Definition 11 ((t, δ)-collision-resistant hash function family)**.** *A family of functions* $\mathcal{H}$ *is said to be* (t, δ)-collision-resistant *if for every size-t adversary* $\mathcal{A}$ *it holds that*

$$\Pr[X_1 \neq X_2, H(X_1) = H(X_2)] \leq \delta,$$

where $H \leftarrow \mathcal{H}$ *and* $(X_1, X_2) \leftarrow \mathcal{A}(H)$.

Definition 12 (Seed-dependent condenser)**.** *A function* $\mathsf{Cond} : \{0, 1\}^n \times \{0, 1\}^d \to \{0, 1\}^m$ *is said to be a* $(k \to_\varepsilon k', t)$-seed-dependent condenser *if for every* $X \leftarrow \mathcal{G}(S)$*, where* $S \leftarrow \{0, 1\}^d$*,* $\mathcal{G}$ *is a size-t circuit, and* $\widetilde{\mathbf{H}}_\infty(X|S) \geq k$*, it holds that*

$$\mathsf{Cond}(X, S), S \approx_\varepsilon Z, S,$$

where $\widetilde{\mathbf{H}}_\infty(Z|S) \geq k'$.

Dodis, Ristenpart, and Vadhan [DRV12] showed that collision-resistant hash functions with strong security are good seed-dependent condensers.

Lemma 7 ([DRV12])**.** *Suppose* $\mathcal{H}$ *is a family of* $(t, 2^{\beta-1-m})$*-collision-resistant hash functions* $h : \{0, 1\}^n \to \{0, 1\}^m$ *for some* $\beta > 0$*. Then, the function* $\mathsf{Cond}(X, H) = H(X)$ *where* $H \leftarrow \mathcal{H}$ *is an* $(m - \beta + 1 \to_\varepsilon m - \beta - \log(1/\varepsilon), t)$*-seed-dependent condenser.*

We will also require the following notion of a family of lossy functions, first introduced and constructed by Peikert and Waters [PW11].

Definition 13 ((t,n,ω)-lossy function family)**.** *A function family $\mathcal{F}=\{\mathcal{F}\}_{\lambda\in\mathbb{N}}$ is a (t,n,ω)-lossy function family if the following conditions hold:*

- *There are two PPT seed generation algorithms $\mathcal{G}_{\mathsf{inj}}$ and $\mathcal{G}_{\mathsf{loss}}$ such that for any size-*$\mathrm{poly}(t)$ *adversary $\mathcal{A}$ it holds that*

$$|\mathsf{Pr}_{s\leftarrow\mathcal{G}_{\mathsf{inj}}(1^\lambda)}[\mathcal{A}(s)=1]-\mathsf{Pr}_{s\leftarrow\mathcal{G}_{\mathsf{los}}(1^\lambda)}[\mathcal{A}(s)=1]|=\mathsf{negl}(t);$$

- *For every $\lambda\in\mathbb{N}$ and every $f\in\mathcal{F}_\lambda$, $f:\{0,1\}^n\to\{0,1\}^n$.*
- *For every $\lambda\in\mathbb{N}$ and every $s\in\mathcal{G}_{\mathsf{inj}}$, $f_s\in\mathcal{F}_\lambda$ is injective.*
- *For every $\lambda\in\mathbb{N}$ and every $s\in\mathcal{G}_{\mathsf{los}}$, $f_s\in\mathcal{F}_\lambda$ is* lossy*, i.e., its image size is at most $2^{n-\omega}$.*
- *There exists a PPT algorithm* Eval *such that $\mathsf{Eval}(s,x)=f_s(x)$ for very $\lambda\in\mathbb{N}$, every s in the support of $\mathcal{G}_{\mathsf{inj}}(1^\lambda)\cup\mathcal{G}_{\mathsf{los}}(1^\lambda)$, and every $x\in\{0,1\}^n$.*

Lemma 8 ([PW11, BHK11])**.** *For any constant $\gamma\in(0,1)$ and for every $\Omega(\lambda)\leq n\leq\mathrm{poly}(\lambda)$ there exists a (t,n,ω)-lossy function family with $t=\lambda^{\log\lambda}$ and $\omega=n-n^\gamma$, assuming the quasi-polynomial hardness of the DDH assumption.*

3 From slightly better two-source non-malleable extractors to two-source extractors for low min-entropy

In this section, we show that slight improvements on the state-of-the-art explicit constructions of two-source non-malleable extractors [CGL16, Li17] are enough to obtain low error two-source extractors for low linear min-entropy. More precisely, we have the following result.

Theorem 5. *For every constant $\delta>0$ there exists a constant $C_\delta>0$ such that the following holds: If for m large enough and some $\gamma=\gamma(m)\geq 1/m$ there exists an explicit $(m(1-\gamma)-3\log m,\varepsilon,C_\delta\cdot(1/\gamma)^5)$-non-malleable extractor* $\mathsf{nmExt}:\{0,1\}^m\times\{0,1\}^m\to\{0,1\}$*, then there exists an explicit $(\delta n,\varepsilon')$-extractor* $\mathsf{Ext}:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ *with $\varepsilon'=\varepsilon+2^{-\Omega(\gamma^c n)}$ and $n=\Theta(m\cdot(1/\gamma)^c)$, where c is an absolute constant.*

Proof. Let $\mathsf{nmExt}:\{0,1\}^m\times\{0,1\}^m\to\{0,1\}$ be the non-malleable extractor with the parameters as in the theorem statement, and let $\mathsf{SCond}:\{0,1\}^n\to\{0,1\}^{m'\cdot\ell}$ be the $(\delta\to 1-\gamma,\varepsilon)$-somewhere condenser from Corollary 1, and $m=m'+\lceil\log\ell\rceil$.

Consider the function $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ defined as

$$F(X,Y)=\bigoplus_{i,j\in[\ell]}\mathsf{nmExt}(\mathsf{SCond}(X)_i\|p_i,\mathsf{SCond}(Y)_j\|p_j),$$

where p_i denotes the $\lceil\log\ell\rceil$-bit binary representation of $i\in[\ell]$. We prove that F is an extractor with the desired parameters.

By the properties of SCond, there exist $V,W\in\{0,1\}^{m'\ell}$ independent somewhere-k' sources with $k'=(1-\gamma)m'$ such that $\mathsf{SCond}(X)\approx_{\varepsilon_1}V$ and $\mathsf{SCond}(Y)\approx_{\varepsilon_1}W$ for $\varepsilon_1=2^{-\Omega(\gamma^c n)}$. Therefore, it suffices to show that

$$\bigoplus_{i,j\in[\ell]}\mathsf{nmExt}(V_i\|p_i,W_j\|p_j)\approx_\varepsilon U_1,\tag{2}$$

and the desired result follows by combining the previous observations with the triangle inequality. By the properties of V and W, there exist independent random variables $I, J \in [\ell]$ such that

$$\mathbf{H}_\infty(V_i|I=i), \mathbf{H}_\infty(W_j|J=j) \geq (1-\gamma)m'.$$

Consider arbitrary fixings $I=i$ and $J=j$. We show that (2) holds for all fixings, and hence it holds in general as well. Under such a fixing, it is enough to show that

$$\Delta(\mathsf{nmExt}(V_i\|p_i, W_j\|p_j); U_1|(\mathsf{nmExt}(V_{i'}\|p_{i'}, W_{j'}\|p_{j'})_{(i',j')\neq(i,j)}) \leq \varepsilon. \tag{3}$$

We will now use the properties of nmExt to prove (3). Note that we can jointly simulate all pairs $(V_{i'}\|p_{i'}, W_{j'}\|p_{j'})$ for $(i',j') \neq (i,j)$ as randomized split-state tamperings of $(V_i\|p_i, W_j\|p_j)$. In other words, there exist randomized functions $g_{1i'}, g_{2j'} : \{0,1\}^m \times \mathcal{R} \to \{0,1\}^m$, all sharing the same independent randomness $R \in \mathcal{R}$, such that

$$\begin{aligned}&(V_i\|p_i, W_j\|p_j), (g_{1i'}(V_i\|p_i, R), g_{2j'}(W_j\|p_j, R))_{(i',j')\neq(i,j)}\\ &\qquad\qquad\qquad\qquad \sim (V_i\|p_i, W_j\|p_j), (V_{i'}\|p_{i'}, W_{j'}\|p_{j'})_{(i',j')\neq(i,j)}.\end{aligned}$$

Indeed, on input $(v_i\|p_i, w_j\|p_j)$, this can be done by sampling $V' = (V|I=i, V_i=v_i)$ and $W' = (W|J=j, W_j=w_j)$ using the extra independent randomness R, and setting $g_{1i'}(v_i\|p_i, R) = V'_{i'}\|p_{i'}$ and $g_{2j'}(w_j\|p_j, R) = W'_{j'}\|p_{j'}$ for all $(i',j') \neq (i,j)$. Moreover, since $p_a \neq p_b$ for $a \neq b$, all tampering functions $g_{1i'}$ and $g_{2j'}$ above have no fixed points for every fixing of the randomness. Finally, since $\ell = O_\delta((1/\gamma)^{5/2})$ and $\gamma \geq 1/m$, it follows that

$$\mathbf{H}_\infty(V_i\|p_i), \mathbf{H}_\infty(W_j\|p_j) \geq (1-\gamma)m' \geq (1-\gamma)m - 3\log m$$

and that nmExt handles at least $\ell^2 \leq C_\delta(1/\gamma)^5$ tamperings for a suitable constant $C_\delta > 0$ depending only on δ. Taking into account these observations and noting that $V_i\|p_i$ and $W_j\|p_j$ are independent, we can invoke Lemma 4 to conclude (3) holds, which completes the proof. □ □

We now present two remarkable corollaries of Theorem 5, one of which was already informally presented in Section 1.2.

Corollary 2. *Suppose that for some $r = r(m) = \omega(1)$, $\varepsilon = \varepsilon(m)$, and some constant $c > 0$ there is an explicit $(m(1-\gamma), \varepsilon, r)$-non-malleable extractor for large enough m. Then, for any constant $\delta > 0$ and large enough n there exists an explicit $(\delta n, \varepsilon')$-extractor with $\varepsilon' = \varepsilon(\Omega(n)) + 2^{-\Omega(n)}$.*

Corollary 3. *There exists an absolute constant $\alpha > 0$ such that if for some constant $\beta < \alpha$ there exists an explicit $(m - m^{1-\beta}, \varepsilon, m^{6\beta})$-non-malleable extractor $\mathsf{nmExt} : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$, then for any constant $\delta > 0$ and large enough n there exists an explicit $(\delta n, \varepsilon')$-extractor with $\varepsilon' = \varepsilon(n^{\Omega(1)}) + 2^{-n^{\Omega(1)}}$.*

According to Corollary 3, improving the min-entropy requirement of the CGL extractor in Proposition 4 to $m - m^{c_0}$ for a sufficiently small constant $c_0 > 0$ would immediately yield explicit low error two-source extractors for *any* linear min-entropy rate.

4 Computational two-source non-malleable extractors for low min-entropy from any collision-resistant hash function family

In this section, we show how the construction used to prove Theorem 5 can also be used to obtain computational *non-malleable* extractors for low min-entropy efficiently samplable sources, efficient tampering, and a *computationally unbounded* distinguisher from slight improvements on the state-of-the-art constructions of non-malleable extractors for high min-entropy sources. This can be achieved under the weak hardness assumption that families of collision-resistant hash functions with decent parameters exist.

Theorem 6. *For every constant $\delta > 0$ there exists a constant $C_\delta > 0$ such that the following holds: If for m large enough and some $\gamma = \gamma(m) \geq 1/m$ there exists an explicit $(m(1-\gamma) - 3\log m - m_h, \varepsilon = \mathsf{negl}(m), C_\delta \cdot (1/\gamma)^5)$-non-malleable extractor $\mathsf{nmExt} : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$, then there exists an explicit $(\mathrm{poly}(n), \mathrm{poly}(n), \infty, k = \delta n, \varepsilon = \mathsf{negl}(m), r = 1)$-non-malleable extractor $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ in the CRS model without auxiliary information with $n = \Theta(m \cdot (1/\gamma)^c)$, where c is an absolute constant, provided that there exists a family $\mathcal{H}$ of $(\mathrm{poly}(n), \mathsf{negl}(n))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{m_h}$ with $m_h = o(n)$.*

Proof. Towards proving the desired statement, we modify the construction used to prove Theorem 5 by including the hashes of the sources in the input to nmExt. More precisely, we set $\mathsf{CRS} = H$ for $H \leftarrow \mathcal{H}$, and consider the function $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \times \mathcal{H} \to \{0,1\}$ defined as

$$\mathsf{cnmExt}(X, Y, H) = \bigoplus_{i,j \in [\ell]} \mathsf{nmExt}(\mathsf{SCond}(X)_i \| p_i \| H(X), \mathsf{SCond}(Y)_j \| p_j \| H(Y)),$$

where nmExt is as in the theorem statement, $\mathsf{SCond} : \{0,1\}^n \to \{0,1\}^{m' \cdot \ell}$ is the $(\delta/2 \to 1-\gamma, \varepsilon_1)$-somewhere condenser from Corollary 1, p_i denotes the $\lceil \log \ell \rceil$-bit binary representation of i, and $m = m' + \lceil \log \ell \rceil + m_h$.

Fix $(t, \delta n)$-samplable sources X and Y and size-$\mathrm{poly}(t)$ deterministic tampering functions $g_1, g_2 : \{0,1\}^n \times \mathcal{H} \to \{0,1\}^n$ such that for each $h \in \mathcal{H}$, one of $g_1(\cdot, h)$ and $g_2(\cdot, h)$ has no fixed points. Our goal is to show that

$$\mathsf{cnmExt}(X, Y, H), \mathsf{cnmExt}(X', Y', H), H \approx_\varepsilon U_1, \mathsf{cnmExt}(X', Y', H), H, \tag{4}$$

where $X' = g_1(X, H)$ and $Y' = g_2(Y, H)$, and $\varepsilon = \mathsf{negl}(n)$. We begin by claiming that the collision-resistance of $\mathcal{H}$ ensures that

$$\Pr_H[X \neq X', H(X) = H(X')] = \mathsf{negl}(n),$$
$$\Pr_H[Y \neq Y', H(Y) = H(Y')] = \mathsf{negl}(n).$$

Indeed, if this does not hold, then we can break the collision-resistance of $\mathcal{H}$ by considering the size-$\mathrm{poly}(t)$ adversary that on input $H \leftarrow \mathcal{H}$ first samples (X, Y), and then outputs either (X, X') or (Y, Y') with probability $1/2$. Since one of $g_1(\cdot, h)$ and $g_2(\cdot, h)$ has no fixed points for each fixing $H = h$, this adversary succeeds with non-negligible probability. With this in mind, with probability $1 - \mathsf{negl}(n)$ over the fixing $H = h$, we have $\Pr[X \neq X', h(X) = h(X')] = \mathsf{negl}(n)$ and $\Pr[Y \neq Y', h(Y) = h(Y')] = \mathsf{negl}(n)$. Throughout the remainder of the proof we can fix such $h \in \mathcal{H}$ and assume that $g_1(\cdot, h)$ has no fixed points without loss of generality. Moreover, we will also condition X on the events $h(X) \neq h(X')$ and $h(X) = z_1$ and Y on the event $h(Y) = z_2$ from now on. Since $h(X) \neq h(X')$ holds with probability $1 - \mathsf{negl}(n)$, by Lemmas 1 and 2 we have

$$\mathbf{H}_\infty(X | h(X) \neq h(X'), h(X) = z_1) \geq \delta n - 1 - m_h - \mathsf{negl}(n) \geq \delta n/2$$

with probability $1 - \mathsf{negl}(n)$ over the choice of z_1. Likewise, we have $\mathbf{H}_\infty(Y | h(Y) = z_2) \geq \delta n/2$ with probability $1 - \mathsf{negl}(n)$ over the choice of z_2. From here onwards, fix such z_1 and z_2.

Given the fixings in the previous paragraph, by the properties of SCond there exist independent somewhere-k' sources $V, W \in \{0,1\}^{m'\ell}$ with $k' = (1-\gamma)m'$ and independent random variables $I, J \in [\ell]$ such that $\mathsf{SCond}(X) \approx_{\varepsilon_1} V$ and $\mathsf{SCond}(Y) \approx_{\varepsilon_1} W$, and

$$\mathbf{H}_\infty(V_i|I=i) \geq (1-\gamma)m' \geq (1-\gamma)m - 3\log m - m_h, \tag{5}$$
$$\mathbf{H}_\infty(W_j|J=j) \geq (1-\gamma)m' \geq (1-\gamma)m - 3\log m - m_h. \tag{6}$$

for all valid fixings $I = i$ and $J = j$. We now wish to proceed by replacing $\mathsf{SCond}(X)$ and $\mathsf{SCond}(Y)$ by V and W, respectively, in our analysis. Observe that we can write $A(\mathsf{SCond}(X)) = (\mathsf{SCond}(X')_i\|i\|h(X'))_{i\in[\ell]}$ for a randomized function A that on input v samples x from $(X|\mathsf{SCond}(X) = v)$ and sets $A(v) = (\mathsf{SCond}(g_1(x,h))_i\|i\|h(g_1(x,h)))_{i\in[\ell]}$ (if the sampling of x fails, simply output a fixed bitstring whose suffix differs from z_1). By our conditioning, we may assume that $A(v) \neq v\|i\|z_1$ for all $i \in [\ell]$. Analogously, we can also write $B(\mathsf{SCond}(Y)) = (\mathsf{SCond}(Y')_j\|j\|h(Y'))_{j\in[\ell]}$ for a randomized function B. Therefore, it now suffices to show that

$$\bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(\mathsf{SCond}(X)_i\|i\|z_1, \mathsf{SCond}(Y)_j\|j\|z_2),$$
$$\bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(\mathsf{SCond}(X))_i, B(\mathsf{SCond}(Y))_j)$$
$$\approx_{\varepsilon'} U_1, \bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(\mathsf{SCond}(X))_i, B(\mathsf{SCond}(Y)_j)). \tag{7}$$

Using the fact that $\mathsf{SCond}(X), \mathsf{SCond}(Y) \approx_{2\varepsilon_1} V, W$, the condition in (7) follows if we show that

$$\bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(V_i\|i\|z_1, W_j\|j\|z_2), \bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(V)_i, B(W)_j)$$
$$\approx_{\varepsilon} U_1, \bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(V)_i, B(W)_j). \tag{8}$$

Consider arbitrary fixings $I = i^\star$ and $J = j^\star$. We show that then we have

$$\Delta(\mathsf{nmExt}(V_{i^\star}\|i\|z_1, W_{j^\star}\|j\|z_2); U_1$$
$$|(\mathsf{nmExt}(V_i\|i\|z_1, W_j\|j\|z_2))_{(i,j)\neq(i^\star,j^\star)}, (\mathsf{nmExt}(A(V)_i, B(W)_j))_{i,j\in[\ell]}) \leq \varepsilon, \tag{9}$$

which implies (8) and concludes the proof. Analogously to the proof of Theorem 5, we can write $g^1_{1i}(V_{i^\star}\|i\|z_1, R) = V_i\|i\|z_1$ and $g^1_{2j}(W_{j^\star}\|j\|z_2, R) = W_j\|j\|z_2$ for randomized tampering functions $g^1_{1i}, g^1_{2j} : \{0,1\}^m \times \mathcal{R} \to \{0,1\}^m$ for $i \neq i^\star$ and $j \neq j^\star$. Observe that the g^1_{1i}'s and g^1_{2j}'s have no fixed points, since $p_i \neq p_{i^\star}$ and $p_j \neq p_{j^\star}$. Moreover, we can also write $g^2_{1i}(V_{i^\star}\|i\|z_1, R) = A(V)_i$ and $g^2_{2j}(W_{j^\star}\|j\|z_2, R) = B(W)_j$ for randomized tampering functions $g^2_{1i}, g^2_{2j} : \{0,1\}^m \times \mathcal{R} \to \{0,1\}^m$ for $i \neq i^\star$. By our previous conditioning, we know that g^2_{1i} has no fixed points, i.e., $g^2_{1i}(V_{i^\star}\|i\|z_1, r) \neq V_{i^\star}\|i\|z_1$ for all r. Finally, since there are at most $2\ell^2 \leq C_\delta(1/\gamma)^5$ tamperings for a suitably large constant C_δ depending only on δ, and since $V_{i^\star}\|p_{i^\star}\|z_1$ and $W_{j^\star}\|p_{j^\star}\|z_2$ are independent and $\mathbf{H}_\infty(V_{i^\star}\|p_{i^\star}\|z_2), \mathbf{H}_\infty(W_{j^\star}\|p_{j^\star}\|z_2) \geq (1-\gamma)m - 3\log m - m_h$ by (5) and (6), we can invoke Lemma 4 to conclude (9) holds, which completes the proof. □ □

Similarly to the previous section, we present two corollaries that are especially meaningful given the current state-of-the-art constructions of two-source non-malleable extractors [CGL16, Li17], one of which was already informally presented in Section 1.2.

Corollary 4. *Suppose that for some $r = r(m) = \omega(1)$, $\varepsilon = \mathsf{negl}(m)$, and some constant $c > 0$ there is an explicit $(m(1-\gamma), \varepsilon, r)$-non-malleable extractor for large enough m. Then, for any constant $\delta > 0$ and large enough n there exists an explicit $(\mathrm{poly}(n), \mathrm{poly}(n), \infty, k = \delta n, \varepsilon = \mathsf{negl}(n), r = 1)$-non-malleable extractor in the CRS model without auxiliary information, provided that there exists a family $\mathcal{H}$ of $(\mathrm{poly}(n), \mathsf{negl}(n))$-collision resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{m_h}$ with $m_h = o(n)$.*

Corollary 5. *There exists an absolute constant $\alpha > 0$ such that if for some constant $\beta < \alpha$ there exists an explicit $(m - m^{1-\beta}, \varepsilon = \mathsf{negl}(m), m^{6\beta})$-non-malleable extractor $\mathsf{nmExt} : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$, then for any constant $\delta > 0$ and large enough n there exists an explicit $(\mathrm{poly}(n), \mathrm{poly}(n), \infty, k = \delta n, \varepsilon = \mathsf{negl}(n), r = 1)$-non-malleable extractor in the CRS model without auxiliary information, provided that there exists a family $\mathcal{H}$ of $(\mathrm{poly}(n), \mathsf{negl}(n))$-collision resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{m_h}$ with $m_h \leq n^\rho$ for a small enough constant $\rho > 0$.*

5 Computational non-malleable extractors from quasi-polynomial hardness assumptions

In this section, we construct computational two-source non-malleable extractors in the CRS model assuming the quasi-polynomial hardness of the DDH assumption. We begin by constructing such a non-malleable extractor for relatively high min-entropy that handles many tamperings. Then, we use this construction as a stepping stone to obtain a non-malleable extractor in the CRS model for low min-entropy.

Theorem 7. *Suppose the following objects exist:*

- *A family $\mathcal{H}$ of $(\mathrm{poly}(t_1), \mathsf{negl}(t_1))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{k_2}$;*
- *A family of $(\mathrm{poly}(t_2), n, \omega)$-lossy functions $\mathcal{F}$, where $t_2 \geq 2^{2k_2} = t_1^{\omega(1)}$ and $\omega = n - n^\gamma$ for some constant $\gamma \in (0,1)$.*
- *A strong average-case (k, ε)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, where $\Omega(t_1) \leq n \leq \mathrm{poly}(t_1)$.*

Then, there exists an explicit $(\mathrm{poly}(t_1), \mathrm{poly}(t_1), \mathrm{poly}(t_2), k' = k + r(4k_2 + 3n^\gamma + 1), \varepsilon + \mathsf{negl}(t_1), r)$-non-malleable extractor $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ in the CRS model.

We instantiate Theorem 7 with the best known explicit statistical two-source extractors in Appendix 5.2.

5.1 Proof of Theorem 7

Our candidate construction is as follows: First, to define CRS, begin by sampling $b \leftarrow \{0,1\}^{2k_2}$, and then sample functions f_{ij} from $\mathcal{F}$ for $i \in [2k_2]$ and $j \in \{0,1\}$ such that f_{ib_i} is injective and f_{i1-b_i} is lossy for each i. Finally, sample $h \leftarrow \mathcal{H}$ and set

$$\mathsf{CRS} = (h, (f_{ij})_{i \in [2k_2], j \in \{0,1\}}) \in \{0,1\}^c.$$

Our function $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^c \to \{0,1\}$ is defined as

$$\mathsf{cnmExt}(x, y, \mathsf{CRS}) = \mathsf{Ext}(f_{h(x)\|h(y)}(x), f_{h(x)\|h(y)}(y)),$$

where for $a \in \{0,1\}^{2k_2}$ we denote $f_a(x) = f_{1a_1}(f_{2a_2}(\cdots(f_{2k_2 a_{2k_2}}(x))\cdots))$.

For the sake of exposition, we present the proof for the case $r = 1$ only. The extension to $r > 1$ tamperings is straightforward. In order to show Theorem 7, we must argue that, for arbitrary $(\mathrm{poly}(t_1), k')$-samplable sources (X, AUX, Y), valid size-poly(t_1) tampering functions $g_1 : \{0,1\}^n \times \{0,1\}^a \times \{0,1\}^c \to \{0,1\}^n$ and $g_2 : \{0,1\}^n \times \{0,1\}^c \to \{0,1\}^n$, and every size-poly$(t_2)$ distinguisher $\mathcal{A}$ it holds that

$$\begin{aligned}|\Pr[&\mathcal{A}(\mathsf{cnmExt}(X, Y, \mathsf{CRS}), \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{AUX}, \mathsf{CRS}) = 1] \\ &- \Pr[\mathcal{A}(U_1, \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{AUX}, \mathsf{CRS}) = 1]| \leq \varepsilon + \mathsf{negl}(t_1), \quad (10)\end{aligned}$$

where $\overline{X} = g_1(X, \mathsf{AUX}, \mathsf{CRS})$ and $\overline{Y} = g_2(Y, \mathsf{CRS})$. As a first step, we prove that it suffices to consider cases where $h(X)\|h(Y) \neq h(\overline{X})\|h(\overline{Y})$ and $h(X)\|h(Y) = b$, where b denotes the indices of the injective functions $(f_{ib_i})_{i\in[2k_2]}$.

Lemma 9. *Let E denote the event that $h(X)\|h(Y) \neq h(\overline{X})\|h(\overline{Y})$ and $h(X)\|h(Y) = b$ hold simultaneously. Then, if*

$$\begin{aligned}|\Pr[&\mathcal{A}(\mathsf{cnmExt}(X, Y, \mathsf{CRS}), \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{AUX}, \mathsf{CRS}) = 1|E] \\ &- \Pr[\mathcal{A}(U_1, \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{AUX}, \mathsf{CRS}) = 1|E]| \leq \varepsilon + \mathsf{negl}(t_1), \quad (11)\end{aligned}$$

it follows that (10) *holds.*

Proof. We proceed similarly to the proof of the analogous claim in [GKK19]. Suppose that (11) holds for every tuple of $(\mathrm{poly}(t_1), k')$-samplable sources (X, AUX, Y), tampering functions g_1 and g_2, and size-poly(t_2) adversary $\mathcal{A}$, but

$$\begin{aligned}|\Pr[&\mathcal{A}(\mathsf{cnmExt}(X, Y, \mathsf{CRS}), \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{AUX}, \mathsf{CRS}) = 1] \\ &- \Pr[\mathcal{A}(U_1, \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS}), \mathsf{AUX}, \mathsf{CRS}) = 1]| > \varepsilon + 1/p(t_1), \quad (12)\end{aligned}$$

where $\overline{X} = g_1(X, \mathsf{AUX}, \mathsf{CRS})$ and $\overline{Y} = g_2(Y, \mathsf{CRS})$, for some pair of $(\mathrm{poly}(t_1), k')$-samplable sources (X, AUX, Y), some tampering functions g_1 and g_2, some size-poly(t_2) adversary $\mathcal{A}$, and some polynomial p. We show that this breaks the t_2-security of the family of lossy functions $\mathcal{F}$. By the t_2-security of $\mathcal{F}$, we know that for every size-poly(t_2) adversary $\mathcal{B}$ we have

$$2^{-2k_2} - \mathsf{negl}(t_2) \leq \Pr[\mathcal{B}(\mathsf{CRS}) = b] \leq 2^{-2k_2} + \mathsf{negl}(t_2). \quad (13)$$

Consider the size-poly(t_2) adversary $\mathcal{B}$ that on input CRS samples (X, AUX, Y), and first checks whether $h(X)\|h(Y) \neq h(\overline{X})\|h(\overline{Y})$. If that is the case, then $\mathcal{B}$ outputs $b' = h(X)\|h(Y)$ as a guess for b, else it outputs $b' \leftarrow \{0,1\}^{2k_2}$. Since $\Pr[h(X)\|h(Y) = h(\overline{X})\|h(\overline{Y})] = \mathsf{negl}(t_1)$ by the collision-resistance of $\mathcal{H}$ and the fact that $\overline{X} \neq X$ or $\overline{Y} \neq Y$ by hypothesis, using (13) we have that

$$\begin{aligned}(1 - \mathsf{negl}(t_1))2^{-2k_2} - \mathsf{negl}(t_2) &\leq \Pr[h(X)\|h(Y) = b, h(X)\|h(Y) \neq h(\overline{X})\|h(\overline{Y})] \\ &\leq (1 + \mathsf{negl}(t_1))2^{-2k_2} + \mathsf{negl}(t_2). \quad (14)\end{aligned}$$

We now proceed to construct a size-poly(t_2) adversary $\mathcal{B}'$ such that

$$\Pr[\mathcal{B}'(\mathsf{CRS}) = b] \geq 1.5 \cdot 2^{-2k_2}.$$

This contradicts (13), which concludes the proof. On input CRS and for $N = p(t_1)^3$, $\mathcal{B}'$ proceeds as follows:

1. Sample (X, AUX, Y) from CRS. If $h(X)\|h(Y) = h(\overline{X})\|h(\overline{Y})$, then re-sample (note that this takes time $\mathrm{poly}(t_1)$). Otherwise, set $z = h(X)\|h(Y)$.
2. For $i \in [N]$: Sample $(X_i, \mathsf{AUX}_i, Y_i)$ from CRS conditioned on $h(X_i)\|h(Y_i) = z$ and $h(X_i)\|h(Y_i) = h(\overline{X_i})\|h(\overline{Y_i})$. By (14), this takes time $\mathrm{poly}(t_2)$. Set
$$\begin{aligned}\delta_i = |\mathcal{A}(\mathsf{cnmExt}(X_i, Y_i, \mathsf{CRS}), \mathsf{cnmExt}(\overline{X_i}, \overline{Y_i}, \mathsf{CRS}), \mathsf{AUX}_i, \mathsf{CRS}) \\ - \mathcal{A}(U_1, \mathsf{cnmExt}(\overline{X_i}, \overline{Y_i}, \mathsf{CRS}), \mathsf{AUX}_i, \mathsf{CRS})|,\end{aligned}$$
where $\overline{X_i} = g_1(X_i, \mathsf{AUX}_i, \mathsf{CRS})$ and $\overline{Y_i} = g_2(Y_i, \mathsf{CRS})$. Note that $\mathcal{A}$ has size $\mathrm{poly}(t_2)$.
3. Compute $\delta = \frac{1}{N}\sum_{i=1}^{N} \delta_i$. If $\delta < \varepsilon + \frac{1}{4p(t_1)}$, then output $b' = z$. Else, output $b' \leftarrow \{0,1\}^{2k_2}$.

We now show that $\Pr[b' = b] \geq 1.5 \cdot 2^{-2k_2}$. It holds that $\mathbb{E}[\delta|z = b] \leq \varepsilon + \mathsf{negl}(t_1) < \varepsilon + \frac{1}{8p(t_1)}$. On the other hand, by (12) and (14) we have $\mathbb{E}[\delta|z \neq b] \geq \varepsilon + \frac{1}{2p(t_1)}$. By the Chernoff bound and the choice of $N = p(t_1)^3$, we then have

$$\Pr[b' = b|z = b] = \Pr\left[\delta < \varepsilon + \frac{1}{4p(t_1)}\middle| z = b\right] \geq 1 - \exp(-\Omega(p(t_1))) = 1 - \mathsf{negl}(t_1),$$

and

$$\Pr\left[\delta \geq \varepsilon + \frac{1}{4p(t_1)}\middle| z \neq b\right] \geq 1 - \exp(-\Omega(p(t_1))) = 1 - \mathsf{negl}(t_1).$$

The latter inequality then implies that $\Pr[b' = b|z \neq b] \geq (1 - \mathsf{negl}(t_1))2^{-2k_2}$. Combining these observations with (16) yields

$$\Pr[b' = b] \geq (2 - \mathsf{negl}(t_1))2^{-2k_2} \geq 1.5 \cdot 2^{-2k_2},$$

which contradicts (13), as desired. □ □

Based on Lemma 9, we can now work under the assumption that the event E holds and show (11). First, combining Lemma 3, (14), and the definition of E, we have that

$$\begin{aligned}\widetilde{\mathbf{H}}_\infty(f_b(X)|E, \mathsf{CRS}) = \widetilde{\mathbf{H}}_\infty(X|E, \mathsf{CRS}) \geq k' - 2k_2 - 1 \geq k + 2k_2 + 3n^\gamma \\ \widetilde{\mathbf{H}}_\infty(f_b(Y)|E, \mathsf{CRS}) = \widetilde{\mathbf{H}}_\infty(Y|E, \mathsf{CRS}) \geq k' - 2k_2 - 1 \geq k + 2k_2 + 3n^\gamma.\end{aligned}$$

As a result, by Lemma 2 it holds that with probability at least $1 - 2^{-n^\gamma} = 1 - \mathsf{negl}(t_1)$ over the fixing $\mathsf{CRS} = \mathsf{crs}$ it holds that

$$\mathbf{H}_\infty(f_b(X)|E, \mathsf{CRS} = \mathsf{crs}), \mathbf{H}_\infty(f_b(Y)|E, \mathsf{CRS} = \mathsf{crs}) \geq k + 2k_2 + 2n^\gamma. \tag{15}$$

For the remainder of the proof, we fix such a good choice $\mathsf{CRS} = \mathsf{crs}$. According to the definition of cnmExt, in order to prove that (11) holds it is now enough to show that for every size-$\mathrm{poly}(t_2)$ distinguisher $\mathcal{A}'$ we have

$$\begin{aligned}|\Pr[\mathcal{A}'(\mathsf{Ext}(f_b(X), f_b(Y)), \mathsf{SideInfo}, \mathsf{AUX}, \mathsf{crs}) = 1|E] \\ - \Pr[\mathcal{A}'(U_1, \mathsf{SideInfo}, \mathsf{AUX}, \mathsf{crs}) = 1|E]| \leq \varepsilon + \mathsf{negl}(t_1),\end{aligned} \tag{16}$$

where

$$\mathsf{SideInfo} = (h(X), h(\overline{X}), h(Y), h(\overline{Y}), f_{h(\overline{X})\|h(\overline{Y})}(\overline{X}), f_{h(\overline{X})\|h(\overline{Y})}(\overline{Y})).$$

Note that X and Y become independent under E and crs when we fix $h(X)\|h(Y) = b = b_1\|b_2$ and $h(\overline{X})\|h(\overline{Y}) = b' = b'_1\|b'_2$ for $b \neq b'$. Since the output length of h is k_2, by (15) and Lemmas 1 and 2 we have

$$\mathbf{H}_\infty(f_b(X)|E, \mathsf{CRS} = \mathsf{crs}, h(X) = b_1, h(\overline{X}) = b'_1) \geq k + n^\gamma, \tag{17}$$

$$\mathbf{H}_\infty(f_b(Y)|E, \mathsf{CRS} = \mathsf{crs}, h(Y) = b_2, h(\overline{Y}) = b'_2) \geq k + n^\gamma, \tag{18}$$

with probability $1 - \mathsf{negl}(t_1)$ over the choices of b and b'. Under E and the good fixings $\mathsf{CRS} = \mathsf{crs}$, $h(X)\|h(Y) = b$, and $h(\overline{X})\|h(\overline{Y}) = b'$, the inequality in (16) follows if we show that

$$\mathsf{Ext}(f_b(X), f_b(Y)), f_b(X), f_{b'}(\overline{Y}) \approx_\varepsilon U_1, f_b(X), f_{b'}(\overline{Y}), \tag{19}$$

because $(f_{b'}(\overline{X}), \mathsf{AUX})$ can be written as a (possibly inefficient) randomized function of $f_b(X)$ only. Since $f_b(X)$ and $(Y, f_{b'}(\overline{Y}))$ are independent under the conditionings above, and since Ext is an strong average-case (k, ε)-extractor, (19) follows from (17) and the fact that

$$\widetilde{\mathbf{H}}_\infty(f_b(Y)|E, \mathsf{CRS} = \mathsf{crs}, h(Y) = b_2, h(\overline{Y}) = b'_2, f_{b'}(\overline{Y})) \geq k,$$

which in turn holds by (18) and Lemma 1, noting that $f_{b'}$ contains a lossy function at a fixed index and hence has image size at most $2^{n-\omega} = 2^{n^\gamma}$. Since the fixings of $\mathsf{CRS} = \mathsf{crs}$, $h(X)\|h(Y)$, and $h(\overline{X})\|h(\overline{Y})$ are good with probability $1 - \mathsf{negl}(t_1)$, this implies (16) and concludes the proof.

5.2 Instantiations of Theorem 7

In this section, we instantiate Theorem 7 with the explicit statistical two-source extractors presented in Section 2. Throughout this section, we set the following parameters

$$\Omega(\lambda) \leq n \leq \mathrm{poly}(\lambda), t_1 = \lambda, t_2 = \lambda^{\log \lambda},$$

where λ is the security parameter. Then, the quasi-polynomial hardness of the DDH assumption allows us to assume the existence of the following objects:

- A family $\mathcal{H}$ of $(\mathrm{poly}(t_1), \mathsf{negl}(t_1))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{k_2}$, where $k_2 = \log \lambda \cdot \log \log \lambda$.
- A family of (t_2, n, ω)-lossy functions $\mathcal{F}$, where $t_2 \geq 2^{2k_2} = t_1^{\omega(1)}$ and $\omega = n - n^\gamma$ for some constant $\gamma \in (0, 1)$.

Using Bourgain's extractor (Proposition 1), we immediately obtain the following corollary.

Corollary 6. *Assuming quasi-polynomial hardness of the DDH assumption and for any n, t_1, and t_2 satisfying*

$$\Omega(\lambda) \leq n \leq \mathrm{poly}(\lambda), t_1 = \lambda, t_2 = \lambda^{\log \lambda},$$

there exists an explicit $(\mathrm{poly}(t_1), \mathrm{poly}(t_1), \mathrm{poly}(t_2), k' = 0.46n, \varepsilon = \mathsf{negl}(t_1), r = \Omega(n^{1-\gamma}))$*-non-malleable extractor* $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *in the CRS model.*

Although Theorem 7 is stated for sources with the same min-entropy only, it can be easily generalized to sources with different min-entropies. Using Raz's extractor (Proposition 2), we obtain the following corollary.

Corollary 7. *Assuming quasi-polynomial hardness of the DDH assumption and for any n, t_1, and t_2 satisfying*

$$\Omega(\lambda) \leq n \leq \mathrm{poly}(\lambda), t_1 = \lambda, t_2 = \lambda^{\log \lambda},$$

for all constants $\delta > 0$ and $1 > c > \gamma$ there exists an explicit $(\mathrm{poly}(t_1), \mathrm{poly}(t_1), \mathrm{poly}(t_2), k_1' = O(n^c), k_2' = (1/2 + \delta)n, \varepsilon = \mathsf{negl}(t_1), r = \Omega(n^{c-\gamma}))$*-non-malleable extractor* $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *in the CRS model.*

Finally, using the Chattopadhyay-Zuckerman extractor (Proposition 3), we obtain the following corollary.

Corollary 8. *Assuming quasi-polynomial hardness of the DDH assumption and for any n, t_1, and t_2 satisfying*

$$\Omega(\lambda) \leq n \leq \mathrm{poly}(\lambda), t_1 = \lambda, t_2 = \lambda^{\log \lambda},$$

for every constant $1 > c > \gamma$ there exists an explicit $(\mathrm{poly}(t_1), \mathrm{poly}(t_1), \mathrm{poly}(t_2), k' = O(n^c), \varepsilon = {t_1}^{-\Omega(1)}, r = \Omega(n^{c-\gamma}))$*-non-malleable extractor* $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *in the CRS model.*

6 A simple non-malleable extractor in the CRS model from nearly optimal collision-resistant hash functions

In this section, we present a simple construction of a non-malleable extractor in the CRS model against computationally bounded samplers and tamperings and against a *computationally unbounded* distinguisher that can be instantiated from families of nearly optimal collision-resistant hash functions and high min-entropy information-theoretic non-malleable extractors. To be precise, we have the following result.

Theorem 8. *Suppose $\mathcal{H}$ is a family of $(3t, 2^{\beta-1-m} = \mathsf{negl}(n))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^m$, and suppose $\mathsf{nmExt} : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$ is an explicit strong $(m - \beta - 2\log^2 n, \varepsilon = \mathsf{negl}(n), r = 1)$-non-malleable extractor. Then, there exists an explicit $(t, t, \infty, k = m - \beta + 1, \varepsilon = \mathsf{negl}(n), r = 1)$-non-malleable extractor $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ in the CRS model.*

Moreover, if nmExt is not strong, then cnmExt is a $(t, t, \infty, m - \beta + 1, \varepsilon = \mathsf{negl}(n), r = 1)$-non-malleable extractor in the CRS model without auxiliary information.

Remark 1. Note that, in Theorem 8, the underlying nmExt for m-bit sources and the resulting cnmExt for n-bit sources have similar min-entropy requirements. When $n \gg m$, this means that we start with an extractor nmExt for m-bit sources with high min-entropy rate, and construct a new extractor cnmExt for n-bit sources with very low min-entropy rate.

Theorem 8. We set $\mathsf{CRS} = H$ for $H \leftarrow \mathcal{H}$ and consider the function

$$\mathsf{cnmExt}(x, y, H) = \mathsf{nmExt}(H(x), H(y)).$$

For the sake of clarity, we present the proof for the case $r = 1$ only. The generalization to $r > 1$ tamperings is straightforward. Fix $(k = m - \beta + 1, t)$-samplable sources (X, AUX, Y) and size-poly(t) deterministic tampering functions $g_1, g_2 : \{0,1\}^n \times \mathcal{H} \to \{0,1\}^n$. Our goal is to show that

$$\Delta(\mathsf{nmExt}(H(X), H(Y)); U_1 | H, \mathsf{nmExt}(H(g_1(X, \mathsf{AUX}, H)), H(g_2(Y, H)))) = \mathsf{negl}(n). \qquad (20)$$

Consider an arbitrary fixing $H = h$. Making use of the collision-resistance properties of $\mathcal{H}$, with probability $1 - \mathsf{negl}(n)$ over the fixing $H = h$ it either holds that

$$\Pr[h(X) = h(g_1(X, \mathsf{AUX}, h))] = \mathsf{negl}(n) \tag{21}$$

or

$$\Pr[h(Y) = h(g_2(Y, h))] = \mathsf{negl}(n),$$

since either $g_1(\cdot, \mathsf{aux}, h)$ has no fixed points for any aux or $g_2(\cdot, h)$ has no fixed points. We now assume that $g_1(\cdot, \mathsf{aux}, h)$ has no fixed points for any aux, in which case (21) holds. The proof for the case where $g_2(\cdot, h)$ has no fixed points is analogous. Additionally, by Lemma 7 coupled with Lemma 2, with probability $1 - \mathsf{negl}(n)$ over the fixing $H = h$ we also have

$$h(X), h(Y) \approx_{\mathsf{negl}(n)} V, W, \tag{22}$$

where $V, W \in \{0,1\}^m$ are independent random variables satisfying

$$\mathbf{H}_\infty(V), \mathbf{H}_\infty(W) \geq m - \beta - \log^2 n.$$

After such a fixing, it now suffices to show that

$$\Delta(\mathsf{nmExt}(h(X), h(Y)); U_1 | \mathsf{nmExt}(h(X'), h(Y')), \mathsf{AUX}) = \mathsf{negl}(n), \tag{23}$$

where $X' = g_1(X, \mathsf{AUX}, h) \neq X$ and $Y' = g_2(Y, h)$. We can see $(h(X'), \mathsf{AUX})$ and $h(Y')$ as randomized functions of $h(X)$ and $h(Y)$, respectively. In other words, there exist randomized functions A, B, and C with shared randomness such that

$$\begin{aligned}&\mathsf{nmExt}(h(X), h(Y)), \mathsf{nmExt}(h(X'), h(Y')), \mathsf{AUX}\\ &\qquad\sim \mathsf{nmExt}(h(X), h(Y)), \mathsf{nmExt}(A(h(X)), B(h(Y))), C(h(X)),\end{aligned}$$

where $\Pr[A(h(X)) = h(X)] = \mathsf{negl}(n)$. Therefore, using (22), in order to prove (23) it is enough to show that

$$\Delta(\mathsf{nmExt}(V, W); U_1 | \mathsf{nmExt}(A(V), B(W)), C(V)) = \mathsf{negl}(n). \tag{24}$$

By (22) and the properties of A, it also holds that $\Pr[A(V) = V] = \mathsf{negl}(n)$. Therefore, we can condition on the event $A(V) \neq V$ and invoke Lemma 4 with nmExt, V, and W (which stay independent and have enough min-entropy after this conditioning) to conclude that (24) holds. The last statement of Theorem 8 follows by an analogous proof with a non-strong nmExt. □ □

Using the non-malleable extractor from Proposition 5 in the statement of Theorem 8, we immediately obtain the following corollary.

Corollary 9. *Suppose $\mathcal{H}$ is a family of $(3t, 2^{\beta-1-m})$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^m$ for $\beta = c \cdot m$, where $c > 0$ is a small enough constant. Then, there exists an explicit $(t, t, \infty, k = m-\beta+1, \varepsilon = \mathsf{negl}(n), r = 1)$-non-malleable extractor $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ in the CRS model.*

Note that the hash output length m in Corollary 9 controls the min-entropy requirement of cnmExt. In particular, if $m = \text{polylog}(n)$, then we obtain a low-error two-source non-malleable extractor for $\text{polylog}(n)$ min-entropy.

The birthday bound tells us that the best possible security for a hash function with m-bit outputs we can hope for is $(t, t^2/2^m)$-collision-resistant. In practice, there are several candidates for which brute-force is the best possible attack. Among them are the widely deployed hash functions SHA-256, SHA-512, SHA-3, and discrete logarithm (over elliptic curves) based constructions. Using any of these hash functions in Theorem 8 allows us to obtain a practical low-error two-source non-malleable extractor for sources with polylogarithmic min-entropy.

References

[BCD$^+$18] Avraham Ben-Aroya, Eshan Chattopadhyay, Dean Doron, Xin Li, and Amnon Ta-Shma. A new approach for constructing low-error, two-source extractors. In *Proceedings of the 33rd Computational Complexity Conference*, CCC '18, pages 3:1–3:19, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[BCDT19] Avraham Ben-Aroya, Gil Cohen, Dean Doron, and Amnon Ta-Shma. Two-source condensers with low error and small entropy gap via entropy-resilient functions. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 43:1–43:20, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[BDT16] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Low-error two-source extractors for polynomial min-entropy. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, page 106, 2016.

[BDT17] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. An efficient reduction from two-source to non-malleable extractors: Achieving near-logarithmic min-entropy. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, page 11851194, New York, NY, USA, 2017. Association for Computing Machinery.

[BHK11] Mark Braverman, Avinatan Hassidim, and Yael Tauman Kalai. Leaky pseudo-entropy functions. In *Innovations in Computer Science (ICS)*, January 2011.

[BKS$^+$10] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. *J. ACM*, 57(4):20:1–20:52, May 2010.

[Bou05] Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

[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.

[CG17] Mahdi Cheraghchi and Venkatesan Guruswami. Non-malleable coding against bit-wise and split-state tampering. *Journal of Cryptology*, 30(1):191–241, Jan 2017.

[CGGL19] Eshan Chattopadhyay, Jesse Goodman, Vipul Goyal, and Xin Li. Extractors for adversarial sources via extremal hypergraphs. Cryptology ePrint Archive, Report 2019/1450, 2019. `https://eprint.iacr.org/2019/1450`, to appear in STOC 2020.

[CGL16] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 285–298. ACM, 2016.

[CZ19] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. *Annals of Mathematics*, 189(3):653–705, 2019.

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 38(1):97–139, 2008.

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

[DRV12] Yevgeniy Dodis, Thomas Ristenpart, and Salil Vadhan. Randomness condensers for efficiently samplable, seed-dependent sources. In Ronald Cramer, editor, *Theory of Cryptography*, pages 618–635, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[DW09] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 601–610, New York, NY, USA, 2009. ACM.

[GK18] Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 685698, New York, NY, USA, 2018. Association for Computing Machinery.

[GKK19] Ankit Garg, Yael Tauman Kalai, and Dakshita Khurana. Computational extractors with negligible error in the CRS model. Cryptology ePrint Archive, Report 2019/1116, 2019. `https://eprint.iacr.org/2019/1116`, to appear in Eurocrypt 2020.

[GKP$^+$18] Vipul Goyal, Ashutosh Kumar, Sunoo Park, Silas Richelson, and Akshayaram Srinivasan. Non-malleable commitments from non-malleable extractors. 2018. unpublished.

[KLR09] Yael Tauman Kalai, Xin Li, and Anup Rao. 2-source extractors under computational assumptions and cryptography with defective randomness. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 617–626, Oct 2009.

[Lew19] Mark Lewko. An explicit two-source extractor with min-entropy rate near 4/9. *Mathematika*, 65(4):950957, 2019.

[Li11] Xin Li. A new approach to affine extractors and dispersers. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 137–147, June 2011.

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 168–177, Oct 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*, pages 1144–1156. ACM, 2017.

[MW97] Ueli Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In Burton S. Kaliski, editor, *Advances in Cryptology — CRYPTO '97*, pages 307–321, Berlin, Heidelberg, 1997. Springer Berlin Heidelberg.

[PW11] Chris Peikert and Brent Waters. Lossy trapdoor functions and their applications. *SIAM Journal on Computing*, 40(6):1803–1844, 2011.

[Rao08] Anup Rao. A 2-source almost-extractor for linear entropy. In Ashish Goel, Klaus Jansen, José D. P. Rolim, and Ronitt Rubinfeld, editors, *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 549–556, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[Raz05] Ran Raz. Extractors with weak random seeds. In *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pages 11–20, New York, NY, USA, 2005. ACM.

[TV00] Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 32–42, Nov 2000.

[ZB11] Noga Zewi and Eli Ben-Sasson. From affine to two-source extractors via approximate duality. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing*, STOC '11, pages 177–186, New York, NY, USA, 2011. ACM.

[Zuc06] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing*, STOC '06, pages 681–690, New York, NY, USA, 2006. ACM.