Two-Source Non-Malleable Extractors and Applications to Privacy Amplification with Tamperable Memory

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

Abstract

We study two-source non-malleable extractors, which extract randomness from weak sources even when an adversary is allowed to learn the output of the extractor on correlated inputs. First, we study consequences of improving the best known constructions of such objects. 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. Moreover, we show the resulting extractor can be made *non-malleable* for samplable sources in the computational CRS model introduced by Garg, Kalai, and Khurana (Eurocrypt 2020) under standard hardness assumptions, against an *unbounded* distinguisher. Remarkably, previous constructions of similar extractors require much stronger assumptions.

To complement the above, we study unconditional explicit constructions of computational two-source non-malleable extractors for samplable sources in the CRS model with significantly better parameters than their information-theoretic counterparts by exploiting stronger hardness assumptions. Under a quasipolynomial hardness assumption, we achieve security against bounded distinguishers, while assuming the existence of nearly optimal collision-resistant hash functions allows us to achieve security against unbounded distinguishers.

Finally, we introduce the setting of *privacy amplification resilient against memory-tampering active adversaries*. Here, we aim to design privacy amplification protocols that are resilient against an active adversary that can additionally choose one honest party at will and arbitrarily corrupt its memory (i.e., its shared secret and randomness tape) before the execution of the protocol. We show how to design such protocols using two-source non-malleable extractors.

1 Introduction

Two-source extractors. 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 30 years, with fundamental connections to combinatorics and many applications in computer science. The first non-trivial explicit construction was given by Chor and Goldreich [CG88], who 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. A standard application of the probabilistic method shows that (inefficient) low-error two-source extractors exist for polylogarithmic min-entropy. While several attempts were made to improve the construction of [CG88] to allow

*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

for sources with smaller min-entropy, the major breakthrough results were obtained after almost two decades. 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. In an incomparable result, Bourgain [Bou05] gave an explicit low-error two-source extractor for sources with min-entropy $(1/2 - \gamma)n$, where $\gamma > 0$ is a small constant. Recently, an improved analysis by Lewko [Lew19] showed that Bourgain's extractor can handle sources with min-entropy $4n/9$. In another groundbreaking work, 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, Coh17, Li17], with the state-of-the-art currently found in [Li19]).

Seeded non-malleable extractors. The key ingredient in these recent constructions [CZ19, Li16, BDT17, Coh17, Li17, Li19] of two-source extractors are the so called seeded non-malleable extractors. In a breakthrough result, Dodis and Wichs [DW09] introduced the notion of seeded non-malleable extractors as a natural tool towards achieving privacy amplification against active adversaries [MW97] with optimal number of rounds and small entropy loss. Roughly speaking, the output of a seeded non-malleable extractor with a uniformly random seed and a source X with some min-entropy should look uniformly random to an adversary who can tamper the seed and obtain the output of the non-malleable extractor on a tampered seed. There has been a long line of work constructing explicit seeded non-malleable extractors (and hence such privacy amplification protocols) with significantly improved parameters (see [ACLV19, Li19] and references therein).

Non-malleable two-source extractors. A natural strengthening of both seeded non-malleable extractors, and two-source extractors are two-source *non-malleable* extractors (also known as seedless non-malleable extractors). Two-source non-malleable extractors were introduced by Cheraghchi and Guruswami [CG17] in the single-tampering setting and by Chattopadhyay, Goyal, and Li [CGL16] in the multi-tampering setting. 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 on several inputs correlated with the original sources. In other words, we require that

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

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 original motivation for studying efficient two-source non-malleable extractors stems from the fact that they directly yield efficient split-state non-malleable codes [DPW18] (provided the extractor also supports efficient preimage sampling).

The first constructions of non-malleable codes [DKO13, ADL14] relied heavily on the (limited) non-malleability of the inner-product two-source extractor. Subsequent improved constructions of non-malleable codes in the split-state model relied on both the inner-product two-source extractor [ADKO15, AO20], and on more sophisticated constructions of two source non-malleable extractors [CGL16, Li17, Li19]. This problem has been extensively studied in the literature, and for a more exhaustive list of works on non-malleable codes in the split-state model, see [AO20] and the references therein. Non-malleable codes and two-source non-malleable extractors have subsequently found other applications such as non-malleable secret sharing [GK18a, GK18b, BS19, ADN+19], randomness extraction from adversarial sources [CGGL20], and network extraction protocols [GSZ20].

1.1 Open questions

One of the main applications of two-source extractors is in cryptography, where it is crucial that the extractor has negligible error. The results of [CZ19, BDT17, Coh17, Li17, Li19] described earlier are unfortunately not appropriate for these applications, because the error of these extractors is non-negligible. 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, BACD+18], and constructing low-error two-source extractors for low min-entropy in the computational setting under different hardness assumptions [KLR09, GKK20]. Other works have focused on computational deterministic extractors [TV00] and seeded extractors [DGKM12]. The question of constructing low-error two-source extractors for low min-entropy still remains open. Thus, it is desirable to develop new promising approaches to this problem. A line of works has focused on reducing this open problem to improving the parameters of known constructions of other pseudorandom objects. Zewi and Ben-Sasson [ZB11] showed that certain improvements to affine extractors would lead to low-error two-source extractors for low min-entropy assuming the Polynomial Freiman-Ruzsa conjecture. Ben-Aroya et al. [BACD+18], adapting techniques from [Li12, CZ19], showed that seeded non-malleable extractors with improved seed length would also settle this question. In light of these connections, one might ask the following natural question

Question 1. *Do slight improvements to state-of-the-art explicit two-source non-malleable extractors lead to explicit low-error two-source extractors for low min-entropy?*

The state-of-the-art construction of a two-source non-malleable extractor by Li [Li19] requires 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 small constant $\gamma > 0$. On the other hand, the probabilistic method shows that there exist two-source non-malleable extractors for min-entropy δn, with $\delta > 0$ an arbitrarily small constant, handling $n^{\Omega(1)}$ tamperings with error $2^{-\Omega(n)}$ and output length $\Omega(n)$ [CGGL20]. This state of affairs naturally raises the following question.

Question 2. *Is there an explicit construction of a low-error two-source non-malleable extractor for min-entropy δn with δ much smaller than 1?*

Finally, all prior applications of non-malleable two-source extractors outside randomness extraction are via efficient constructions of non-malleable codes in the split-state model, and hence require efficient preimage sampling. Notwithstanding, non-malleable two-source extractors seem interesting in their own right, and it would be desirable to have an application of two-source non-malleable extractors that doesn't require efficient preimage sampling.

Question 3. *Is there a natural application of non-malleable two-source extractors that does not require efficient preimage sampling?*

In this work, we make progress on all the questions above.

1.2 Our contributions

Implications of improved two-source non-malleable extractors. We begin by answering Question 1 in the positive, and show an even stronger result in the computational setting: We prove that a small improvement to the parameters of [CGL16, Li17, Li19] leads to explicit low-error two-source extractors for min-entropy δn with a small constant $\delta > 0$. Put differently, the constructions of [CGL16, Li17, Li19] are almost the best we can hope for without solving a longstanding open problem in pseudorandomness along the way.

Then, we show that the reduction above can be strengthened *in the computational setting*. More precisely, we consider the Common Reference String (CRS) model introduced by Garg, Kalai, and Khurana [GKK20]. At a high-level, in this model a CRS is sampled once and for all, and we consider three adversaries 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 source samples, 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. Assuming the existence of collision-resistant hash functions, we prove that the low-error two-source extractor from our reduction above can be made *non-malleable* if the sampler and tamperer are computationally bounded, while the distinguisher may be unbounded. In contrast, existing computational extractors require much stronger hardness assumptions.

Constructions of two-source non-malleable extractors in the CRS model. We study Question 2 further in the CRS model. While our previous result requires yet unattained statistical non-malleable extractors, we present two explicit constructions of two-source non-malleable extractors in the CRS model with significantly improved parameters.

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 best statistical counterparts [CGL16, Li17, Li19], against a computationally bounded distinguisher. This construction achieves essentially the same parameters as the extractor from [GKK20], which only handles *one-sided* tampering, under the same hardness assumption. While the previous construction requires a bounded distinguisher, we also give a simple low-error two-source non-malleable extractor in the CRS model for very low min-entropy against a computationally *unbounded* distinguisher, assuming the existence of nearly optimal collision-resistant hash functions.

Novel application of non-malleable extractors: Privacy amplification resilient against memory-tampering active adversaries. To complement our previous results on two-source non-malleable extractors and taking into account Question 3, we introduce a natural extension of the well-known problem of privacy amplification against active adversaries originally considered by Maurer and Wolf [MW97], where the active adversary also has memory-tampering capabilities. Remarkably, we show that two-source non-malleable extractors (even without efficient preimage sampling) can be used to design privacy amplification protocols in this stronger adversarial setting.

More precisely, we extend the classical problem of Maurer and Wolf [MW97] to a setting where the active adversary, Eve, is also allowed to *fully corrupt* the internal memory of one of the honest parties, Alice and Bob, before the execution of the protocol. Informally, in an initial phase we assume that Alice and Bob share a secret W with sufficient min-entropy, and that they have access to local independent randomness tapes A and B, respectively, which may also be weak sources.

[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 say (W, A) (resp. (W, B)) is Alice's *memory* (resp. Bob's). Before the execution of the privacy amplification protocol between Alice and Bob, we allow Eve to specify a tampering function F and one of Alice and Bob to be corrupted (e.g., by infecting either Alice's or Bob's storage device with a virus). If, say, Alice is chosen, then Alice's memory (W, A) is replaced by $(\widetilde{W}, \widetilde{A}) = F(W, A)$. Eve does not learn the output of F, and Alice and Bob do not know whether (or which) memory was corrupted. Crucially, note that the secret W and the randomness tape *are tampered together*, which means that the corrupted secret and tape may be arbitrarily correlated. Finally, Alice and Bob run some interactive protocol where Eve is allowed to tamper all messages between the honest parties. The goal of the privacy amplification protocol is twofold:

1. **Eve is passive:** If Eve does not tamper neither of Alice's or Bob's memories nor does she tamper any of the messages between them, then Alice and Bob must agree on a shared key that is (statistically or computationally) indistinguishable from the uniform distribution to Eve.

2. **Eve is active:** In this case, with high probability either one of Alice or Bob detects tampering, or they agree on a shared key that is indistinguishable from the uniform distribution to Eve.

In the information-theoretic setting, we show that information-theoretic low-error two-source non-malleable extractors with sufficiently low min-entropy requirement yield 4-round privacy amplification protocols resilient against memory-tampering active adversaries with good parameters, completing our study of Question 2. As mentioned before, it is known that such (inefficient) extractors exist even with much better parameters than required [CGGL20], but, as we show in this paper, obtaining such *explicit* extractors is an extremely challenging problem. To complement this result, we show that, assuming the subexponential hardness of the DDH assumption, explicit computational two-source non-malleable extractors in the CRS model can be used to give such explicit computational 4-round privacy amplification protocols in the CRS model.

We put our results in context of previous work in Section 1.3, and provide a more technical overview in Section 1.4.

1.3 Comparison to previous work

Computational extractors. 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. Computational seeded extractors were also studied by Dachman-Soled, Gennaro, Krawczyk, and Malkin [DGKM12], who considered the standard approach of composing an information-theoretic extractor with a pseudorandom generator.

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 [GKK20] 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 [BACD$^+$18] to construct an extractor of the same type for unbalanced sources with lower min-entropy. We remark that the assumption in [GKK20] 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 [GKK20]. To go below such min-entropy requirements, a subexponential hardness assumption appears to be necessary.

Our computational two-source non-malleable extractors are constructed in the CRS model of [GKK20]. Consequently, our results are incomparable to those of [TV00, KLR09, DGKM12]. As mentioned before, [GKK20] construct two-source extractors that handle one-sided tampering In contrast, we focus on constructing two-source non-malleable extractors, which handle two-sided tampering. Moreover, there are other key differences with respect to [GKK20]. 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, previous results on low-error computational (even malleable) extractors for low min-entropy in the CRS model require at least subexponential hardness assumptions. For our second construction, we make use of a quasi-polynomial hardness assumption, and similarly to [GKK20] consider a computationally bounded distinguisher. We are able to essentially match the parameters of the one-sided tampering extractor obtained in [GKK20] under the same hardness assumption. 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). A comparison of our constructions with previous work can be found in Table 1.

Privacy amplification. The setting of privacy amplification is fundamental in cryptography, and it has deep connections to randomness extractors. Strong seeded extractors yield non-interactive privacy amplification protocols against a passive eavesdropper [BBR88, BBCM95], while strong seeded non-malleable extractors were introduced by Dodis and Wichs [DW09] to obtain 2-round (which is optimal) privacy amplification protocols against active adversaries, a setting originally introduced in [MW97]. This has led to a deep line of work constructing explicit seeded non-malleable extractors (and hence such privacy amplification protocols) with significantly improved parameters (see [Li19] and references therein). In a different direction, the setting of *fuzzy* privacy amplification has also received significant attention. Here, the secrets of Alice and Bob are not necessarily equal, but may only be close in some metric. *Fuzzy* extractors [DORS08] yield non-interactive fuzzy privacy amplification protocols, effectively showing that information reconciliation and regular privacy amplification can be accomplished together in a single round. When the adversary is active, *robust fuzzy* extractors can be used to obtain such fuzzy privacy amplification

[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 at most 2^{n^c} solves the Decisional Diffie-Hellman problem with non-negligible (in n) advantage.

protocols [BDK+05, CDF+08, DKK+12]. Similar problems have been studied in the computational setting [DHP+18, EHOR20].

Privacy amplification with tamperable memory is harder than regular privacy amplification against a passive or active adversary, and is incomparable to fuzzy privacy amplification. In our setting, there is no guarantee that the secrets held by Alice and Bob are close according to some distance after tampering, and, unlike other privacy amplification settings, the tampered secret may even be correlated with the party's randomness. On the other hand, in fuzzy privacy amplification one requires that privacy be achieved even when Alice's and Bob's secrets are different (but close enough), provided the adversary remains passive during the protocol. In our setting, we must allow the parties to abort if either Alice's or Bob's memory is tampered, as the task is impossible otherwise.

Other works on cryptography with tamperable memory. Besides the concept of non-malleability, the extension of cryptographic problems to settings with tamperable memory has also been considered in various ways, e.g., [GLM+04, IPSW06, KKS11]. Most relevant to our privacy amplification problem, Austrin, Chung, Mahmoody, Pass, and Seth [ACM+14] study key agreement protocols, which are intimately connected to privacy amplification, in a setting where the adversary is allowed to tamper the randomness of both parties via an *online p-tampering attack* before the protocol starts. This setting and the associated result are incomparable to ours. Indeed, there are two key differences: On the one hand, we consider arbitrary tampering attacks that jointly target the randomness and shared secret of a party, which are much stronger than the online p-tampering attacks considered in [ACM+14]. On the other hand, we must restrict tampering to one of the parties, as otherwise privacy amplification is impossible.

1.4 Technical Overview

1.4.1 Slightly better non-malleable extractors imply great two-source extractors

To show that small improvements to the best known two-source non-malleable extractors low-error lead to explicit low-error two-source extractors for low min-entropy sources, we consider 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 uniform seed, 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. Then, computing $\mathsf{nmExt}(X', Y')$ would lead to nearly uniform randomness without even exploiting the non-maleability of nmExt. Although deterministic condensers do not exist, there does exist a deterministic object with related properties, called a *somewhere-condenser*. Such an object 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 suffixes added to ensure that the tamperings induced by the somewhere-condenser 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 $\ell^2 - 1$ 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 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 $2^{-\Omega(n)}$ [CGGL20].

1.4.2 Slightly better non-malleable extractors imply great computational non-malleable extractors under standard assumptions

Given our reduction above, it is natural to wonder whether Ext defined in (1) 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 suffices to break the (information-theoretic) non-malleability of Ext. We move this problem to the CRS model [GKK20], and ask instead whether Ext can be made non-malleable in this computational model.

The CRS model. 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 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. Formal definitions can be found in Section 2.4.

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.

Modifying the extractor. Intuitively, the only way to break non-malleability of Ext is to proceed as 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. In the CRS model, we overcome this problem by sampling a collision-resistant hash function H from *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), 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 modified function

$$\mathsf{cnmExt}(X, Y, H) = \bigoplus_{i,j \in [\ell]} \mathsf{nmExt}(X'_i \| p_i \| H(X), Y'_j \| p_j \| H(Y)) \tag{2}$$

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. Remarkably, we obtain an exact analogue of Theorem 1 in the CRS model from *standard assumptions* and *with added non-malleability.* More details can be found in Section 4.

1.4.3 Non-malleable extractors in the CRS model

Since our previous result is conditional on small improvements on the information-theoretic two-source non-malleable extractors from [CGL16, Li19], we turn to obtaining unconditional explicit constructions of two-source non-malleable extractors in the CRS model. Table 1 compares our constructions with previous results on computational two-source extractors.

Two-source non-malleable extractors in the CRS model from quasi-polynomial hardness. Building on techniques developed in [BHK11, GKK20], we construct an explicit two-source non-malleable extractor against a computationally bounded distinguisher assuming the quasipolynomial hardness of DDH with essentially the same parameters as the corresponding extractor from [GKK20], which only handles one-sided tampering.

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 $f : \{0,1\}^n \to \{0,1\}^n$

[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.

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 families of collision-resistant hash functions $\mathcal{H}_1$ and $\mathcal{H}_2$ with output lengths not too large.

We show that a simple modification of the extractor from [GKK20] for one-sided tampering is enough to obtain a two-source non-malleable extractor cnmExt in the CRS model for error and min-entropy requirements matching those of the best statistical *malleable* two-source extractors under the quasi-polynomial hardness of the DDH assumption. This construction is quite flexible, and we shall see that upgrading the hardness assumption to the subexponential hardness of the DDH assumption also allows us to assume that one of the sources can be sampled in subexponential time.

For simplicity, we illustrate only the case where $\mathcal{H}_1 = \mathcal{H}_2$. To set the CRS, first we sample hash functions $h \leftarrow \mathcal{H}$ with output length ℓ. Then, we sample $b \leftarrow \{0,1\}^{2\ell}$, and sample f_{ij} from $\mathcal{F}$ for $i \in [2\ell]$ 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 as

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

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

$$f_a(x) = f_{1a_1}(f_{2a_2}(\cdots(f_{2\ell a_{2\ell}}(x))\cdots)).$$

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 are still independent do not lose much min-entropy when we reveal $f_{h(\overline{X})\|h(\overline{Y})}(\overline{X})$, $\mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS})$, and all the hashes. This allows us to invoke the statistical properties of Ext to obtain the desired result.

As an example, instantiating Ext with the best known statistical two-source extractors [Bou05, Raz05, Lew19, CZ19] yields the following informal result. Formal statements and more details can be found in Section 5.

Theorem 2 (Informal)**.** *Assuming the quasi-polynomial hardness of the DDH assumption, there exist explicit two-source non-malleable extractors in the CRS model with essentially the same parameters as the best explicit information-theoretic two-source (malleable) extractors.*

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 the sense that a birthday attack is essentially the best possible). In practice, this is not a far-fetched assumption: For most widely deployed hash functions such as SHA-256, SHA-512, and SHA-3 we currently cannot do better than a birthday attack.

Our construction is given by $\mathsf{cnmExt}(X, Y, H) = \mathsf{nmExt}(H(X), H(Y))$, where H is sampled from a family of nearly optimal collision-resistant hash functions $\mathcal{H}$, and $\mathsf{CRS} = H$. Intuitively, the security of this construction follows not only from the collision-resistance of H, but also from the fact that both $H(X)$ and $H(Y)$ are statistically close to high min-entropy sources [DRV12]. Formal statements and more details can be found in Section 6.

Theorem 3 (Informal)**.** *If there exist nearly optimal collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^\ell$ for some $\ell = \Omega(\mathrm{polylog}(n))$, then there exists an explicit low-error two-source non-malleable extractor for n-bit sources with min-entropy ℓ in the CRS model.*

1.4.4 Privacy amplification resilient against memory-tampering active adversaries

We introduce the setting of privacy amplification resilient against memory-tampering active adversaries (a high-level description of the problem can already be found in Section 1.2), and show that low-error two-source non-malleable extractors with good min-entropy requirements can be used to design privacy amplification protocols in this setting, both in the information-theoretic and computational settings. To be more precise, we consider the 4-round protocol illustrated in Figure 1 instantiated with an appropriate strong two-source non-malleable extractor nmExt.

$$
\begin{array}{lcl}
\text{Alice} & & \text{Bob} \\
\text{Memory: } (W, A) & & \text{Memory: } (W, B) \\
 & A \longrightarrow A' & \\
 & B' \longleftarrow B & \\
R_A = \mathsf{nmExt}(W, A \| B') & & R_B = \mathsf{nmExt}(W, A' \| B) \\
 & [R_A]_\alpha \longrightarrow [R_A]'_\alpha & \\
 & [R_B]'_{\alpha\cdot 2\alpha} \longleftarrow [R_B]_{\alpha:2\alpha} & \\
\text{If } [R_A]_{\alpha\cdot 2\alpha} = [R_B]'_{\alpha:2\alpha} & & \text{If } [R_B]_\alpha = [R_A]'_\alpha \\
\quad \text{then } S_A = [R_A]_{2\alpha:} & & \quad \text{then } S_B = [R_B]_{2\alpha:} \\
\text{Otherwise } S_A = \perp & & \text{Otherwise } S_B = \perp
\end{array}
$$

Figure 1: Privacy amplification protocol against memory-tampering active adversaries. In the above, for an n-bit string x we define $[x]_i = (x_1, x_2, \ldots, x_i)$, $[x]_{i:j} = (x_{i+1}, \ldots, x_j)$, and $[x]_{j:} = (x_{j+1}, \ldots, x_n)$.

We provide an informal argument for why the protocol from Figure 1 works in the information-theoretic setting when the randomness tapes A and B of Alice and Bob, respectively, are sufficiently shorter than the shared secret W. If Eve is passive and simply eavesdrops on the communication between Alice and Bob, then the fact that nmExt is a strong extractor implies that Alice and Bob output strings $S_A \neq \perp$ and $S_B \neq \perp$, respectively, such that $S_A = S_B$ and which are close to uniform with respect to Eve's view. Assume now that Eve is active and corrupts Alice. This means that Alice's memory, (W, A), is replaced by an arbitrary function $(\widetilde{W}, \widetilde{A}) = F(W, A)$. We show that, with high probability over the fixings $A = a$ and $\widetilde{A} = \widetilde{a}$, with high probability either one of Alice or Bob aborts (i.e., outputs $\perp$), or we have $S_A = S_B \neq \perp$ and S_A, S_B are close to uniform from Eve's view. Observe that, after fixing $A = a$ and $\widetilde{A} = \widetilde{a}$, we have that $\widetilde{W}$ is now a tampering of W only. Moreover, assuming that A is appropriately shorter than W, with high probability it holds that W still has enough min-entropy after conditioning on $A = a$ and $\widetilde{A} = \widetilde{a}$. For the sake of exposition, suppose that $\widetilde{W} \neq W$ always. We claim that, in this case, Bob detects tampering and aborts. Indeed, note that $R_A = \mathsf{nmExt}(\widetilde{W}, a \| B')$ can be written as $R_A = \mathsf{nmExt}(g_1(W), g_2(\widetilde{a} \| B))$

for tampering functions g_1 and g_2, where g_1 has no fixed points. Therefore, by the properties of nmExt, we conclude that, even after revealing $(a, \widetilde{a}, B, R_A)$, we have that $R_B = \mathsf{nmExt}(W, \widetilde{a} \| B)$ is close to uniform. As a result, in the third round of the protocol, Eve can only guess a (long enough) prefix of R_B with very small probability. Therefore, Bob aborts with high probability, as desired. As an example, we can obtain the following informal result.

Theorem 4 (Informal)**.** *If* nmExt *is a low-error two-source non-malleable extractor for sources with min-entropy* $0.05n$*, then there exists a 4-round privacy amplification protocol resilient against memory-tampering active adversaries when* $\mathbf{H}_\infty(W) \geq 0.3n$ *and* $\mathbf{H}_\infty(A), \mathbf{H}_\infty(B) \geq 0.05n$ *whenever* $|A|, |B| \leq 0.1n$*.*

Since (inefficient) two-source non-malleable extractors are known to exist with much better parameters than those required above [CGGL20], we readily conclude that 4-round privacy amplification protocols resilient against memory-tampering active adversaries exist with very good parameters. However, we do not know any explicit constructions of two-source non-malleable extractors with sufficiently good min-entropy requirements, and we have given evidence in Section 1.4.1 that obtaining such extractors appears to be extremely challenging. This novel application provides an additional motivation for obtaining improved two-source non-malleable extractors.

Nevertheless, we show that, extending the notion of privacy amplification resilient against memory-tampering active adversaries to the CRS model, we can exploit our explicit constructions of two-source non-malleable extractors in the CRS model to obtain *explicit* 4-round privacy amplification protocols resilient against memory-tampering active adversaries in the CRS model with very good parameters. We remark that this task is not trivial in the CRS model, because the shared secret W and randomness tapes A and B are arbitrarily correlated with the CRS, and Eve also has full knowledge of the CRS at all times. Moreover, care is needed in this computational setting because we need to ensure that sources remain samplable by appropriately sized circuits even after some conditionings. We overcome this by using a strong two-source non-malleable extractor that allows the left source to be sampled in subexponential time.

Theorem 5 (Informal)**.** *Assuming the sub-exponential hardness of the DDH assumption, there exists an explicit 4-round privacy amplification protocol resilient against memory-tampering active adversaries in the CRS model for* $W \in \{0,1\}^n$ *such that* $\mathbf{H}_\infty(W) \geq 0.52n$ *and* A, B *with sublinear min-entropy.*

Formal statements and more details can be found in Section 7.

1.5 Organization

In Section 2, we introduce notation and preliminary concepts and results. Section 3 discusses the information-theoretic 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. Simple non-malleable extractors in the CRS model obtained from nearly optimal collision-resistant hash functions are analyzed in Section 6. Finally, we discuss our novel privacy amplification setting and protocols using two-source non-malleable extractors in Section 7.

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 $x\|y$. Additionally, given an n-symbol string x, we define $[x]_i = (x_1, x_2 \dots, x_i)$, $[x]_{i:j} = (x_{i+1}, x_{i+2}, \dots, x_j)$, and $[x]_{i:} = (x_{i+1}, x_{i+2}, \dots, x_n)$. 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 $|\mathsf{supp}(Z)| \le 2^\lambda$, we have*

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

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

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

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

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

Proof. We have

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

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

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 $\mathbf{H}_\infty(X)\ge k$. Moreover, X is said to be* flat *if it is uniformly distributed over a set of size at least 2^k.*

Definition 5 ((k_1,k_2,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,W) and Y such that $\widetilde{\mathbf{H}}_\infty(X|W)\ge k_1$ and Y is an (n,k_2)-source we have*

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

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 $\mathsf{Ext}:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^m$ with $k_1=(1/2+\gamma)n$, $k_2=O(\log n)$, $\varepsilon=2^{-\Omega(n)}$, and $m=\Omega(n)$.*

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_1,k_2,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_1,k_2,ε,r)-non-malleable extractor *if for every pair of independent distributions (X,W) and Y such that $\widetilde{\mathbf{H}}_\infty(X|W)\ge k_1$ and Y is an (n,k_2)-source, 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 $i=1,\dots,r$, we have*

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

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

Proposition 4 ([CGL16], [GSZ20, Appendix A])**.** *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], [GSZ20, Appendix A])**.** *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* $m = \Omega(n/\log n)$.

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 $i = 1, \dots, r$ either $g_{1i}(\cdot, \mathsf{rand})$ or $g_{2i}(\cdot, \mathsf{rand})$ has no fixed points, it holds that*

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

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*

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

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 $F(Y, \mathsf{rand})$ is a function of Y only). □

Definition 7 (Somewhere-k sources)**.** *A distribution $Y = (Y_1, \dots, Y_\ell) \in \{0,1\}^{m\cdot\ell}$ 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 (Samplable sources in the CRS model)**.** *A tuple*

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

is said to be a tuple of (t_1, t_2, k_1, k_2)-samplable sources in the CRS model *if there exists* $\mathsf{CRS} \in \{0,1\}^c$ *such that the following hold:*

- *There exist a size-t_1 circuit $\mathcal{G}_1$ and a size-t_2 circuit $\mathcal{G}_2$ such that $X \leftarrow \mathcal{G}_1(\mathsf{CRS})$ and $(Y, \mathsf{AUX}) \leftarrow \mathcal{G}_2(\mathsf{CRS})$.*
- *X and (Y, AUX) are conditionally independent given CRS.*
- *$\mathbf{H}_\infty(X | \mathsf{CRS} = \mathsf{crs}) \geq k_1$ and $\mathbf{H}_\infty(Y | \mathsf{CRS} = \mathsf{crs}) \geq k_2$ for every fixing $\mathsf{CRS} = \mathsf{crs}$.*

When AUX is the empty string, we say (X, Y) are (t_1, t_2, k_1, k_2)-samplable sources without auxiliary information.

For simplicity, when $t_1 = t_2 = t$ we say that (X, Y, AUX) are (t, k_1, k_2)-samplable, when $k_1 = k_2 = k$ we say that (X, Y, AUX) are (t_1, t_2, k)-samplable, and when both hold we say that (X, Y, AUX) are (t, k)-samplable.

Definition 10 (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* $(t_1, t_2, t_1', t_2', t'', k_1, k_2, \varepsilon, r)$-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_1, t_2, k_1, k_2)-samplable sources from CRS, every tuple of deterministic size-t_1' circuits $g_{11}, \ldots, g_{1r} : \{0,1\}^n \times \{0,1\}^c \to \{0,1\}^n$ and size-t_2' $g_{21}, \ldots, g_{2r} : \{0,1\}^n \times \{0,1\}^a \times \{0,1\}^c \to \{0,1\}^n$ such that for every $i \in [r]$ and every fixing $\mathsf{CRS} = \mathsf{crs}$ either $g_{1i}(\cdot, \mathsf{crs})$ has no fixed points or $g_{2i}(\cdot, \mathsf{aux}, \mathsf{crs})$ has no fixed points for every fixing $\mathsf{AUX} = \mathsf{aux}$, and every size-$t''$ adversary $\mathcal{A}$ we have

[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.

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

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

We say cnmExt *is a* $(t_1,t_2,t_1',t_2',t'',k,\varepsilon,r)$-non-malleable extractor without auxiliary information *if the above holds for all* (t_1,t_2,k_1,k_2)*-samplable sources* (X,Y) *without auxiliary information.*

For simplicity, when $t_1=t_2=t$*,* $t_1'=t_2'=t'$*, and* $k_1=k_2=k$*, we say that* cnmExt *is a* $(t,t',t'',k,\varepsilon,r)$*-non-malleable extractor in the CRS model.*

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* $\Omega(\lambda) \le n \le \mathrm{poly}(\lambda)$ *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 non-malleable extractors to great two-source extractors

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 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* $\gamma = \gamma(m) \ge 1/m$ *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

$$\bigoplus_{i,j \in [\ell]} \mathsf{nmExt}(V_i \| p_i, W_j \| p_j) \approx_\varepsilon U_1, \tag{3}$$

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

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

Consider arbitrary fixings $I = i$ and $J = j$. We show that (3) holds for all fixings, and hence it holds in general as well. Under such a fixing, it is enough to show that

$$\Delta(\mathsf{nmExt}(V_i \| p_i, W_j \| p_j); U_1 | (\mathsf{nmExt}(V_{i'} \| p_{i'}, W_{j'} \| p_{j'})_{(i',j') \ne (i,j)}) \le \varepsilon. \tag{4}$$

We will now use the properties of nmExt to prove (4). Note that we can jointly simulate all pairs $(V_{i'} \| p_{i'}, W_{j'} \| p_{j'})$ for $(i',j') \ne (i,j)$ as randomized split-memory-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

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

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 (4) holds, which completes the proof. □

We now present two remarkable corollaries of Theorem 6, one of which was already informally presented in Section 1.4.

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 From slightly better non-malleable extractors to great computational non-malleable extractors under standard assumptions

In this section, we show how the construction used to prove Theorem 6 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 7. *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 6 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-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

$$\mathsf{cnmExt}(X,Y,H), \mathsf{cnmExt}(X',Y',H), H \approx_\varepsilon U_1, \mathsf{cnmExt}(X',Y',H), H, \tag{5}$$

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-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

$$\mathbf{H}_\infty(V_i | I = i) \geq (1-\gamma)m' \geq (1-\gamma)m - 3\log m - m_h, \tag{6}$$
$$\mathbf{H}_\infty(W_j | J = j) \geq (1-\gamma)m' \geq (1-\gamma)m - 3\log m - m_h. \tag{7}$$

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

$$\bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(\mathsf{SCond}(X)_i\|i\|z_1, \mathsf{SCond}(Y)_j\|j\|z_2),$$
$$\bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(\mathsf{SCond}(X))_i, B(\mathsf{SCond}(Y))_j)$$
$$\approx_{\varepsilon'} U_1, \bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(\mathsf{SCond}(X))_i, B(\mathsf{SCond}(Y)_j). \quad (8)$$

Using the fact that $\mathsf{SCond}(X), \mathsf{SCond}(Y) \approx_{2\varepsilon_1} V, W$, the condition in (8) follows if we show that

$$\bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(V_i\|i\|z_1, W_j\|j\|z_2), \bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(V)_i, B(W)_j)$$
$$\approx_{\varepsilon} U_1, \bigoplus_{i,j\in[\ell]} \mathsf{nmExt}(A(V)_i, B(W)_j). \quad (9)$$

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

$$\Delta(\mathsf{nmExt}(V_{i^\star}\|i\|z_1, W_{j^\star}\|j\|z_2); U_1$$
$$|(\mathsf{nmExt}(V_i\|i\|z_1, W_j\|j\|z_2))_{(i,j)\neq(i^\star,j^\star)}, (\mathsf{nmExt}(A(V)_i, B(W)_j))_{i,j\in[\ell]}) \leq \varepsilon, \quad (10)$$

which implies (9) and concludes the proof. Analogously to the proof of Theorem 6, 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 (6) and (7), we can invoke Lemma 4 to conclude (10) 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 1.4.

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 8. *Suppose the following objects exist:*

- *A family $\mathcal{H}_1$ of $(\mathrm{poly}(t_{11}), \mathsf{negl}(t_{11}))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{\ell_1}$;*
- *A family $\mathcal{H}_2$ of $(\mathrm{poly}(t_{12}), \mathsf{negl}(t_{12}))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{\ell_2}$;*
- *A family of $(\mathrm{poly}(t_2), n, \omega)$-lossy functions $\mathcal{F}$, where $t_2 \geq 2^{\ell_1+\ell_2}$, with $2^{\ell_1} = t_{11}^{\omega(1)}$, $2^{\ell_2} = t_{12}^{\omega(1)}$, and $\omega = n - n^\gamma$ for some constant $\gamma \in (0,1)$.*
- *A strong (k_1, k_2, ε)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, where $\Omega(t_1) \leq n \leq \mathrm{poly}(t_1)$ for $t_1 = \min(t_{11}, t_{12})$.*

Then, there exists an explicit $(\mathrm{poly}(t_{11}), \mathrm{poly}(t_{12}), \mathrm{poly}(t_1), \mathrm{poly}(t_2), k_1' = k_1 + r(2\ell_1 + 2n^\gamma), k_2' = k_2 + r(2\ell_2 + m + \log^2 n), \varepsilon + \mathsf{negl}(n), r)$-non-malleable extractor $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ in the CRS model.

We instantiate Theorem 8 with the best known explicit statistical two-source extractors in Section 5.2.

5.1 Proof of Theorem 8

Our candidate construction is as follows: First, to define CRS, begin by sampling $b \leftarrow \{0,1\}^\ell$, where $\ell = \ell_1 + \ell_2$, and then sample functions f_{ij} from $\mathcal{F}$ for $i \in [\ell]$ and $j \in \{0,1\}$ such that f_{ib_i} is injective and f_{i1-b_i} is lossy for each i. Finally, sample $h_1 \leftarrow \mathcal{H}_1$ and $h_2 \leftarrow \mathcal{H}_2$, and set

$$\mathsf{CRS} = (h_1, h_2, (f_{ij})_{i\in[\ell], j\in\{0,1\}}) \in \{0,1\}^c.$$

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

$$\mathsf{cnmExt}(x, y, \mathsf{CRS}) = \mathsf{Ext}(f_{h_1(x)\|h_2(y)}(x), y),$$

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

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 8, we must argue that, for arbitrary $(\mathrm{poly}(t_{11}), \mathrm{poly}(t_{12}), k_1', k_2')$-samplable sources (X, Y, AUX), valid size-$\mathrm{poly}(t_1)$ tampering functions $g_1 : \{0,1\}^n \times \{0,1\}^c \to \{0,1\}^n$ and $g_2 : \{0,1\}^n \times \{0,1\}^a \times \{0,1\}^c \to \{0,1\}^n$, and every size-$\mathrm{poly}(t_2)$ distinguisher $\mathcal{A}$ it holds that

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

where $\overline{X} = g_1(X, \mathsf{CRS})$ and $\overline{Y} = g_2(Y, \mathsf{AUX}, \mathsf{CRS})$. 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 $(f_{ib_i})_{i\in[\ell]}$.

Lemma 9. *Let E denote the event that $h_1(X)\|h_2(Y) \neq h_1(\overline{X})\|h_2(\overline{Y})$ and $h_1(X)\|h_2(Y) = b$ hold simultaneously. Then, if*

$$|\Pr[\mathcal{A}(\mathsf{cnmExt}(X,Y,\mathsf{CRS}),\mathsf{cnmExt}(\overline{X},\overline{Y},\mathsf{CRS}),\mathsf{AUX},\mathsf{CRS})=1|E] \\ -\Pr[\mathcal{A}(U_1,\mathsf{cnmExt}(\overline{X},\overline{Y},\mathsf{CRS}),\mathsf{AUX},\mathsf{CRS})=1|E]| \leq \varepsilon + \mathsf{negl}(t_1), \quad (12)$$

it follows that (11) *holds.*

Proof. We proceed similarly to the proof of the analogous claim in [GKK20]. Suppose that (12) holds for every tuple of $(\mathrm{poly}(t_{11}), \mathrm{poly}(t_{12}), k_1', k_2')$-samplable sources (X, Y, AUX), tampering functions g_1 and g_2, and size-$\mathrm{poly}(t_2)$ adversary $\mathcal{A}$, but

$$|\Pr[\mathcal{A}(\mathsf{cnmExt}(X,Y,\mathsf{CRS}),\mathsf{cnmExt}(\overline{X},\overline{Y},\mathsf{CRS}),\mathsf{AUX},\mathsf{CRS})=1] \\ -\Pr[\mathcal{A}(U_1,\mathsf{cnmExt}(\overline{X},\overline{Y},\mathsf{CRS}),\mathsf{AUX},\mathsf{CRS})=1]| > \varepsilon + 1/p(t_1), \quad (13)$$

where $\overline{X} = g_1(X, \mathsf{CRS})$ and $\overline{Y} = g_2(Y, \mathsf{AUX}, \mathsf{CRS})$, for some pair of $(\mathrm{poly}(t_{11}), \mathrm{poly}(t_{12}), k_1', k_2')$-samplable sources (X, Y, AUX), some tampering functions g_1 and g_2, some size-$\mathrm{poly}(t_2)$ 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-$\mathrm{poly}(t_2)$ adversary $\mathcal{B}$ we have

$$2^{-\ell} - \mathsf{negl}(t_2) \leq \Pr[\mathcal{B}(\mathsf{CRS}) = b] \leq 2^{-\ell} + \mathsf{negl}(t_2). \quad (14)$$

Consider the size-$\mathrm{poly}(t_2)$ adversary $\mathcal{B}$ that on input CRS samples (X, Y, AUX), and first checks whether $h_1(X)\|h_2(Y) \neq h_1(\overline{X})\|h_2(\overline{Y})$. If that is the case, then $\mathcal{B}$ outputs $b' = h_1(X)\|h_2(Y)$ as a guess for b, else it outputs $b' \leftarrow \{0,1\}^\ell$. Since $\Pr[h_1(X)\|h_2(Y) = h_1(\overline{X})\|h_2(\overline{Y})] = \mathsf{negl}(t_1)$ by the collision-resistance of $\mathcal{H}$ and the fact that $\overline{X} \neq X$ or $\overline{Y} \neq Y$ by hypothesis, using (14) we have that

$$(1-\mathsf{negl}(t_1))2^{-\ell} - \mathsf{negl}(t_2) \leq \Pr[h(X)\|h(Y) = b, h(X)\|h(Y) \neq h(\overline{X})\|h(\overline{Y})] \\ \leq (1+\mathsf{negl}(t_1))2^{-\ell} + \mathsf{negl}(t_2). \quad (15)$$

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

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

This contradicts (14), which concludes the proof. On input CRS and for $N = p(t_{11} + t_{12})^3$, $\mathcal{B}'$ proceeds as follows:

1. Sample (X, Y, AUX) from CRS. If $h_1(X)\|h_2(Y) = h_1(\overline{X})\|h_2(\overline{Y})$, then re-sample. Otherwise, set $z = h_1(X)\|h_2(Y)$. Note that this takes time $\mathrm{poly}(t_{11} + t_{12})$.

2. For $i \in [N]$: Sample $(X_i, Y_i, \mathsf{AUX}_i)$ from CRS conditioned on $h_1(X_i)\|h_2(Y_i) = z$ and $h_1(X_i)\|h_2(Y_i) \neq h(\overline{X_i})\|h(\overline{Y_i})$. By (15) and the fact that $2^\ell = \mathrm{poly}(t_2)$, this takes time $\mathrm{poly}(t_2)$. Set

$$\delta_i = |\mathcal{A}(\mathsf{cnmExt}(X_i,Y_i,\mathsf{CRS}),\mathsf{cnmExt}(\overline{X_i},\overline{Y_i},\mathsf{CRS}),\mathsf{AUX}_i,\mathsf{CRS}) \\ -\mathcal{A}(U_m,\mathsf{cnmExt}(\overline{X_i},\overline{Y_i},\mathsf{CRS}),\mathsf{AUX}_i,\mathsf{CRS})|,$$

   where $\overline{X_i} = g_1(X_i, \mathsf{CRS})$ and $\overline{Y_i} = g_2(Y_i, \mathsf{AUX}_i, \mathsf{CRS})$. Note that $\mathcal{A}$ has size $\mathrm{poly}(t_2)$.

3. Compute $\delta = \frac{1}{N}\sum_{i=1}^N \delta_i$. If $\delta < \varepsilon + \frac{1}{4p(t_{11}+t_{12})}$, then output $b' = z$. Else, output $b' \leftarrow \{0,1\}^{2\ell}$.

We now show that $\Pr[b' = b] \geq 1.5 \cdot 2^{-\ell}$. 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 (13) and (15) we have $\mathbb{E}[\delta | z \neq b] \geq \varepsilon + \frac{1}{2p(t_1)}$. By the Chernoff bound and the choice of $N = p(t_{11} + t_{12})^3$, 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 $\Pr[b' = b | z \neq b] \geq (1 - \mathsf{negl}(t_1))2^{-\ell}$. Combining these observations with (15) yields

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

which contradicts (14), as desired. $\square$

Based on Lemma 9, we can now work under the assumption that the event E holds and show (12). According to the definition of cnmExt, in order to prove that (12) holds it is now enough to show that for every size-poly(t_2) distinguisher $\mathcal{A}'$ we have

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

where

$$\mathsf{SideInfo} = (h_1(X), h_1(\overline{X}), h_2(Y), h_2(\overline{Y}), f_{h_1(\overline{X}) \| h_2(\overline{Y})}(\overline{X}), \mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{CRS})).$$

With this in mind, consider an arbitrary fixing of the side information $\mathsf{CRS} = \mathsf{crs}$, $h_1(X) \| h_2(Y) = b = b_1 \| b_2$, $h_1(\overline{X}) \| h_2(\overline{Y}) = b' = b'_1 \| b'_2$ with $b' \neq b$, $f_{b'}(\overline{X}) = z'$, and $\mathsf{cnmExt}(\overline{X}, \overline{Y}, \mathsf{crs}) = y'$. Observe that, under such a fixing, the event E holds and X and Y are independent. This is because, after fixing CRS, $h_1(\overline{X}) \| h_2(\overline{Y})$, and $f_{b'}(\overline{X})$, we have that $\mathsf{cnmExt}(\overline{X}, \overline{Y}) = \mathsf{Ext}(f_{b'}(\overline{X}), \overline{Y})$ is a deterministic function of $\overline{Y}$. Therefore, it is now enough to show that

$$|\Pr[\mathcal{A}'(\mathsf{Ext}(f_b(X), Y), b, b', z', y', \mathsf{AUX}, \mathsf{crs}) = 1] - \Pr[\mathcal{A}'(U_m, b, b', z', y', \mathsf{AUX}, \mathsf{crs}) = 1]| \leq \varepsilon \quad (17)$$

for arbitrary $\mathcal{A}'$ with probability $1 - \mathsf{negl}(t_1)$ over the choice of fixings above.

Note that, by Lemmas 1 and 2, with probability at least $1 - \mathsf{negl}(t_1)$ over the fixings we have

$$\begin{aligned} \mathbf{H}_\infty(f_b(X) | \mathsf{CRS} = \mathsf{crs}, h_1(X) = b_1, h_1(\overline{X}) = b'_1, f_{b'}(\overline{X}) = z') &\geq k'_1 - 2\ell_1 - n^\gamma - \log^2 n \\ &\geq k_1, \end{aligned}$$

since f_b is injective, $|h_1(X)| = |h_1(\overline{X})| = \ell_1$, $n = \Omega(t_1)$, and $f_{b'}(\overline{X})$ takes on at most 2^{n^γ} values because $b' \neq b$ (and so at least one index in $f_{b'}$ corresponds to a lossy function). Moreover, we also have

$$\begin{aligned} \mathbf{H}_\infty(Y | \mathsf{CRS} = \mathsf{crs}, h_2(Y) = b_2, h_2(\overline{Y}) = b'_2, \mathsf{Ext}(z', \overline{Y}) = y') &\geq k'_2 - 2\ell_2 - m - \log^2 n \\ &\geq k_2, \end{aligned}$$

since $|h(Y)| = |h(\overline{Y})| = \ell_2$, $|\mathsf{Ext}(z', \overline{Y})| = m$, and $n = \Omega(t_1)$. The desired inequality in (17) now follows immediately by noting that Ext is a strong (k_1, k_2, ε)-extractor and that, after fixing CRS, AUX is a (possibly randomized) function of Y only.

5.2 Instantiations of Theorem 8

In this section, we instantiate Theorem 8 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 $h : \{0,1\}^n \to \{0,1\}^\ell$, where $\ell = \log \lambda \cdot \log\log \lambda$. Then, we set $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$ and $t_{11} = t_{12} = t_1$ in Theorem 8.
- A family of (t_2, n, ω)-lossy functions $\mathcal{F}$, where $t_2 \geq 2^{2\ell} = t_1^{\omega(1)}$ 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.*

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 $\delta, r > 0$ there exists an explicit $(\mathrm{poly}(t_1), \mathrm{poly}(t_1), \mathrm{poly}(t_2), k'_1 = (1/2 + \delta)n, k'_2 = \log^3 n, \varepsilon = \mathsf{negl}(t_1), r)$*-non-malleable extractor* $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *in the CRS model.*

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 9. *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 9, 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 9. 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, Y, AUX) 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

$$\Delta(\mathsf{nmExt}(H(X), H(Y)); U_1 | H, \mathsf{nmExt}(H(g_1(X, H)), H(g_2(Y, \mathsf{AUX}, H)))) = \mathsf{negl}(n). \tag{18}$$

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

$$\Pr[h(X) = h(g_1(X, h))] = \mathsf{negl}(n) \tag{19}$$

or

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

since either $g_1(\cdot, h)$ has no fixed points for any aux or $g_2(\cdot, \mathsf{aux}, h)$ has no fixed points. We now assume that $g_1(\cdot, h)$ has no fixed points, in which case (19) holds. The proof for the case where $g_2(\cdot, \mathsf{aux}, h)$ has no fixed points for any aux is analogous. Additionally, by Lemma 7 coupled with Lemma 2, with probability $1 - \mathsf{negl}(n)$ over the fixing $H = h$ we also have

$$h(X), h(Y) \approx_{\mathsf{negl}(n)} V, W, \tag{20}$$

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

$$\Delta(\mathsf{nmExt}(h(X), h(Y)); U_1 | \mathsf{nmExt}(h(X'), h(Y')), \mathsf{AUX}) = \mathsf{negl}(n), \tag{21}$$

where $X' = g_1(X, h) \neq X$ and $Y' = g_2(Y, \mathsf{AUX}, h)$. We can see $h(X')$ and $(h(Y'), \mathsf{AUX})$ 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

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

where $\Pr[A(h(X)) = h(X)] = \mathsf{negl}(n)$. Therefore, using (20), in order to prove (21) it is enough to show that

$$\Delta(\mathsf{nmExt}(V, W); U_1 | \mathsf{nmExt}(A(V), B(W)), C(W)) = \mathsf{negl}(n). \tag{22}$$

By (20) 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 (22) holds. The last statement of Theorem 9 follows by an analogous proof with a non-strong nmExt. $\square$

Using the non-malleable extractor from Proposition 5 in the statement of Theorem 9, 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 $m = \mathrm{polylog}(n)$, then we obtain a low-error two-source non-malleable extractor for $\mathrm{polylog}(n)$ 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 9 allows us to obtain a practical low-error two-source non-malleable extractor for sources with polylogarithmic min-entropy.

7 Privacy amplification resilient against memory-tampering active adversaries

In the following sections, we formalize the notion of privacy amplification resilient against memory-tampering active adversaries, and show that two-source non-malleable extractors are natural tools for designing such privacy amplification protocols in both the information-theoretic and computational settings.

7.1 The information-theoretic setting

We begin by formally defining with we mean by a privacy amplification protocol against memory-tampering active adversaries.

Definition 14 (Protocol against memory-tampering active adversaries)**.** *An $(r, \ell_1, k_1, \ell_2, k_2, m)$-*protocol against memory-tampering active adversaries *is a protocol between Alice and Bob, with a man-in-the-middle Eve, that proceeds in r rounds. Initially, we assume that Alice and Bob have access to random variables (W, A) and (W, B), respectively, where W is an (ℓ_1, k_1)-source (the* secret*), and A, B are (ℓ_2, k_2)-sources (the* randomness tapes*) independent of each other and of W. The protocol proceeds as follows:*

In the first stage, Eve submits an arbitrary function $F : \{0,1\}^{\ell_1} \times \{0,1\}^{\ell_2} \to \{0,1\}^{\ell_1} \times \{0,1\}^{\ell_2}$ and chooses one of Alice and Bob to be corrupted, so that either (W, A) is replaced by $F(W, A)$ (if Alice is chosen), or (W, B) is replaced by $F(W, B)$ (if Bob is chosen).

In the second stage, Alice and Bob exchange messages $(C_1, C_2, \dots, C_r)$ over a non-authenticated channel, with Alice sending the odd-numbered messages and Bob the even-numbered messages, and Eve is allowed to replace each message C_i by C'_i based on $(C_1, C'_1, \dots, C_{i-1}, C'_{i-1}, C_i)$ and independent random coins, so that the recipient of the i-th message observes C'_i. Messages C_i sent by Alice are deterministic functions of (W, A) and $(C'_2, C'_4, \dots, C'_{i-1})$, and messages C_i sent by Bob are deterministic functions of (W, B) and $(C'_1, C'_3, \dots, C'_{i-1})$.

In the third stage, Alice outputs $S_A \in \{0,1\}^m \cup \{\perp\}$ as a deterministic function of (W, A) and $(C'_2, C'_4, \dots)$, and Bob outputs $S_B \in \{0,1\}^m \cup \{\perp\}$ as a deterministic function of (W, B) and $(C'_2, C'_4, \dots)$.

Definition 15 (Privacy amplification protocol against memory-tampering active adversaries)**.** *An $(r, \ell_1, k_1, \ell_2, k_2, m, \varepsilon, \delta)$-privacy amplification protocol against memory-tampering active adversaries is an $(r, \ell_1, k_1, \ell_2, k_2, m)$-protocol against memory-tampering active adversaries with the following additional properties:*

- **If Eve is passive:** *In this case, F is the identity function and Eve only wiretaps. Then, $S_A = S_B \neq \perp$ with S_A satisfying*
$$S_A, C \approx_\varepsilon U_m, C, \tag{23}$$
where $C = (C_1, C'_1, C_2, C'_2, \dots, C_r, C'_r)$ denotes Eve's view.

- **If Eve is active:** *Then, with probability at least $1 - \delta$ either $S_A = \perp$ or $S_B = \perp$ (i.e., one of Alice and Bob detects tampering), or $S_A = S_B \neq \perp$ with S_A satisfying* (23)*.*

We are now ready to state our main result in the information-theoretic setting, which states that every strong two-source non-malleable extractor with appropriate parameters yields a 4-round privacy amplification protocol resilient against memory-tampering active adversaries via the protocol illustrated in Figure 1.

Theorem 10. *Let $\mathsf{nmExt} : \{0,1\}^{\ell_1} \times \{0,1\}^{2\ell_2} \to \{0,1\}^{m+2\alpha}$ be a strong $(k_1 - \ell_2 - 2\gamma - 1, k_2 - \gamma - 1, \varepsilon)$-two-source non-malleable extractor. Then, there exists an $(r = 4, \ell_1, k_1, \ell_2, k_2, m, \varepsilon, \delta = \varepsilon + 2^{-\alpha} + 2 \cdot 2^{-\gamma})$-privacy amplification protocol against memory-tampering active adversaries. Moreover, the protocol is explicit if nmExt is explicit.*

Proof. We consider the 4-round protocol from Figure 1. Without loss of generality, we may assume that Eve is deterministic. We proceed by cases:

1. **Eve is passive:** Then, we have $R_A = R_B$ (and hence $S_A = S_B \neq \perp$), and the desired result follows by noting that
$$R_A = \mathsf{nmExt}(W, A\|B), A\|B \approx_\varepsilon U_{m+2\alpha}, A\|B,$$
since W is independent of $A\|B$, $\mathbf{H}_\infty(W) \geq k_1$, $\mathbf{H}_\infty(A\|B) \geq 2k_2$, and nmExt is a strong extractor. This implies that $S_A, C \approx_\varepsilon U_m, C$, where $C = (A, B, [R_A]_{2\alpha})$ denotes Eve's view.

2. **Eve is active and Alice is corrupted:** Denote $(\widetilde{W}, \widetilde{A}) = F(W, A)$, and consider arbitrary fixings $A = a$ and $\widetilde{A} = \widetilde{a}$. Note that $\widetilde{W}$ is now a deterministic tampering of W, and that, by Lemmas 1 and 2 and the fact that $|\widetilde{A}| = \ell_2$, it holds that
$$\begin{aligned}\mathbf{H}_\infty(W|A = a, \widetilde{A} = \widetilde{a}) &\geq \widetilde{\mathbf{H}}_\infty(W|A, \widetilde{A}) - \gamma \\ &\geq \widetilde{\mathbf{H}}_\infty(W|A) - \ell_2 - \gamma\end{aligned}$$

$$= k_1 - \ell_2 - \gamma \tag{24}$$

with probability at least $1-2^{-\gamma}$ over the fixings. We assume that the fixings above satisfy (24), and simply add a $2^{-\gamma}$ term to δ via a union bound. We now have

$$R_A = \mathsf{nmExt}(\widetilde{W}, \widetilde{a}\|B')$$

and

$$R_B = \mathsf{nmExt}(W, \widetilde{a}'\|B),$$

where $\widetilde{a}'$ is a deterministic function of $\widetilde{a}$ (hence it is fixed), and B' is a deterministic function of B since A and $\widetilde{A}$ are fixed. As a result, W and $\widetilde{a}'\|B$ are independent, and we can write $\widetilde{W} = f(W)$ and $\widetilde{a}\|B' = g(\widetilde{a}'\|B)$ for deterministic tampering functions f and g. Let $\mathcal{L} = \{w : f(w) = w\}$ and $\mathcal{R} = \{b : g(\widetilde{a}'\|b) = \widetilde{a}'\|b\}$. We now argue differently depending on whether $W \in \mathcal{L}$ and $B \in \mathcal{R}$ hold or not. We begin by noting that if either $\Pr[W \in \mathcal{L} \wedge B \in \mathcal{R}] < 2^{-\gamma}$ or $\Pr[W \notin \mathcal{L} \vee B \notin \mathcal{R}] < 2^{-\gamma}$, then we can add a $2^{-\gamma}$ term to δ via a union bound and assume that the opposite event holds. We are thus reduced to the two cases below:

(a) If $\Pr[W \in \mathcal{L} \wedge B \in \mathcal{R}] = \Pr[W \in \mathcal{L}] \cdot \Pr[B \in \mathcal{R}] \geq 2^{-\gamma}$, by Lemma 3 it holds that

$$\mathbf{H}_\infty(W | A = a, \widetilde{A} = \widetilde{a}, W \in \mathcal{L}) \geq k_1 - \ell_2 - 2\gamma \tag{25}$$

and

$$\mathbf{H}_\infty(B | A = a, \widetilde{A} = \widetilde{a}, B \in \mathcal{R}) \geq k_2 - \gamma. \tag{26}$$

Therefore, since under this conditioning we still have that W and $\widetilde{a}'\|B$ are independent and they both have enough min-entropy by (25) and (26), it is the case that $R_A = R_B$ and

$$R_A = \mathsf{nmExt}(W, \widetilde{a}\|B), \widetilde{a}\|B \approx_\varepsilon U_{m+2\alpha}, \widetilde{a}\|B.$$

If $[R_A]'_\alpha = [R_A]_\alpha$ and $[R_B]'_{\alpha:2\alpha} = [R_B]_{\alpha:2\alpha}$, then $S_A = S_B \neq \perp$ and $S_A, C \approx_\varepsilon U_m, C$, where $C = (\widetilde{a}, \wr a', B, B', [R_A]_{2\alpha}, [R_A]'_{2\alpha})$ denotes Eve's view. Otherwise, we have either $S_A = \perp$ or $S_B = \perp$ with probability 1.

(b) On the other hand, if $\Pr[W \notin \mathcal{L} \vee B \notin \mathcal{R}] \geq 2^{-\gamma}$, by Lemma 3 it either holds that

$$\mathbf{H}_\infty(W | A = a, \widetilde{A} = \widetilde{a}, W \notin \mathcal{L}) \geq k_1 - \ell_2 - 2\gamma - 1 \tag{27}$$

or

$$\mathbf{H}_\infty(B | A = a, \widetilde{A} = \widetilde{a}, B \notin \mathcal{R}) \geq k_2 - \gamma - 1. \tag{28}$$

Assume that (27) holds and condition on $W \notin \mathcal{L}$. The proof when (28) holds and we condition on $B \notin \mathcal{R}$ is analogous. Then, we have that f has no fixed points over the support of W under this conditioning, and so, by the fact that nmExt is a strong $(k_1-\ell_2-2\gamma-1, k_2-\gamma-1, \varepsilon)$-non-malleable extractor, W and $\widetilde{a}'\|B$ are independent, (27) and that $\mathbf{H}_\infty(\widetilde{a}'\|B) \geq k_2$, it holds that

$$\Delta(R_B = \mathsf{nmExt}(W, \widetilde{a}'\|B); U_{m+2\alpha} | R_A = \mathsf{nmExt}(f(W), g(\widetilde{a}'\|B)), \widetilde{a}'\|B) \leq \varepsilon.$$

This implies that the probability that $[R_A]'_\alpha = [R_B]_\alpha$, and hence $S_B \neq \perp$, is at most $\varepsilon + 2^{-\alpha}$, which we add to δ via a union bound.

3. **Eve is active and Bob is corrupted:** The reasoning follows analogously to the previous case, but we set $(\widetilde{W}, \widetilde{B}) = F(W, B)$ and fix B and $\widetilde{B}$ instead.

□

We present a corollary of Theorem 10 to showcase that strong two-source non-malleable extractors with good parameters yield privacy amplification protocols against memory-tampering active adversaries with likewise good parameters.

Corollary 10. *Suppose that for some constant $\delta > 0$ there exists a strong $(k = \delta n, \varepsilon, r = 1)$-non-malleable extractor* $\mathsf{nmExt} : \{0,1\}^n \to \{0,1\}^n \to \{0,1\}^{m'}$ *with $m' = \Omega(n)$. Then, there exists an $(r = 4, \ell_1 = n, k_1 = 3\delta n, \ell_2 = 1.5\delta n, k_2 = 1.1\delta n, m = \Omega(n), \varepsilon, \delta = \varepsilon + 2^{-\Omega(n)})$-privacy amplification protocol against memory-tampering active adversaries. Moreover, the protocol is explicit if* nmExt *is explicit.*

We currently do not know explicit constructions of two-source non-malleable extractors with parameters matching those required for Corollary 10, although it is known that there exist such (inefficient) extractors with significantly better parameters [CGGL20]. Thus, we leave this connection between two-source non-malleable extractors and privacy amplification with a very strong adversary as an interesting motivation for further study of such extractors with lower min-entropy requirement in the information-theoretic setting. In the next section, we show that we can construct *explicit* privacy amplification protocols against memory-tampering active adversaries in the computational setting from computational strong two-source non-malleable extractors.

7.2 An efficient privacy amplification protocol in the CRS model

In this section, we focus on designing *efficient* privacy amplification protocols resilient against memory-tampering active adversaries in the CRS model. Informally, we consider an analogous setting to Section 7.1, but assuming that there is a public common reference string CRS, the sources W, A, and B are efficiently samplable given the CRS, and that the memory-tampering active adversary Eve is computationally bounded. A formal definition follows below.

Definition 16 (Protocol against memory-tampering active adversaries in the CRS model)**.** *An* $(r, \lambda, \ell_1, k_1, \ell_2, k_2, m)$-protocol against memory-tampering active adversaries in the CRS model *is a protocol between Alice and Bob, with a man-in-the-middle Eve, that proceeds in r rounds. Initially, we assume that Alice and Bob have access to random variables (W, A) and (W, B), respectively, where $W \in \{0,1\}^{\ell_1}$, $A, B \in \{0,1\}^{\ell_2}$ are independent, and W is $(\mathrm{poly}(\lambda), k_1)$-samplable from* CRS*, and A and B are both $(\mathrm{poly}(\lambda), k_2)$-samplable from* CRS*. The protocol proceeds as follows:*

In the first stage, Eve submits a size-$\mathrm{poly}(\lambda)$ circuit $F : \{0,1\}^{\ell_1} \times \{0,1\}^{\ell_2} \times \{0,1\}^c \to \{0,1\}^{\ell_1} \times \{0,1\}^{\ell_2}$ and chooses one of Alice and Bob to be corrupted, so that either (W, A) is replaced by $F(W, A, \mathsf{CRS})$ (if Alice is chosen), or (W, B) is replaced by $F(W, B, \mathsf{CRS})$ (if Bob is chosen).

In the second stage, Alice and Bob exchange messages $(C_1, C_2, \ldots, C_r)$ over a non-authenticated channel, with Alice sending the odd-numbered messages and Bob the even-numbered messages, and Eve is allowed to replace each message C_i by $C'_i \leftarrow \mathcal{A}(C_1, C'_1, \ldots, C_{i-1}, C'_{i-1}, C_i, \mathsf{CRS})$, where $\mathcal{A}$ is a size-$\mathrm{poly}(\lambda)$ circuit, so that the recipient of the i-th message observes C'_i. Messages C_i sent by Alice are deterministic functions of (W, A), CRS, and $(C'_2, C'_4, \ldots, C'_{i-1})$, and messages C_i sent by Bob are deterministic functions of (W, B), CRS, and $(C'_1, C'_3, \ldots, C'_{i-1})$.

In the third stage, Alice outputs $S_A \in \{0,1\}^m \cup \{\perp\}$ as a deterministic function of (W, A), CRS, and $(C'_2, C'_4, \ldots)$, and Bob outputs $S_B \in \{0,1\}^m \cup \{\perp\}$ as a deterministic function of (W, B), CRS, and $(C'_2, C'_4, \ldots)$.

Definition 17 (Privacy amplification protocol against memory-tampering active adversaries in the CRS model)**.** *An* $(r, \lambda, \ell_1, k_1, \ell_2, k_2, m)$-privacy amplification protocol against memory-tampering active adversaries in the CRS model *is an* $(r, \lambda, \ell_1, k_1, \ell_2, k_2, m)$*-protocol against memory-tampering active adversaries in the CRS model with the following additional properties:*

- **Eve is passive:** *In this case, F is the identity function and Eve only wiretaps. Then, $S_A = S_B \neq \perp$ with S_A satisfying*

$$S_A, C, \mathsf{CRS} \approx^c_\lambda U_m, C, \mathsf{CRS}, \tag{29}$$

where $C = (C_1, C_1', C_2, C_2', \ldots, C_r, C_r')$ denotes Eve's view, U_m is independent of C and CRS*, and $\approx^c_\lambda$ denotes computational indistinguishability for all distinguishers running in time* $\text{poly}(\lambda)$.

- **Eve is active:** *Then, with probability at least* $1 - \mathsf{negl}(\lambda)$ *either $S_A = \perp$ or $S_B = \perp$ (i.e., one of Alice and Bob detects tampering), or $S_A = S_B \neq \perp$ with S_A satisfying* (29).

We construct an efficient privacy amplification protocol against memory-tampering active adversaries in the CRS model under a subexponential hardness assumption by combining the protocol illustrated in Figure 1 with a careful instantiation of the two-source non-malleable extractor in the CRS model from Section 5. Combining Theorem 8 with Raz's extractor (Proposition 2), we have the following explicit two-source non-malleable extractor in the CRS model, which allows the left source to be sampled in subexponential time.

Corollary 11. *Assuming the subexponential hardness of the DDH assumption, there exist constants $0 < \gamma, \eta, c < 1$ such that for any $0 < \nu < \eta < 1$ and n large enough there exists an explicit* $(\text{poly}(2^{n^\nu}), \text{poly}(n), \text{poly}(n), \text{poly}(2^{n^\eta}), k_1 = 0.51n, k_2 = m + \log^3 n, \varepsilon = \mathsf{negl}(n), r = 1)$*-non-malleable extractor in the CRS model* $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^c \to \{0,1\}^m$ *for any $m \leq cn$ and $\ell \leq n$.*

Proof. We set the parameters

$$t_{11} = 2^{n^\nu}, t_{12} = n, t_2 = 2^{n^\eta},$$

where $0 < \nu < \eta < 1$. Therefore, we also have $t_1 = \min(t_{11}, t_{12}) = n$. The subexponential hardness of the DDH assumption ensures that there exist constants $0 < \eta, \gamma < 1$ such that for all $0 < \nu < \eta$ the following primitives exist:

- A family $\mathcal{H}_1$ of $(\text{poly}(t_{11}), \mathsf{negl}(t_{11}))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{\ell_1}$, where $\ell_1 = \lambda^\nu \cdot \log \lambda$;

- A family $\mathcal{H}_2$ of $(\text{poly}(t_{12}), \mathsf{negl}(t_{12}))$-collision-resistant hash functions $h : \{0,1\}^n \to \{0,1\}^{\ell_2}$, where $\ell_1 = \log \lambda \cdot \log \log \lambda$;

- A family of (t_2, n, ω)-lossy functions $\mathcal{F}$ with $\omega = n - n^\gamma$. Note that $t_2 \geq 2^{\ell_1 + \ell_2}$ and $2^{\ell_1} = t_{11}^{\omega(1)}$, $2^{\ell_2} = t_{12}^{\omega(1)}$ by the choice of parameters above.

- The explicit strong $(k_1 = 0.501n, k_2 = O(\log n), \varepsilon = 2^{-\Omega(n)})$-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ for any $m \leq cn$ from Proposition 2, for some constant $c > 0$.

Invoking Theorem 8 with the primitives above, we conclude that there exists an explicit $(\text{poly}(2^{n^\nu}), \text{poly}(n), \text{poly}(n), \text{poly}(2^{n^\eta}), k_1', k_2', \mathsf{negl}(n), r = 1)$-non-malleable extractor $\mathsf{cnmExt}' : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^c \to \{0,1\}^m$, where $k_1' = k_1 + 2\ell_1 + 2n^\gamma \leq 0.51n$ and $k_2' = k_2 + 2\ell_2 + m + \log^2 n \leq m + \log^3 n$, provided n is large enough. The desired statement now follows by considering, for $\ell \leq n$, the extractor $\mathsf{cnmExt} : \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^c \to \{0,1\}^m$ defined as

$$\mathsf{cnmExt}(X, Y, \mathsf{CRS}) = \mathsf{cnmExt}'(X, Y \| 0^{n-\ell}, \mathsf{CRS}),$$

and observing that $Y \| 0^{n-\ell}$ is samplable by a size-poly(n) circuit if Y is too. $\square$

We can use the extractor from Corollary 11 to design an explicit 4-round privacy amplification protocol resilient against efficient memory-tampering active adversaries in the CRS model.

Theorem 11. *Assuming the subexponential hardness of the DDH assumption, for a small enough constant $\nu > 0$ and every $\log^2 \lambda \leq m \leq \lambda^\nu$, $m + 2\log^2 \lambda + 2\log^3 \lambda + 1 \leq k_2 \leq \ell_2 \leq \lambda^\nu$ there exists an explicit $(r = 4, \lambda, \ell_1 = \lambda, k_1 = 0.52\lambda, \ell_2, k_2, m, \varepsilon = \mathsf{negl}(\lambda), \delta = \mathsf{negl}(\lambda))$-privacy amplification protocol against memory-tampering active adversaries in the CRS model.*

Remark 2. *In fact, we may even assume that the distinguisher for the privacy amplification of Theorem 11 is allowed to run in time 2^{λ^ν}.*

Proof of Theorem 11. We consider the 4-round protocol from Figure 1 with the explicit two-source non-malleable extractor in the CRS model $\mathsf{cnmExt} : \{0,1\}^\lambda \times \{0,1\}^{2\ell_2} \times \{0,1\}^c \to \{0,1\}^{m+2\alpha}$ from Corollary 11 with $\alpha = \log^2 \lambda$. The proof is similar to the proof of Theorem 10, but we present it for completeness. Without loss of generality, we may assume that Eve is deterministic. We proceed by cases:

1. **Eve is passive:** Then, we have $R_A = R_B$ (and hence $S_A = S_B \neq \perp$), and the desired result follows by noting that

$$R_A = \mathsf{cnmExt}(W, A\|B, \mathsf{CRS}), A\|B, \mathsf{CRS} \approx^c_\lambda U_{m+2\alpha}, A\|B, \mathsf{CRS},$$

   since, for every fixing $\mathsf{CRS} = \mathsf{crs}$ W is independent of $A\|B$, W is $(\text{poly}(\lambda), k_1)$-samplable from CRS, and $(A\|B, \mathsf{AUX} = A\|B)$ is $(\text{poly}(\lambda), 2k_2)$-samplable from CRS. This implies that $S_A, C \approx^c_\lambda U_m, C$, where $C = (A, B, [R_A]_{2\alpha})$ denotes Eve's view.

2. **Eve is active and Alice is corrupted:** Denote $(\widetilde{W}, \widetilde{A}) = F(W, A)$, and consider arbitrary fixings $A = a$ and $\widetilde{A} = \widetilde{a}$. Note that $\widetilde{W}$ is now obtained from W by a deterministic size-poly(λ) circuit, $(W|A = a, \widetilde{A} = \widetilde{a})$ is samplable from CRS to within statistical error $\mathsf{negl}(\lambda)$ by a circuit of size $\text{poly}(\lambda, 2^{\ell_2}) = \text{poly}(2^{\lambda^\nu})$ with probability at least $1 - \mathsf{negl}(\lambda)$ over the fixings, and that, by Lemmas 1 and 2 and the fact that $|\widetilde{A}| = \ell_2$, it holds that

$$\begin{aligned}\mathbf{H}_\infty(W|A = a, \widetilde{A} = \widetilde{a}) &\geq \widetilde{\mathbf{H}}_\infty(W|A, \widetilde{A}) - \log^2 \lambda \\ &\geq \widetilde{\mathbf{H}}_\infty(W|A) - \ell_2 - \log^2 \lambda \\ &\geq 0.51\lambda \end{aligned} \tag{30}$$

   with probability at least $1 - \mathsf{negl}(\lambda)$ over the fixings. We assume that the fixings above satisfy (30) and the property that $(W|A = a, \widetilde{A} = \widetilde{a})$ is samplable to within statistical error $\mathsf{negl}(n)$ by a circuit of size $\text{poly}(2^{\lambda^\nu})$, and simply add a $\mathsf{negl}(\lambda)$ term to the final error δ via a union bound. We now have

$$R_A = \mathsf{cnmExt}(\widetilde{W}, \widetilde{a}\|B', \mathsf{CRS})$$

   and

$$R_B = \mathsf{cnmExt}(W, \widetilde{a}'\|B, \mathsf{CRS}),$$

where $\widetilde{a}'$ is a deterministic function of $\widetilde{a}$ (hence it is fixed), and B' is a deterministic function of B since A and $\widetilde{A}$ are fixed. As a result, W and $\widetilde{a}'\|B$ are independent, and we can write $\widetilde{W} = f(W)$ and $\widetilde{a}\|B' = g(\widetilde{a}'\|B)$ for size-poly(λ) deterministic circuits f and g. Let $\mathcal{L} = \{w : f(w) = w\}$ and $\mathcal{R} = \{b : g(\widetilde{a}'\|b) = \widetilde{a}'\|b\}$. We now argue differently depending on whether $W \in L$ and $B \in R$ hold or not. We begin by noting that if either $\Pr[W \in \mathcal{L} \wedge B \in \mathcal{R}] = \mathsf{negl}(\lambda)$ or $\Pr[W \notin \mathcal{L} \vee B \notin \mathcal{R}] = \mathsf{negl}(\lambda)$, then we can add a $\mathsf{negl}(\lambda)$ term to δ via a union bound and assume that the opposite event holds. We are thus reduced to the two cases below:

(a) If $\Pr[W \in \mathcal{L} \wedge B \in \mathcal{R}] = \Pr[W \in \mathcal{L}] \cdot \Pr[B \in \mathcal{R}] \geq \frac{1}{\text{poly}(\lambda)}$, by Lemma 3 it holds that

$$\mathbf{H}_\infty(W|A = a, \widetilde{A} = \widetilde{a}, W \in \mathcal{L}, \mathsf{CRS} = \mathsf{crs}) \geq k_1 - \ell_2 - \log^2 \lambda \geq 0.51\lambda \tag{31}$$

and

$$\mathbf{H}_\infty(B|A = a, \widetilde{A} = \widetilde{a}, B \in \mathcal{R}, \mathsf{CRS} = \mathsf{crs}) \geq k_2 - \log^2 \lambda \geq m + \log^3 \lambda. \tag{32}$$

Moreover, by the hypothesis above we have that $(W|A = a, \widetilde{A} = \widetilde{a}, W \in \mathcal{L})$ is samplable to within statistical error $\mathsf{negl}(\lambda)$ by a size-poly(2^{λ^ν}) circuit from crs, and that $(B|A = a, \widetilde{A} = \widetilde{a}, B \in \mathcal{R})$ is samplable to within statistical error $\mathsf{negl}(\lambda)$ by a size-poly(λ) circuit from crs. Therefore, since under this conditioning we still have that W and $\widetilde{a}'\|B$ are independent and they both have enough min-entropy by (31) and (32), it is the case that $R_A = R_B$ and

$$R_A = \mathsf{cnmExt}(W, \widetilde{a}\|B, \mathsf{crs}), \widetilde{a}\|B \approx^c_\lambda U_{m+2\alpha}, \widetilde{a}\|B.$$

If $[R_A]'_\alpha = [R_A]_\alpha$ and $[R_B]'_{\alpha:2\alpha} = [R_B]_{\alpha:2\alpha}$, then $S_A = S_B \neq \perp$ and $S_A, C, \mathsf{CRS} \approx^c_\lambda U_m, C, \mathsf{CRS}$, where $C = (\widetilde{a}, \widetilde{a}', B, B', [R_A]_{2\alpha}, [R_A]'_{2\alpha})$ denotes Eve's view. Otherwise, we have either $S_A = \perp$ or $S_B = \perp$ with probability 1.

(b) On the other hand, if $\Pr[W \notin \mathcal{L} \vee B \notin \mathcal{R}] \geq \frac{1}{\text{poly}(\lambda)}$, it either holds that $\Pr[W \notin \mathcal{L}] \geq \frac{1}{\text{poly}(\lambda)}$ or $\Pr[B \notin \mathcal{R}] \geq \frac{1}{\text{poly}(\lambda)}$. Therefore, by Lemma 3 it either holds that

$$\mathbf{H}_\infty(W|A = a, \widetilde{A} = \widetilde{a}, W \notin \mathcal{L}, \mathsf{CRS} = \mathsf{crs}) \geq k_1 - 2\ell_2 - \log^2 \lambda - 1 \geq 0.51\lambda \tag{33}$$

or

$$\mathbf{H}_\infty(B|A = a, \widetilde{A} = \widetilde{a}, B \notin \mathcal{R}, \mathsf{CRS} = \mathsf{crs}) \geq k_2 - \log^2 \lambda - 1 \geq m + \log^3 \lambda. \tag{34}$$

Assume that $\Pr[W \notin \mathcal{L}] \geq \frac{1}{\text{poly}(\lambda)}$ and condition on $W \notin \mathcal{L}$. The proof when $\Pr[B \notin \mathcal{R}] \geq \frac{1}{\text{poly}(\lambda)}$ and we condition on $B \notin \mathcal{R}$ is analogous. Then, we have that f has no fixed points over the support of W under this conditioning and that, by the hypothesis above, $(W|A = a, \widetilde{A} = \widetilde{a}, W \notin \mathcal{L})$ is samplable by a size-poly(2^{λ^ν}) circuit from crs. Moreover, $(B|A = a, \widetilde{A} = \widetilde{a})$ is still samplable by a size-poly(λ) circuit from crs and B still has min-entropy at least $m + \log^3 \lambda$ after these fixings. Therefore, we conclude that

$$\begin{aligned} R_B = \mathsf{cnmExt}(W, \widetilde{a}'\|B, \mathsf{crs}), R_A = \mathsf{cnmExt}(f(W), g(\widetilde{a}'\|B), \mathsf{crs}), \widetilde{a}'\|B \\ \approx^c_\varepsilon U_{m+2\alpha}, R_A = \mathsf{cnmExt}(f(W), g(\widetilde{a}'\|B), \mathsf{crs}), \widetilde{a}'\|B \end{aligned}$$

By the choice of $\alpha = \log^2 \lambda$, this implies that the probability that $[R_A]'_\alpha = [R_B]_\alpha$, and hence $S_B \neq \perp$, is at most $\mathsf{negl}(\lambda)$, which we add to δ via a union bound.

3. **Eve is active and Bob is corrupted:** The reasoning follows analogously to the previous case, but we set $(\widetilde{W}, \widetilde{B}) = F(W, B)$ and fix B and $\widetilde{B}$ instead.

□

References

[ACLV19] Divesh Aggarwal, Kai-Min Chung, Han-Hsuan Lin, and Thomas Vidick. A quantum-proof non-malleable extractor - with application to privacy amplification against active quantum adversaries. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part II*, volume 11477 of *Lecture Notes in Computer Science*, pages 442–469. Springer, 2019.

[ACM+14] Per Austrin, Kai-Min Chung, Mohammad Mahmoody, Rafael Pass, and Karn Seth. On the impossibility of cryptography with tamperable randomness. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology – CRYPTO 2014*, pages 462–479, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[ADKO15] Divesh Aggarwal, Yevgeniy Dodis, Tomasz Kazana, and Maciej Obremski. Non-malleable reductions and applications. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 459–468. ACM, 2015.

[ADL14] Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 774–783. ACM, 2014.

[ADN+19] Divesh Aggarwal, Ivan Damgård, Jesper Buus Nielsen, Maciej Obremski, Erick Purwanto, João L. Ribeiro, and Mark Simkin. Stronger leakage-resilient and non-malleable secret sharing schemes for general access structures. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part II*, volume 11693 of *Lecture Notes in Computer Science*, pages 510–539. Springer, 2019.

[AO20] Divesh Aggarwal and Maciej Obremski. A constant rate non-malleable code in the split state model. *FOCS*, 2020.

[BACD+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 Rocco A. Servedio, editor, *33rd Computational Complexity Conference (CCC 2018)*, volume 102 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:19, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[BBCM95] Charles H. Bennett, Gilles Brassard, Claude Crépeau, and Ueli M. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, 1995.

[BBR88] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM J. Comput.*, 17(2):210–229, April 1988.

[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.

[BDK+05] Xavier Boyen, Yevgeniy Dodis, Jonathan Katz, Rafail Ostrovsky, and Adam Smith. Secure remote authentication using biometric data. In Ronald Cramer, editor, *Advances in Cryptology – EUROCRYPT 2005*, pages 147–163, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[BDT16] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Low-error two-source extractors for polynomial min-entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 23(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 1185–1194, 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.

[BS19] Saikrishna Badrinarayanan and Akshayaram Srinivasan. Revisiting non-malleable secret sharing. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part I*, volume 11476 of *Lecture Notes in Computer Science*, pages 593–622. Springer, 2019.

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

[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.

[CGGL20] Eshan Chattopadhyay, Jesse Goodman, Vipul Goyal, and Xin Li. Extractors for adversarial sources via extremal hypergraphs. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, page 1184–1197, New York, NY, USA, 2020. ACM.

[CGL16] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 285–298, New York, NY, USA, 2016. ACM.

[Coh17] Gil Cohen. Towards optimal two-source extractors and Ramsey graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1157–1170, New York, NY, USA, 2017. ACM.

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

[DGKM12] Dana Dachman-Soled, Rosario Gennaro, Hugo Krawczyk, and Tal Malkin. Computational extractors and pseudorandomness. In Ronald Cramer, editor, *Theory of Cryptography*, pages 383–403, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[DHP+18] Pierre-Alain Dupont, Julia Hesse, David Pointcheval, Leonid Reyzin, and Sophia Yakoubov. Fuzzy password-authenticated key exchange. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2018*, pages 393–424, Cham, 2018. Springer International Publishing.

[DKK+12] Yevgenyi Dodis, Bhavana Kanukurthi, Jonathan Katz, Leonid Reyzin, and Adam Smith. Robust fuzzy extractors and authenticated key agreement from close secrets. *IEEE Transactions on Information Theory*, 58(9):6207–6222, 2012.

[DKO13] Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, volume 8043 of *Lecture Notes in Computer Science*, pages 239–257. Springer, 2013.

[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 41st Annual ACM Symposium on Theory of Computing*, pages 601–610, New York, NY, USA, 2009. ACM.

[EHOR20] Andreas Erwig, Julia Hesse, Maximilian Orlt, and Siavash Riahi. Fuzzy asymmetric password-authenticated key exchange. Cryptology ePrint Archive, Report 2020/987, 2020.

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

[GK18b] Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing for general access structures. In Hovav Shacham and Alexandra Boldyreva, editors, *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part I*, volume 10991 of *Lecture Notes in Computer Science*, pages 501–530. Springer, 2018.

[GKK20] Ankit Garg, Yael Tauman Kalai, and Dakshita Khurana. Low error efficient computational extractors in the CRS model. In Anne Canteaut and Yuval Ishai, editors, *Advances in Cryptology – EUROCRYPT 2020*, pages 373–402, Cham, 2020. Springer International Publishing.

[GLM+04] Rosario Gennaro, Anna Lysyanskaya, Tal Malkin, Silvio Micali, and Tal Rabin. Algorithmic tamper-proof (ATP) security: Theoretical foundations for security against hardware tampering. In Moni Naor, editor, *Theory of Cryptography*, pages 258–277, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[GSZ20] Vipul Goyal, Akshayaram Srinivasan, and Chenzhi Zhu. Multi-source non-malleable extractors and applications. Cryptology ePrint Archive, Report 2020/157, 2020.

[IPSW06] Yuval Ishai, Manoj Prabhakaran, Amit Sahai, and David Wagner. Private circuits II: Keeping secrets in tamperable circuits. In Serge Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, pages 308–327, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[KKS11] Yael Tauman Kalai, Bhavana Kanukurthi, and Amit Sahai. Cryptography with tamperable and leaky memory. In Phillip Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, pages 373–390, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[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):950–957, 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.

[Li12] Xin Li. Non-malleable extractors, two-source extractors and privacy amplification. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 688–697, 2012.

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science*, 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, New York, NY, USA, 2017. ACM.

[Li19] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. In Amir Shpilka, editor, *34th Computational Complexity Conference (CCC 2019)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:49, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[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 37th Annual ACM Symposium on Theory of Computing*, pages 11–20, New York, NY, USA, 2005. ACM.

[TV00] Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *Proceedings of the 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 43rd Annual ACM Symposium on Theory of Computing*, 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 38th Annual ACM Symposium on Theory of Computing*, pages 681–690, New York, NY, USA, 2006. ACM.

Work	Min-Entropy Requirements	Sources	Non-Malleability	Error	Distinguisher	Assumptions
[KLR09]	$k \geq \delta n$ for any constant $\delta > 0$	General (without assumptions on sampling efficiency)	No	Negligible	Computationally bounded	Nearly optimal exponentially secure one-way permutations
[GKK20]	$k_1 = \Omega(n^c)$ and $k_2 = \text{poly}(\log n)$	Efficiently samplable	Polynomial number of tamperings, but only of one source	Negligible	Computationally bounded	Quasi-polynomial hardness of DDH, and CRS.
This work	$k \geq 0.46n$. Or, $k_1 \geq (1/2 + \delta)n$ and $k_2 = \text{poly}(\log n)$	Efficiently samplable	Multiple tamperings	Negligible	Computationally bounded	Quasi-polynomial hardness of DDH, and CRS.
This work	$k = \Omega(n^c)$	Efficiently samplable	Multiple tamperings	$n^{-\alpha}$	Computationally bounded	Quasi-polynomial hardness of DDH, and CRS.
This work	$k = \text{poly}(\log n)$	Efficiently samplable	Multiple tamperings	Negligible	Unbounded	Nearly optimal collision-resistant hash functions

Table 1: Comparison with previous work.