Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 52 (2016)

Making the Most of Advice: New Correlation Breakers and Their Applications *

Gil Cohen†

July 4, 2021

Abstract

A typical obstacle one faces when constructing pseudorandom objects is undesired correlations between random variables. Identifying this obstacle and constructing certain types of "correlation breakers" was central for recent exciting advances in the construction of multi-source and non-malleable extractors. One instantiation of correlation breakers is *correlation breakers with advice*. These are algorithms that break the correlation a "bad" random variable Y' has with a "good" random variable Y using an "advice" – a fixed string α that is associated with Y which is guaranteed to be distinct from the corresponding string α' associated with Y'. Prior to this work, explicit constructions of correlation breakers with advice require the entropy of the involved random variables to depend linearly on the advice length.

In this work, building on *independence-preserving mergers*, a pseudorandom primitive that was recently introduced by Cohen and Schulman, we devise a new construction of correlation breakers with advice that has optimal, logarithmic, dependence on the advice length. This enables us to obtain the following results.

- We construct an extractor for 5 independent n-bit sources with min-entropy $(\log n)^{1+o(1)}$. This result puts us tantalizingly close to the goal of constructing extractors for 2 sources with min-entropy $O(\log n)$, which would have exciting implications to Ramsey theory.
- We construct non-malleable extractors with error guarantee ε for n-bit sources, with seed length $d = O(\log n) + (\log(1/\varepsilon))^{1+o(1)}$ for any min-entropy $k = \Omega(d)$. Prior to this work, all constructions require either very high min-entropy or

*This paper subsumes the technical report Non-Malleable Extractors with Logarithmic Seeds [Coh16].

†Computing and Mathematical Sciences Department, Caltech. Supported by a Walter S. Baer and Jeri Weiss CMI Postdoctoral Fellowship. Email: coheng@caltech.edu.

ISSN 1433-8092

otherwise have seed length $\omega(\log n)$ for any ε. Further, our extractor has near-optimal output length. Prior constructions that achieve comparable output length work only for very high min-entropy $k \approx n/2$.

- By instantiating the Dodis-Wichs framework with our non-malleable extractor, we obtain near-optimal privacy amplification protocols against active adversaries, improving upon all (incomparable) known protocols.

Contents

1 Introduction

When constructing pseudorandom objects, such as various types of extractors, mergers, condensers, and so forth, one often faces undesired correlations between random variables. At some point in the construction and its analysis, a pair of random variables $X_{\mathsf{good}}, X_{\mathsf{bad}}$ is obtained. Although one can show that X_{good} is uniform (or, more generally, that X_{good} is "well behaved"), X_{bad} may correlate arbitrarily with X_{good}, preventing one from proceeding with the construction and analysis. In some cases, working around the undesired correlation between X_{good} and X_{bad} can be done by exploiting more information about the nature of the correlation [GRS06, Li11b, CS15]. More typically, one is careful enough to avoid the presence of correlations to begin with, though such a cautious strategy is sometimes costly and may rule out what could have been a natural and direct construction.

Although a recurring theme, the problem of efficiently breaking arbitrary correlations a random variable has with a uniformly distributed random variable, using (unavoidably) an auxiliary source of randomness, was first explicitly studied by [Coh15a] in the form of an object called a *local correlation breaker*. The construction of the latter is based on the alternating extraction technique [DP07, DW09], and is influenced by [Li13b]. [1] By adapting the construction of local correlation breakers, Chattopadhyay *et al.* [CGL15] gave a construction for a different type of correlation breakers, which we call a *correlation breaker with advice*. This primitive is the main component, both conceptually and in terms of technical effort, in existing constructions of non-malleable extractors [CGL15, Coh15b, Coh16]. Although only recently introduced, correlation breakers with advice already found further applications [CS16].

We turn to give the formal definition of correlation breakers with advice. We assume familiarity with standard notions from the literature such as statistical distance, min-entropy, weak-sources, and seeded extractors. The unfamiliar reader may consult the Preliminaries (Section 5).

Definition 1.1 (Correlation breakers with advice)**.** *A function*

$$\mathsf{AdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

is called a (k,ε)-correlation breaker with advice *if the following holds. Let* Y *be a random variable that is uniformly distribution over* ℓ*-bit strings, and let* Y' *be an* ℓ*-bit random variable that may be arbitrarily correlated with* Y*. Let* X *be an* (n,k)*-source that is arbitrarily correlated with an* n*-bit random variable* X'*. Assume that the joint distribution of* X, X' *is independent of the joint distribution of* Y, Y'*. Then, for any pair of distinct*

[1] In his pioneer work on multi-source extractors [Li13b], Li developed a technique for breaking correlations between a pair of random variables assuming *both* of which are uniform.

a-bit strings α, α',

$$(\mathsf{AdvCB}(X, Y, \alpha), \mathsf{AdvCB}(X', Y', \alpha')) \approx_\varepsilon (U, \mathsf{AdvCB}(X', Y', \alpha')).$$

Although every effort was made to keep Definition 1.1 to its most succinct form, the definition is still somewhat involved. Thus, we proceed by providing some informal remarks that are meant to clarify the definition. We think of Y in Definition 1.1 as being the good random variable. By "good" we mean that Y is uniformly distributed. The role of the "bad" random variable is formalized by Y' that, according to the definition, is allowed to correlate with Y in an arbitrary manner. The third random variable X is a weak-source of randomness that, as it turns out, is required for the purpose of breaking the arbitrary correlation Y' may have with Y. We think of X as an auxiliary, or external, source of randomness as it is independent of the joint distribution of Y, Y'. Note that one does not have to use the same source X when applying AdvCB to Y' with α'. In fact, X' can be arbitrarily correlated with X, as long as their joint distribution is independent of the joint distribution of Y, Y'. For the sake of simplicity, in the remaining of this section we put less emphasis on the output length m.

We think of the fixed a-bit string α as the advice that is given to the correlation breaker, with the guarantee that the a-bit string α' that is associated with Y' is different than α. Such an advice, of course, is unavoidable (think of $Y = Y'$, $X = X'$). We remark that in some cases, the variables Y and Y' are explicitly computed by our algorithm and in such case a, a' can be taken to be some labeling of these variables. In other cases, one consider Y' only in the analysis, in which case a, a' are sometimes computed, or generated, by applying some function f to Y and (in the analysis) to Y', respectively.

The quality of a correlation breaker with advice is determined by the min-entropy k that it requires from its auxiliary weak-source of randomness X, and by the length ℓ of Y (which can be thought of as the entropy of Y). Thus, given n, a, and a desired error guarantee $\varepsilon > 0$, the goal is to construct (k, ε)-correlation breakers with advice with k, ℓ as small as possible.

A straightforward probabilistic argument can be used to show that for all integers n, a, and for any $\varepsilon > 0$, there exists a (k, ε)-correlation breaker with advice for

$$\begin{aligned} k &= 2\log(1/\varepsilon) + O(1), \\ \ell &= \log a + \log(n-k) + 2\log(1/\varepsilon) + O(1). \end{aligned}$$

By adapting the construction of local correlation breakers [Coh15a], Chattopadhyay *et al.* [CGL15] gave an explicit construction of a (k, ε)-correlation breaker with advice with

$$k, \ell = O\left(a \cdot \log\left(\frac{an}{\varepsilon}\right)\right).$$

Note that both k, ℓ grow linearly with the advice length a (in fact, the dependence on a is super-linear). Moreover, a is multiplied by (rather than added to) $\log(n/\varepsilon)$, which turns out to be the bottleneck for applications to non-malleable extractors.

2 Our Contribution

The main technical contribution of this work is an explicit construction of a correlation breaker with advice that has *logarithmic* and *additive* dependence on the advice length, significantly improving upon known results.

Theorem 2.1 (Main technical result)**.** *For all integers n, a, m, for any $\varepsilon > 0$, and for any constant $\alpha > 0$ such that $a < 2^{n/\varepsilon}$ [2], there exists an explicit (k, ε)-correlation breaker with advice, with*

$$\ell = O\left(\log a + \log n\right) + \left(\log(1/\varepsilon)\right)^{1+o(1)},$$
$$k = (2+\alpha)m + O(\ell).$$

Note that the requirement from ℓ in our construction is optimal (up to constant factors) but for the slight sub-optimal dependence on ε. Note further that the dependencies on each parameter n, a, ε add rather than multiply. Moreover, our correlation breaker supports a very low min-entropy k. Building on Theorem 2.1, we obtain the following results:

5-source extractors for near-logarithmic entropy. We construct an extractor for 5 independent n-bit sources with min-entropy $(\log n)^{1+o(1)}$. This result puts us tantalizingly close to the goal of constructing extractors for 2 independent sources with min-entropy $O(\log n)$, which would have exciting implications to Ramsey theory. See Theorem 2.2.

Near-optimal non-malleable extractors. We construct non-malleable extractors with error guarantee ε for n-bit sources, having seed length $d = O(\log n) + (\log(1/\varepsilon))^{1+o(1)}$ for any min-entropy $k = \Omega(d)$. Prior to our work, all known constructions require either very high min-entropy or otherwise have seed length $\omega(\log n)$ regardless of the error guarantee. Furthermore, our extractor can be set to have output length $k/(2+\alpha)$ for any constant $\alpha > 0$, which is very close to the optimal $k/2$ bound.

[2] This assumption is made mainly for the sake of presentation. We stress that this restriction is (more than) flexible enough so as to capture the advice length of current applications. Furthermore, one can easily relax this bound further to, say, double or triple-exponential in n/ε. We remark that in some cases, n is not proportional to the input size of the problem, and so one may want to consider such advice length.

Prior to this work, constructions that achieve output length $\Omega(k)$ work only for very high min-entropy $k \approx n/2$. See Theorem 2.4.

Near-optimal privacy amplification protocols. By instantiating the Dodis-Wichs framework [DW09] with our non-malleable extractor, we obtain near-optimal privacy amplification protocols in the active setting. In particular, the entropy-loss of the induced two-round protocol is $O(\log n) + \lambda^{1+o(1)}$ for security parameter λ. See Theorem 2.5.

In the following sections we elaborate on our results, put them in context, and compare with known results.

2.1 A 5-source extractor for near-logarithmic min-entropy

For an integer $s \geq 1$, a (k, ε) s-source extractor [CG88, BIW06] is a function $\mathsf{Ext}\colon (\{0,1\}^n)^s \to \{0,1\}^m$ with the following property. For any s independent (n,k)-sources $X_1, \ldots, X_s$, the random variable $\mathsf{Ext}(X_1, \ldots, X_s)$ is ε-close to uniform. In this paper we focus on constant error guarantee ε, and $m = 1$ output bits. It is easy to see that a one-source extractor does not exist even for min-entropy as high as $k = n - 1$. On the other hand, there exists a two-source extractor for min-entropy as low as $k = \log n + O(1)$ [CG88]

Besides being a natural problem, finding explicit two-source extractors for logarithmic min-entropy would resolve a classical problem in combinatorics, namely, matching Erdős proof for the existence of Ramsey graphs [Erd47] with a constructive proof. In fact, it suffices to construct a two-source disperser for the same min-entropy, where a disperser is a weakening of an extractor in which the output is only required to be non-constant, as apposed to being close to uniform.

Recall that an undirected graph on n vertices is called k-*Ramsey* if it contains no clique or independent set of size k. Ramsey [Ram28] proved that there does not exist a $0.5 \log n$-Ramsey graph on n vertices. This result was later complemented by Erdős [Erd47], who proved that most graphs on n vertices are $(2 + o(1)) \log n$-Ramsey.

In a long line of research [CG88, BIW06, Bou05, Raz05, Rao09, BRSW12, BSZ11, Li11a, Li13b, Li13a] that has accumulated to [Li15b] (see Table **??** in Appendix **??**), Li constructed a three-source extractor for min-entropy $\operatorname{polylog} n$. Based on this extractor and on the challenge-response mechanism [BKS+05, BRSW12], the first two-source disperser for min-entropy $\operatorname{polylog} n$ was constructed in [Coh15c]. Subsequently, Chattopadhyay and Zuckerman [CZ15] (followed by some improvements [Li15a, Mek15]) constructed a two-source extractor for min-entropy $\operatorname{polylog} n$.

Although exciting, these results are still polynomially far from optimal. This fairly modest gap is far more significant when considering the implications to Ramsey theory.

Indeed, the constructions of [Coh15c, CZ15] induce explicit k-Ramsey graph on n vertices with $k = 2^{\text{poly}\log\log n}$, as apposed to the desired $k = O(\log n)$. The most natural goal today is to obtain k-Ramsey graphs on n vertices with $k = \text{polylog}\, n$. Such graphs correspond to two-source dispersers that support min-entropy $O(\log n)$.

Aiming towards this goal, Cohen and Schulman [CS16] observed that previous techniques for constructing s-source extractors and dispersers break below min-entropy $(\log n)^2$ for any constant s. Based on a new primitive they introduced, called independence-preserving mergers, a construction of an extractor for $O(s)$-sources with min-entropy $(\log n)^{1+1/s}$ was obtained [CS16], breaking the "$\log^2 n$ barrier", and paving the way towards constructing extractors for near-logarithmic min-entropy. Continuing this research path, based on Theorem 1.1 and on revising the framework for constructing multi-source extractor set by [CS16], we obtain a 5-source extractor for near-logarithmic min-entropy.

Theorem 2.2. *For all n there exists an explicit extractor* $\mathsf{5Ext}\colon (\{0,1\}^n)^5 \to \{0,1\}$ *for min-entropy* $(\log n)^{1+o(1)}$.

Building on the framework set by [CS16], Chattopadhyay and Li [CL16] obtained, independently of our work and using different ideas, an extractor for s-sources, each with min-entropy $(\log n)^{1+o(1)}$, where s is some universal constant. The number of sources s, although constant, is quite large – it is proportional to the inverse of the constant $\beta > 0$ for which Bourgain's two-source extractor [Bou05] can support min-entropy rate $1/2 - \beta$. [3]

2.2 Non-malleable extractors

As mentioned, correlation breakers with advice were introduced in the context of non-malleable extractors [CGL15]. As we improve upon previous constructions of correlation breakers with advice, we readily obtain improved constructions of non-malleable extractors. In this section we recall the definition of non-malleable extractors, give an account for explicit constructions of non-malleable extractors from the literature, and state our result.

A non-malleable extractor is a seeded extractor with a very strong guarantee concerning the correlations (or, more precisely, the lack thereof) of the outputs of the extractor when fed with different seeds. The notion of a non-malleable extractor was introduced by Dodis and Wichs [DW09], motivated by the problem of devising privacy amplification protocols against active adversaries (see Section 2.3). More recently, non-malleable extractors played a key role in the construction of two-source extractors [CZ15].

[3] To the best of our knowledge, taking into account recent points-lines incidence theorems over finite fields [Jon12, RNRS14], together with Bourgain's application of this result [Bou05, Rao07], and the way Bourgain's extractor is applied by [CS16], $s \geq 1000$.

Definition 2.3 (Non-malleable extractors [DW09])**.** *A function* $\mathsf{nmExt}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is called a* (k,ε)-non-malleable extractor *if for any* (n,k)*-source* X *and any function* $\mathcal{A}\colon \{0,1\}^d \to \{0,1\}^d$ *with no fixed points, it holds that*

$$(\mathsf{nmExt}(X,Y), \mathsf{nmExt}(X,\mathcal{A}(Y)), Y) \approx_\varepsilon (U, \mathsf{nmExt}(X,\mathcal{A}(Y)), Y),$$

where Y *is uniformly distributed over* $\{0,1\}^d$ *independently of* X*.*

Computational aspects aside, for any integer n and $\varepsilon > 0$, Dodis and Wichs [DW09] proved the existence of (k,ε)-non-malleable extractors having m output bits and seed length $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$ for any $k > 2m + 2\log(1/\varepsilon) + \log d + O(1)$. Although the mere existence of non-malleable extractors, and with such great parameters, is somewhat surprising, [4] explicit constructions are far more desirable.

Constructing non-malleable extractors gained a significant attention in the literature (see Table **??** in Appendix **??**). However, up until this work, all constructions require very high min-entropy or otherwise have seed length $\omega(\log n)$ regardless of the error guarantee. More precisely, to support min-entropy $\operatorname{polylog} n$ (or even min-entropy $n^{0.99}$), the seed length of all prior constructions is super logarithmic in n and, at best, has dependence of the order of $\log^2(1/\varepsilon)$ on the error guarantee ε. [5]

Building on Theorem 2.1 we obtained the following result.

Theorem 2.4. *For any integer* n*, any* $\varepsilon > 0$*, and any constant* $\alpha > 0$*, there exists an efficiently-computable* (k,ε)*-non-malleable extractor* $\mathsf{nmExt}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k/(2+\alpha)}$ *with seed length* $d = O(\log n) + (\log(1/\varepsilon))^{1+o(1)}$ *for any* $k = \Omega(d)$*.*

Using different techniques, and independently of our work, Chattopadhyay and Li [CL16] obtained a non-malleable extractor with seed length $(\log(n/\varepsilon))^{1+o(1)}$ and output length $k/2^{\sqrt{\log\log(n/\varepsilon)}}$. Note that the seed length of [CL16] is super-logarithmic in n even for constant ε (and, in fact, the dependence on n is not as good as previously known results [Coh15b]). Further, the number of output bits of their construction is $o(k)$, whereas we obtain output length which can be taken arbitrarily close to the optimal $k/2$ bound.

The precise dependence of the seed length on the error guarantee, as well as the dependence obtained by [CL16], is $\log(1/\varepsilon) \cdot c^{\sqrt{\log\log(1/\varepsilon)}}$ for some constant $c > 1$. Interestingly, this similar dependence is due to different reasons. In fact, it seems that by combining ideas from both works, one can slightly improve the result and obtain a construction with

[4] In fact, atypically, the proof is non-trivial. We observe that an alternative existential proof follows by the straightforward existential proof for correlation breakers with advice combined with the [CGL15] framework and the "switch" idea of [Coh15b].

[5] This work subsumes a technical report by the author [Coh16] in which a non-malleable extractor with seed length $O(\log n + \log^3(1/\varepsilon))$ is obtained.

seed length $O(\log n) + \log(1/\varepsilon) \cdot c^{(\log\log(1/\varepsilon))^{1/3}}$. [6] Although an insignificant quantitative improvement, it does suggest that different ideas are used in both works. Indeed, the strategy taken by [CL16] is to construct a variant of independence-preserving mergers, and in particular, Chattopadhyay and Li do not attempt to obtain improved correlation breakers with advice.

2.3 Privacy amplification protocols

In the classical problem of privacy amplification [BBR85, Mau93, BBCM95] two parties, Alice and Bob, share a secret that is "somewhat random" from the point of view of an adversary Eve. Formally, the secret is modeled by an (n, k)-source X. In the classical setting, Alice and Bob can communicate over an authenticated channel that is eavesdropped by Eve. Put differently, Eve is a passive adversary, and cannot tamper with the communication. Throughout the paper we consider only the information-theoretic setting in which Eve is computationally unbounded. Further, we assume that both Alice and Bob have local (that is, non-shared) randomness that is independent of X.

The goal of Alice and Bob is to agree on an m-bit string R that is ε-close to uniform even conditioned on the transcript of the protocol, that is visible to Eve. We refer to $\lambda = \log(1/\varepsilon)$ as the security parameter of the protocol. The quality of a privacy amplification protocol is measured by the following parameters:

Round complexity. The number of rounds required by the protocol.

Entropy-loss. The amount of min-entropy that is lost during the protocol, namely, $k - m$.

Communication complexity. The total number of bits that are communicated.

Supported min-entropy. The least value k for which the protocol is secure.

A strong seeded extractor yields a one-round privacy amplification protocol: Given a (k, ε)-strong seeded extractor $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, Alice samples $s \sim U_d$ and sends s to Bob. Alice and Bob then compute $R = \mathsf{Ext}(X, s)$. As Ext is strong, R is ε-close to uniform even conditioned on s, the transcript of the protocol.

Note that the entropy-loss of the protocol equals the entropy-loss of the extractor. The communication complexity is the seed length d, and the supported min-entropy of the protocol is that supported by Ext. By instantiating the protocol with explicit near-optimal strong seeded extractors [GUV09], for any $k = \Omega(\lambda)$, one obtains an explicit

[6] This assertion was not verified as carefully as the proofs in this paper, and should be trusted accordingly.

one-round protocol with entropy-loss $O(\lambda)$ and communication complexity $O(\log n + (\lambda + \log k) \cdot \log k)$.

2.3.1 Privacy amplification protocols against an active adversary

A significantly more challenging problem is to devise privacy amplification protocols when Eve has full control over the communication channel, and can therefore tamper with the communication to her liking. That is, when the channel is unauthenticated. We allow ourselves to be somewhat informal regarding the exact requirement from a protocol in this setting and refer the reader to, say, [DW09] for a formal treatment. We only emphasize the difficulty that a protocol for this model has to overcome: not only R must be close to uniform from Eve's point of view, but also Alice and Bob must agree on the *same* string R, if possible, and otherwise (and only if necessary) declare that the communication has been tampered with.

The problem of devising privacy amplification protocols in the active setting was first studied by Maurer and Wolf [MW97] who constructed a one-round protocol for $k > 2n/3$. Subsequently, Dodis *et al.* [DKRS06] relaxed the bound to $k > n/2$. The entropy-loss and communication complexity of these protocols is $n - k$, which is significantly larger than what was obtained in the passive setting. Unfortunately, a one-round protocol cannot support min-entropy $k < n/2$ [DW09], and so unlike in the passive adversary model, to avoid high entropy-loss and to save on communication, in the active adversary setting, one must resort to multiple round protocols, even for large k. Following [MW97], a long line of research studied the problem of devising privacy amplification protocols against active adversaries (see [MW97, DKRS06, RW03, DW09, KR09, CKOR14, Li12a, Li12b] and references therein), though until the work of Dodis and Wichs [DW09], all protocols required more than two rounds.

Dodis and Wichs [DW09] suggested an elegant framework for constructing two-round privacy amplification protocols in the active setting. Much like the framework for the passive setting, the Dodis-Wichs framework is also instantiated with an extractor, and the parameters of the protocol are determined by those of the extractor. However, to handle active adversaries, the extractor used by the Dodis-Wichs framework is required to be non-malleable.

To be more precise, the Dodis-Wichs framework also requires a strong seeded extractor, though we "hardwire" a known construction of almost optimal strong seeded extractors [GUV09]. Further, the protocol can in fact be instantiated with a weaker object than a full-blown non-malleable extractor, called a look-ahead extractor, though then one has to use other more sophisticated primitives (a special type of message authenticated codes) which results in a protocol with weaker parameters.

The Dodis-Wichs framework motivated the study of non-malleable extractors. Al-

though Dodis and Wichs proved the existence of such extractors, an explicit construction was not obtained in [DW09] and, as covered in Section 2.2, a significant attention was given for matching the existential result with a constructive proof. Further, Li [Li12a, Li12b] devised a privacy amplification protocol for the active setting which only requires a *non-malleable condenser* – a weaker object than a non-malleable extractor. Li was able to construct such condensers and thus, due to the lack of good enough non-malleable extractors at the time, obtained improved privacy amplification protocols.

We do not present the Dodis-Wichs protocol in this paper, and are content with relating the parameters of the protocol with that of the non-malleable extractor that is being used. The entropy-loss is $O(d)$, where d is the seed length of the non-malleable extractor applied to n-bit strings and set with error guarantee ε. The communication complexity of the protocol is $O(d + (\lambda + \log k) \cdot \log k)$, and the supported min-entropy is that supported by the extractor.

Prior to this work, the best explicit non-malleable extractor [Coh15b] (which has better parameters than known non-malleable condensers [Li12a, Li12b]) requires a seed of length $\Omega(\log(n/\varepsilon) \cdot \log(\log(n)/\varepsilon))$ and thus induces a protocol with entropy-loss $\Omega(\lambda^2 + \lambda \log n + \log n \cdot \log \log n)$. By instantiating the Dodis-Wichs framework with the non-malleable extractor that is given by Theorem 2.4, we obtain the following near-optimal protocol that, in particular, has entropy-loss of the order of $\lambda^{1+o(1)} + \log n$.

Theorem 2.5. *For all n, λ, there exists an explicit two-round privacy amplification protocol against active adversaries that supports min-entropy $k = \Omega(d)$, with entropy-loss $O(d)$, and communication complexity $O(d + (\lambda + \log k) \cdot \log k)$, where $d = \lambda^{1+o(1)} + O(\log n)$.*

Based on their non-malleable extractor [CL16], Chattopadhyay and Li obtained privacy amplification protocols with higher entropy-loss and communication complexity as a function of n.

3 Proof Outline

Our construction of correlation breakers with advice that is given by Theorem 2.1 heavily relies on the notion of *independence-preserving mergers* – a pseudorandom primitive that was introduced recently by [CS16] for the construction of multi-source extractors. We also make use of the notion of *hierarchy of independence*. To outline our proof, we must first present these two concepts.

3.1 Independence-preserving mergers

Informally speaking, an independence-preserving merger is a function that mergers a sequence of random variables to a single random variable while preserving some form of

independence that sequence has with a second sequence of random variables. To be more precise, we make use of the notion of *somewhere-independent matrices* [CS16].

We say that a sequence of random variables $X_1, \ldots, X_r$ is somewhere-independent of the sequence $Y_1, \ldots, Y_r$ if the following two conditions are met:

- There exists $g \in [r]$ such that X_g is close to uniform even conditioned on Y_g;
- For all $i \in [r]$, the random variable X_i is close to uniform.

We typically consider random variables on, say, ℓ-bit strings, and stack the random variables in the sequence $X_1, \ldots, X_r$ (resp. $Y_1, \ldots, Y_r$) as the rows of an $r \times \ell$ matrix X (resp. Y). We say that X is somewhere-independent of Y.

An independence-preserving merger is a function of the form

$$\mathsf{IPMerg}\colon \{0,1\}^{r\times\ell} \times \{0,1\}^n \to \{0,1\}^\ell,$$

that has the following property: if X is somewhere-independent of Y then IPMerg applied to X is close to uniform even conditioned on the corresponding application to Y. Note that IPMerg has two arguments. The first is the matrix whose rows we want to merger. The second argument is fed with a sample from an auxiliary weak-source of randomness that is required for the purpose of the merging process. The actual construction that we use (see Theorem 6.4) requires a sample from a second weak-source that is allowed to correlate with the matrix. This is done for technical reasons that we prefer to avoid delving into in this section.

In [CS16], a construction of independence-preserving mergers was given that requires the row length ℓ, as well as the min-entropy of the auxiliary source of randomness, to be of order $r \cdot \log(n/\varepsilon)$, where ε measures the statistical closeness of the output of IPMerg applied to X from the uniform distribution conditioned on the corresponding output applied to Y (see Theorem 6.4). We make an extensive black-box use of this construction.

3.2 Hierarchy of independence

The notion of hierarchy of independence is captured by a pair of functions

$$\mathsf{a}(y,x)\colon \{0,1\}^\ell \times \{0,1\}^n \to \{0,1\}^\ell,$$
$$\mathsf{b}(y,x)\colon \{0,1\}^\ell \times \{0,1\}^n \to \{0,1\}^\ell,$$

that has the following property. Let Y, Y' be arbitrarily correlated ℓ-bit random variables such that Y is uniform. Let X, X' be arbitrarily correlated n-bit random variables such that X has sufficiently high min-entropy. Assume further that the joint distribution (X, X') is independent of the joint distribution (Y, Y'). Then, the following holds:

- $\mathsf{a}(Y, X)$ is close to uniform, and
- $\mathsf{b}(Y, X)$ is close to uniform even conditioned on $\mathsf{a}(Y, X), \mathsf{a}(Y', X')$.

That is, the variable $\mathsf{b}(Y, X)$, which we think of as being in the higher level of the hierarchy, is uniform even conditioned on the random variables $\mathsf{a}(Y, X), \mathsf{a}(Y', X')$ in the lower level of the hierarchy. Thus, in this hierarchic-sense, the pair (a, b) allows one to break correlations between random variables. Based on the alternating extraction technique, one can efficiently construct such a pair of functions. In fact, for technical reasons, the function b requires one more argument. We refer the reader to Section 6.2 for more details.

3.3 The general strategy and context

With independence-preserving mergers and the notion of hierarchy of independence in hand, we are ready to outline the proof of Theorem 2.1. Our construction can be divided to three modular steps. We give a short description for each of these steps, and elaborate further in Section 4.

Step 1 – Constructing a base correlation breaker. First, we construct a correlation breaker with advice to which we refer to as the *base correlation breaker with advice*. More precisely, we construct a (k, ε)-correlation breaker with advice

$$\mathsf{BaseAdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

with $\ell = O\left(\log n + a \cdot \log\left(a/\varepsilon\right)\right)$ and $k = 3m + O(\ell)$.

Note that $\mathsf{BaseAdvCB}$ already modestly improves upon the existing construction as the advice length a is added to $\log n$, rather than being multiplied by $\log n$, [7] though it is not the reason we bother constructing a new correlation breaker with advice. We do so mainly for the sake of completeness. Indeed, the existing construction of correlation breakers with advice is only implicit in [CGL15]. The explicit definition was coined only subsequently in [Coh15b], referring to [CGL15] for a proof. Having the definition and formal statement in one source and the proof, implicitly, in a second source is far from ideal. Moreover, one needs to be careful when adopting the proof of [CGL15] due to the dependence on the error guarantee. We give a completely different construction which we believe to be simpler and more direct given independence-preserving mergers.

An informal description of the construction of the base correlation breaker with advice is given in Section 4.1.

[7] By applying a more careful analysis, one can show that the construction of the base correlation breaker with advice only requires $\ell = O\left(\log n + a + \frac{a}{\log a} \cdot \log(1/\varepsilon)\right)$. However, we do not benefit from this given the stepping-up algorithm that we present in Step 2.

Step 2 – Stepping-up correlation breakers with advice. The base correlation breaker with advice that we construct in the Step 1 requires min-entropy that is linear in the advice length a. In the second step, we design an efficient algorithm that transforms, in a black-box manner, one correlation breaker with advice $\mathsf{AdvCB}_{\mathsf{in}}$ to another $\mathsf{AdvCB}_{\mathsf{out}}$, with better dependence on the advice length (as long as the dependence is not "too good" to begin with). By applying this transformation repeatedly, each time with the previously generated correlation breaker, we obtain a correlation breaker with advice that only requires min-entropy $2^{O(\sqrt{\log a})} \cdot \log(n/\varepsilon)$. The stepping-up algorithm, as well as the construction of the base correlation breaker, makes use of independence-preserving mergers. For more details, see Section 4.2.

Step 3 – Condensing the advice. Although the correlation breaker with advice that is obtained in Step 2 already significantly improves upon the known construction, it still has a super logarithmic, and multiplicative, dependence on the advice length a. Unfortunately, we do not know how to improve the dependence on the advice length, at least not without having access to a better independence-preserving merger. Instead, we take a completely different approach – we make the advice shorter!

More precisely, in the third step we devise an efficient algorithm that is given as input an advice of length a, and outputs a new advice of length $\log(1/\varepsilon) + \tau(a)$. Here, ε is a bound on the probability that the new (allegedly) advice will fail to remain a valid advice – namely, be distinct from the value obtained by applying the same procedure to any different a-bit string. The function $\tau(a)$ is an extremely slowly growing function of a. In particular, if a is bounded above by, say, $2^{n/\varepsilon}$ or even by an expression which is double or triple exponential in n/ε, the affect $\tau(a)$ has is negligible.

Putting it all together. The seed length and min-entropy required for the advice condenser is $O(\log(na/\varepsilon))$. Thus, by condensing the advice prior to the application of the correlation breaker with advice that is obtained in Step 2, one only requires ℓ, k of order

$$\log(an) + 2^{\sqrt{\log(\tau(a)+\log(1/\varepsilon))}} \cdot \log(n/\varepsilon). \tag{3.1}$$

This is not quite what is stated in Theorem 2.1. In particular, note the undesired multiplicative dependence. This dependence can be removed by applying the "switch" idea [Coh15b]. More precisely, before applying the correlation breaker with advice, and after condensing the advice, we apply some transformation to the source and the seed so to obtain a new source and a new seed, both of length $\approx \log(n/\varepsilon)$. Thus, informally speaking, the quantitative affect this transformation has on the seed length is that any appearance of n on the right summand of Equation (3.1) is replaced by $\log(n/\varepsilon)$. Using the fact that $\tau(a)$ is a very slowly growing function of n, a short calculation shows that

the multiplicative dependence is bounded by the additive terms as stated in Theorem 2.1.

4 A More Detailed Proof Outline

In this section we elaborate a bit further on each of the three steps that were presented in the previous section.

4.1 Step 1 – the base correlation breaker

As mentioned, in the first step we construct the base (k, ε)-correlation breaker with advice

$$\mathsf{BaseAdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m,$$

where $\ell = O\left(\log n + a \cdot \log\left(a/\varepsilon\right)\right)$ and $k = 3m + O(\ell)$. In this section we give an informal description of the construction.

On input $x \in \{0,1\}^n$, $y \in \{0,1\}^\ell$, and $\alpha \in \{0,1\}^a$, the first step for computing $\mathsf{BaseAdvCB}(x, y, \alpha)$ is constructing a matrix $m = m(x, y, \alpha)$ with $2a$ rows, as follows. For every $i \in [a]$, if $\alpha_i = 0$ we set

$$\begin{aligned} m_{2i-1} &= a(y, x), \\ m_{2i} &= b(y, x). \end{aligned}$$

Otherwise, if $\alpha_i = 1$, we set

$$\begin{aligned} m_{2i-1} &= b(y, x), \\ m_{2i} &= a(y, x). \end{aligned}$$

Observe that by doing so, one is guaranteed that $M = m(X, Y, \alpha)$ is somewhere-independent of $M' = m(X', Y', \alpha')$ for any $\alpha \neq \alpha'$ and any random variables X, Y, X', Y' as described above. Indeed, by construction, one of the rows of M contains $b(Y, X)$ while the corresponding row of M' contains $a(Y', X')$. By the hierarchy of independence, that row of M is close to uniform even conditioned on the corresponding row of M'. Moreover, note that every row of M is close to uniform, and so indeed M is somewhere-independent of M'.

At this point, one can apply the independence-preserving merger to M so to obtain the output Z. The corresponding application to M' will result in a random variable Z' with the guarantee that Z is close to uniform even conditioned on Z', as desired.

To be more precise, one should use only, say, a prefix of Y for the construction of the matrix M so not to exhaust all the min-entropy of Y, and leaving some for the independence-preserving merger. We consider this issue to be a technicality and prefer not to delve into the details in this section. The precise statement and proof appear in Section 7.

4.2 Step 2 – stepping-up correlation breakers with advice

The base correlation breaker with advice that was constructed in Step 1 requires entropy that is linear in the advice length. In the second step, we devise an efficient algorithm that transforms, in a black-box manner, one correlation breaker with advice $\mathsf{AdvCB}_{\mathsf{in}}$ to another $\mathsf{AdvCB}_{\mathsf{out}}$, with a better dependence of the required entropy on the advice length.

The high-level idea is as follows. Given $x \in \{0,1\}^n$, $y \in \{0,1\}^\ell$, and $\alpha \in \{0,1\}^a$, we partition the a-bit advice string α to b substrings, or blocks, of equal length which we denote by $\alpha_1, \ldots, \alpha_b$. We then apply $\mathsf{AdvCB}_{\mathsf{in}}$ to each of the b blocks and stack all the results in a matrix with b rows. To that matrix we apply the independence-preserving merger so to obtain the final output.

Again, we are being intentionally blur regarding the exact way we use x, y. Indeed, one must leave enough entropy in the random variables after one computation so to meet the requirement of the next. Further, the independence between (X, X') and (Y, Y') must always be preserved.

Let us consider this transformation when applied to the base correlation breaker with advice, with $b = \sqrt{a}$. As the advice passed to $\mathsf{AdvCB}_{\mathsf{in}}$ is of length $a/b = \sqrt{a}$, the required entropy for computing the matrix is roughly of order $\sqrt{a} \cdot \log(n/\varepsilon)$. The obtained matrix has $\sqrt{a}$ rows, and so the independence-preserving merger requires only $\sqrt{a} \cdot \log(n/\varepsilon)$ entropy from its sources. All in all, we have managed to reduce the entropy dependence on the advice length from linear in a to $O(\sqrt{a})$.

Applying the transformation again, now to the newly obtained correlation breaker with advice, this time with $b = a^{1/3}$, one obtains a third correlation breaker with advice that only requires entropy $O(a^{1/3} \cdot \log(n/\varepsilon))$. We continue to apply this sequence of improvements where on the j'th iteration, we set $b = a^{1/(j+1)}$. After v iterations, the required entropy is roughly of the form $2^{O(v)} \cdot a^{1/v} \cdot \log(n/\varepsilon)$. By setting $v = \sqrt{\log a}$, we obtain a correlation breaker with advice that only requires entropy $2^{\sqrt{\log a}} \cdot \log(n/\varepsilon)$. For more details, we refer the reader to Section 8.

4.3 Step 3 – condensing the advice

Somewhat orthogonally to the ideas presented so far, in the third step we show how to shorten, or condense, a given advice, at least as long the advice is longer than $\log(1/\varepsilon)$. More precisely, we devise an algorithm to which we call an *advice condenser*

$$\mathsf{AdvCond}\colon \{0,1\}^{a_{\mathsf{in}}} \times \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{a_{\mathsf{out}}}$$

that has the following property. For any X, X', Y, Y' as above, and for any distinct fixed a_{in}-bit strings α, α', it holds that

$$\mathbf{Pr}\left[\mathsf{AdvCond}(\alpha, X, Y) = \mathsf{AdvCond}(\alpha', X', Y')\right] \leq \varepsilon.$$

Moreover, $a_{\mathsf{out}} = O(\log(1/\varepsilon) + \tau(a_{\mathsf{in}}))$, where $\tau(a_{\mathsf{in}}) = \log^{(c)}(a_{\mathsf{in}})$ is the c-iterated log function applied to a_{in}. [8]

For any constant c, the entropy required from X, Y for the purpose of condensing the advice is only $O(\log(na/\varepsilon))$, which we are willing to pay. Hence, informally speaking, by having our advice condenser, one can always assume that the advice has length of order $\min(a, \log(1/\varepsilon) + \tau(n))$. In particular, the advice length can be assumed to grow extremely slowly as a function of n. This is quite a strong assumption whereas, somewhat surprisingly, the construction of our advice condenser, which is influenced by the advice generator of [CGL15], is fairly simple (see Section 9).

5 Preliminaries

In this section we set some notations that will be used throughout the paper and recall some of the more standard results from the literature that we apply frequently. In Section 6 we present some more recent results from the literature that we make use of.

Setting some standard notations. Unless stated otherwise, the logarithm in this paper is always taken base 2. For every natural number $n \geq 1$, define $[n] = \{1, 2, \ldots, n\}$. Throughout the paper, whenever possible, we avoid the use of floor and ceiling in order not to make the equations cumbersome. Whenever we say that a function is efficiently-computable we mean that the corresponding family of functions can be computed by a (uniform) algorithm that runs in polynomial-time in the input length.

Random variables and distributions. We sometimes abuse notation and syntactically treat random variables and their distribution as equal, specifically, we denote by U_m a random variable that is uniformly distributed over $\{0,1\}^m$. Furthermore, if U_m appears in a joint distribution (U_m, X) then U_m should be understood as being independent of X. When m is clear from context, we omit it from the subscript and write U. The support of a random variable X is denoted by $\mathsf{supp}(X)$. Let X, Y be two random variables. We say that Y is a *deterministic function of* X if the value of X determines the value of Y. Namely, there exists a function f such that $Y = f(X)$.

Statistical distance. The *statistical distance* between two distributions X, Y on the same domain D is defined by

$$\mathsf{SD}(X, Y) = \max_{A \subseteq D} \{ |\, \mathbf{Pr}[X \in A] - \mathbf{Pr}[Y \in A] \,| \}\,.$$

[8] The c-iterated log function is defined as follows. First, $\log^{(0)}(x) = x$, and for any integer $c > 0$, define $\log^{(c)}(x) = \log(\log^{(c-1)}(x))$ recursively.

If $\mathsf{SD}(X,Y) \leq \varepsilon$ we write $X \approx_\varepsilon Y$ and say that X and Y are ε-close.

We make frequent use of the following lemma.

Lemma 5.1. *Let X, X' be two random variables on the same domain. Let Y, Z be a random variables such that for any $y \in \mathsf{supp}(Y)$, the random variables $X \mid Y = y$, $Z \mid Y = y$ are independent and the random variables $X' \mid Y = y$, $Z \mid Y = y$ are independent. Then,*

$$\mathsf{SD}\left((X,Y),(X',Y)\right) = \mathsf{SD}\left((X,Z,Y),(X',Z,Y)\right).$$

Min-entropy. The *min-entropy* of a random variable X, denoted by $H_\infty(X)$, is defined by

$$H_\infty(X) = \min_{x \in \mathsf{supp}(X)} \log_2(1/\mathbf{Pr}[X = x]).$$

If X is supported on $\{0,1\}^n$, we define the *min-entropy rate* of X by $H_\infty(X)/n$. In such case, if X has min-entropy k or more, we say that X is an (n,k)-source. When wish to refer to an (n,k)-source without specifying the quantitative parameters, we sometimes use the standard terms *source* or *weak-source*.

Average conditional min-entropy. Let X, W be two random variables. The *average conditional min-entropy* of X given W is defined as

$$\widetilde{H}_\infty(X \mid W) = -\log_2\left(\mathop{\mathbf{E}}_{w \sim W}\left[2^{-H_\infty(X|W=w)}\right]\right).$$

We make frequent use of the following lemmas.

Lemma 5.2 ([DORS08])**.** *Let X, Y, Z be random variables such that Y has support size at most 2^ℓ. Then,*

$$\widetilde{H}_\infty(X \mid (Y,Z)) \geq \widetilde{H}_\infty((X,Y) \mid Z) - \ell \geq \widetilde{H}_\infty(X \mid Z) - \ell.$$

In particular, $\widetilde{H}_\infty(X \mid Y) \geq H_\infty(X) - \ell$.

Lemma 5.3 ([DORS08])**.** *For any two random variables X, Y and any $\varepsilon > 0$, it holds that*

$$\mathop{\mathbf{Pr}}_{y \sim Y}\left[H_\infty(X \mid Y = y) < \widetilde{H}_\infty(X \mid Y) - \log(1/\varepsilon)\right] \leq \varepsilon.$$

Lemma 5.4. *Let X, Y, Z be random variables such that for any $y \in \mathsf{supp}(Y)$ it holds that $X \mid Y = y$ and $Z \mid Y = y$ are independent. Then, $\widetilde{H}_\infty(X \mid (Y,Z)) = \widetilde{H}_\infty(X \mid Y)$. In particular, if X and Z are independent then $\widetilde{H}_\infty(X \mid Z) = H_\infty(X)$.*

Extractors. We provide standard definitions of extractors and state some of the results we use.

Definition 5.5 (Seeded extractors)**.** *A function* $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is called a* (k,ε)-seeded extractor *if for any* (n,k)*-source* X *it holds that* $\mathsf{Ext}(X,S) \approx_\varepsilon U_m$*, where* S *is uniformly distributed over* $\{0,1\}^d$ *and is independent of* X*. We say that* Ext *is a* strong *if* $(\mathsf{Ext}(X,S),S) \approx_\varepsilon U_{m+d}$*.*

We sometimes say that an extractor Ext *supports* min-entropy k. By that we mean that Ext is an extractor for min-entropy k. Throughout the paper we make use of the following explicit strong seeded extractors.

Theorem 5.6 ([GUV09])**.** *There exists a universal constant* $c_{\mathsf{GUV}} > 0$ *such that the following holds. For all positive integers* n,k *and* $\varepsilon > 0$*, there exists an efficiently-computable* (k,ε)*-strong seeded-extractor* $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *having seed length* $d = c_{\mathsf{GUV}} \cdot \log(n/\varepsilon)$ *and* $m = k/2$ *output bits.*

The following theorem readily follows by Theorem 4 in [Raz05] and Theorem 5.6.

Theorem 5.7. *There exist universal constants* $c_{\mathsf{Raz}}, c'_{\mathsf{Raz}}$ *such that the following holds. Let* n,k *be integers and let* $\varepsilon > 0$*. Set* $d = c_{\mathsf{Raz}} \cdot \log(n/\varepsilon)$*. For all* $k \geq c'_{\mathsf{Raz}} d$*, there exists an efficiently-computable function*

$$\mathsf{Raz}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k/2}$$

with the following property. Let X *be an* (n,k)*-source, and let* Y *be an independent* $(d,0.6d)$*-source. Then,* $(\mathsf{Raz}(X,Y),Y) \approx_\varepsilon (U,Y)$*.*

We make use of the following lemma. A proof and a short discussion regarding this lemma can be found in [CS16].

Lemma 5.8. *Let* $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a* (k,ε)*-strong seeded extractor. Let* W,S *be random variables over* n*-bit strings and* d*-bit strings, respectively. Let* $\mathcal{H}$ *be some random variable such that*

$$\widetilde{H}_\infty(W \mid \mathcal{H}) \geq k + \log(1/\varepsilon),$$
$$(S,\mathcal{H}) \approx_\delta (U,\mathcal{H})\,.$$

Assume further that conditioned on $\mathcal{H}$*, the random variables* W,S *are independent. Then,*

$$(\mathsf{Ext}(W,S),S,\mathcal{H}) \approx_{\delta+2\varepsilon} (U,S,\mathcal{H})\,.$$

Error correcting codes. We also need the following standard definition of an error correcting code.

Definition 5.9. *Let Σ be some set. A mapping $\mathsf{ECC}\colon \Sigma^k \to \Sigma^n$ is called an error correcting code with relative-distance δ if for any $x, y \in \Sigma^k$, it holds that the Hamming distance between $\mathsf{ECC}(x)$ and $\mathsf{ECC}(y)$ is at least δn. The rate of the code, denoted by ρ, is defined by $\rho = k/n$. We say that the alphabet size of the code is $|\Sigma|$.*

Theorem 5.10 ([GS95] (see also [Sti09]))**.** *Let p be any prime number and let m be an even integer. Set $q = p^m$. For every $\rho \in [0,1]$ and for any large enough integer n, there exists an efficiently-computable rate ρ linear error correcting code $\mathsf{ECC}\colon \mathbb{F}_q^{\rho n} \to \mathbb{F}_q^n$ with relative distance δ such that*

$$\rho + \delta \geq 1 - \frac{1}{\sqrt{q} - 1}.$$

6 Less Familiar Preliminaries

In this section we cover two components that we use throughout the paper. Unlike error correcting codes and extractors, that were covered in the previous section, and with which the reader is probably familiar, in this section we recall newer and less familiar concepts such as somewhere-independent matrices and independence-preserving mergers [CS16] as well as the notion of hierarchy of independence which is based on alternating extraction. We start this section by giving a formal definition for correlation breakers with advice, which is slightly more complete than the one given in the introduction.

6.1 Correlation breakers with advice

The following definition was first used, implicity, in [CGL15] and was later formalized explicitly in [Coh15b].

Definition 6.1 (Correlation breakers with advice)**.** *A function*

$$\mathsf{AdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

*is called a (k,ε)-*correlation breaker with advice *if the following holds. Let Y, Y' be ℓ-bit random variables such that Y uniform. Let X, X' be n-bit random variables with $H_\infty(X) \geq k$, and such that (X, X') is independent of (Y, Y'). Then, for any pair of distinct a-bit strings α, α',*

$$\left(\mathsf{AdvCB}\left(X, Y, \alpha\right), \mathsf{AdvCB}\left(X', Y', \alpha'\right)\right) \approx_\varepsilon \left(U, \mathsf{AdvCB}\left(X', Y', \alpha'\right)\right).$$

Further, we say that AdvCB *is* strong *if*

$$\left(\mathsf{AdvCB}\left(X, Y, \alpha\right), \mathsf{AdvCB}\left(X', Y', \alpha'\right), Y, Y'\right) \approx_\varepsilon \left(U, \mathsf{AdvCB}\left(X', Y', \alpha'\right), Y, Y'\right).$$

6.2 Hierarchy of independence

Let c_{GUV} be the constant that is given by Theorem 5.6. Let n_1, n_2, b be some integers, and let $\varepsilon > 0$. Let $s_1 = c_{\mathsf{GUV}} \cdot \log(n_1/\varepsilon)$ be a length that suffices for a seed of the strong seeded extractor that is given by Theorem 5.6 when fed with a sample from an n_1-bit source and when set with error guarantee ε. Similarly, we define $s_2 = c_{\mathsf{GUV}} \cdot \log(n_2/\varepsilon)$ to be a length that suffices for a seed of the extractor that is given by Theorem 5.6 when fed with a sample from an n_2-bit source and when set with error guarantee ε. We further assume that $b \geq s_1$. Let

$$\mathsf{Ext}_1 \colon \{0,1\}^{n_1} \times \{0,1\}^{s_1} \to \{0,1\}^{s_2}$$

be the $(2s_2, \varepsilon)$-strong seeded extractor that is given by Theorem 5.6. Let

$$\mathsf{Ext}_2 \colon \{0,1\}^{n_2} \times \{0,1\}^{s_2} \to \{0,1\}^{b}$$

be the $(2b, \varepsilon)$-strong seeded extractor that is given by Theorem 5.6. We let

$$\mathsf{Ext}_3 \colon \{0,1\}^{n_2} \times \{0,1\}^{s_2} \to \{0,1\}^{s_1}$$

be the function that is obtained by applying Ext_2 and taking only the length s_1 prefix of the output (recall that $b \geq s_1$). We define a pair of functions

$$\begin{aligned} \mathsf{a} &\colon \{0,1\}^{s_2} \times \{0,1\}^{n_2} \to \{0,1\}^{b}, \\ \mathsf{b} &\colon \{0,1\}^{s_2} \times \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^{b}, \end{aligned}$$

as follows. For $v \in \{0,1\}^{s_2}$, $z \in \{0,1\}^{n_1}$, and $w \in \{0,1\}^{n_2}$,

$$\begin{aligned} \mathsf{a}(v,w) &= \mathsf{Ext}_2(w,v), \\ \mathsf{b}(v,z,w) &= \mathsf{Ext}_2(w, \mathsf{Ext}_1(z, \mathsf{Ext}_3(w,v))). \end{aligned}$$

The following lemma appears with different twists in several previous works [DP07, DW09, Li13a, Li15b, Coh15a, CS16].

Lemma 6.2. *Let $\mathcal{Y} = (Y, Y')$ be a pair of s_2-bit random variables such that*

$$(Y, Y', \mathcal{H}) \approx_\delta (U, Y', \mathcal{H}).$$

Let $\mathcal{Z} = (Z, Z')$ be a pair of n_1-bit random variables, and let $\mathcal{W} = (W, W')$ be a pair of n_2-bit random variables, such that

- $\widetilde{H}_\infty(Z \mid \mathcal{H}) \geq 5s_2$.
- $\widetilde{H}_\infty(W \mid \mathcal{H}) \geq 5b$.

- *For any $h \in \mathsf{supp}(\mathcal{H})$, the random variable $\mathcal{W} \mid (\mathcal{H} = h)$ is independent of the joint distribution of $\mathcal{Y} \mid (\mathcal{H} = h)$, $\mathcal{Z} \mid (\mathcal{H} = h)$.*

Write

$$\begin{aligned}\widehat{\mathcal{A}} &= \mathsf{a}(Y, W), \mathsf{a}(Y', W'),\\ \widehat{\mathcal{Z}} &= \mathsf{Ext}_1(Z, \mathsf{Ext}_3(W, Y)), \mathsf{Ext}_1(Z', \mathsf{Ext}_3(W', Y')).\end{aligned} \tag{6.1}$$

Then, the following holds:

1. $(\mathsf{a}(Y, W), \mathcal{Z}, \mathcal{Y}, \mathcal{H}) \approx_{\delta+2\varepsilon} (U, \mathcal{Z}, \mathcal{Y}, \mathcal{H})$,
2. $\left(\mathsf{b}(Y, Z, W), \mathsf{a}(Y', W'), \mathcal{Z}, \widehat{\mathcal{Z}}, \widehat{\mathcal{A}}, \mathcal{Y}, \mathcal{H}\right) \approx_{\delta+6\varepsilon} \left(U, \mathsf{a}(Y', W'), \mathcal{Z}, \widehat{\mathcal{Z}}, \widehat{\mathcal{A}}, \mathcal{Y}, \mathcal{H}\right)$.

Furthermore,

3. $\widetilde{H}_\infty\left(Z \mid \widehat{\mathcal{Z}}, \widehat{\mathcal{A}}, \mathcal{Y}, \mathcal{H}\right) \geq \widetilde{H}_\infty(Z \mid \mathcal{H}) - 5s_2$,
4. $\widetilde{H}_\infty\left(W \mid \widehat{\mathcal{Z}}, \widehat{\mathcal{A}}, \mathcal{Y}, \mathcal{H}\right) \geq \widetilde{H}_\infty(W \mid \mathcal{H}) - 5b$.

6.3 Independence-preserving mergers

Definition 6.3 (Somewhere-independent matrices [CS16])**.** *Let M, M' be random variables in the form of $r \times \ell$ matrices. Let $\mathcal{H}$ be a random variable and let $\delta > 0$. We say that M is $(\delta, \mathcal{H})$-*somewhere independent *of M' if the following holds:*

- *There exists $g \in [r]$ such that $\left(M_g, M'_g, \mathcal{H}\right) \approx_\delta \left(U, M'_g, \mathcal{H}\right)$. For any such g we say that M is $(\delta, \mathcal{H})$-*independent *of M' at g.*
- *For any $i \in [r]$, $(M_i, \mathcal{H}) \approx_\delta (U, \mathcal{H})$.*

Theorem 6.4 ([CS16])**.** *There exists a universal constant $c_{\mathsf{IPM}} \geq 1$ such that the following holds. Let n, r, m be integers such that r is an even power of 2. Let $\varepsilon, \delta > 0$, and set $s = c_{\mathsf{GUV}} \cdot \log(n/\varepsilon)$, where c_{GUV} is the constant given by Theorem 5.6. Then, there exists an efficiently-computable function*

$$\mathsf{IPMerg}\colon \{0,1\}^{r \times s} \times \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$$

with the following property. Let $\mathcal{M} = (M, M')$ be a pair of random variables in the form of $r \times s$ matrices such that M is $(\delta, \mathcal{H})$-somewhere independent of M'. Denote $v = c_{\mathsf{IPM}} \cdot r \cdot \log(n/\varepsilon)$. Let $\mathcal{Z} = (Z, Z')$ and $\mathcal{W} = (W, W')$ be two pairs of n-bit random variables such that

- $\widetilde{H}_\infty(Z \mid \mathcal{H}) \geq v$.

- $\widetilde{H}_\infty(W \mid \mathcal{H}) \geq v + 2m$.
- *For any* $h \in \mathsf{supp}(\mathcal{H})$, *the random variable* $\mathcal{W} \mid (\mathcal{H} = h)$ *is independent of the joint distribution of* $\mathcal{M} \mid (\mathcal{H} = h)$, $\mathcal{Z} \mid (\mathcal{H} = h)$.

Set $N = \mathsf{IPMerg}(M, Z, W)$, $N' = \mathsf{IPMerg}(M', Z', W')$, *and set* $\mathcal{N} = (N, N')$. *Then, there exists a random variable* $\mathcal{H}'$ *such that the following holds:*

- $(N, N', \mathcal{H}') \approx_{(\delta+12\varepsilon)r} (U, N', \mathcal{H}')$.
- $\widetilde{H}_\infty(Z \mid \mathcal{H}') \geq \widetilde{H}_\infty(Z \mid \mathcal{H}) - v$.
- $\widetilde{H}_\infty(W \mid \mathcal{H}') \geq \widetilde{H}_\infty(W \mid \mathcal{H}) - v$.
- *For any* $h \in \mathsf{supp}(\mathcal{H}')$, *the random variable* $\mathcal{Z} \mid (\mathcal{H}' = h)$ *is independent of the joint distribution of* $\mathcal{N} \mid (\mathcal{H}' = h)$, $\mathcal{W} \mid (\mathcal{H}' = h)$.

7 The Base Correlation Breaker with Advice

The main result of this section is given by the following lemma. We remark that one can easily relax the condition $k = 3m + c'_{\mathsf{base}}\ell$ in the statement of the lemma to $k = (2+\alpha)m + c'_{\mathsf{base}}\ell$ for any constant $\alpha > 0$.

Lemma 7.1. *There exist universal constants* $c_{\mathsf{base}}, c'_{\mathsf{base}} \geq 1$ *such that the following holds. For all integers* n, a, m, *and for any* $\varepsilon > 0$, *set*

$$\ell = c_{\mathsf{base}} \cdot (\log n + a \cdot \log(a/\varepsilon)).$$

Then, there exists an explicit (k, ε)*-strong correlation breaker with advice*

$$\mathsf{BaseAdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m,$$

with $k = 3m + c'_{\mathsf{base}}\ell$.

Proof of Lemma 7.1. Let c_{GUV} be the constant that is given by Theorem 5.6, and let c_{Raz} be the constant that is given by Theorem 5.7. Set

$$\begin{aligned} b &= c_{\mathsf{GUV}} \cdot \log(\ell/\varepsilon), \\ \ell_1 &= c_{\mathsf{GUV}} \cdot \log(n/\varepsilon), \\ \ell_2 &= 5\ell_1, \\ \ell_3 &= \max(10\ell_2, c_{\mathsf{Raz}} \cdot \log(n/\varepsilon)). \end{aligned}$$

We further assume that $\ell \geq 10\ell_3$. For an ℓ-bit string y, and for $i = 1, 2, 3$, we denote by y_i the ℓ_i-bit prefix of y. Note that y_1 is a prefix of y_2 which, in turn, is a prefix of y_3. Further, observe that the constant c_{base} can be taken large enough (with respect to c_{GUV} and c_{Raz}) so that ℓ_1, ℓ_2, ℓ_3, and ℓ could be set the way we described above.

Building blocks. For the construction of $\mathsf{BaseAdvCB}$ we make use of the following building blocks:

- Let $\mathsf{a}\colon \{0,1\}^{\ell_1} \times \{0,1\}^n \to \{0,1\}^b$ and $\mathsf{b}\colon \{0,1\}^{\ell_1} \times \{0,1\}^{\ell_2} \times \{0,1\}^n \to \{0,1\}^b$ be the pair of functions that are defined in Section 6.2, set with error guarantee ε. Note that ℓ_1 was set according to Lemma 6.2. Further, as $\ell \geq \ell_2$, the parameter b is large enough as required by Lemma 6.2.
- Let $\mathsf{Raz}\colon \{0,1\}^n \times \{0,1\}^{\ell_3} \to \{0,1\}^{\ell}$ be the $(2\ell, \varepsilon)$-extractor with weak seeds that is given by Theorem 5.7. Note that ℓ_3 was chosen to be sufficiently large as required by Theorem 5.7.
- Let $\mathsf{IPMerg}\colon \{0,1\}^{(2a)\times b} \times \{0,1\}^{\ell} \times \{0,1\}^{\ell} \to \{0,1\}^{\ell_1}$ be the independence-preserving merger that is given by Theorem 6.4, set with error guarantee ε. Note that b was set according to Theorem 6.4.
- Let $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^{\ell_1} \to \{0,1\}^m$ be the $(2m, \varepsilon)$-strong seeded extractor that is given by Theorem 5.6. Note that ℓ_1 is sufficiently large are required by Theorem 5.6.

The construction. On input $x \in \{0,1\}^n$, $y \in \{0,1\}^{\ell}$, $\alpha \in \{0,1\}^a$, the first step in the computation of $\mathsf{BaseAdvCB}(x, y, \alpha)$ is constructing a $(2a) \times b$ matrix $m = m(x, y_2, \alpha)$ that is defined as follows [9]. For $i \in [2a]$, the i'th row of the matrix m is given by

$$m(x, y_2, \alpha)_i = \begin{cases} \mathsf{a}(y_1, x), & i \neq \alpha_{\lceil i/2 \rceil} \pmod 2; \\ \mathsf{b}(y_1, y_2, x), & i = \alpha_{\lceil i/2 \rceil} \pmod 2. \end{cases}$$

We then compute

$$z = \mathsf{IPMerg}\left(m\left(x, y_2, \alpha\right), \mathsf{Raz}\left(x, y_3\right), y\right).$$

Finally, we define

$$\mathsf{BaseAdvCB}(x, y, \alpha) = \mathsf{Ext}\left(x, z\right).$$

Analysis. Let X, X' be a pair of n-bit random variables such that $H_\infty(X) \geq k$. Let Y, Y' be a pair of ℓ-bit random variables that are jointly independent of the joint distribution (X, X'), and such that Y is uniformly distributed. Let α, α' be distinct, fixed, a-bit strings.

Set $M = m(X, Y_2, \alpha)$ and $M' = m(X', Y_2', \alpha')$. Let $i \in [a]$ be such that $\alpha_i \neq \alpha_i'$. Set $g = 2i - \alpha_i$, and observe that $M_g = \mathsf{b}(Y_1, Y_2, X)$ whereas $M_g' = \mathsf{a}(Y_1', X')$. By our choice

[9] By convention, the output length of $\mathsf{BaseAdvCB}$ is also denoted by m though this should cause no confusion.

of ℓ_2, and since $k \geq 5b$, we can apply Lemma 6.2 with $\mathcal{Y} = (Y_1, Y_1')$, $\mathcal{Z} = (Y_2, Y_2')$, and $\mathcal{W} = (X, X')$, and conclude, by Item 2 of the lemma, that

$$\left(M_g, M_g', Y_2, Y_2'\right) \approx_{6\varepsilon} \left(U, M_g', Y_2, Y_2'\right), \tag{7.1}$$

Moreover, by Item 1 and Item 2 of Lemma 6.2 it follows that for every $i \in [2a]$,

$$\left(M_i, Y_2, Y_2'\right) \approx_{6\varepsilon} \left(U, Y_2, Y_2'\right). \tag{7.2}$$

Equation (7.1) and Equation (7.2) then imply that M is $(6\varepsilon, (Y_2, Y_2'))$-independent of M' at g.

By Item 4 of Lemma 6.2,

$$\widetilde{H}_\infty\left(X \mid Y_2, Y_2'\right) \geq k - 5b \geq \max\left(2\ell, c'_{\mathsf{Raz}}\ell_3\right) + \log(1/\varepsilon),$$

where the last inequality follows by setting the constant c'_{base} large enough. Further, by Lemma 5.2 and Lemma 5.4, and since the random variables M, M' are deterministic functions of Y_2, Y_2' and X, X',

$$\widetilde{H}_\infty\left(Y_3 \mid Y_2, Y_2'\right) \geq \ell_3 - 2\ell_2 \geq 0.8\ell_3. \tag{7.3}$$

Equation (7.3) and Equation (7.3), together with the fact that X is independent of Y_3 conditioned on any fixing of Y_2, Y_2' allows us to apply Theorem 5.7 and Lemma 5.8, and conclude that

$$\left(\mathsf{Raz}(X, Y_3), Y_3, Y_2, Y_2'\right) \approx_{2\varepsilon} \left(U, Y_3, Y_2, Y_2'\right).$$

Note that conditioned on Y_3, Y_2, Y_2', the random variable $\mathsf{Raz}(X, Y_3)$ is independent of Y_3'. Thus, we may apply Lemma 5.4 to conclude that

$$\left(\mathsf{Raz}(X, Y_3), Y_3, Y_3'\right) \approx_{2\varepsilon} \left(U, Y_3, Y_3'\right),$$

where we have used the fact that Y_2 and Y_2' are prefixes of Y_3 and Y_3', respectively.

Recall that M is $(6\varepsilon, (Y_2, Y_2'))$-somewhere independent of M'. As the joint distribution of M, M' is independent of the joint distribution of Y_3, Y_3' conditioned on any fixing of Y_2, Y_2', it holds that M is $(6\varepsilon, (Y_3, Y_3'))$-somewhere independent of M'. By Lemma 5.2 and by the fact that $\ell \geq 10\ell_3$, it is possible to set the constant c_{base} large enough with respect to the constant c_{IPM} that is given by Theorem 6.4, so to guarantee that

$$\begin{aligned}
\widetilde{H}_\infty\left(Y \mid Y_3, Y_3'\right) &\geq \ell - 2\ell_3 \\
&\geq 0.8\ell \\
&\geq c_{\mathsf{IPM}} \cdot 2a \cdot \log\left(\ell/\varepsilon\right),
\end{aligned}$$

Note further that

$$\begin{aligned}\widetilde{H}_\infty(X \mid Y_3, Y_3') &= \widetilde{H}_\infty(X \mid Y_2, Y_2') \\ &\geq k - 5b \\ &\geq c_{\mathsf{IPM}} \cdot 2a \cdot \log(\ell/\varepsilon) + 2\ell_1.\end{aligned}$$

Thus, we are in a position to apply Theorem 6.4 to conclude that there exists a random variable $\mathcal{H}$ such that

$$(Z, Z', \mathcal{H}) \approx_{O(a\varepsilon)} (U, Z', \mathcal{H}), \tag{7.4}$$

where

$$\begin{aligned}Z &= \mathsf{IPMerg}(M, \mathsf{Raz}(X, Y_3), Y), \\ Z' &= \mathsf{IPMerg}(M', \mathsf{Raz}(X', Y_3'), Y').\end{aligned}$$

Furthermore, conditioned on $\mathcal{H}$, both Z and Z' are deterministic functions of Y, Y' which, in turn, are jointly independent of the joint distribution (X, X'). Thus, conditioned on $Z', \mathcal{H}$, the random variable Z is independent of $\mathsf{Ext}(X, Z')$, and so by Equation (7.4) and Lemma 5.1,

$$(Z, \mathsf{Ext}(X', Z'), Z', \mathcal{H}) \approx_{O(a\varepsilon)} (U, \mathsf{Ext}(X', Z'), Z', \mathcal{H}). \tag{7.5}$$

Now, as $\mathsf{Ext}(X', Z')$ consists of m bits, $\mathsf{Raz}(X, Y_3)$ and $\mathsf{Raz}(X', Y_3')$ each consists of ℓ bits, and since each of M, M' has only 2 distinct random variables that occupy its length b rows, we have that

$$\begin{aligned}\widetilde{H}_\infty(X \mid \mathsf{Ext}(X', Z'), Z', \mathcal{H}) &\geq \widetilde{H}_\infty(X \mid Z', \mathcal{H}) - m \\ &= \widetilde{H}_\infty(X \mid \mathcal{H}) - m \\ &\geq \widetilde{H}_\infty(X \mid \mathcal{H}) - m - 4b - 2\ell \\ &\geq 2m + \log(1/\varepsilon).\end{aligned}$$

Therefore, by Theorem 5.6, Lemma 5.8, and by Equation (7.5), it holds that

$$(\mathsf{Ext}(X, Z), \mathsf{Ext}(X', Z'), Z, Z', \mathcal{H}) \approx_{O(a\varepsilon)} (U, \mathsf{Ext}(X', Z'), Z, Z', \mathcal{H}).$$

Conditioned on $\mathsf{Ext}(X', Z'), Z, Z', \mathcal{H}$, the random variables $\mathsf{Ext}(X, Z)$ is independent of the joint distribution of Y, Y'. Thus, by Lemma 5.3,

$$(\mathsf{BaseAdvCB}(X, Y, \alpha), \mathsf{BaseAdvCB}(X', Y', \alpha'), Y, Y') \approx_{O(a\varepsilon)} (U, \mathsf{BaseAdvCB}(X', Y', \alpha'), Y, Y').$$

The proof then follows by setting the error guarantee of all building blocks we have used to $\beta\varepsilon/a$ for some small enough constant $\beta > 0$ so to reduce the total error from $O(a\varepsilon)$ to ε. □

8 Stepping-Up Correlation Breakers with Advice

In this section we devise an algorithm that transforms, in a black-box manner, a given correlation breaker with advice to one with better dependence on the advice length. This is the content of Lemma 8.1. Then, in Lemma 8.3 we apply this transformation, repeatedly, starting with the base correlation breaker with advice that is given by Lemma 7.1.

Lemma 8.1. *There exists a universal constant $c_{\mathsf{SU}} \geq 1$ such that the following holds. Let $n, a_{\mathsf{out}}, m_{\mathsf{out}}, b, \ell_{\mathsf{in}}, k_{\mathsf{in}}$ be integers and let $\varepsilon_{\mathsf{in}} > 0$. Set*

$$\begin{aligned} a_{\mathsf{in}} &= a_{\mathsf{out}}/b, \\ \ell_{\mathsf{out}} &= 3m_{\mathsf{out}} + c_{\mathsf{SU}} \cdot \left(\ell_{\mathsf{in}} + \log n + b \cdot \log\left(b/\varepsilon_{\mathsf{in}}\right)\right), \\ m_{\mathsf{in}} &= c_{\mathsf{GUV}} \cdot \log(\ell_{\mathsf{out}}/\varepsilon_{\mathsf{in}}), \end{aligned}$$

where c_{GUV} is the constant that is given by Theorem 5.6. Let

$$\mathsf{AdvCB}_{\mathsf{in}} \colon \{0,1\}^n \times \{0,1\}^{\ell_{\mathsf{in}}} \times \{0,1\}^{a_{\mathsf{in}}} \to \{0,1\}^{m_{\mathsf{in}}}$$

be an explicit $(k_{\mathsf{in}}, \varepsilon_{\mathsf{in}})$-strong correlation breaker with advice. Then, there exists an explicit $(k_{\mathsf{out}}, \varepsilon_{\mathsf{out}})$-strong correlation breaker with advice

$$\mathsf{AdvCB}_{\mathsf{out}} \colon \{0,1\}^n \times \{0,1\}^{\ell_{\mathsf{out}}} \times \{0,1\}^{a_{\mathsf{out}}} \to \{0,1\}^{m_{\mathsf{out}}}$$

with

$$\begin{aligned} k_{\mathsf{out}} &= \max\left(k_{\mathsf{in}}, 4\ell_{\mathsf{out}}\right), \\ \varepsilon_{\mathsf{out}} &= c_{\mathsf{SU}} \cdot b\sqrt{\varepsilon_{\mathsf{in}}}. \end{aligned}$$

Proof. For a string $y \in \{0,1\}^{\ell_{\mathsf{out}}}$, let y_1, y_2 denote the length $\ell_{\mathsf{in}}, \ell_{\mathsf{Raz}}$ prefixes of y, respectively, where we recall that ℓ_{in} is the length of the second argument of $\mathsf{AdvCB}_{\mathsf{in}}$, and define

$$\ell_{\mathsf{Raz}} = \max\left(10\ell_{\mathsf{in}}, c_{\mathsf{Raz}} \cdot \log(n/\varepsilon_{\mathsf{in}})\right),$$

where c_{Raz} is the constant that is given by Theorem 5.7. We further assume that $\ell_{\mathsf{out}} \geq 10\ell_{\mathsf{Raz}}$. This assumption can be met by taking c_{SU} to be a large enough constant (compared to the constant c_{Raz}).

Building blocks. On top of $\mathsf{AdvCB}_{\mathsf{in}}$, for the construction of $\mathsf{AdvCB}_{\mathsf{out}}$ we make use of the following building blocks:

- Let $\mathsf{Raz} \colon \{0,1\}^n \times \{0,1\}^{\ell_{\mathsf{Raz}}} \to \{0,1\}^{\ell_{\mathsf{out}}}$ be the $(2\ell_{\mathsf{out}}, \varepsilon_{\mathsf{in}})$-extractor with weak-seeds that is given by Theorem 5.7. Note that ℓ_{Raz} was set to have sufficient length as required by Theorem 5.7.

- Set $d = c_{\mathsf{GUV}} \cdot \log(n/\varepsilon_{\mathsf{in}})$, and let $\mathsf{IPMerg}\colon \{0,1\}^{b \times m_{\mathsf{in}}} \times \{0,1\}^{\ell_{\mathsf{out}}} \times \{0,1\}^{\ell_{\mathsf{out}}} \to \{0,1\}^d$ be the independence-preserving merger that is given by Theorem 6.4, set with error guarantee $\varepsilon_{\mathsf{in}}$. Recall that $m_{\mathsf{in}} = c_{\mathsf{GUV}} \cdot \log(\ell_{\mathsf{out}}/\varepsilon_{\mathsf{in}})$, as required by Theorem 6.4.
- Let $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{m_{\mathsf{out}}}$ be the $(2m_{\mathsf{out}}, \varepsilon_{\mathsf{in}})$-strong seeded extractor that is given by Theorem 5.6. Note that d suffices so as to be used as a seed length for Ext.

The construction. Let $x \in \{0,1\}^n$, $y \in \{0,1\}^{\ell_{\mathsf{out}}}$, and let α be an a_{out}-bit string. We partition α to b consecutive equal length substrings, or blocks, $\alpha = \alpha_1 \circ \cdots \circ \alpha_b$. Note that for any $i \in [b]$, $|\alpha_i| = a_{\mathsf{in}}$. Define the $b \times m_{\mathsf{in}}$ matrix $m(x, y_1)$ as follows. For $i \in [b]$, the i'th row of $m(x, y_1)$ is given by

$$m(x, y_1)_i = \mathsf{AdvCB}_{\mathsf{in}}(x, y_1, \alpha_i).$$

Define

$$z = \mathsf{IPMerg}\left(m(x, y_1), \mathsf{Raz}(x, y_2), y\right).$$

Finally, define

$$\mathsf{AdvCB}_{\mathsf{out}}(x, y, \alpha) = \mathsf{Ext}\left(x, z\right).$$

Analysis. Let X, X' be n-bit random variables such that $H_\infty(X) \geq k_{\mathsf{out}}$. Let Y, Y' be ℓ_{out}-bit random variables such that Y is uniform, and such that (X, X') is independent of (Y, Y'). Let $\alpha \neq \alpha'$ be distinct, fixed, a_{out}-bit strings. As $\alpha \neq \alpha'$, there exists $g \in [b]$ such that $\alpha_g \neq \alpha'_g$. As $k_{\mathsf{out}} \geq k_{\mathsf{in}}$ and since $\mathsf{AdvCB}_{\mathsf{in}}$ is a $(k_{\mathsf{in}}, \varepsilon_{\mathsf{in}})$-strong correlation breaker with advice,

$$\left(\mathsf{AdvCB}_{\mathsf{in}}(X, Y_1, \alpha_g), \mathsf{AdvCB}_{\mathsf{in}}(X', Y'_1, \alpha'_g), Y_1, Y'_1\right) \approx_{\varepsilon_{\mathsf{in}}} \left(U, \mathsf{AdvCB}_{\mathsf{in}}(X', Y'_1, \alpha'_g), Y_1, Y'_1\right).$$

Further, note that for any $i \in [b]$,

$$\left(\mathsf{AdvCB}_{\mathsf{in}}(X, Y_1, \alpha_i), Y_1, Y'_1\right) \approx_{\varepsilon_{\mathsf{in}}} \left(U, Y_1, Y'_1\right). \tag{8.1}$$

Indeed, pick any string $\beta \neq \alpha_i$ and observe that

$$\left(\mathsf{AdvCB}_{\mathsf{in}}(X, Y_1, \alpha_i), \mathsf{AdvCB}_{\mathsf{in}}(X', Y'_1, \beta), Y_1, Y'_1\right) \approx_{\varepsilon_{\mathsf{in}}} \left(U, \mathsf{AdvCB}_{\mathsf{in}}(X', Y'_1, \beta), Y_1, Y'_1\right),$$

which, in particular implies Equation (8.1). Thus, $m(X, Y_1)$ is $(\varepsilon_{\mathsf{in}}, \mathcal{H})$-somewhere independent of $m(X', Y'_1)$, where $\mathcal{H} = Y_1, Y'_1$.

By Lemma 5.2,

$$\widetilde{H}_\infty\left(Y_2 \mid \mathcal{H}\right) \geq \ell_{\mathsf{Raz}} - 2\ell_{\mathsf{in}} \geq 0.8\ell_{\mathsf{Raz}},$$

and so, by Lemma 5.3, except with probability $O(\varepsilon_{\mathsf{in}})$ over the fixing of $\mathcal{H}$, the random variable Y_2 has min-entropy rate 0.7. Thus, by Theorem 5.7, and since

$$k_{\mathsf{out}} \geq \max\left(2\ell_{\mathsf{out}}, c'_{\mathsf{Raz}}\ell_{\mathsf{Raz}}\right) + \log(1/\varepsilon_{\mathsf{in}}),$$

we have that

$$(\mathsf{Raz}(X, Y_2), Y_2, \mathcal{H}) \approx_{O(\varepsilon_{\mathsf{in}})} (U, Y_2, \mathcal{H}).$$

Note that conditioned on any fixing of $Y_2, \mathcal{H}$, the random variable $\mathsf{Raz}(X, Y_2)$ is independent of Y'_2. Therefore, by Lemma 5.1,

$$(\mathsf{Raz}(X, Y_2), \mathcal{H}') \approx_{O(\varepsilon_{\mathsf{in}})} (U, \mathcal{H}'),$$

where $\mathcal{H}' = Y_2, Y'_2$. Note further that by Lemma 5.2,

$$\widetilde{H}_\infty\left(Y \mid \mathcal{H}'\right) \geq \ell_{\mathsf{out}} - 2\ell_{\mathsf{Raz}} \geq 0.8\ell_{\mathsf{out}}.$$

Thus, by Lemma 5.3, except with probability $O(\varepsilon_{\mathsf{in}})$ over the fixing of $\mathcal{H}'$, the random variable Y has min-entropy $0.7\ell_{\mathsf{out}}$. Note further that $m(X, Y_1)$ is $(\varepsilon_{\mathsf{in}}, \mathcal{H}')$-somewhere independent of $m(X, Y'_1)$ (as apposed to just being $(\varepsilon_{\mathsf{in}}, \mathcal{H})$-somewhere independent).

One can now apply Markov's inequality and the union bound over all rows of $m(X, Y_1)$ to conclude that except with probability $O(b\sqrt{\varepsilon_{\mathsf{in}}})$ over the fixing of $\mathcal{H}'$, the following holds.

- $m(X, Y_1)$ is $O(\sqrt{\varepsilon_{\mathsf{in}}})$-somewhere independent of $m(X', Y'_1)$.
- $\mathsf{Raz}(X, Y_2)$ is $O(\sqrt{\varepsilon_{\mathsf{in}}})$-close to uniform.
- The random variables $m(X, Y_1)$, $m(X', Y'_1)$, $\mathsf{Raz}(X, Y_2)$, and $\mathsf{Raz}(X', Y'_2)$ are all deterministic functions of (X, X'), and in particular are jointly independent of (Y, Y').
- $H_\infty(Y) \geq 0.7\ell_{\mathsf{out}}$.

This, together with our choice of ℓ_{out}, allows us to apply Theorem 6.4, and conclude that

$$(Z, Z', \mathcal{H}'') \approx_{O(b\sqrt{\varepsilon_{\mathsf{in}}})} (U, Z', \mathcal{H}''),$$

where

$$Z = \mathsf{IPMerg}\left(m(X, Y_1), \mathsf{Raz}(X, Y_2), Y\right),$$
$$Z' = \mathsf{IPMerg}\left(m(X', Y'_1), \mathsf{Raz}(X', Y'_2), Y'\right),$$

and $\mathcal{H}''$ is a random variable such that conditioned on any fixing of $\mathcal{H}''$, each of the random variables Z, Z' is a deterministic function of Y, Y', respectively. Further, conditioned

on $\mathcal{H}''$ the joint distribution of the random variables X, X' is independent of the joint distribution of Y, Y'. Thus, by Lemma 5.1,

$$(Z, \mathsf{Ext}(X', Z'), Z', \mathcal{H}'') \approx_{O(b\sqrt{\varepsilon_{\mathsf{in}}})} (U, \mathsf{Ext}(X', Z'), Z', \mathcal{H}'') \,. \tag{8.2}$$

Now, as each of $m(X, Y_1)$, $m(X', Y_1')$ consists of bm_{in} bits and since Raz, Ext have output lengths ℓ_{out}, m_{out}, respectively, Lemma 5.2 implies that

$$\begin{aligned} \widetilde{H}_\infty (X \mid \mathsf{Ext}(X', Z'), Z', \mathcal{H}'') &\geq k_{\mathsf{out}} - 2(bm_{\mathsf{in}} + \ell_{\mathsf{out}}) - m_{\mathsf{out}} \\ &\geq 2m_{\mathsf{out}} + \log(1/\varepsilon_{\mathsf{in}}), \end{aligned} \tag{8.3}$$

where the last inequality holds by taking the constant c_{SU} large enough. Equation (8.2), Equation (8.3), together with the fact that X is independent of Z conditioned on any fixing of $\mathsf{Ext}(X', Z')$, Z', $\mathcal{H}''$ imply that

$$(\mathsf{Ext}(X, Z), Z, \mathsf{Ext}(X', Z'), Z', \mathcal{H}'') \approx_{O(b\sqrt{\varepsilon_{\mathsf{in}}})} (U, Z, \mathsf{Ext}(X', Z'), Z', \mathcal{H}'') \,.$$

Conditioned on the fixings of $Z, \mathsf{Ext}(X', Z'), Z', \mathcal{H}''$, the random variable $\mathsf{Ext}(X, Z)$ is independent of the joint distribution of Y, Y'. Hence, by Lemma 5.1,

$$(\mathsf{AdvCB}_{\mathsf{out}}(X, Y, \alpha), \mathsf{AdvCB}_{\mathsf{out}}(X', Y', \alpha'), Y, Y') \approx_{O(b\sqrt{\varepsilon})} (U, \mathsf{AdvCB}_{\mathsf{out}}(X', Y', \alpha'), Y, Y') \,.$$

This concludes the proof. □

The following corollary readily follows by Lemma 8.1. The proof is by a straightforward induction. The base case is obtained by taking AdvCB_1 to be the base correlation breaker with advice $\mathsf{BaseAdvCB}$ that is given by Lemma 7.1. For the inductive step, we apply Lemma 8.1 to AdvCB_v as $\mathsf{AdvCB}_{\mathsf{in}}$ to obtain $\mathsf{AdvCB}_{v+1} = \mathsf{AdvCB}_{\mathsf{out}}$, with $b = a^{1/v}$. We omit the tedious technical details.

Corollary 8.2. *There exists universal constants $c, c' \geq 1$ such that the following holds. For all n, a, m, all $\varepsilon > 0$, and for any $v = o(\log a)$, there exists an explicit (k_v, ε)-strong correlation breaker with advice*

$$\mathsf{AdvCB}_v \colon \{0,1\}^n \times \{0,1\}^{\ell_v} \times \{0,1\}^a \to \{0,1\}^m,$$

with

$$\begin{aligned} \ell_v &= c^v \cdot \left(\log n + a^{1/v} \cdot \log\left(a/\varepsilon\right)\right), \\ k_v &= c'\ell_v + 3m. \end{aligned}$$

By setting $v = O(\sqrt{\log a})$, Corollary 8.2 readily implies the following lemma.

Lemma 8.3. *There exist universal constants $c'_{\mathsf{SU}}, c''_{\mathsf{SU}} \geq 1$ such that the following holds. For all integers n, a, and for any $\varepsilon > 0$, there exists an efficiently-computable (k, ε)-strong correlation breaker with advice*

$$\mathsf{AdvCB} \colon \{0,1\}^n \times \{0,1\}^{\ell} \times \{0,1\}^a \to \{0,1\}^m,$$

where $\ell = (c'_{\mathsf{SU}})^{\sqrt{\log a}} \cdot \log(n/\varepsilon)$ and $k = 3m + c''_{\mathsf{SU}}\ell$.

9 Advice Condensers

In this section we formally define, and construct, advice condensers.

Definition 9.1 (Advice condensers)**.** *A function*

$$\mathsf{AdvCond}\colon \{0,1\}^{a_{\mathsf{in}}} \times \{0,1\}^{n} \times \{0,1\}^{d} \to \{0,1\}^{a_{\mathsf{out}}}$$

is called a (k,ε)-advice condenser *if the following holds. Let* X, X' *be* n*-bit random variables such that* $H_\infty(X) \geq k$*. Let* Y, Y' *be* d*-bit random variables such that* Y *is uniform. Assume further that the joint distribution of* X, X' *is independent of the joint distribution of* Y, Y'*. Then, for any pair of distinct* a*-bit strings* α, α'*, it holds that*

$$\mathbf{Pr}\left[\mathsf{AdvCond}(\alpha, X, Y) = \mathsf{AdvCond}(\alpha', X', Y')\right] \leq \varepsilon.$$

The main result of this section is the following lemma. Note that the lemma makes use of the iterated logarithm function. Recall that for an integer $c \geq 0$, the function $\log^{(c)}(x)$ stands for the c-iterated $\log(\cdot)$ function. That is, $\log^{(0)}(x) = x$, and for an integer $c > 0$, $\log^{(c)}(x) = \log(\log^{(c-1)}(x))$. All of our logarithms are taken base 2.

Lemma 9.2. *There exists a universal constant* $c_{\mathsf{AC}} \geq 1$ *such that the following holds. For all integers* $n, a_{\mathsf{in}}, c \geq 1$*, and for any* $\varepsilon > 0$*, there exists an efficiently-computable* (k,ε)*-advice condenser*

$$\mathsf{AdvCond}_c\colon \{0,1\}^{a_{\mathsf{in}}} \times \{0,1\}^{n} \times \{0,1\}^{d} \to \{0,1\}^{a_{\mathsf{out}}},$$

with

$$\begin{aligned} a_{\mathsf{out}} &= \log^{(c)}(a_{\mathsf{in}}) + c_{\mathsf{AC}}^{c} \cdot \log(1/\varepsilon), \\ d &= c_{\mathsf{AC}}^{c} \cdot \log(a_{\mathsf{in}} n/\varepsilon), \\ k &= c_{\mathsf{AC}}^{c} \cdot \log(a_{\mathsf{in}} n/\varepsilon). \end{aligned}$$

For the proof of Lemma 9.2 we start by constructing what we call the base advice condenser, which will be used in our final construction. Technically, this function does not have a third input, though we still refer to it as an advice condenser. More precisely, naturally, we call it a $(0,\varepsilon)$-advice condenser as one may think of the third source as being some fixed constant string, having no entropy.

Lemma 9.3. *For any integer* a_{in} *and for any* $\varepsilon > 0$*, there exists an efficiently-computable* $(0,\varepsilon)$*-advice condenser*

$$\mathsf{BaseAdvCond}\colon \{0,1\}^{a_{\mathsf{in}}} \times \{0,1\}^{d} \to \{0,1\}^{a_{\mathsf{out}}},$$

with

$$\begin{aligned} d &= \log(a_{\mathsf{in}}/\varepsilon) + 2, \\ a_{\mathsf{out}} &= \log a_{\mathsf{in}} + 3\log(1/\varepsilon) + 7, \end{aligned}$$

Proof of Lemma 9.3. Set q to be the least even power of two that is larger than $(2/\varepsilon+1)^2$. Identify $\{0,1\}^{a_{\mathsf{in}}}$ with some arbitrary subset of $\mathbb{F}_q^{a_{\mathsf{in}}}$. Let $\mathsf{ECC}\colon \mathbb{F}_q^{a_{\mathsf{in}}} \to \mathbb{F}_q^b$ be the error correcting code that is given by Theorem 5.10 set with relative distance $\delta = 1-\varepsilon$. By Theorem 5.10, such an explicit code exists with rate $\rho \geq \varepsilon/2$, and so $b \leq 2a_{\mathsf{in}}/\varepsilon$.

For $\alpha \in \{0,1\}^{a_{\mathsf{in}}}$ and $y \in \{0,1\}^d$, define

$$\mathsf{BaseAdvCond}(\alpha, y) = \mathsf{ECC}(\alpha)_y \circ y.$$

Note that d was chosen large enough so as to be used as an index for entries in the codeword $\mathsf{ECC}(\alpha)$. Furthermore, as $q \leq 4 \cdot (2/\varepsilon+1)^2 \leq 32/\varepsilon^2$, the output length of $\mathsf{BaseAdvCond}$ is

$$|\mathsf{ECC}(\alpha)_y| + |y| \leq \log q + \log b \leq \log a_{\mathsf{in}} + 3\log(1/\varepsilon) + 7 = m.$$

Further, $d \leq \log b \leq \log(a_{\mathsf{in}}/\varepsilon) + 2$, as stated. Now,

$$\begin{aligned}\mathbf{Pr}\left[\mathsf{BaseAdvCond}(\alpha, Y) = \mathsf{BaseAdvCond}(\alpha', Y')\right] &= \mathbf{Pr}\left[\mathsf{ECC}(\alpha)_Y \circ Y = \mathsf{ECC}(\alpha')_{Y'} \circ Y'\right] \\ &= \sum_y \mathbf{Pr}\left[\mathsf{ECC}(\alpha)_y \circ y = \mathsf{ECC}(\alpha')_{Y'_y} \circ Y'_y\right],\end{aligned}$$

where Y'_y denotes the random variable $Y' \mid (Y = y)$. Note that for every y,

$$\begin{aligned}\mathbf{Pr}\left[\mathsf{ECC}(\alpha)_y \circ y = \mathsf{ECC}(\alpha')_{Y'_y} \circ Y'_y\right] &\leq \mathbf{Pr}\left[\mathsf{ECC}(\alpha)_y \circ y = \mathsf{ECC}(\alpha')_{Y'_y} \circ Y'_y \mid Y'_y = y\right] \\ &= \mathbf{Pr}\left[\mathsf{ECC}(\alpha)_y = \mathsf{ECC}(\alpha')_y\right].\end{aligned}$$

Thus,

$$\mathbf{Pr}\left[\mathsf{BaseAdvCond}(\alpha, Y) = \mathsf{BaseAdvCond}(\alpha', Y')\right] \leq \sum_y \mathbf{Pr}\left[\mathsf{ECC}(\alpha)_y = \mathsf{ECC}(\alpha')_y\right] \leq \varepsilon.$$

□

The proof of Lemma 9.2 readily follows by the following lemma.

Lemma 9.4. *Let*

$$\mathsf{AdvCond}_{\mathsf{in}}\colon \{0,1\}^a \times \{0,1\}^n \times \{0,1\}^{d_{\mathsf{in}}} \to \{0,1\}^{a_{\mathsf{in}}}$$

be an explicit $(k_{\mathsf{in}}, \varepsilon_{\mathsf{in}})$-advice condenser. Then, there exists an explicit $(k_{\mathsf{out}}, \varepsilon_{\mathsf{out}})$-advice condenser

$$\mathsf{AdvCond}_{\mathsf{out}}\colon \{0,1\}^a \times \{0,1\}^n \times \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}^{a_{\mathsf{out}}}$$

with

$$\begin{aligned}d_{\mathsf{out}} &= \max\left(10 \cdot (d_{\mathsf{in}} + \log(1/\varepsilon_{\mathsf{in}})),\ c_{\mathsf{Raz}} \cdot \log(n/\varepsilon_{\mathsf{in}})\right), \\ k_{\mathsf{out}} &= \max\left(k_{\mathsf{in}},\ 2a_{\mathsf{in}} + c'_{\mathsf{Raz}} d_{\mathsf{out}} + \log(1/\varepsilon_{\mathsf{in}}),\ 2a_{\mathsf{in}} + 2\log(a_{\mathsf{in}}) + 3\log(1/\varepsilon_{\mathsf{in}}) + 4\right), \\ a_{\mathsf{out}} &= \log a_{\mathsf{in}} + 3\log(1/\varepsilon_{\mathsf{in}}) + 7, \\ \varepsilon_{\mathsf{out}} &= O(\sqrt{\varepsilon_{\mathsf{in}}}),\end{aligned}$$

where the constant c_{Raz} in the definition of d_{out} is the constant that appears in the statement of Theorem 5.7.

Proof of Lemma 9.4. Recall that $d_{\mathsf{out}} \geq d_{\mathsf{in}}$. Given a d_{out}-bit string y, we denote its d_{in}-bit prefix by y_1. On top of $\mathsf{AdvCond}_{\mathsf{in}}$, for the construction of $\mathsf{AdvCond}_{\mathsf{out}}$ we make use of the following components.

Building blocks.

- Set $d_{\mathsf{base}} = \log(a_{\mathsf{in}}/\varepsilon_{\mathsf{in}}) + 2$, and let $\mathsf{BaseAdvCond}\colon \{0,1\}^{a_{\mathsf{in}}} \times \{0,1\}^{d_{\mathsf{base}}} \to \{0,1\}^{a_{\mathsf{out}}}$ be the $(0, \varepsilon_{\mathsf{in}})$-advice condenser that is given by Lemma 9.3. Note that both d_{base} and a_{out} were set according to Lemma 9.3.
- Let $\mathsf{Raz}\colon \{0,1\}^n \times \{0,1\}^{d_{\mathsf{out}}} \to \{0,1\}^{d_{\mathsf{base}}}$ be the $(2d_{\mathsf{base}}, \varepsilon_{\mathsf{in}})$-extractor with weak-seeds that is given by Theorem 5.7. Note that d_{out} is sufficiently large, as required by Theorem 5.7.

The construction. On input $\alpha \in \{0,1\}^a$, $x \in \{0,1\}^n$, and $y \in \{0,1\}^{d_{\mathsf{out}}}$, we define

$$\mathsf{AdvCond}_{\mathsf{out}}(\alpha, x, y) = \mathsf{BaseAdvCond}\left(\mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, x, y_1\right), \mathsf{Raz}\left(x, y\right)\right).$$

Analysis. Let X, X' be n-bit random variables such that $H_\infty(X) \geq k_{\mathsf{out}}$. Let Y, Y' be d_{out}-bit random variables such that Y is uniform. Assume that the joint distribution (X, X') is independent of the joint distribution (Y, Y'). Let α, α' be distinct, fixed, a-bit strings.

As $k_{\mathsf{out}} \geq k_{\mathsf{in}}$ and since $\mathsf{AdvCB}_{\mathsf{in}}$ is a $(k_{\mathsf{in}}, \varepsilon_{\mathsf{in}})$-advice condenser, we have that

$$\mathbf{Pr}\left[\mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, X, Y_1\right) = \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha', X', Y_1'\right)\right] \leq \varepsilon_{\mathsf{in}}.$$

Thus, by Markov's inequality, except with probability $\sqrt{\varepsilon_{\mathsf{in}}}$ over $(y_1, y_1') \sim (Y_1, Y_1')$, it holds that

$$\mathbf{Pr}\left[\mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, X, y_1\right) = \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha', X', y_1'\right)\right] \leq \sqrt{\varepsilon_{\mathsf{in}}}.$$

We further condition on the fixings of $\beta \sim \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, X, y_1\right)$ and $\beta' \sim \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha', X', y_1'\right)$. Note that conditioned on the fixings of Y_1, Y_1', the random variables $\mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, X, Y_1\right)$, $\mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, X', Y_1'\right)$ are deterministic functions of (X, X') and so this further conditioning does not introduce dependencies between (X, X') and (Y, Y'). Set

$$\mathcal{H} = \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, X, Y_1\right), \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha', X', Y_1'\right), Y_1, Y_1'.$$

By Lemma 5.2, Lemma 5.4, and by our choice of $d_{\mathsf{out}}, d_{\mathsf{in}}$,

$$\begin{aligned}\widetilde{H}_\infty\left(Y \mid \mathcal{H}\right) &= \widetilde{H}_\infty\left(Y \mid Y_1, Y_1'\right)\\ &\geq d_{\mathsf{out}} - 2d_{\mathsf{in}}\\ &\geq 0.8d_{\mathsf{out}}\\ &\geq 0.7d_{\mathsf{out}} + \log(1/\varepsilon_{\mathsf{in}}).\end{aligned}$$

Thus, by Lemma 5.3, except with probability $\varepsilon_{\mathsf{in}}$ over the fixings of $\mathcal{H}$, the random variable Y has min-entropy rate 0.7. Furthermore, by Lemma 5.2 and Lemma 5.4,

$$\begin{aligned}\widetilde{H}_\infty\left(X \mid \mathcal{H}\right) &= \widetilde{H}_\infty\left(X \mid \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha, X, Y_1\right), \mathsf{AdvCond}_{\mathsf{in}}\left(\alpha', X', Y_1'\right)\right)\\ &\geq k_{\mathsf{out}} - 2a_{\mathsf{in}}\\ &\geq \max\left(2d_{\mathsf{base}}, c'_{\mathsf{Raz}}d_{\mathsf{out}}\right) + \log(1/\varepsilon_{\mathsf{in}}).\end{aligned}$$

This, together with the fact that $d_{\mathsf{out}} \geq c_{\mathsf{Raz}} \cdot \log(n/\varepsilon)$, implies that

$$\left(\mathsf{Raz}(X,Y), \mathcal{H}\right) \approx_{2\varepsilon_{\mathsf{in}}} \left(U, \mathcal{H}\right).$$

By Markov's inequality, except with probability $O(\sqrt{\varepsilon_{\mathsf{in}}})$ over the fixing of $\mathcal{H}$, it holds that $\mathsf{Raz}(X,Y) \approx_{\sqrt{\varepsilon_{\mathsf{in}}}} U$.

To summarize, by the union bound, except with probability $O(\sqrt{\varepsilon_{\mathsf{in}}})$ over the fixing of $\mathcal{H}$, we have that $\beta \neq \beta'$ and that $\mathsf{Raz}(X,Y)$ is $O(\sqrt{\varepsilon_{\mathsf{in}}})$-close to uniform. Thus, by Lemma 9.3,

$$\mathbf{Pr}\left[\mathsf{AdvCond}_{\mathsf{out}}\left(\alpha, X, Y\right) = \mathsf{AdvCond}_{\mathsf{out}}\left(\alpha', X', Y'\right)\right] = O\left(\sqrt{\varepsilon_{\mathsf{in}}}\right).$$

This concludes the proof. □

10 Proof of Theorem 2.1

After developing all the required components in previous sections, in this section we finally turn to prove Theorem 2.1. The proof of Theorem 2.1 follows by the following lemma applied with $c = 4$. Note that we prove the theorem with $k = 3m$. This can be easily relaxed to $k = (2+\alpha)m$ for any constant $\alpha > 0$. We state the result as is for simplicity.

Lemma 10.1. *There exists a universal constant $c' > 1$ such that the following holds. For all integers n, a, for any $\varepsilon > 0$, and for any constant integer $c \geq 1$, there exists an efficiently-computable (k, ε)-strong correlation breaker with advice*

$$\mathsf{AdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

with

$$\ell = c' \log(an) + (c')^{\sqrt{\log\left(\log^{(c)}(a)+\log(1/\varepsilon)\right)}} \cdot \left(\log^{(c)}(a) + \log(1/\varepsilon)\right),$$
$$k = 3m + c'\ell.$$

Proof of Lemma 10.1. Let $c_{\mathsf{GUV}}, c_{\mathsf{Raz}}, c_{\mathsf{SU}}, c_{\mathsf{AC}}$ be the constants that appear in the statements of Theorem 5.6, Theorem 5.7, Lemma 8.3, and Lemma 9.2, respectively. Set

$$\begin{aligned}
a' &= \log^{(c)}(a) + c_{\mathsf{AC}}^{c} \cdot \log(1/\varepsilon),\\
m' &= c_{\mathsf{GUV}} \cdot \log(n/\varepsilon),\\
\ell' &= (c'_{\mathsf{SU}})^{\sqrt{\log a'}} \cdot \log(\ell/\varepsilon),\\
\ell_1 &= c_{\mathsf{AC}}^{c} \cdot \log(an/\varepsilon),\\
\ell_2 &= \max\left(10\ell_1 + 10\log(1/\varepsilon), c_{\mathsf{Raz}} \cdot \log(n/\varepsilon)\right).
\end{aligned}$$

Note that our hypothesis of ℓ implies that $\ell \geq \ell_2$. Given an ℓ-bit string y, we denote its length ℓ_1 prefix by y_1 and its length ℓ_2 prefix by y_2. For the proof of Lemma 10.1 we make use of the following building blocks.

Building blocks.

- Let $\mathsf{AdvCB}' \colon \{0,1\}^{\ell} \times \{0,1\}^{\ell'} \times \{0,1\}^{a'} \to \{0,1\}^{m'}$ be the $(\ell/2, \varepsilon)$-strong correlation breaker with advice that is given by Lemma 8.3. Note that ℓ' was set according to Lemma 8.3. Moreover, the min-entropy $\ell/2$ is large enough as required by Lemma 8.3 as it can be easily verified that, by our hypothesis, $\ell = \Omega(\ell')$.
- Let $\mathsf{Raz} \colon \{0,1\}^{n} \times \{0,1\}^{\ell_2} \to \{0,1\}^{\ell'}$ be the $(2\ell', \varepsilon)$-extractors with weak-seeds that is given by Theorem 5.7. Note that ℓ_2 was set large enough as required by Theorem 5.7.
- Let $\mathsf{AdvCond} \colon \{0,1\}^{a} \times \{0,1\}^{n} \times \{0,1\}^{\ell_1} \to \{0,1\}^{a'}$ be the (k, ε)-advice condenser that is given by Lemma 9.2. Note that ℓ_1 is large enough as required by Lemma 9.2. Furthermore, a' was set according to Lemma 9.2.
- Let $\mathsf{Ext} \colon \{0,1\}^{n} \times \{0,1\}^{m'} \to \{0,1\}^{m}$ be the $(2m, \varepsilon)$-strong seeded extractor that is given by Theorem 5.6. Note that m' has sufficient length as required by Theorem 5.6.

The construction. On input $x \in \{0,1\}^{n}$, $y \in \{0,1\}^{\ell}$, and $\alpha \in \{0,1\}^{a}$, we define

$$\mathsf{AdvCB}(x, y, \alpha) = \mathsf{Ext}\left(x, z\right),$$

where

$$z = \mathsf{AdvCB}'\left(y, \mathsf{Raz}(x, y_2), \mathsf{AdvCond}(\alpha, x, y_1)\right).$$

Analysis. Let X, X' be n-bit random variables such that $H_\infty(X) \geq k$. Let Y, Y' be ℓ-bit random variables such that Y is uniform. We further assume that the joint distribution (X, X') is independent of the joint distribution (Y, Y'). Let α, α' be two distinct, fixed, a-bit strings.

As $k \geq c_{\mathsf{AC}} \cdot \log a + c^c_{\mathsf{AC}} \cdot \log(1/\varepsilon)$, we can apply Lemma 9.2 and conclude that

$$\mathbf{Pr}\left[\mathsf{AdvCond}(\alpha, X, Y_1) = \mathsf{AdvCond}(\alpha', X', Y_1')\right] \leq \varepsilon.$$

By Markov's inequality, except with probability $\sqrt{\varepsilon}$ over the fixings $(y_1, y_1') \sim (Y_1, Y_1')$, it holds that

$$\mathbf{Pr}\left[\mathsf{AdvCond}(\alpha, X, y_1) = \mathsf{AdvCond}(\alpha', X', y_1')\right] \leq \sqrt{\varepsilon}.$$

We now further condition on the fixings of $\beta = \mathsf{AdvCond}(\alpha, X, y_1)$, $\beta' = \mathsf{AdvCond}(\alpha', X', y_1')$. Note that conditioned on the fixings of Y_1, Y_1', the random variables $\mathsf{AdvCond}(\alpha, X, Y_1)$, $\mathsf{AdvCond}(\alpha', X', Y_1')$ are deterministic functions of (X, X'), and so the latter conditioning does not introduce dependencies between (X, X') and (Y, Y').

Set

$$\mathcal{H} = Y_1, Y_1', \mathsf{AdvCond}(\alpha, X, Y_1), \mathsf{AdvCond}(\alpha', X', Y_1').$$

By Lemma 5.4 and Lemma 5.2,

$$\begin{aligned} \widetilde{H}_\infty\left(Y_2 \mid \mathcal{H}\right) &= \widetilde{H}_\infty\left(Y_2 \mid Y_1, Y_1'\right) \\ &\geq \ell_2 - 2\ell_1 \\ &\geq 0.8\ell_2. \end{aligned} \tag{10.1}$$

Hence, by Lemma 5.3, except with probability ε over the fixings of (Y_1, Y_1'), it holds that Y_2 has min-entropy rate 0.7. Moreover, as $\mathsf{AdvCond}$ has output length a', Lemma 5.2 together with Lemma 5.4 imply that

$$\begin{aligned} \widetilde{H}_\infty\left(X \mid \mathcal{H}\right) &= \widetilde{H}_\infty\left(X \mid \mathsf{AdvCond}(\alpha, X, Y_1), \mathsf{AdvCond}(\alpha', X', Y_1')\right) \\ &\geq k - 2a' \\ &\geq \max\left(2\ell', c'_{\mathsf{Raz}}\ell_2\right) + \log(1/\varepsilon). \end{aligned} \tag{10.2}$$

By Equation (10.1), Equation (10.2), and Theorem 5.7, and since X and Y_2 are independent conditined on $\mathcal{H}$,

$$\left(\mathsf{Raz}(X, Y_2), Y_2, \mathcal{H}\right) \approx_{O(\varepsilon)} \left(U, Y_2, \mathcal{H}\right).$$

As conditioned on $Y_2, \mathcal{H}$, the random variables $\mathsf{Raz}(X, Y_2)$ and Y_2' are independent, Lemma 5.1 implies that

$$\left(\mathsf{Raz}(X, Y_2), Y_2', Y_2, \mathcal{H}\right) \approx_{O(\varepsilon)} \left(U, Y_2', Y_2, \mathcal{H}\right).$$

Now, by Lemma 5.2,

$$\begin{aligned}\widetilde{H}_\infty\left(Y \mid Y_2', Y_2, \mathcal{H}\right) &= \widetilde{H}_\infty\left(Y \mid Y_2', Y_2\right) \\ &\geq \ell - 2\ell_2 \\ &\geq \ell/2 + \log(1/\varepsilon).\end{aligned}$$

Thus, by Lemma 5.3, except with probability ε over the fixings of $Y_2', Y_2, \mathcal{H}$, the random variable Y has min-entropy $\ell/2$.

To recap, by the union bound, except with probability $O(\sqrt{\varepsilon})$ over the fixings of $\mathcal{H}, Y_2, Y_2'$, the following holds:

- β, β' are distinct, fixed, strings.
- $\mathsf{Raz}(X, Y_2)$ is $O(\sqrt{\varepsilon})$-close to uniform.
- $H_\infty(Y) \geq \ell/2$.
- The joint distribution (Y, Y') is independent of the joint distribution of $\mathsf{Raz}(X, Y_2)$ and $\mathsf{Raz}(X', Y_2')$.

Hence, as $\ell/2 \geq c''_{\mathsf{SU}}\ell' + 3m'$, we are in a position to apply Lemma 8.3, which implies that

$$(Z, \mathcal{H}') \approx_{O(\sqrt{\varepsilon})} (U, \mathcal{H}'),$$

where

$$\begin{aligned}Z &= \mathsf{AdvCB}'\left(Y, \mathsf{Raz}(X, Y_2), \mathsf{AdvCond}(\alpha, X, Y_1)\right), \\ Z' &= \mathsf{AdvCB}'\left(Y', \mathsf{Raz}(X', Y_2'), \mathsf{AdvCond}(\alpha', X', Y_1')\right), \\ \mathcal{H}' &= Z', \mathsf{Raz}(X, Y_2), \mathsf{Raz}(X', Y_2'), Y_2', Y_2, \mathcal{H}.\end{aligned}$$

Conditioned on $\mathcal{H}'$, the random variable Z is a deterministic function of Y whereas $\mathsf{Ext}(X', Z')$ is independent of Y. Thus, by Lemma 5.1,

$$(Z, \mathsf{Ext}(X', Z'), \mathcal{H}') \approx_{O(\sqrt{\varepsilon})} (U, \mathsf{Ext}(X', Z'), \mathcal{H}').$$

By applying Lemma 5.2, Lemma 5.4, together with Equation (10.2), we have that

$$\begin{aligned}\widetilde{H}_\infty\left(X \mid \mathsf{Ext}(X', Z'), \mathcal{H}'\right) &\geq \widetilde{H}_\infty\left(X \mid \mathcal{H}\right) - m - 2\ell' \\ &\geq k - 2a' - m - 2\ell' \\ &\geq 2m + \log(1/\varepsilon).\end{aligned}$$

Thus, by Theorem 5.6,

$$(\mathsf{Ext}(X, Z), Z, \mathsf{Ext}(X', Z'), \mathcal{H}') \approx_{O(\sqrt{\varepsilon})} (U, Z, \mathsf{Ext}(X', Z'), \mathcal{H}').$$

To conclude the proof, recall that $\mathsf{Ext}(X,Z) = \mathsf{AdvCB}(X,Y,\alpha)$, $\mathsf{Ext}(X',Z') = \mathsf{AdvCB}(X',Y',\alpha')$, and that conditioned on $Z, \mathsf{Ext}(X',Z'), \mathcal{H}'$, the random variable $\mathsf{Ext}(X,Z)$ is independent of Y, Y'. Thus,

$$(\mathsf{AdvCB}(X,Y,\alpha), \mathsf{AdvCB}(X',Y',\alpha'), Y, Y') \approx_{O(\sqrt{\varepsilon})} (U, \mathsf{AdvCB}(X',Y',\alpha'), Y, Y')\,.$$

As usual, the error can be reduced to ε by taking primitives with a lower error guarantee. This has no affect on the theorem statement. □

11 A 5-Source Extractor for Near Logarithmic Min-Entropy

In this section we prove Theorem 2.2. The proof relies on a generalization of correlation breakers with advice.

Definition 11.1 (t-Correlation breakers with advice)**.** *A function*

$$\mathsf{AdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$$

is called a (k,ε) t-correlation breaker with advice *if the following holds. Let Y be a random variable that is uniformly distribution over ℓ-bit strings, and let $Y_1,\ldots,Y_t$ be ℓ-bit random variables that are arbitrarily correlated with Y. Let X be an (n,k)-source that is arbitrarily correlated with the n-bit random variables $X_1,\ldots,X_t$. Assume that the joint distribution of $X, X_1,\ldots,X_t$ is independent of the joint distribution of $Y, Y_1,\ldots,Y_t$. Then, for any a-bit strings $\alpha, \alpha_1,\ldots,\alpha_t$ such that $\alpha \notin \{\alpha_i\}_{i=1}^t$,*

$$\left(\mathsf{AdvCB}\left(X,Y,\alpha\right), \left\{\mathsf{AdvCB}\left(X_i,Y_i,\alpha_i\right)\right\}_{i=1}^t\right) \approx_\varepsilon \left(U, \left\{\mathsf{AdvCB}\left(X_i,Y_i,\alpha_i\right)\right\}_{i=1}^t\right).$$

The proof of Theorem 2.1 can be extended in a straightforward manner to prove the following theorem and so we omit the details.

Theorem 11.2. *For any constant integer $t \geq 1$ there exist constants $c_{\mathsf{ACB}}, c'_{\mathsf{ACB}} \geq 1$ such that the following holds. For all integers n, m, a, and for any $\varepsilon > 0$ such that $a < 2^{n/\varepsilon}$, there exists an explicit (k,ε) t-correlation breaker with advice $\mathsf{AdvCB}\colon \{0,1\}^n \times \{0,1\}^\ell \times \{0,1\}^a \to \{0,1\}^m$, with*

$$\begin{aligned} \ell &= c_{\mathsf{ACB}} \cdot \log(an) + \log(1/\varepsilon) \cdot c_{\mathsf{ACB}}^{\sqrt{\log\log(1/\varepsilon)}}, \\ k &= c'_{\mathsf{ACB}} \cdot (m+\ell). \end{aligned}$$

The proof of Theorem 2.2 makes use of a fair number of components and results from the literature. We begin by introducing them, starting with seeded condensers.

Definition 11.3 (Seeded condensers)**.** *A function* $\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is said to be a* $k \to_\varepsilon k'$ condenser *if for any* (n,k)*-source* X *and for any independent random variable* S *that is uniformly distributed over* d*-bit strings, it holds that* $\mathsf{Cond}(X,S)$ *is* ε*-close to a random variable with min-entropy* k'*. The function* Cond *is called a* lossless condenser *if* $k' = k + d$*.*

Theorem 11.4 ([GUV09])**.** *For any constant* $\tau > 0$ *(*τ *can be taken to be larger than* 1*), all integers* n, k *such that* $k \leq n$*, and for any* $\varepsilon > 0$*, there exists an efficiently-computable* $k \to_\varepsilon k + d$ *(lossless) condenser* $\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *having seed length* $d = (1 + 1/\tau)\log(nk/\varepsilon) + O(1)$ *and* $m = (1+\tau)k + 2d$ *output bits.*

We make use of the following lemma that, informally speaking, states that any seeded condenser is also strong.

Lemma 11.5 ([Li11a, CS16])**.** *Let* $\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *be a* $k \to_\varepsilon k'$ *condenser. Let* X *be an* (n,k)*-source and let* S *be an independent random variable that is uniformly distributed over* d*-bit strings. Then, for any* $\delta > 0$*, with probability* $1 - \delta$ *over* $s \sim S$ *it holds that* $\mathsf{Cond}(X,s)$ *is* $(2\varepsilon/\delta)$*-close to having min-entropy* $k' - d - \log(2/\delta)$*.*

Barak *et al.* [BIW06] proved a useful lemma which roughly states that taking the bitwise XOR of t independent n-bit random variables, each is ε-close to uniform, results in a random variable that is ε^t-close to uniform. We need to generalize this lemma.[10]

Lemma 11.6. *Let* $X_1, \ldots, X_m$ *be independent* n*-bit random variables, each is* ε*-close to having min-entropy* k*, where* $\varepsilon \leq \frac{1}{4}$*. Then,* $X_1 \oplus \cdots \oplus X_m$ *is* $(2\varepsilon)^m$*-close to having min-entropy* $k - 2$*.*

The proof of Lemma 11.6 can be found in Appendix A.

We also make use of the following lemma.

Lemma 11.7 ([Li12b])**.** *Let* X, X' *be random variables with a common range such that* $\mathsf{SD}(X,X') \leq \varepsilon$*. Let* (X,Y) *be a joint distribution. Then, there exists a joint distribution* (X',Y) *such that* $\mathsf{SD}((X,Y),(X',Y)) \leq \varepsilon$*.*

With these components and results from the literature, we are ready to prove the following lemma, which we later use for the proof of Theorem 2.2.

Lemma 11.8. *For all integers* n, ℓ*, there exists an efficiently-computable function*

$$f\colon (\{0,1\}^n)^3 \to \{0,1\}^{r \times \ell},$$

[10] The original version of this paper had a flaw in the proof of this lemma that was communicated to us by Dori Medini. In this version we fix the flaw (obtaining slightly weaker parameters) joint with Alon Frydberg.

with $r = \mathrm{poly}(n)$, having the following property. For any three independent (n,k)-sources X_1, X_2, X_3, with $k = \Omega(\ell + \log n)$, there exists $B \subseteq [r]$ of size $|B| \leq 3r^{0.4}$, and a random variable M in the form of an $r \times \ell$ matrix, such that the following holds:

- *$f(X_1, X_2, X_3)$ is $8r^{-0.05}$-close to M.*
- *For any $i \in [r] \setminus B$, the random variable M_i has min-entropy rate $1/300$.*

Proof. Set $\tau = 100$ and $\varepsilon = n^{-100}$. Let $r = 2^d$ where $d = (1 + 1/\tau)\log(nk/\varepsilon) + O(1)$ is the seed length of the seeded condenser $\mathsf{Cond}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ that is given by Theorem 11.4 when applied with τ, ε as chosen above to (n,k)-sources. Let $\ell = (1+\tau)k + 2d$ be the output length of the condenser from Theorem 11.4 when applied with these parameters. Set $\delta = r^{-0.6}$.

The construction. On input $x_1, x_2, x_3 \in \{0,1\}^n$, for each $i \in [r]$ we define

$$f(x_1, x_2, x_3)_i = \mathsf{Cond}(x_1, i) \oplus \mathsf{Cond}(x_2, i) \oplus \mathsf{Cond}(x_3, i),$$

where we identify $[r]$ with $\{0,1\}^d$ (and so we think of i as a seed for Cond).

Analysis. Let X_1, X_2, X_3 be independent (n,k)-sources. Let $j \in [3]$. By Theorem 11.4 and by Lemma 11.5, there exists a set $B_j \subset [r]$, of size $|B_j| \leq \delta r = r^{0.4}$ such that for any $i \notin B_j$ it holds that $\mathsf{Cond}(X_j, i)$ is $(2\varepsilon/\delta)$-close to having min-entropy $k - d - \log(2/\delta) \geq k/2$, where the last inequality follows by taking k to be larger than a large enough constant multiple of $\log n$. One can verify that by our choice of parameters $2\varepsilon/\delta < r^{-0.35}$. Thus, for any $i \notin B_j$ it holds that $\mathsf{Cond}(X_j, i)$ is $r^{-0.35}$-close to having min-entropy rate $k/(2\ell) \geq 1/200$.

Set $B = B_1 \cup B_2 \cup B_3$, and note that $|B| \leq 3r^{0.4}$. By Lemma 11.6, for each $i \notin B$, we have that $f(X_1, X_2, X_3)_i$ is $(2r^{-0.35})^3 = 8r^{-1.05}$-close to having min-entropy rate $\frac{1}{200} - \frac{2}{\ell} \geq \frac{1}{300}$. Thus, by Lemma 11.7, applied one by one to all rows in $[r] \setminus B$, it holds that $f(X_1, X_2, X_3)$ is $r \cdot 8r^{-1.05} = 8r^{-0.05}$-close to a random variable M in the form of an $r \times \ell$ matrix such that for any $i \notin B$ it holds that M_i has min-entropy rate $1/300$. This concludes the proof. □

For the proof of Theorem 2.2 we make use of the following lemmas.

Lemma 11.9 ([Vio14])**.** *Let $Y_1, \ldots, Y_r$ be a sequence of $\{0,1\}$ random variables. Let $\alpha > 0$ be some constant. Assume that there exists $B \subseteq [r]$, with $|B| \leq r^{1/2-\alpha}$, such that the random variables $\{Y_i \mid i \in [r] \setminus B\}$ are t-wise independent and uniform. Then,*

$$\mathrm{bias}\left(\mathsf{Majority}(Y_1, \ldots, Y_r)\right) = O\left(\frac{\log t}{\sqrt{t}} + r^{-\alpha}\right). \tag{11.1}$$

Lemma 11.10 ([AGM03])**.** *Let $X_1, \ldots, X_n$ be $\{0,1\}$ random variables. Assume that for any $\emptyset \neq I \subseteq [n]$, with size $|I| \leq t$, the joint distribution of $\{X_i\}_{i \in I}$ is ε-close to uniform. Then, $X_1, \ldots, X_n$ is $(n^t \cdot \varepsilon)$-close to a t-wise independent distribution.*

Theorem 11.11 ([Raz05, BKS+05, Zuc07])**.** *For every constant $\delta > 0$, there exists an efficiently-computable function $\mathsf{Cond}\colon \{0,1\}^n \to \{0,1\}^{r \times \ell}$, with $r = O(1)$ and $\ell = \Theta(n)$ such that the following holds. If X is an $(n, \delta n)$-source, then $\mathsf{Cond}(X)$ is $2^{-\Omega(n)}$-close to a convex combination of distributions, each of which has some row with min-entropy rate* 0.9.

We are now ready to prove Theorem 2.2.

Proof of Theorem 2.2. For a large enough constant c to be fixed later on, set $\ell = \log n \cdot c^{\sqrt{\log \log n}}$. For the construction of $\mathsf{5Ext}$ we make use of the following building blocks.

Building blocks.

- Let $f\colon (\{0,1\}^n)^3 \to \{0,1\}^{r \times \ell}$ be the function that is given by Lemma 11.8. Recall that by Lemma 11.8, $r = \mathrm{poly}(n)$.
- Let $\mathsf{Cond}\colon \{0,1\}^\ell \to \{0,1\}^{r' \times \ell'}$ be the somewhere-condenser that is given by Theorem 11.11 set with $\delta = 1/300$. By Theorem 11.11, $r' = O(1)$ and $\ell' = \Theta(\ell)$.
- Set $\ell'' = (c'_{\mathsf{Raz}}/2) \cdot \ell'$. Let $\mathsf{Raz}\colon \{0,1\}^n \times \{0,1\}^{\ell'} \to \{0,1\}^{\ell''}$ be the $(2\ell'', r^{-2})$-extractors with weak-seeds that is given by Theorem 5.7. Note that by our choice of ℓ'', such an explicit extractor does indeed exist.
- Let $t \geq 1$ be a constant whose value we choose later on. Set $a = \log(rr')$ and let $\mathsf{AdvCB}\colon \{0,1\}^n \times \{0,1\}^{\ell''} \times \{0,1\}^a \to \{0,1\}$ be the (k, ε) (tr')-correlation breaker with advice that is given by Theorem 11.2 set with error guarantee $\varepsilon = r^{-(t+1)}$. Note that by setting the constant c in the definition of ℓ to be large enough, ℓ'' is large enough as required by Theorem 11.2.

The construction. On input $x_1, \ldots, x_5 \in \{0,1\}^n$, we define $\mathsf{5Ext}(x_1, \ldots, x_5)$ as follows:

1. Let m be the $r \times \ell$ matrix whose i'th row is given by $m_i = f(x_1, x_2, x_3)_i$.
2. For each $i \in [r]$ define the $r' \times \ell'$ matrix $m^i = \mathsf{Cond}(m_i)$.
3. For each $i \in [r]$ define the $r' \times \ell''$ matrix $(m')^i$ whose j'th row is given by $(m')^i_j = \mathsf{Raz}(x_4, m^i_j)$.
4. For each $i \in [r]$ let $v^i \in \{0,1\}^{r'}$ be the vector whose j'th entry is defined as $(v^i)_j = \mathsf{AdvCB}(x_5, (m')^i_j, (i,j))$, where we think of (i,j) as a $\log(rr')$-bit string.

5. For each $i \in [r]$ define $y_i = \oplus_{j=1}^{r'} (v^i)_j$.

6. Finally, we define $\mathsf{5Ext}(x_1, \ldots, x_5) = \mathsf{Majority}(y_1, \ldots, y_r)$.

Analysis. Let $X_1, \ldots, X_5$ be independent (n, k)-sources, with $k = c''\ell$, where c'' is some large enough constant. By Lemma 11.8, which is applicable by our assumption on k, there exists a set $B \subseteq [r]$, of size $|B| \leq 3r^{0.4}$ and a random variable M in the form of an $r \times \ell$ matrix, such that

- $f(X_1, X_2, X_3)$ is $8r^{-0.05}$-close to M.
- For any $i \notin B$, the random variable M_i has min-entropy rate $1/300$.

From here on we assume that $f(X_1, X_2, X_3)$ itself has the second the property above, namely, for any $i \notin B$, the random variable $f(X_1, X_2, X_3)_i$ has min-entropy rate $1/300$. We do so for the ease of reading, and aggregate the negligible expression $8r^{-0.05}$ to the total error.

For $i \in [r]$, let $M^i = \mathsf{Cond}(f(X_1, X_2, X_3)_i)$. Fix $i \in [r] \setminus B$. By Theorem 11.11, M^i is $2^{-\Omega(\ell)}$-close to a convex combination of distributions over $r' \times \ell'$ matrices, each of which has some row with min-entropy rate 0.9. For ease of reading, we make the simplifying assumption that there exists $g(i) \in [r']$ such that $(M^i)_{g(i)}$ is $2^{-\Omega(\ell)}$-close to a having min-entropy rate 0.9. A more formal treatment would require as to introduce cumbersome notation for the different summands of the convex combination, which we prefer to avoid.

By Lemma 11.7 applied repeatedly to all $i \in [r] \setminus B$, the sequence of matrices $M^1, \ldots, M^r$ is $(r \cdot 2^{-\Omega(\ell)})$-close to a second sequence of matrices $\bar{M}^1, \ldots, \bar{M}^r$ such that for any $i \in [r] \setminus B$ it holds that $(\bar{M}^i)_{g(i)}$ has min-entropy rate 0.9. To avoid introducing further notations, we abuse notation and identify M^i with $\bar{M}^i$. This simplifying assumption can be made at the cost of aggregating $r \cdot 2^{-\Omega(\ell)} = o(1)$ to the total error.

For $i \in [r]$ and $j \in [r']$, let $(M')^i_j = \mathsf{Raz}(X_4, M^i_j)$. Let $i \in [r] \setminus B$. As $M^i_{g(i)}$ has min-entropy rate 0.9 and since $H_\infty(X_4) = k \geq 2\ell''$, Theorem 5.7 implies that for any $i \in [r] \setminus B$ it holds that $(M')^i_{g(i)} \approx_{r^{-2}} U$. Thus, by Lemma 11.7, up to a negligible error of r^{-1} which we aggregate, we have that for all $i \in [r] \setminus B$ it holds that $(M')^i_{g(i)}$ is uniform.

For $i \in [r]$ and $j \in [r']$, let $(V^i)_j = \mathsf{AdvCB}(X_5, (M')^i_j, (i, j))$. By Theorem 11.2 and as AdvCB is a (tr')-correlation breaker with advice set with error guarantee $r^{-(t+1)}$, we have that for all $1 \leq v \leq t$ and for any distinct $i_1, \ldots, i_v \in [r] \setminus B$,

$$\left((V^{i_1})_{g(i_1)}, \ldots, (V^{i_v})_{g(i_v)}, \left\{(V^{i_1})_j\right\}_{j \in [r'] \setminus g(i_1)}, \ldots, \left\{(V^{i_v})_j\right\}_{j \in [r'] \setminus g(i_v)}\right) \approx_{O(r^{-(t+1)})}$$
$$\left(U_v, \left\{(V^{i_1})_j\right\}_{j \in [r'] \setminus g(i_1)}, \ldots, \left\{(V^{i_v})_j\right\}_{j \in [r'] \setminus g(i_v)}\right).$$

Hence,

$$(Y_{i_1}, \ldots, Y_{i_v}) \approx_{O(r^{-(t+1)})} U.$$

Therefore, by Lemma 11.10, the sequence of random variables $\{Y_i\}_{i\in[r]\setminus B}$ is $O(r^{-1})$-close to a sequence of uniform and t-wise independent random variables. As $|B| \leq 3r^{0.4}$, Lemma 11.9 completes the proof.

□

12 Non-Malleable Extractors

In this section we prove the following theorem which readily implies Theorem 2.4. As mentioned, the proof follows by plugging-in our correlation breaker with advice, that is given by Theorem 2.1, to the [CGL15] framework for constructing non-malleable extractors applied together with the "switch" idea [Coh15b]. The theorem is stated with extractors that have $k/4$ output bits. This can be easily improved to $k/(2+\alpha)$ for any constant $\alpha > 0$.

Theorem 12.1. *There exist constants $c, c' \geq 1$ such that the following holds. For any integer n and for any $\varepsilon > 0$, there exists an efficiently-computable (k,ε)-non-malleable extractor $\mathsf{nmExt}\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{k/4}$ with seed length $d = c\log n + \log(1/\varepsilon)\cdot c^{\sqrt{\log\log(1/\varepsilon)}}$ for any $k \geq c'd$.*

Proof. Let $c_{\mathsf{GUV}}, c_{\mathsf{Raz}}$ be the constants that are given by Theorem 5.6 and Theorem 5.7, respectively. Set

$$d_1 = c_{\mathsf{GUV}} \cdot \log(n/\varepsilon),$$
$$d_2 = \max\left(10d_1, c_{\mathsf{Raz}} \cdot \log(n/\varepsilon)\right).$$

For a d-bit string y, let y_1 denote the length d_1 prefix of y. Similarly, let y_2 denote the length d_2 prefix of y. We further assume that $d \geq 10d_2$. Note that this assumption can be met by taking the constant c large enough with respect to the constants $c_{\mathsf{GUV}}, c_{\mathsf{Raz}}$. For the construction of nmExt we make use of the following building blocks.

Building blocks.

- Let q be the least even prime power of 2 that is larger or equal than $5/\varepsilon^2$. Note that $q \leq 20/\varepsilon^2$. Let r be the least integer such that $q^r \geq d$. We identify $[d]$ with an arbitrary subset of $\mathbb{F}_q^r$. Set $v = 2r/\varepsilon$ and let $\mathsf{ECC}\colon \mathbb{F}_q^r \to \mathbb{F}_q^v$ be the error correcting code that is given by Theorem 5.10, set with relative distance $\delta = 1-\varepsilon$. By Theorem 5.10, an explicit code with these parameters (namely, relative distance $1-\varepsilon$, rate $2/\varepsilon$, and alphabet size $q \leq 20/\varepsilon^2$) exists.
- Let $\mathsf{Ext}\colon \{0,1\}^n \times \{0,1\}^{d_1} \to \{0,1\}^{\log v}$ be the $(2\log v, \varepsilon)$-strong seeded extractor that is given by Theorem 5.6. Note that d_1 was defined to be of sufficient length so as to be used as a seed for Ext. We identify the output of Ext as an element of $[v]$.

- Let $0 < \alpha < 1$ be a constant whose value we set later on. Let $\mathsf{Raz}\colon \{0,1\}^n \times \{0,1\}^{d_2} \to \{0,1\}^{\alpha d}$ be the $(2\alpha d, \varepsilon)$-extractor with weak-seeds that is given by Theorem 5.7. Note that d_2 was chosen large enough as required by Theorem 5.7.
- Set $a = d_1 + \log q$. Let $\mathsf{AdvCB}\colon \{0,1\}^d \times \{0,1\}^{\alpha d} \times \{0,1\}^a \to \{0,1\}^{d_1}$ be the (k,ε)-correlation breaker with advice that is given by Theorem 2.1.
- Let $\mathsf{Ext}'\colon \{0,1\}^n \times \{0,1\}^{d_1} \to \{0,1\}^{k/4}$ be the $(k/2,\varepsilon)$-strong seeded extractor that is given by Theorem 5.6. Note that d_1 is large enough as required by Theorem 5.6.

The construction. On input $x \in \{0,1\}^n$, $y \in \{0,1\}^d$, we define $\mathsf{nmExt}(x,y)$ as follows. First we compute the function

$$\mathsf{AdvGen}(x,y) = y_1 \circ \mathsf{ECC}(y)_{\mathsf{Ext}(x,y_1)}.$$

In the expression above, by $\mathsf{ECC}(y)_{\mathsf{Ext}(x,y_1)}$ we mean the following – we interpret the output of Ext as an index $i \in [v]$ of the codeword $\mathsf{ECC}(y)$. Then, $\mathsf{ECC}(y)_i$ refers to the content in that i'th entry interpreted as a $(\log q)$-bit string. Define

$$z = \mathsf{AdvCB}\left(y, \mathsf{Raz}(x,y_2), \mathsf{AdvGen}(x,y)\right).$$

Finally, we define

$$\mathsf{nmExt}(x,y) = \mathsf{Ext}'(x,z).$$

Analysis. Let X be an (n,k)-source, let Y be a random variable that is uniformly distributed over d-bit strings, independently of X, and let $\mathcal{A}\colon \{0,1\}^d \to \{0,1\}^d$ be a function with no fixed points. Denote $y' = \mathcal{A}(y)$. We start by proving the following claim.

Claim 12.2.

$$\Pr_{(x,y)\sim(X,Y)}\left[\mathsf{AdvGen}(x,y) = \mathsf{AdvGen}(x,y')\right] = O(\sqrt{\varepsilon}).$$

Proof. As Ext is a (k,ε)-strong seeded extractor,

$$\left(\mathsf{Ext}(X,Y_1), Y_1\right) \approx_\varepsilon \left(U, Y_1\right).$$

Conditioned on any fixing of Y_1, the random variables X, Y_1' are independent. Thus, by Lemma 5.1,

$$\left(\mathsf{Ext}(X,Y_1), Y_1', Y_1\right) \approx_\varepsilon \left(U, Y_1', Y_1\right).$$

Therefore, by Markov's inequality, except with probability $\sqrt{\varepsilon}$ over the fixings of $(y_1, y_1') \sim (Y_1, Y_1')$, it holds that

$$\mathsf{Ext}(X, y_1) \approx_{\sqrt{\varepsilon}} U. \tag{12.1}$$

By aggregating an error of $\sqrt{\varepsilon}$ to the total error, we condition on the event $(Y_1, Y_1') = (y_1, y_1')$ for which Equation (12.1) holds.

Observe that whenever $y_1 \neq y_1'$, it holds that $\mathsf{AdvGen}(X, Y) \neq \mathsf{AdvGen}(X, Y')$, and so, to bound the probability that $\mathsf{AdvGen}(X, Y) = \mathsf{AdvGen}(X, Y')$, we only need to consider the case $y_1 = y_1'$.

Recall that $Y \neq Y'$ and so, as ECC has relative distance $\delta = 1 - \varepsilon$, the codewords $\mathsf{ECC}(Y), \mathsf{ECC}(Y') \in \mathbb{F}_q^v$ agree on at most ε fraction of the coordinates. Let

$$I = \{i \in [v] \mid \mathsf{ECC}(Y)_i = \mathsf{ECC}(Y')_i\}$$

be the random variable that consists of all indices on which the two codewords agree. By the above, $|I| \leq \varepsilon v$. As $\mathsf{Ext}(X, y_1)$ is $\sqrt{\varepsilon}$-close to uniform and since $\mathsf{Ext}(X, y_1)$ is independent of I (as I is a deterministic function of Y), we have that

$$\mathop{\mathbf{Pr}}_{X,Y}[\mathsf{Ext}(X, y_1) \in I] \leq \varepsilon.$$

We can now conclude the proof of the claim by taking back into account the event that y_1 is not a good seed for X. □

By Lemma 5.2, by our choice of parameters, and as ECC has alphabet size q,

$$\widetilde{H}_\infty(Y_2 \mid \mathsf{AdvGen}(X, Y), \mathsf{AdvGen}(X, Y')) \geq d_2 - 2(d_1 + \log q) \geq 0.6 d_2.$$

Further,

$$\begin{aligned}\widetilde{H}_\infty(X \mid \mathsf{AdvGen}(X, Y), \mathsf{AdvGen}(X, Y')) &\geq k - 2\log v \\ &\geq \max(2d, c'_{\mathsf{Raz}} d_2) + \log(1/\varepsilon),\end{aligned}$$

where the last inequality holds for a large enough constant c'.

Observe that one can condition on the fixings of $\mathsf{AdvGen}(X, Y)$, $\mathsf{AdvGen}(X, Y')$ while maintaining the independence between X and Y. Indeed, after conditioning on Y_1, Y_1', the random variables $\mathsf{Ext}(X, Y_1)$, $\mathsf{Ext}(X, Y_1')$ are deterministic functions of X, and so one can further condition on these random variables without introducing dependencies between X and Y. Conditioned on Y_1, Y_1', $\mathsf{Ext}(X, Y_1)$, $\mathsf{Ext}(X, Y_1')$, the random variables $\mathsf{AdvGen}(X, Y)$, $\mathsf{AdvGen}(X, Y')$ are deterministic functions of Y, and so condition on these variables does not introduce dependencies between X, Y.

By the above, we can apply Theorem 5.7 and conclude that

$$(\mathsf{Raz}(X, Y_2), Y_2, \mathsf{AdvGen}(X, Y), \mathsf{AdvGen}(X, Y')) \approx_{O(\varepsilon)} (U, Y_2, \mathsf{AdvGen}(X, Y), \mathsf{AdvGen}(X, Y')).$$

By Lemma 5.1, and since $\mathsf{Raz}(X, Y_2)$ is independent of Y_2' when conditioned on any fixing of Y_2, $\mathsf{AdvGen}(X, Y)$, $\mathsf{AdvGen}(X, Y')$, we have that

$$(\mathsf{Raz}(X, Y_2), \mathcal{H}) \approx_{O(\varepsilon)} (U, \mathcal{H}),$$

where $\mathcal{H} = Y_2', Y_2, \mathsf{AdvGen}(X, Y), \mathsf{AdvGen}(X, Y')$.

By Lemma 5.2,

$$\widetilde{H}_\infty (Y \mid \mathcal{H}) \geq d - 2(d_2 + \log q) \geq d/2.$$

By applying Lemma 5.3, we have that except with probability $O(\sqrt{\varepsilon})$ over the fixing of $\mathcal{H}$, the following holds.

- The joint distribution of $\mathsf{Raz}(X, Y_2)$ and $\mathsf{Raz}(X, Y_2')$ is independent of the joint distribution of Y, Y'.
- $\mathsf{Raz}(X, Y_2)$ is $O(\sqrt{\varepsilon})$-close to uniform.
- By setting the constant α to be small enough, $d/2$ is sufficient min-entropy as required by Theorem 2.1.
- $\mathsf{AdvGen}(X, Y)$ and $\mathsf{AdvGen}(X, Y')$ are distinct, fixed, strings.

Thus, we can apply Theorem 2.1 to conclude that

$$(Z, \mathcal{H}') \approx_{O(\sqrt{\varepsilon})} (U, \mathcal{H}'),$$

where

$$\begin{aligned} Z &= \mathsf{AdvCB}(Y, \mathsf{Raz}(X, Y_2), \mathsf{AdvGen}(X, Y)), \\ Z' &= \mathsf{AdvCB}(Y', \mathsf{Raz}(X, Y_2'), \mathsf{AdvGen}(X, Y')), \\ \mathcal{H}' &= Z', \mathsf{Raz}(X, Y_2), \mathsf{Raz}(X, Y_2'), \mathcal{H}. \end{aligned}$$

Note that conditioned on the fixing of $\mathcal{H}'$, the random variables Z and $\mathsf{Ext}'(X, Z')$ are independent. Thus, by Lemma 5.1,

$$(Z, \mathsf{Ext}'(X, Z'), \mathcal{H}') \approx_{O(\sqrt{\varepsilon})} (U, \mathsf{Ext}'(X, Z'), \mathcal{H}').$$

By Lemma 5.2 and since Ext' has $k/4$ output bits,

$$\widetilde{H}_\infty(X \mid \mathsf{Ext}'(X, Z'), \mathcal{H}') \geq k - k/4 - 2(d + \log v) \geq k/2 + \log(1/\varepsilon).$$

Thus, by Lemma 5.8,

$$(\mathsf{Ext}'(X, Z), Z, \mathsf{Ext}'(X, Z'), \mathcal{H}') \approx_{O(\sqrt{\varepsilon})} (U, Z, \mathsf{Ext}'(X, Z'), \mathcal{H}').$$

By the definition of nmExt and since conditioned on Z, $\mathsf{Ext}'(X, Z')$, $\mathcal{H}'$, the random variable $\mathsf{Ext}'(X, Z)$ is independent of Y, we have that

$$(\mathsf{nmExt}(X, Y), \mathsf{nmExt}(X, Y'), Y) \approx_{O(\sqrt{\varepsilon})} (U, \mathsf{nmExt}(X, Y'), Y),$$

as desired. □

Acknowledgement

I wish to thank Leonard Schulman for insightful and highly enjoyable discussions regarding this work and, generally, on the fascinating problem of randomness extraction. I wish to thank Ben Lund and Adam Sheffer for referring me to [Jon12, RNRS14].

References

[AGM03] N. Alon, O. Goldreich, and Y. Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003.

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

[BBR85] C. H. Bennett, G. Brassard, and J. M. Robert. How to reduce your enemy's information. In *Advances in Cryptology (CRYPTO)*, volume 218, pages 468–476. Springer, 1985.

[BIW06] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. *SIAM Journal on Computing*, 36(4):1095–1118, 2006.

[BKS+05] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. In *Proceedings of the thirty-seventh annual ACM Symposium on Theory of Computing*, pages 1–10. ACM, 2005.

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

[BRSW12] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson. 2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl-Wilson construction. *Annals of Mathematics*, 176(3):1483–1544, 2012.

[BSZ11] E. Ben-Sasson and N. Zewi. From affine to two-source extractors via approximate duality. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, pages 177–186. ACM, 2011.

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

[CGL15] E. Chattopadhyay, V. Goyal, and X. Li. Non-malleable extractors and codes, with their many tampered extensions. *arXiv preprint arXiv:1505.00107*, 2015.

[CKOR14] N. Chandran, B. Kanukurthi, R. Ostrovsky, and L. Reyzin. Privacy amplification with asymptotically optimal entropy loss. *Journal of the ACM (JACM)*, 61(5):29, 2014.

[CL16] E. Chattopadhyay and X. Li. Explicit non-malleable extractors, multi-source extractors and almost optimal privacy amplification protocols. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 36, 2016.

[Coh15a] G. Cohen. Local correlation breakers and applications to three-source extractors and mergers. In *IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 845–862. IEEE, 2015.

[Coh15b] G. Cohen. Non-malleable extractors – new tools and improved constructions. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 183, 2015.

[Coh15c] G. Cohen. Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs. *arXiv preprint arXiv:1506.04428*, 2015.

[Coh16] G. Cohen. Non-malleable extractors with logarithmic seeds. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 30, 2016.

[CS15] G. Cohen and I. Shinkar. Zero-fixing extractors for sub-logarithmic entropy. In *Automata, Languages, and Programming*, pages 343–354. Springer, 2015.

[CS16] G. Cohen and L. Schulman. Extractors for near logarithmic min-entropy. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 14, 2016.

[CZ15] E. Chattopadhyay and D. Zuckerman. Explicit two-source extractors and resilient functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

[DKRS06] Y. Dodis, J. Katz, L. Reyzin, and A. Smith. Robust fuzzy extractors and authenticated key agreement from close secrets. In *Advances in Cryptology-CRYPTO 2006*, pages 232–250. Springer, 2006.

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

[DP07] S. Dziembowski and K. Pietrzak. Intrusion-resilient secret sharing. In *48th Annual IEEE Symposium on Foundations of Computer Science*, pages 227–237, 2007.

[DW09] Y. Dodis and D. Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the forty-first annual ACM Symposium on Theory of Computing*, pages 601–610. ACM, 2009.

[Erd47] P. Erdős. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[GRS06] A. Gabizon, R. Raz, and R. Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM Journal on Computing*, 36(4):1072–1094, 2006.

[GS95] A. Garcia and H. Stichtenoth. A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound. *Inventiones Mathematicae*, 121(1):211–222, 1995.

[GUV09] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Journal of the ACM*, 56(4):20, 2009.

[Jon12] T. Jones. Further improvements to incidence and Beck-type bounds over prime finite fields. *arXiv preprint arXiv:1206.4517*, 2012.

[KR09] B. Kanukurthi and L. Reyzin. Key agreement from close secrets over unsecured channels. In *Advances in Cryptology-EUROCRYPT 2009*, pages 206–223. Springer, 2009.

[Li11a] X. Li. Improved constructions of three source extractors. In *IEEE 26th Annual Conference on Computational Complexity*, pages 126–136, 2011.

[Li11b] X. Li. A new approach to affine extractors and dispersers. In *IEEE 26th Annual Conference on Computational Complexity*, pages 137–147, 2011.

[Li12a] X. Li. Design extractors, non-malleable condensers and privacy amplification. In *Proceedings of the forty-fourth annual ACM Symposium on Theory of Computing*, pages 837–854, 2012.

[Li12b] X. Li. Non-malleable condensers for arbitrary min-entropy, and almost optimal protocols for privacy amplification. *arXiv preprint arXiv:1211.0651*, 2012.

[Li13a] X. Li. Extractors for a constant number of independent sources with polylogarithmic min-entropy. In *IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 100–109, 2013.

[Li13b] X. Li. New independent source extractors with exponential improvement. In *Proceedings of the forty-fifth annual ACM Symposium on Theory of Computing*, pages 783–792. ACM, 2013.

[Li15a] X. Li. Improved constructions of two-source extractors. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 125, 2015.

[Li15b] X. Li. Three-source extractors for polylogarithmic min-entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

[Mau93] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993.

[Mek15] R. Meka. Explicit resilient functions matching Ajtai-Linial. *arXiv preprint arXiv:1509.00092*, 2015.

[MW97] U. Maurer and S. Wolf. Privacy amplification secure against active adversaries. In *Advances in Cryptology—CRYPTO'97*, pages 307–321. Springer, 1997.

[Ram28] F. P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, 30(4):338–384, 1928.

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

[Rao09] A. Rao. Extractors for a constant number of polynomially small min-entropy independent sources. *SIAM Journal on Computing*, 39(1):168–194, 2009.

[Raz05] R. Raz. Extractors with weak random seeds. In *Proceedings of the thirty-seventh annual ACM Symposium on Theory of Computing*, pages 11–20, 2005.

[RNRS14] O. Roche-Newton, M. Rudnev, and I. D. Shkredov. New sum-product type estimates over finite fields. *arXiv preprint arXiv:1408.0542*, 2014.

[RRV99] Ran Raz, Omer Reingold, and Salil Vadhan. Error reduction for extractors. In *40th Annual Symposium on Foundations of Computer Science (New York, 1999)*, pages 191–201. IEEE Computer Soc., Los Alamitos, CA, 1999.

[RW03] R. Renner and S. Wolf. Unconditional authenticity and privacy from an arbitrarily weak secret. In *Advances in Cryptology-CRYPTO 2003*, pages 78–95. Springer, 2003.

[Sti09] H. Stichtenoth. *Algebraic function fields and codes*, volume 254. Springer Science & Business Media, 2009.

[Vio14] E. Viola. Extractors for circuit sources. *SIAM Journal on Computing*, 43(2):655–672, 2014.

[Zuc07] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3:103–128, 2007.

A Proof of Lemma 11.6

In this section we prove Lemma 11.6. We make use of the standard notation in which a random variable X is identified with its probability vector, i.e., $\mathbf{Pr}[X = x]$ is denoted by $X(x)$. For a random variable X with an ambient support Ω and $i \geq 0$ we define

$$\mathsf{Bad}_i(X) = \{x \in \Omega \mid X(x) \geq 2^{-i}\}.$$

We prove the following lemma following an argument due to Raz, Reingold and Vadhan [RRV99].

Lemma A.1. *Let X be a random variable that is ε-close to a random variable X_k having min-entropy k. Denote $\mathsf{Bad} = \mathsf{Bad}_{k-1}(X)$. Then, $\mathbf{Pr}[X \in \mathsf{Bad}] \leq 2\varepsilon$.*

To prove Lemma A.1 we prove a sequence of simple claims.

Claim A.2. *With the notation of Lemma A.1, for every set B,*

$$\mathbf{Pr}[X \in B] \leq 2^{-k}|B| + \varepsilon.$$

Proof. As $\mathsf{SD}(X, X_k) \leq \varepsilon$,

$$|\mathbf{Pr}[X \in B] - \mathbf{Pr}[X_k \in B]| \leq \varepsilon.$$

Since $H_\infty(X_k) \geq k$, we have

$$\mathbf{Pr}[X_k \in B] = \sum_{x \in B} \mathbf{Pr}[X_k = x] \leq 2^{-k}|B|,$$

and so

$$\mathbf{Pr}[X \in B] \leq \mathbf{Pr}[X_k \in B] + \varepsilon \leq 2^{-k}|B| + \varepsilon.$$

□

Claim A.3. *With the notation of Lemma A.1,*

$$\mathbf{Pr}\left[X \in \mathsf{Bad}\right] \geq 2^{-k+1}|\mathsf{Bad}|.$$

Proof. By the definition of Bad, for every $x \in \mathsf{Bad}$, $\mathbf{Pr}\left[X = x\right] \geq 2^{-(k-1)}$ and so

$$\mathbf{Pr}\left[X \in \mathsf{Bad}\right] = \sum_{x \in \mathsf{Bad}} \mathbf{Pr}[X = x] \geq 2^{-k+1}|\mathsf{Bad}|.$$

□

Claim A.4. *With the notation of Lemma A.1,* $|\mathsf{Bad}| \leq 2^k\varepsilon$.

Proof. By Claim A.2 and Claim A.3, we have

$$2^{-k+1}|\mathsf{Bad}| \leq \mathbf{Pr}\left[X \in \mathsf{Bad}\right] \leq 2^{-k}|\mathsf{Bad}| + \varepsilon,$$

and so $2|\mathsf{Bad}| \leq |\mathsf{Bad}| + 2^k\varepsilon$ which implies $|\mathsf{Bad}| \leq 2^k\varepsilon$. □

Lemma A.1 now readily follows.

Proof of Lemma A.1. By Claim A.2 and Claim A.4,

$$\mathbf{Pr}\left[X \in \mathsf{Bad}\right] \leq 2^{-k}|\mathsf{Bad}| + \varepsilon \leq 2^{-k} \cdot 2^k\varepsilon + \varepsilon = 2\varepsilon.$$

□

Next we prove that a random variable that is close to having min-entropy k can be written as a convex combination of distributions where most of its weight is on a random variable with min-entropy $k - 2$.

Lemma A.5. *Let X be a random variable that is ε-close to a random variable X_k having min-entropy k, and assume $\varepsilon \leq \frac{1}{4}$. Then, X can be decomposed as*

$$X = (1 - \alpha)X_{k-2} + \alpha E,$$

where $\alpha \leq 2\varepsilon$ and E, X_{k-2} are random variables such that X_{k-2} has min-entropy $k - 2$.

Proof. For every x,

$$\mathbf{Pr}\left[X = x \mid X \notin \mathsf{Bad}\right] = \frac{\mathbf{Pr}\left[X = x \wedge X \notin \mathsf{Bad}\right]}{\mathbf{Pr}\left[X \notin \mathsf{Bad}\right]} \leq \frac{2^{-(k-1)}}{\mathbf{Pr}\left[X \notin \mathsf{Bad}\right]}.$$

Invoking Lemma A.1, we get

$$\mathbf{Pr}\left[X = x \mid X \notin \mathsf{Bad}\right] \leq \frac{2^{-(k-1)}}{1 - 2\varepsilon} \leq 2^{-(k-2)}$$

per our assumption $\varepsilon \leq \frac{1}{4}$. Thus, the random variable $X \mid (X \notin \mathsf{Bad})$ has min-entropy $k-2$. By the law of total probability, we can write

$$X = (1-\alpha) \cdot (X \mid X \notin \mathsf{Bad}) + \alpha \cdot (X \mid X \in \mathsf{Bad})$$

where $\alpha = \mathbf{Pr}[X \in \mathsf{Bad}] \leq 2\varepsilon$ as follows by Lemma A.1. This concludes the proof.

□

We also need the following simple claim relating convex combinations and statistical distance.

Claim A.6. *For all random variables X, Y, Z satisfying $X = (1-\epsilon)Y + \epsilon Z$ it holds that $\mathsf{SD}(X,Y) \leq \epsilon$.*

Proof.

$$\begin{aligned}
\mathsf{SD}(X,Y) &= \frac{1}{2}\sum_a |X(a) - Y(a)| \\
&= \frac{1}{2}\sum_a |(1-\varepsilon)Y(a) + \varepsilon Z(a) - Y(a)| \\
&= \frac{1}{2}\sum_a |-\varepsilon Y(a) + \varepsilon Z(a)| \\
&\leq \frac{1}{2}\sum_a |\varepsilon Y(a)| + |\varepsilon Z(a)|.
\end{aligned}$$

As Y, Z are random variables, we have that

$$\begin{aligned}
\mathsf{SD}(X,Y) &\leq \frac{1}{2}\sum_a \varepsilon Y(a) + \varepsilon Z(a) \\
&= \frac{1}{2}\varepsilon\left(\sum_a Y(a) + \sum_a Z(a)\right) \\
&= \varepsilon.
\end{aligned}$$

□

For ease of readability, we first prove Lemma 11.6 for the parity of $m = 2$ random variables. The proof of the general case follows using a similar argument as sketched afterwards.

Lemma A.7. *Let X, Y be a pair of independent random variables, each is ε-close to having min-entropy k, where $\varepsilon \leq \frac{1}{4}$. Then, $X \oplus Y$ is $4\varepsilon^2$-close to having min-entropy $k-2$.*

Proof. By Lemma A.5, we can write

$$X = (1-\alpha_X)X_{k-2} + \alpha_X E_X,$$
$$Y = (1-\alpha_Y)Y_{k-2} + \alpha_Y E_Y,$$

where $\alpha_X \leq 2\varepsilon$ and $\alpha_Y \leq 2\varepsilon$ and X_{k-2}, Y_{k-2} are random variables, each with min-entropy $k-2$. Furthermore, E_X, E_Y are random variables such that (X_{k-2}, E_X) is independent of (Y_{k-2}, E_Y).

Observe that sampling from X can be done as follows. Let I_X be a random variable that is 1 with probability $1-\alpha_X$ and 0 otherwise. To sample from X, one can first sample $i_X \sim I_X$. If $i_X = 1$, sample from X_{k-2}; otherwise, sample from E_X. The same holds for Y. Note that the indicators I_X, I_Y are independent. Therefore, sampling from $X \oplus Y$ can be done by first sampling (i_X, i_Y) from (I_X, I_Y), then proceeding as described above, and finally performing XOR. Hence $X \oplus Y$ can be written as

$$\begin{aligned} X \oplus Y = &(1-\alpha_X)(1-\alpha_Y)(X_{k-2} \oplus Y_{k-2}) + \\ &(1-\alpha_X)\alpha_Y(X_{k-2} \oplus E_Y) + \\ &\alpha_X(1-\alpha_Y)(E_X \oplus Y_{k-2}) + \\ &\alpha_X\alpha_Y(E_X \oplus E_Y). \end{aligned}$$

Note that each of the three random variables $(X_{k-2} \oplus Y_{k-2}), (X_{k-2} \oplus E_Y), (E_X \oplus Y_{k-2})$ has min-entropy at least $k-2$. Let us denote the random variable $(X \oplus Y \mid (i_X, i_Y) \neq (0,0))$ by Z_{k-2} which, note, has min-entropy $k-2$. The random variable $X \oplus Y$ can then be written as

$$X \oplus Y = (1-\alpha_X\alpha_Y)Z_{k-2} + \alpha_X\alpha_Y(E_X \oplus E_Y). \tag{A.1}$$

Invoking Claim A.6, we have that $X \oplus Y$ is $\alpha_X\alpha_Y$-close to Z_{k-2}, a random variable with min-entropy $k-2$. The proof follows as $\alpha_X \leq 2\varepsilon$ and $\alpha_Y \leq 2\varepsilon$. □

Lastly, we give the proof sketch of Lemma 11.6.

Proof sketch of Lemma 11.6. As in the proof of Lemma A.7, we can decompose each $X_i = (1-\alpha_i)X_{i,k-2} + \alpha_i E_i$ where $\alpha_i \leq 2\varepsilon$ and $X_{i,k-2}$ has min-entropy $k-2$. As in Equation (A.1), we can decompose

$$\bigoplus_{i=1}^{m} X_i = (1-\alpha)\, Z_{k-2} + \alpha E$$

where $E = E_1 \oplus \cdots \oplus E_m$ and Z_{k-2} has min-entropy $k-2$ (being a convex combination of $2^m - 1$ such sources), and

$$\alpha = \prod_{i=1}^{m} \alpha_i \leq (2\varepsilon)^m.$$

□

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il