2016 IEEE 57th Annual Symposium on Foundations of Computer Science

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

Gil Cohen
Computing and Mathematical Sciences Department
California Institute of Technology
Pasadena, CA, USA
Email: coheng@caltech.edu

***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 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.**

***Keywords*-extractors; non-malleable; privacy amplification; correlation breakers; independence-preserving mergers**

I. 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 of the full version of this paper.

Definition I.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 fol-*

[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.

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.28

lowing 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 a-bit strings α, α',

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

Although every effort was made to keep Definition I.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 I.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

$$k = 2\log(1/\varepsilon) + O(1),$$
$$\ell = \log a + \log(n-k) + 2\log(1/\varepsilon) + O(1).$$

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.

II. 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 II.1 (Main technical result)**.** *For all integers n, a, m, for any $\varepsilon > 0$, and for any constant $\alpha > 0$ such that $a < 2^{n/\varepsilon}$ there exists an explicit (k,ε)-correlation breaker with advice, with*

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

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 II.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 II.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. Prior to this work, constructions that achieve output length $\Omega(k)$ work only for very high min-entropy $k \approx n/2$. See Theorem II.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 II.5.

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

A. 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], Li constructed a three-source extractor for min-entropy $\mathrm{polylog}\, n$. Based on this extractor and on the challenge-response mechanism [BKS$^+$05], [BRSW12], the first two-source disperser for min-entropy $\mathrm{polylog}\, n$ was constructed in [Coh15c]. Subsequently, Chattopadhyay and Zuckerman [CZ15] (followed by some improvements [Li15a], [Mek15]) constructed a two-source extractor for min-entropy $\mathrm{polylog}\, n$.

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 $k = 2^{\mathrm{poly} \log \log n}$, as apposed to the desired $k = O(\log n)$. The most natural goal today is to obtain k-Ramsey graphs on n vertices with $k = \mathrm{polylog}\, n$. 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 I.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 II.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 $1/2 - \beta$. [2]

B. *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 II-C). More recently, non-malleable extractors played a key role in the construction of two-source extractors [CZ15].

Definition II.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

[2] 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], $s \geq 1000$.

existence of non-malleable extractors, and with such great parameters, is somewhat surprising, explicit constructions are far more desirable.

Constructing non-malleable extractors gained a significant attention in the literature. 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 $\mathrm{polylog}\, n$ (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. [3]

Building on Theorem II.1 we obtained the following result.

Theorem II.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. We remark that in [CL16] there are several other incomparable constructions.

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 seed length $O(\log n) + \log(1/\varepsilon) \cdot c^{(\log\log(1/\varepsilon))^{1/3}}$. [4] 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.

C. 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: (1) Round complexity – the number of rounds required by the protocol; (2) Entropy-loss – the amount of min-entropy that is lost during the protocol, namely, $k - m$; (3) Communication complexity – the total number of bits that are communicated; and (4) 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 one-round protocol with entropy-loss $O(\lambda)$ and communication complexity $O(\log n + (\lambda + \log k) \cdot \log k)$.

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

[3] 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.

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

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. Although Dodis and Wichs proved the existence of such extractors, an explicit construction was not obtained in [DW09] and, as covered in Section II-B, 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 II.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 II.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.

III. Proof Outline

Due to lack of space we do not give any formal proof in this version of the paper and refer the reader to its full version. Instead, in the sequel we present the outline of our proofs.

Our construction of correlation breakers with advice that is given by Theorem II.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.

A. 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 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. We make an extensive black-box use of this construction.

B. Hierarchy of independence

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

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

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.

C. 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 II.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 IV.

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 $\log n$, [5] 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 IV-A.

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 IV-B.

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

[5] 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.

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

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

This is not quite what is stated in Theorem II.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 (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 II.1.

IV. 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.

A. 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.

B. 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 $2^{\sqrt{\log a}} \cdot \log(n/\varepsilon)$.

C. 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

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

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}. [6]

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.

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

[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.

[6] 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.

[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.

[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.

[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 CryptologyCRYPTO'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 Bourgains 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.

[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.