Explicit Non-Malleable Extractors, Multi-Source Extractors and Almost Optimal Privacy Amplification Protocols

Eshan Chattopadhyay
Department of Computer Science
UT Austin
Austin, USA
Email: eshanc@cs.utexas.edu

Xin Li
Department of Computer Science
JHU
Baltimore, USA
Email: lixints@cs.jhu.edu

***Abstract*—We make progress in the following three problems: 1. Constructing optimal seeded non-malleable extractors; 2. Constructing optimal privacy amplification protocols with an active adversary, for any possible security parameter; 3. Constructing extractors for independent weak random sources, when the min-entropy is extremely small (i.e., near logarithmic).**

For the first two problems, the best known non-malleable extractors by Chattopadhyay, Goyal and Li, and by Cohen all require seed length and min-entropy with quadratic loss in parameters. As a result, the best known explicit privacy amplification protocols with an active adversary, which achieve two rounds of communication and optimal entropy loss was sub-optimal in the min-entropy of the source.

In this paper we give an explicit non-malleable extractor that works for nearly optimal seed length and min-entropy, and yields a two-round privacy amplification protocol with optimal entropy loss for almost all ranges of the security parameter.

For the third problem, we improve upon a very recent result by Cohen and Schulman and give an explicit extractor that uses an absolute constant number of sources, each with almost logarithmic min-entropy.

The key ingredient in all our constructions is a generalized, and much more efficient version of the independence preserving merger introduced by Cohen, which we call non-malleable independence preserving merger. Our construction of the merger also simplifies that of Cohen and Schulman, and may be of independent interest.

***Keywords*-privacy amplification; non-malleable extractors; independent source extractors;**

I. Introduction

The theory of *randomness extractors* is a broad area and a fundamental branch of the more general study of pseudorandomness. Informally, randomness extractors are functions that transform biased probability distributions (weak random sources) into almost uniform probability distributions. Here we measure the entropy of a weak random source by the standard min-entropy. A source $\mathbf{X}$ is said to have min-entropy k if for any x, $\Pr[\mathbf{X} = x] \leq 2^{-k}$. An (n, k)-source $\mathbf{X}$ is a distribution on n bits with min-entropy at least k.

It is well known that it is impossible to construct deterministic randomness extractors when the input is just one (arbitrary) weak random source, even if the min-entropy is as large as $n-1$. A natural relaxation is then to give the extractor a short independent uniform seed, and such extractors are called *seeded extractors*. With this relaxation it is indeed possible to construct extractors that work for any weak random source with essentially any min-entropy. We now formally define such extractors.

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

Definition I.2 ([41])**.** *A function* Ext : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a seeded extractor for min-entropy k and error ϵ if for any source $\mathbf{X}$ of min-entropy k, $|\mathrm{Ext}(\mathbf{X}, \mathbf{U}_d) - \mathbf{U}_m| \leq \epsilon$.* Ext *is strong if in addition $|(\mathrm{Ext}(\mathbf{X}, \mathbf{U}_d), \mathbf{U}_d) - (\mathbf{U}_m, \mathbf{U}_d)| \leq \epsilon$, where $\mathbf{U}_m$ and $\mathbf{U}_d$ are independent.*

Through a long line of research we now have explicit constructions of seeded extractors with almost optimal parameters [25], [26], [37].

In recent years, there has been much interest in the study of two other kinds of randomness extractors. The first one, known as *non-malleable extractors* (introduced by Dodis and Wichs [23]), is a generalization of strong seeded extractors.

Definition I.3 (Non-malleable extractor)**.** *A function* nmExt : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a (k, ϵ)-non-malleable extractor if the following holds: For any (n, k)-source $\mathbf{X}$, an independent uniform seed $\mathbf{Y}$ on d bits and any function $\mathcal{A} : \{0,1\}^d \to \{0,1\}^d$ with no fixed*

The second author is partially supported by NSF Grant CCF-1617713.

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.25

points,[1]

$$|(\text{nmExt}(\mathbf{X},\mathbf{Y}),\text{nmExt}(X,\mathcal{A}(\mathbf{Y})),\mathbf{Y})- \\ (\mathbf{U}_m,\text{nmExt}(\mathbf{X},\mathcal{A}(\mathbf{Y})),\mathbf{Y})| \leq \epsilon.$$

The second one, known as *multi-source extractors* (first studied by Chor and Goldreich [12]), is another natural relaxation of deterministic extractors for one weak random source, in the sense that now the input to the extractor are several (at least two) independent weak random sources. Curiously, although this problem was first studied around 30 years ago, it was not until recently that significant progress has been achieved.

The above two kinds of extractors are closely related, and in many cases techniques used for one can also be used to improve the constructions of the other. These connections have been demonstrated in a number of works (e.g., [11], [30]–[32], [34]).

We now briefly discuss the motivations for these two kinds of extractors.

A. Non-malleable extractors and privacy amplification

The initial motivation for non-malleable extractors comes from the problem of privacy amplification with an active adversary [6], [7], [39]. As a basic problem in information theoretic cryptography, privacy amplification deals with the case where two parties want to communicate with each other to convert their shared secret weak random source $\mathbf{X}$ into shared secret nearly uniform random bits. On the other hand, the communication channel is watched by an adversary Eve, who has unlimited computational power. To make this task possible, we assume two parties have local (non-shared) uniform random bits.

If Eve is passive (i.e., can only see the messages but cannot change them), this problem can be solved easily by applying the aforementioned strong seeded extractors. However, in the case where Eve is active (i.e., can arbitrarily change, delete and reorder messages), the problem becomes much more complicated. The major challenge here is to design a protocol that uses as few number of interactions as possible, and outputs a uniform random string $\mathbf{R}$ that has length as close to $H_\infty(\mathbf{X})$ as possible (the difference is called *entropy loss*). A bit more formally, we pick a security parameter s, and if the adversary Eve remains passive during the protocol then the two parties should achieve shared secret random bits that are 2^{-s}-close to uniform. On the other hand, if Eve is active, then the probability that Eve can successfully make the two parties output two different strings without being detected should be at most 2^{-s}. We refer the readers to [22] for a formal definition.

There has been a long line of work on this problem [1], [9], [19], [21]–[23], [27], [29], [30], [33], [38], [45]. When the entropy rate of $\mathbf{X}$ is large, i.e., bigger than $1/2$, there are known protocols that take only one round (e.g., [21], [38]). However these protocols all have very large entropy loss. When the entropy rate of $\mathbf{X}$ is smaller than $1/2$, [23] showed that no one round protocol exists; furthermore the length of $\mathbf{R}$ has to be at least $O(s)$ smaller than $H_\infty(\mathbf{X})$. Thus, the natural goal is to design a two-round protocol with such optimal entropy loss. However, all protocols before the work of [22] either need to use $O(s)$ rounds, or need to incur an entropy loss of $O(s^2)$.

In [23], Dodis and Wichs showed that explicit constructions of the aforementioned non-malleable extractors can be used to give two-round privacy amplification protocols with optimal entropy loss. Using the probabilistic method, they also showed that non-malleable extractors exist when $k > 2m + 2\log(1/\varepsilon) + \log d + 6$ and $d > \log(n-k+1) + 2\log(1/\varepsilon) + 5$. However, they were not able to give explicit constructions even for min-entropy $k = n-1$. The first explicit construction of non-malleable extractors appeared in [22], with subsequent improvements in [1], [19], [24], [29], [30]. All these constructions require the min-entropy of the weak source to be bigger than $0.49n$, and thus only give two-round privacy amplification protocols with optimal entropy loss for such min-entropy. Together with some other ideas, [22] also gives $\text{poly}(1/\delta)$ round protocols with optimal entropy loss for min-entropy $k \geq \delta n$, any constant $\delta > 0$. This was subsequently improved by one of the authors in [30] to obtain a two-round protocol with optimal entropy loss for min-entropy $k \geq \delta n$, any constant $\delta > 0$. In the general case, using a relaxation of non-malleable extractors called non-malleable condensers, one of the authors [33] also obtained a two-round protocol with optimal entropy loss for min-entropy $k \geq C\log^2 n$, some constant $C > 1$, as long as the security parameter s satisfies $k \geq Cs^2$. For larger security parameter, the best known protocol with optimal entropy loss in [30] still takes $O(s/\sqrt{k})$ rounds.

In a recent work, Chattopadhyay, Goyal and Li [10] constructed explicit non-malleable extractors with error ε, for min-entropy $k = \Omega(\log^2(n/\epsilon))$ and seed-length $d = O(\log^2(n/\epsilon))$. This gives an alternative protocol matching that of [30]. Subsequently, Cohen [15] improved this result, and constructed non-malleable extractors with seed length $d = O(\log(n/\epsilon)\log((\log n)/\epsilon))$ and min-entropy $k = \Omega(\log(n/\epsilon)\log((\log n)/\epsilon))$. In this work, he also gave another construction that worked for $k = n/(\log n)^{O(1)}$ with seed-length $O(\log n)$. In

[1] i.e., for any x, $\mathcal{A}(x) \neq x$

a follow up, Cohen [16] constructed non-malleable extractors with seed length $d = O(\log n + \log^3(1/\epsilon))$ and min-entropy $k = \Omega(d)$. However, in terms of the general error parameter ε, all of these results require min-entropy and seed length at least $\log^2(1/\varepsilon)$, thus none of them can be used to improve the privacy amplification protocols in [33].

A recent work by Aggarwal, Hosseini and Lovett [2] obtained some conditional results. In particular, they used a weaker variant of non-malleable extractors to construct privacy amplification protocols with optimal entropy loss for $k = \Omega(\log(1/\epsilon)\log n)$ assuming a conjecture in additive combinatorics.

B. Multi-source extractors for independent sources

As mentioned before, Chor and Goldreich [12] introduced the problem of designing extractors for two or more independent sources. Explicit constructions of such extractors can also be used in explicit constructions of Ramsey graphs ([5], [11], [17]). A simple probabilistic argument shows the existence of two-source extractors for min-entropy $k \geq \log n + O(1)$. However, explicit constructions of such functions are extremely challenging.

Chor and Goldreich [12] proved that the inner-product function is a two-source extractor for min-entropy greather than $n/2$. It was not until 20 years later when Bourgain [8] broke the entropy rate $1/2$ barrier and constructed a two-source extractor for min-entropy $0.49n$. Raz [44] obtained another construction which requires one source with min-entropy more than $n/2$ and the other source with min-entropy $O(\log n)$. Recently, Chattopadhyay and Zuckerman [11] improved the situation substantially by constructing two-source extractors for min-entropy $k \geq \mathrm{polylog}(n)$, with subsequent improvements obtained by Li [36] and Meka [40]. The ultimate goal here is to obtain two-source extractors matching the entropy bound given by the probabilistic method.

If we allow the extractor to have a constant number of sources instead of just two sources, then an exciting line of work [3]–[5], [13], [28], [31], [32], [34], [42], [43] constructed extractors with excellent parameters. However, the smallest entropy these constructions can achieve is $\log^{2+\delta} n$ for any constant $\delta > 0$ [31], which uses $O(1/\delta) + O(1)$ sources. In a very recent work, Cohen and Schulman [20] managed to break this "quadratic" barrier, and constructed extractors for $O(1/\delta) + O(1)$ sources, each with min-entropy $\log^{1+\delta} n$.

C. Our results

Non-Malleable Extractors Our first result is a new construction of non-malleable extractors that breaks the $\log^2(1/\varepsilon)$ barrier for min-entropy and seed length. Specifically, we have the following theorem.

Theorem 1. *There exists a constant* $C > 0$ *s.t for all* $n, k \in \mathbb{N}$ *and any* $\epsilon > 0$*, with* $k \geq \log(n/\epsilon)2^{C\sqrt{\log\log(n/\epsilon)}}$*, there exists an explicit* (k, ϵ)*-non-malleable extractor* $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, where* $d = \log(n/\epsilon)2^{C\sqrt{\log\log(n/\epsilon)}}$ *and* $m = k/2^{\sqrt{\log\log(n/\epsilon)}}$.

We also construct a non-malleable extractor with seed-length $O(\log n)$ for min-entropy $k = \Omega(\log n)$ and $\epsilon \geq 2^{-\log^{1-\beta}(n)}$ for any $\beta > 0$. Prior to this, explicit non-malleable extractors with seed-length $O(\log n)$ either requires min-entropy at least $n/\mathrm{poly}(\log n)$ [15] or requires $\epsilon \geq 2^{-\log^{1/3}(n)}$ [16].

Theorem 2. *There exists a constant* $C > 0$ *s.t for and all* $n, k \in \mathbb{N}$ *with* $k \geq C\log n$*, any constant* $0 < \beta < 1$*, and any* $\epsilon \geq 2^{-\log^{1-\beta}(n)}$*, there exists an explicit* (k, ϵ)*-non-malleable extractor* $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, where* $d = O(\log n)$ *and* $m = \Omega(\log(1/\epsilon))$.

Remark I.4. *A careful examination reveals that our seed length and min-entropy requirement are better than those of [15], [16] in all cases except the case that* ε *is large enough (e.g.,* $\epsilon \geq 2^{-\log^{1/3}(n)}$*), where both [16] and our results require seed length and min-entropy* $O(\log n)$.

Note that given any error parameter ε, our non-malleable extractor in Theorem 1 only requires min-entropy and seed length $\log^{1+o(1)}(n/\varepsilon)$.

We also show how to further lower the min-entropy requirement of the non-malleable extractor in Theorem 1 at the expense of using a larger seed. We complement this result by constructing another non-malleable extractor with shorter seed-length than in Theorem 1 at the expense of larger entropy. We now state these results more formally.

Theorem 3. *There exists a constant* $C > 0$ *such that for all* $n, k \in \mathbb{N}$ *and any* $\epsilon > 0$*, with* $k \geq \log(n/\epsilon)2^{2^{C\sqrt{\log\log\log(n/\epsilon)}}}$*, there exists an explicit* (k, ϵ)*-non-malleable extractor* $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, where* $d = (\log(n/\epsilon))^3 2^{(\log\log\log(n/\epsilon))^{O(1)}}$, $m = \Omega(k)$.

Theorem 4. *There exists a constant* $C > 0$ *such that for all* $n, k \in \mathbb{N}$ *and any* $\epsilon > 0$*, with* $k \geq (\log(n/\epsilon))^3 2^{(\log\log\log(n/\epsilon))^C}$*, there exists an explicit* (k, ϵ)*-non-malleable extractor* $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, where* $d = \log(n/\epsilon)2^{2^{O(\sqrt{\log\log\log(n/\epsilon)})}}$, $m = \frac{k}{\log(n/\epsilon)2^{(\log\log\log(n/\epsilon))^{O(1)}}} - O((\log(n/\epsilon))^2)$.

Table II-E summarizes our new non-malleable extractors compared to previous results.

Independent and Subsequent Work: Independent of our work, Cohen [14] also obtained similar results, where he constructed a non-malleable extractor with seed length and entropy $O(\log n) + \log(1/\epsilon)2^{O(\sqrt{\log\log(1/\epsilon)})}$. Subsequently, Cohen [18] and Li [35] further improved this result, and achieve non-malleable extractors with seed length and entropy $O(\log n) + \log(1/\epsilon)\mathrm{poly}(\log\log(1/\epsilon))$, and $O(\log n) + O(\log(1/\epsilon)\log\log(1/\epsilon))$ respectively.

Privacy Amplification Using Theorem 1 and the protocol in [23], we immediately obtain a two-round privacy amplification protocol with optimal entropy loss, for almost all possible security parameters.

Theorem 5. *There exists a constant $C > 0$ such that for any security parameter s with $k \geq (s + \log n)2^{C\sqrt{\log(s+\log n)}}$, there exists an explicit 2-round privacy amplification protocol for (n, k)-sources with entropy loss $O(\log n+s)$ and communication complexity $(s+\log n)2^{O(\sqrt{\log(s+\log n)})}$, in the presence of an active adversary.*

In particular, this gives us two-round privacy amplification protocols with optimal entropy loss for security parameter $s \leq k^{1-\alpha}$ for any constant $\alpha > 0$.

Instead if we use the non-malleable extractor from Theorem 3, we obtain a two-round privacy amplification protocol with optimal entropy loss, for even smaller min-entropy (at the expense of larger communication complexity). More formally, we have the following theorem.

Theorem 6. *There exists a constant $C > 0$ such that for any security parameter s with $k \geq (s + \log n)2^{2^{C\sqrt{\log\log(s+\log n)}}}$, there exists an explicit 2-round privacy amplification protocol for (n, k)-sources with entropy loss $O(\log n+s)$ and communication complexity $(s + \log n)^3 2^{(\log\log(s+\log n))^{O(1)}}$, in the presence of an active adversary.*

t-Non-Malleable Extractors and 2-Source Extractors Our techniques for constructing non-malleable extractors can be generalized directly to construct t-non-malleable extractors (non-malleable extractors with t tampering functions, see the full version of the paper for more details). Such t-non-malleable extractors were used in [11] to construct two-source extractors. With subsequent improvements [36], [40], the best known 2-source extractor for constant error requires min-entropy $C(\log n)^{10}$, and for polynomially small error requires min-entropy $C(\log n)^{18}$. By plugging in our improved t-non-malleable extractor, we obtain two-source extractors that require min-entropy $(\log n)^8$ for constant error, and $(\log n)^{14}$ for polynomially small error. By a well-known connection to Ramsey graphs (see [5]), the constant error 2-source extractor implies an explicit $2^{(\log\log n)^8}$-Ramsey graph on n vertices.

Multi-Source Extractors Next, we improve the entropy requirement in extractors for a constant number of independent sources. In particular, we give explicit extractors for $O(1)$ sources, each having min-entropy $\log^{1+o(1)}(n)$. More formally, we have the following theorem.

Theorem 7. *There exist constants $C > 0, C' > 0$ s.t for all $n, k \in \mathbb{N}$ with $k \geq \log n 2^{C'\sqrt{\log\log(n)}}$ and any constant $\epsilon > 0$,[2] there exists an explicit function* Ext : $(\{0,1\}^n)^C \to \{0,1\}$*, such that if $\mathbf{X}_1, \ldots, \mathbf{X}_C$ are independent (n, k) sources, then*

$$|\mathrm{Ext}(\mathbf{X}_1, \ldots, \mathbf{X}_C) - \mathbf{U}_1| \leq \epsilon.$$

D. Non-malleable independence preserving merger

The barrier of $\log^2(1/\varepsilon)$ in seed length and min-entropy requirement of non-malleable extractors, as well as the barrier of $\log^2 n$ in min-entropy requirement of multi-source extractors mainly come from the fact that the previous constructions rely heavily on the "alternating extraction" based techniques. In [20], Cohen and Schulman introduced a new object called *independence preserving merger* (IPM for short). This is the key component in their construction, which helps them to obtain the $O(1/\delta) + O(1)$ source extractor for min-entropy $k \geq \log^{1+\delta} n$. The construction of the independence preserving merger in [20] is fairly complicated and takes up a bulk of work.

A key component in all of our constructions is a generalized, and much more efficient version of the independence preserving merger in [20], which we call non-malleable independence preserving merger (NIPM for short). In addition, we believe that our construction of NIPM is simpler than the construction of IPM in [20]. We now define this object below.

Definition I.5. *A $(L, t, d', \epsilon, \epsilon')$-*NIPM : $\{0,1\}^{Lm} \times \{0,1\}^d \to \{0,1\}^{m_1}$ *satisfies the following property. Suppose*

- *$\mathbf{X}, \mathbf{X}^1, \ldots, \mathbf{X}^t$ are r.v's, each supported on boolean $L \times m$ matrices s.t for any $i \in [L]$, $|\mathbf{X}_i - \mathbf{U}_m| \leq \epsilon$,*
- *$\{\mathbf{Y}, \mathbf{Y}^1, \ldots, \mathbf{Y}^t\}$ is independent of $\{\mathbf{X}, \mathbf{X}^1, \ldots, \mathbf{X}^t\}$, s.t $\mathbf{Y}, \mathbf{Y}^1, \ldots, \mathbf{Y}^t$ are each supported on $\{0,1\}^d$ and $H_\infty(\mathbf{Y}) \geq d - d'$,*
- *there exists an $h \in [L]$ such that $|(\mathbf{X}_h, \mathbf{X}_h^1, \ldots, \mathbf{X}_h^t) - (\mathbf{U}_m, \mathbf{X}_h^1, \ldots, \mathbf{X}_h^t)| \leq \epsilon$,*

[2] As in [20], the error can actually be slightly sub-constant.

then

$$\begin{aligned}|(L,t,d',\epsilon,\epsilon')\text{-NIPM}((\mathbf{X},\mathbf{Y}),(L,t,d',\epsilon,\epsilon')\text{-NIPM}\\ (\mathbf{X}^1,\mathbf{Y}^1),\ldots,(L,t,d',\epsilon,\epsilon')\text{-NIPM}(\mathbf{X}^t,\mathbf{Y}^t)\\ -\mathbf{U}_{m_1},(L,t,d',\epsilon,\epsilon')\text{-NIPM}(\mathbf{X}^1,\mathbf{Y}^1),\ldots,\\ (L,t,d',\epsilon,\epsilon')\text{-NIPM}(\mathbf{X}^t,\mathbf{Y}^t)|\le\epsilon'.\end{aligned}$$

We present an explicit construction of an NIPM which requires seed length $d = \log(m/\epsilon)L^{o(1)}$ for the case $t = 1$. More formally, we have the following theorem.

Theorem 8. *For all integers $m, L > 0$, any $\epsilon > 0$, there exists an explicit $(L, 1, 0, \epsilon, \epsilon')$-NIPM : $\{0,1\}^{mL} \times \{0,1\}^d \to \{0,1\}^{m'}$, where $d = 2^{O(\sqrt{\log L})}\log(m/\epsilon)$, $m' = \frac{m}{2^{\sqrt{\log L}}} - 2^{O(\sqrt{\log L})}\log(m/\epsilon)$ and $\epsilon' = O(\epsilon L)$.*

We have a more general version of the above theorem, for which we refer the reader to the full version, that works for general t. This is crucial for us to obtain our results on t-non malleable extractors and extractors for independent sources with near logarithmic min-entropy.

Using our NIPM, we construct a standard IPM introduced in the work of Cohen and Schulman [20]. We first define an IPM.

Definition I.6. *A $(L, k, t, \epsilon, \epsilon')$-IPM : $\{0,1\}^{Lm} \times \{0,1\}^n \to \{0,1\}^{m_1}$ satisfies the following property. Suppose*

- *$\mathbf{X}, \mathbf{X}^1, \ldots, \mathbf{X}^t$ are r.v's, each supported on boolean $L \times m$ matrices s.t for any $i \in [L]$, $|\mathbf{X}_i - \mathbf{U}_m| \le \epsilon$,*
- *$\mathbf{Y}$ is an (n, k)-source, independent of $\{\mathbf{X}, \mathbf{X}^1, \ldots, \mathbf{X}^t\}$.*
- *there exists an $h \in [L]$ such that $|(\mathbf{X}_h, \mathbf{X}_h^1, \ldots, \mathbf{X}_h^t) - (\mathbf{U}_m, \mathbf{X}_h^1, \ldots, \mathbf{X}_h^t)| \le \epsilon$,*

then

$$\begin{aligned}|(L,k,t,\epsilon,\epsilon')\text{-IPM}(\mathbf{X},\mathbf{Y}),(L,k,t,\epsilon,\epsilon')\text{-IPM}(\mathbf{X}^1,\mathbf{Y}),\\ \ldots,(L,k,t,\epsilon,\epsilon')\text{-NIPM}(\mathbf{X}^t,\mathbf{Y})\\ -\mathbf{U}_{m_1},(L,k,t,\epsilon,\epsilon')\text{-IPM}(\mathbf{X}^1,\mathbf{Y}),\ldots,\\ (L,k,t,\epsilon,\epsilon')\text{-IPM}(\mathbf{X}^t,\mathbf{Y})|\le\epsilon'\end{aligned}$$

Theorem 9. *There exists a constant $C > 0$ such that for all integers $m, L > 0$, any $\epsilon > 0$, and any $k \ge 2^{C\sqrt{\log L}}\log(m/\epsilon)$, there exists an explicit $(L, k, 1, \epsilon, \epsilon')$-IPM : $\{0,1\}^{mL} \times \{0,1\}^n \to \{0,1\}^{m'}$, with $m' = \frac{1}{2^{\sqrt{\log L}}}(m - O(\log(n/\epsilon))) - 2^{O(\sqrt{\log L})}\log(m/\varepsilon)$ and $\epsilon' = O(\epsilon L)$.*

As in the case of NIPM, we in fact construct an IPM for general t. The construction of IPM from NIPM is relatively straightforward, and using this explicit IPM we derive our improved results on extractors for independent sources.

We note that there are several important differences between our IPM and the construction in [20]. First, we only require that there exists at least *one* "good" row $\mathbf{X}_h$ in the matrix (i.e., $\mathbf{X}_h$ is uniform even given $\mathbf{X}_h^1, \ldots, \mathbf{X}_h^t$). In contrast, the IPM in [20] requires that 0.99 fraction of the rows are good. Second, the construction of IPM in [20] offers a trade-off between the number of additional sources required and the min-entropy requirement of each source. In particular, they construct an IPM using b additional sources, each having min-entropy $k = \Omega(L^{1/b}\log(n/\epsilon))$. In contrast, we use just *one* additional source with min-entropy $k = \Omega(L^{o(1)}\log(m/\varepsilon))$, and works as long as $m \ge O(\log(n/\epsilon)) + L^{o(1)}\log(m/\varepsilon)$. In typical applications, we will have $m \approx k < n$, so it suffices to set $k = L^{o(1)}\log(n/\varepsilon)$. For all applications in this paper, we will choose $L = O(\log(n/\varepsilon))$ and thus we get $k = \log^{1+o(1)}(n/\varepsilon)$. The fact that our IPM uses only one additional source improves significantly upon the IPM in [20] and is crucial for us to obtain an $O(1)$ source extractor for min-entropy $k = \log^{1+o(1)} n$.

We also present a more involved construction of a NIPM that uses a shorter seed in comparison to the NIPM in Theorem 8, but requires matrices with larger rows (i.e., m is required to be larger). More formally, we have the following result.

Theorem 10. *For all integers $m, L > 0$, any $\epsilon > 0$, there exists an explicit $(L, 1, 0, \epsilon, \epsilon')$-NIPM : $\{0,1\}^{mL} \times \{0,1\}^d \to \{0,1\}^{m'}$, where $d = 2^{O(\sqrt{\log\log L})}\log(m/\varepsilon), m' = \frac{m}{L2^{(\log\log L)^{O(1)}}} - O(L\log(m/\varepsilon))$ and $\epsilon' = 2^{O(\sqrt{\log\log L})}L\varepsilon$.*

We use the NIPM from the above theorem in obtaining the non-malleable extractors in Theorem 3 and Theorem 4.

II. Outline of Constructions

Here we give an informal and high level description of our constructions. Due to lack of space, we refer the reader to the full version of the paper for formal proofs of the results discussed in this section. We start with our non-malleable independence preserving merger (NIPM) with a uniform (or high entropy rate) seed.

A. Non-malleable independence preserving merger with uniform seed

For simplicity we start by describing the case of only one tampering adversary. Here, we have two correlated random variables $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_L)$ and $\mathbf{X}' = (\mathbf{X}_1', \ldots, \mathbf{X}_L')$, each of them is an $L \times m$ matrix. We have another two correlated random variables $\mathbf{Y}, \mathbf{Y}'$. We assume the following conditions: $(\mathbf{X}, \mathbf{X}')$ is independent of $(\mathbf{Y}, \mathbf{Y}')$, each $\mathbf{X}_i$ is uniform and there exists a $j \in [L]$ such that $\mathbf{X}_j$ is uniform even conditioned on $\mathbf{X}'_j$, and $\mathbf{Y}$ is uniform. Our goal is to construct a function NIPM such that NIPM$(\mathbf{X}, \mathbf{Y})$ is uniform conditioned on NIPM$(\mathbf{X}', \mathbf{Y}')$, i.e., using $\mathbf{Y}$ we can merge $\mathbf{X}$ into a uniform random string which keeps the independence property over $\mathbf{X}'$ even with a tampered seed $\mathbf{Y}'$.

Our starting point is the following simple observation. Let $(\mathbf{X}, \mathbf{X}')$ be two correlated weak sources and $(\mathbf{R}, \mathbf{R}')$ be two correlated random variables such that $(\mathbf{X}, \mathbf{X}')$ is independent of $(\mathbf{R}, \mathbf{R}')$. Let $\mathbf{R}$ be uniform and take any strong seeded extractor Ext, consider $\mathbf{Z} = \text{Ext}(\mathbf{X}, \mathbf{R})$ and $\mathbf{Z}' = \text{Ext}(\mathbf{X}', \mathbf{R}')$. Assume the length of the output of Ext is small enough. Then $\mathbf{Z}$ is close to uniform given $\mathbf{Z}'$ if either of the following two conditions holds: $\mathbf{R}$ is uniform given $\mathbf{R}'$ or $\mathbf{X}$ has sufficient min-entropy conditioned on $\mathbf{X}'$. Indeed, in the first case, we can first fix $\mathbf{R}'$, and argue that conditioned on this fixing, $\mathbf{Z}'$ is a deterministic function of $\mathbf{X}'$. We can now further fix $\mathbf{Z}'$, and conditioned on this fixing, $\mathbf{X}$ still has enough entropy left (since the length of $\mathbf{Z}'$ is small). Note that at this point $\mathbf{R}$ is still uniform and independent of $\mathbf{X}$, thus $\mathbf{Z} = \text{Ext}(\mathbf{X}, \mathbf{R})$ is close to uniform given $\mathbf{Z}'$. In the second case, we can first fix $\mathbf{X}'$, and conditioned on this fixing $\mathbf{X}$ still has enough entropy left. Now since Ext is a strong extractor, we know that $\mathbf{Z} = \text{Ext}(\mathbf{X}, \mathbf{R})$ is close to uniform even given $\mathbf{R}$. Since we have already fixed $\mathbf{X}'$ and $(\mathbf{R}, \mathbf{R}')$ is independent of $(\mathbf{X}, \mathbf{X}')$, this means that $\mathbf{Z}$ is also close to uniform even given $\mathbf{R}'$ and $\mathbf{X}'$, which gives us $\mathbf{Z}' = \text{Ext}(\mathbf{X}', \mathbf{R}')$.

Now we can describe our basic NIPM. The construction is actually simple in the sense that it is essentially an alternating extraction process between $\mathbf{X}$ and $\mathbf{Y}$, except that in each alternation we use a *new* row from $\mathbf{X}$. Specifically, we first take a small slice $\mathbf{S}_1$ from $\mathbf{X}_1$, and apply a strong seeded extractor to obtain $\mathbf{R}_1 = \text{Ext}(\mathbf{Y}, \mathbf{S}_1)$; we then use $\mathbf{R}_1$ to extract from $\mathbf{X}_2$ and obtain $\mathbf{S}_2 = \text{Ext}(\mathbf{X}_2, \mathbf{R}_1)$. Now we continue and obtain $\mathbf{R}_2 = \text{Ext}(\mathbf{Y}, \mathbf{S}_2)$ and $\mathbf{S}_3 = \text{Ext}(\mathbf{X}_3, \mathbf{R}_2)$... .The final output of our merger will be $\mathbf{S}_L = \text{Ext}(\mathbf{X}_L, \mathbf{R}_{L-1})$.

To see why this construction works, first assume that the length of each $\mathbf{S}_i, \mathbf{R}_i$ is small enough. Let j be the first index in $[L]$ such that $\mathbf{X}_j$ is uniform even conditioned on $\mathbf{X}'_j$. Then, we can fix all the intermediate random variables $\mathbf{S}_1, \mathbf{S}'_1, \mathbf{R}_1, \mathbf{R}'_1, \mathbf{S}_2, \mathbf{S}'_2, \mathbf{R}_2, \mathbf{R}'_2 \ldots \mathbf{S}_{j-1}, \mathbf{S}'_{j-1}$, and conditioned on these fixings we know that: 1. $(\mathbf{R}_{j-1}, \mathbf{R}'_{j-1})$ are deterministic functions of $(\mathbf{Y}, \mathbf{Y}')$, and thus independent of $(\mathbf{X}, \mathbf{X}')$; 2. $\mathbf{R}_{j-1}$ is close to uniform; 3. $\mathbf{X}_j$ still has enough entropy conditioned on $\mathbf{X}'_j$. Now, by the first case we discussed above, this implies that $\mathbf{S}_j$ is close to uniform given $\mathbf{S}'_j$. From this point on, by using the second case we discussed above and an inductive approach, we can argue that for all subsequent $t \geq j$, we have that $\mathbf{R}_t$ is close to uniform given $\mathbf{R}'_t$ and $\mathbf{S}_t$ is close to uniform given $\mathbf{S}'_t$. Thus the final output $\mathbf{S}_L$ is close to uniform given $\mathbf{S}'_L$.

Note that this construction can work even if $\mathbf{Y}$ is a very weak random source instead of being uniform or having high min-entropy rate. However this basic approach will require the min-entropy of Y to be at least $O(L \log(m/\epsilon))$, which is pretty large if L is large. We next describe a way to reduce this entropy requirement, in the case where $\mathbf{Y}$ is uniform or has high min-entropy rate.

The idea is that, rather than merging the L rows in one step, we merge them in a sequence of steps, with each step merging all the blocks of some ℓ rows. Thus, it will take us roughly $\frac{\log L}{\log \ell}$ steps to merge the entire matrix. Now first assume that $\mathbf{Y}$ is uniform, then in each step we will not use the entire $\mathbf{Y}$ to do the alternating extraction and merging, but just use a *small slice* of $\mathbf{Y}$ for this purpose. That is, we will first take a small slice $\mathbf{Y}_1$ and use this slice to merge L/ℓ blocks of $\mathbf{X}$, where each block has ℓ rows; we then take another slice $\mathbf{Y}_2$ of $\mathbf{Y}$ and use this slice to merge L/ℓ^2 new blocks, where each block has ℓ rows, and so on. The advantage of this approach is that now the entropy consumed in each merging step is contained in the slice $\mathbf{Y}_i$ (and $\mathbf{Y}'_i$), and won't affect the rest of $\mathbf{Y}$ much.

As we discussed before, we need to make sure that each slice $\mathbf{Y}_i$ has min-entropy $O(\ell \log(m/\epsilon))$ conditioned on the fixing of all previous $(\mathbf{Y}_j, \mathbf{Y}'_j)$. As a result, we need to set $|\mathbf{Y}_{i+1}| \geq 2|\mathbf{Y}_i| + O(\ell \log(m/\epsilon))$. It suffices to take $|\mathbf{Y}_i| = c^i \ell \log(m/\epsilon)$ for some constant $c > 2$. We know that the whole merging process is going to take roughly $\frac{\log L}{\log \ell}$ steps, so the total length (or min-entropy) of $\mathbf{Y}$ is something like $c^{\frac{\log L}{\log \ell}} \ell \log(m/\epsilon)$. We just need to choose a proper ℓ to minimize this quantity. A simple calculation shows that the best ℓ is roughly such that $\log \ell = \sqrt{\log L}$, which gives us a seed length of $2^{O(\sqrt{\log L})} \log(m/\epsilon)$. This gives us the NIPM in Theorem 8.

It is not difficult to see that this argument also extends to the case where $\mathbf{Y}$ is not perfectly uniform but has high min-entropy rate (e.g., $1 - o(1)$) where we can still start with a small slice of $\mathbf{Y}$, and the case where we have $t+1$ correlated matrices $\mathbf{X}, \mathbf{X}^1, \ldots, \mathbf{X}^t$ and t tampered seeds $\mathbf{Y}^1, \ldots, \mathbf{Y}^t$ of $\mathbf{Y}$.

B. Non-malleable extractor with almost optimal seed

The NIPM in Theorem 8 is already enough to yield our construction of a non-malleable extractor with almost optimal seed length. Specifically, given an (n, k) source $\mathbf{X}$, an independent seed $\mathbf{Y}$, and a tampered seed $\mathbf{Y}'$, we follow the approach of one of the authors' previous work [10] by first obtaining an advice of length $L = O(\log(n/\epsilon))$. Let the advice generated by $(\mathbf{X}, \mathbf{Y})$ be S and the advice generated by $\mathbf{X}, \mathbf{Y}'$ be $\mathbf{S}'$. We have that with probability $1-\epsilon$, $\mathbf{S} \neq \mathbf{S}'$. Further, conditioned on $(\mathbf{S}, \mathbf{S}')$ and some other random variables, we have that $\mathbf{X}$ is still independent of $(\mathbf{Y}, \mathbf{Y}')$ and $\mathbf{Y}$ has high min-entropy rate.

Now we take a small slice $\mathbf{Y}_1$ of $\mathbf{Y}$, and use $\mathbf{X}$ and $\mathbf{Y}_1$ to generate a random matrix $\mathbf{V}$ with L rows, where the i'th row is obtained by doing a flip-flop alternating extraction (introduced in [13]) using the i'th bit of S. Similarly a matrix $\mathbf{V}'$ is generated using $\mathbf{X}'$ and $\mathbf{Y}_1$. The flip-flop alternating extraction guarantees that each row in $\mathbf{V}$ is close to uniform, and moreover if the i'th bit of S and S' are different, then $\mathbf{V}_i$ is close to uniform even given $\mathbf{V}_i'$. Note that conditioned on the fixing of $(\mathbf{Y}_1, \mathbf{Y}_1')$, we have that $(\mathbf{V}, \mathbf{V}')$ are deterministic functions of $\mathbf{X}$, and are thus independent of $(\mathbf{Y}, \mathbf{Y}')$. Furthermore $\mathbf{Y}$ still has high min-entropy rate.

At this point we can just use our NIPM and $\mathbf{Y}$ to merge $\mathbf{V}$ into a uniform string $\mathbf{Z}$, which is guaranteed to be close to uniform given $\mathbf{Z}'$ (obtained from $(\mathbf{V}'$, $\mathbf{Y}')$). The seed length of $\mathbf{Y}$ will be $O(\log(n/\epsilon)) + 2^{O(\sqrt{\log\log(n/\epsilon)})}\log(k/\epsilon)$. A careful analysis shows that the final error will be $O(\epsilon \log(n/\epsilon))$. Thus we need to set the error parameter ϵ slightly smaller in order to achieve a desired final error ϵ', but that does not affect the seed length much. Altogether this gives us a seed length and entropy requirement of $2^{O(\sqrt{\log\log(n/\epsilon)})}\log(n/\epsilon)$, as in Theorem 1.

C. Further improvements in various aspects

We can further improve the non-malleable independence preserving merger and non-malleable extractor in various aspects. For this purpose, we observe that our NIPM starts with a basic merger for ℓ rows and then use roughly $\frac{\log L}{\log \ell}$ steps to merge the entire L rows. If the basic merger uses a seed length of d, then the whole merger roughly uses seed length $c^{\frac{\log L}{\log \ell}} d$. In our basic and simple merger, we have $d = O(\ell \log(m/\epsilon))$. However, now that we have our improved NIPM, we can certainly use the more involved construction to replace the basic merger, where we only need seed length $d = 2^{O(\sqrt{\log \ell})}\log(m/\epsilon)$. Now we can choose another ℓ to optimize $c^{\frac{\log L}{\log \ell}} d$, which roughly gives $\log \ell = \log^{2/3} L$ and the new seed length is $2^{O(\log^{1/3} L)}\log(m/\epsilon)$. We can now again use this merger to replace the basic merger. By doing this recursively, we can get smaller and smaller seed length. On the other hand, the entropy requirement becomes larger. We can also switch the roles of the seed and source, and achieve smaller entropy requirement at the price of a larger seed. Eventually, we can get $d = 2^{O(\sqrt{\log\log L})}\log(m/\varepsilon)$ and $m = O(L^2 2^{(\log\log L)^{O(1)}})\log(m/\varepsilon)$, or vice versa. This gives us Theorem 10. Applying these NIPMs to non-malleable extractors as outlined above, we get Theorem 3 and Theorem 4.

D. Independence preserving merger with weak random seed

We now use our NIPM from Theorem 8 to construct a standard independence preserving merger with weak random seed, an object introduced in [20]. Suppose we are given $(\mathbf{X}, \mathbf{X}')$ as described above and an independent random variable $\mathbf{Y}$. Here $\mathbf{Y}$ can be a very weak source, so our first step is to convert it to a uniform (or high min-entropy rate) seed.

To do this, our observation is that since we know that each row in $\mathbf{X}$ is uniform, we can just take a small slice $\mathbf{W}$ of the first row $\mathbf{X}_1$, and apply a strong seeded extractor to $\mathbf{Y}$ to obtain $\mathbf{Z} = \mathrm{Ext}(\mathbf{Y}, \mathbf{W})$, which is guaranteed to be close to uniform. However by doing this we also created a correlated $\mathbf{Z}' = \mathrm{Ext}(\mathbf{Y}, \mathbf{W}')$ where $\mathbf{W}'$ is a slice of $\mathbf{X}_1'$. Note that conditioned on the fixing of $(\mathbf{W}, \mathbf{W}')$ we have $(\mathbf{X}, \mathbf{X}')$ is independent of $(\mathbf{Z}, \mathbf{Z}')$, each row of $\mathbf{X}$ still has high min-entropy, and the "good" row $\mathbf{X}_j$ still has high min-entropy even given $\mathbf{X}_j'$. We now take a small slice $\mathbf{V}$ of $\mathbf{Z}$, and use it to extract from each row of $\mathbf{X}$ to obtain another matrix $\overline{\mathbf{X}}$. Similarly we also have a slice $\mathbf{V}'$ from $\mathbf{Z}'$ and obtain $\overline{\mathbf{X}'}$. We can now argue that conditioned on the fixing of $(\mathbf{V}, \mathbf{V}')$, $(\overline{\mathbf{X}}, \overline{\mathbf{X}'})$ is independent of $(\mathbf{Z}, \mathbf{Z}')$, each row of $\overline{\mathbf{X}}$ is close to uniform, and the "good" row $\overline{\mathbf{X}}_j$ is close to uniform even given $\overline{\mathbf{X}'}_j$. Moreover $\mathbf{Z}$ still has high min-entropy rate.

Thus, we have reduced this case to the case of an independence preserving merger with a tampered high min-entropy rate seed. We can therefore apply our NIPM to finish the construction. It is also not difficult to see that our construction can be extended to the case where we have $\mathbf{X}, \mathbf{X}^1, \ldots, \mathbf{X}^t$ instead of having just $\mathbf{X}$ and $\mathbf{X}'$.

E. Improved multi-source extractor

We can now apply our independence preserving merger with weak random seed to improve the multi-source extractor construction in [20]. Our construction follows the framework of that in [20], except that we replace their independence preserving merger with ours.

Essentially, the key step in the construction of [20], and the only step which takes $O(1/\delta)$ independent $(n, \log^{1+\delta} n)$ sources (if we only aim at achieving constant or slightly sub-constant error) is to merge a matrix with $O(\log n)$ rows using $(n, \log^{1+\delta} n)$ sources. For this purpose and since the error of the merger needs to be $1/\text{poly}(n)$, the independence preserving merger in [20] uses two additional sources in each step to reduce the number of rows by a factor of $\log^{\delta} n$. Thus altogether it takes $2/\delta$ sources. Our merger as described above, in contrast, only requires *one* extra independent source with min-entropy at least $O(\log n) + 2^{O(\sqrt{\log\log n})} \log n = \log^{1+o(1)} n$. Therefore, we obtain a multi-source extractor for an absolute constant number of $(n, \log^{1+o(1)} n)$ sources, which outputs one bit with constant (or slightly sub-constant) error.

The improved two-source extractors are obtained directly by plugging in our improved t-non-malleable extractors to the constructions in [11], [40].

References

[1] D. Aggarwal, Y. Dodis, Z. Jafargholi, E. Miles, and L. Reyzin. *Advances in Cryptology – CRYPTO 2014: 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part II*, chapter Amplifying Privacy in Privacy Amplification, pages 183–198. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[2] D. Aggarwal, K. Hosseini, and S. Lovett. Affine-malleable extractors, spectrum doubling, and application to privacy amplification. Cryptology ePrint Archive, Report 2015/1094, 2015. http://eprint.iacr.org/.

[3] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. *SIAM J. Comput.*, 36(4):1095–1118, Dec. 2006.

[4] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. *J. ACM*, 57(4), 2010.

[5] 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):1483–1543, 2012. Preliminary version in STOC '06.

[6] C. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM J. Comput.*, 17:210–229, 1988.

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

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

[9] N. Chandran, B. Kanukurthi, R. Ostrovsky, and L. Reyzin. Privacy amplification with asymptotically optimal entropy loss. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 785–794, 2010.

[10] E. Chattopadhyay, V. Goyal, and X. Li. Non-malleable extractors and codes, with their many tampered extensions. In *STOC*, 2016.

[11] E. Chattopadhyay and D. Zuckerman. Explicit two-source extractors and resilient functions. In *STOC*, 2016.

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

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

[14] G. Cohen. Making the most of advice: New correlation breakers and their applications. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[15] G. Cohen. Non-malleable extractors - new tools and improved constructions. In *CCC*, 2016.

[16] G. Cohen. Non-malleable extractors with logarithmic seeds. Technical Report TR16-030, ECCC, 2016.

[17] G. Cohen. Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs. In *STOC*, 2016.

[18] G. Cohen. Two-source extractors for quasi-logarithmic min-entropy and improved privacy amplification protocols. Technical Report TR16-114, ECCC: Electronic Colloquium on Computational Complexity, 2016.

[19] G. Cohen, R. Raz, and G. Segev. Non-malleable extractors with short seeds and applications to privacy amplification. *SIAM J. Comput.*, 43(2):450–476, 2014.

[20] G. Cohen and L. Schulman. Extractors for near logarithmic min-entropy. Technical Report TR16-014, ECCC, 2016.

[21] Y. Dodis, J. Katz, L. Reyzin, and A. Smith. Robust fuzzy extractors and authenticated key agreement from close secrets. In *Advances in Cryptology — CRYPTO '06, 26th Annual International Cryptology Conference, Proceedings*, pages 232–250, 2006.

[22] Y. Dodis, X. Li, T. D. Wooley, and D. Zuckerman. Privacy amplification and non-malleable extractors via character sums. *SIAM J. Comput.*, 43(2):800–830, 2014.

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

[24] Y. Dodis and Y. Yu. Overcoming weak expectations. In *10th Theory of Cryptography Conference*, 2013.

[25] Z. Dvir, S. Kopparty, S. Saraf, and M. Sudan. Extensions to the method of multiplicities, with applications to Kakeya sets and mergers. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 181–190, 2009.

[26] V. Guruswami, C. Umans, and S. P. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J. ACM*, 56(4), 2009.

[27] B. Kanukurthi and L. Reyzin. Key agreement from close secrets over unsecured channels. In *EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 2009.

[28] X. Li. Improved constructions of three source extractors. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC 2011, San Jose, California, June 8-10, 2011*, pages 126–136, 2011.

[29] X. Li. Design extractors, non-malleable condensers and privacy amplification. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, pages 837–854, 2012.

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

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

[32] X. Li. New independent source extractors with exponential improvement. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 783–792, 2013.

[33] X. Li. Non-malleable condensers for arbitrary min-entropy, and almost optimal protocols for privacy amplification. In *12th Theory of Cryptography Conference*, 2015.

[34] X. Li. Three-source extractors for polylogarithmic min-entropy. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, 2015.

[35] X. Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. Technical Report TR16-115, ECCC: Electronic Colloquium on Computational Complexity, 2016.

[36] X. Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[37] C.-J. Lu, O. Reingold, S. P. Vadhan, and A. Wigderson. Extractors: optimal up to constant factors. In *STOC*, pages 602–611, 2003.

[38] U. Maurer and S. Wolf. Privacy amplification secure against active adversaries. In *Advances in Cryptology — CRYPTO '97*, volume 1294, pages 307–321, Aug. 1997.

[39] U. M. Maurer. Conditionally-perfect secrecy and a provably-secure randomized cipher. *Journal of Cryptology*, 5(1):53–66, 1992.

[40] R. Meka. Explicit resilient functions matching Ajtai-Linial. *CoRR*, abs/1509.00092, 2015.

[41] N. Nisan and D. Zuckerman. Randomness is linear in space. *J. Comput. Syst. Sci.*, 52(1):43–52, 1996.

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

[43] A. Rao and D. Zuckerman. Extractors for three uneven-length sources. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques, 11th International Workshop, APPROX 2008, and 12th International Workshop, RANDOM 2008, Boston, MA, USA, August 25-27, 2008. Proceedings*, pages 557–570, 2008.

[44] R. Raz. Extractors with weak random seeds. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 11–20, 2005.

[45] R. Renner and S. Wolf. Unconditional authenticity and privacy from an arbitrarily weak secret. In *Advances in Cryptology — CRYPTO '03, 23rd Annual International Cryptology Conference, Proceedings*, pages 78–95, 2003.

Reference	Min-Entropy	Seed Length
[23] (non-constructive)	$> 2m + 2\log(1/\epsilon) + \log d + 6$	$> \log(n-k+1) + 2\log(1/\epsilon) + 5$
[22]	$> n/2$	n
[19], [24], [29]	$> n/2$	$O(\log(n/\epsilon))$
[30]	$0.49n$	n
[10]	$\Omega((\log(n/\epsilon))^2)$	$O((\log(n/\epsilon))^2)$
[15]	$\Omega(\log(n/\epsilon)\log((\log n)/\epsilon))$	$O(\log(n/\epsilon)\log((\log n)/\epsilon))$
[16]	$\Omega(\log n + (\log(1/\epsilon))^3)$	$O(\log n + (\log(1/\epsilon))^3)$
Theorem 1	$\log(n/\epsilon)2^{\Omega(\sqrt{\log\log(n/\epsilon)})}$	$\log(n/\epsilon)2^{O(\sqrt{\log\log(n/\epsilon)})}$
Theorem 4	$\log(n/\epsilon)2^{2^{\Omega(\sqrt{\log\log\log(n/\epsilon)})}}$	$(\log(n/\epsilon))^{3+o(1)}$
Theorem 3	$(\log(n/\epsilon))^{3+o(1)}$	$\log(n/\epsilon)2^{2^{O(\sqrt{\log\log\log(n/\epsilon)})}}$

Table I
A SUMMARY OF RESULTS ON NON-MALLEABLE EXTRACTORS