# Two Source Extractors for Asymptotically Optimal Entropy, and (Many) More

Xin Li
*Department of Computer Science*
*Johns Hopkins University*
Baltimore, MD USA
lixints@cs.jhu.edu

***Abstract*—A long line of work in the past two decades or so established close connections between several different pseudorandom objects and applications, including seeded or seedless non-malleable extractors, two source extractors, (bipartite) Ramsey graphs, privacy amplification protocols with an active adversary, non-malleable codes and many more. These connections essentially show that an asymptotically optimal construction of one central object will lead to asymptotically optimal solutions to all the others. However, despite considerable effort, previous works can get close but still lack one final step to achieve truly asymptotically optimal constructions.**

**In this paper we provide the last missing link, thus simultaneously achieving explicit, asymptotically optimal constructions and solutions for various well studied extractors and applications, that have been the subjects of long lines of research. Our results include:**

- **Asymptotically optimal seeded non-malleable extractors, which in turn give two source extractors for asymptotically optimal min-entropy of $O(\log n)$, explicit constructions of $K$-Ramsey graphs on $N$ vertices with $K = \log^{O(1)} N$, and truly optimal privacy amplification protocols with an active adversary.**
- **Two source non-malleable extractors and affine non-malleable extractors for some linear min-entropy with exponentially small error, which in turn give the first explicit construction of non-malleable codes against 2-split state tampering and affine tampering with constant rate and *exponentially* small error.**
- **Explicit extractors for affine sources, sumset sources, interleaved sources, and small space sources that achieve asymptotically optimal min-entropy of $O(\log n)$ or $2s + O(\log n)$ (for space $s$ sources).**
- **An explicit function that requires strongly linear read once branching programs of size $2^{n-O(\log n)}$, which is optimal up to the constant in $O(\cdot)$. Previously, even for standard read once branching programs, the best known size lower bound for an explicit function is $2^{n-O(\log^2 n)}$.**

***Index Terms*—extractor, non-malleable, two-source, Ramsey graph, affine**

## I. Introduction

This paper studies a wide range of pseudorandom objects and applications. We first briefly survey each of them, and then state our main results.

Supported by NSF CAREER Award CCF-1845349 and NSF Award CCF-2127575.

*a) Randomness Extractors.:* Through decades of study, *randomness extractors* have become fundamental objects in the area of pseudorandomness, with intimate connections to other areas such as cryptography, complexity theory, combinatorics and graph theory, and so on. The original motivation of randomness extractors comes from bridging the gap between uniform random strings required in many applications, and poor quality random sources available in practice. We use the following standard definition, where the *min-entropy* of a random variable $X$ is defined as $H_\infty(X) = \min_{x \in \mathsf{supp}(X)} \log_2(1/\Pr[X = x])$. For $X \in \{0,1\}^n$, we call $X$ an $(n, H_\infty(X))$-source, or an $H_\infty(X)$-source when $n$ is clear from context, and we say $X$ has *entropy rate* $H_\infty(X)/n$.

The goal is to extract almost uniform random bits from weak random sources. Unfortunately, no deterministic extractor can exist when the input is a single general weak random source even with min-entropy $k = n - 1$. Hence, the study of randomness extractors has been focusing on several relaxed models. For example, Nisan and Zuckerman [83] introduced the notion of *seeded extractors*, where the extractor has access to an additional independent short uniform random seed. Typically, we require the seeded extractor to be *strong* in the sense that the output of the extractor is close to uniform even conditioned on the seed. It can be shown that there exist strong seeded extractors with excellent parameters, and we now have almost optimal constructions (e.g., [47], [48], [59], [81]) after a long line of research.

Although seeded extractors have proven to be quite useful, in certain applications (e.g., cryptography) even the short uniform random seed is undesirable, thus another relaxed model is to put more restrictions on the weak source, and construct *deterministic* or *seedless* extractors for a certain class of weak sources. We have the following definition.

**Definition I.1.** Let $\mathcal{X}$ be a family of distribution over $\{0,1\}^n$. A function $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$ is a deterministic extractor for $\mathcal{X}$ with error $\epsilon$ if for every distribution $X \in \mathcal{X}$, we have

$$\mathsf{Ext}(X) \approx_\epsilon U_m,$$

where $U_m$ stands for the uniform distribution over $\{0,1\}^m$, and $\approx_\epsilon$ means $\epsilon$ close in statistical distance. We say $\mathsf{Ext}$ is explicit if it is computable by a polynomial-time algorithm.

Historically, the most well studied class of sources is the class of two (or more) independent sources. Here, a simple probabilistic argument shows that there exist two source extractors for $(n, k)$ sources with $k = \log n + O(1)$, which is optimal up to the constant $O(1)$; and the first explicit construction of two source extractors was given by Chor and Goldreich [29] more than 35 years ago, which achieves $k > n/2$. Due to their connections to explicit Ramsey graphs, and applications in distributed computing and cryptography with general weak random sources [62], [63], such extractors have also been the subject of extensive study [7]–[10], [14], [19], [26], [29], [32], [33], [36]–[38], [40], [69], [70], [73]–[75], [77]–[80], [85], [87]. The ultimate goal is to construct explicit two source extractors for $k = \log n + O(1)$, which would also imply an (strongly) explicit Ramsey graph on $N$ vertices with no clique or independent set of size $O(\log N)$, solving a long standing open problem proposed by Erdős [51] in his seminal paper that inaugurated the probabilistic method. Previously, the best explicit construction of two source extractors in terms of entropy is that of [80], which achieves $k = O(\log n \cdot \frac{\log \log n}{\log \log \log n})$ and gives an explicit Ramsey graph on $N$ vertices with no clique or independent set of size $(\log N)^{O(\frac{\log \log \log N}{\log \log \log \log N})}$.

Deterministic extractors for many other classes of sources have been studied. These include for example bit fixing sources [30], [54], [65], [86], which are sources that are obtained by fixing some unknown bits of a uniform random string; affine sources [11], [15], [17], [53], [71], [78], [86], [90], [96], which generalize bit-fixing sources and are the uniform distributions over some unknown affine subspaces of a vector space; samplable sources [93], [94], which are sources that are generated by small circuits or efficient algorithms; interleaved sources [25], [88], which are a generalization of independent sources where the bits of the sources are mixed in some arbitrary order; and small-space sources [64], where the sources are generated by a small width branching program. Deterministic extractors for these sources have applications in areas such as exposure-resilient cryptography [30], [65], Boolean circuit lower bounds [42], [52], and best-partition communication complexity lower bound [88].

In [20], Chattopadhyay and Li introduced the model of sumset sources, which is the sum of two (or more) independent weak random sources. This model generalizes many of the previously studied models, such as independent sources, bit fixing sources, affine sources, interleaved sources, and small space sources. For clarity we defer the formal definitions of these sources to later chapters. Thus, improved constructions of explicit extractors for sumset sources may also lead to improved explicit extractors for many of the above sources. While [20] only constructed explicit extractors for the sum of a constant number of $(n, k)$ sources with $k = \log^{O(1)} n$, a recent improvement by Chattopadhyay and Liao [22] gives explicit extractors for the sum of two independent $(n, k)$ sources with $k = O(\log n \log \log n \log \log \log^3 n)$. This in turn implies explicit extractors for affine sources and interleaved two sources with the same entropy. By an improved reduction from small space sources to sumset sources in [22], this also gives explicit extractors for space $s$-sources with min-entropy $k = 2s + O(\log n \log \log n \log \log \log^3 n)$. These are the previously best known constructions for each corresponding class of sources in terms of entropy.[1] We note that non-explicitly, one can show that with high probability random functions are extractors for affine sources and interleaved two sources with entropy $k = O(\log n)$, and for space $s$-sources with min-entropy $k = 2s + O(\log n)$. Interestingly, it is not clear if a random function is an extractor for the sum of two independent $(n, k)$ sources. However, since sumset sources are a generalization of two independent sources, the entropy lower bound of $\log n + O(1)$ for two source extractors also implies an entropy lower bound of $\log n/2 + O(1)$ for the sum of two independent sources.

*b) Non-malleable extractors.:* Motivated from cryptographic applications, an important variant of seeded/seedless extractors known as *non-malleable extractors* has been the focus of much study in the past 15 years or so. Here, one or more inputs to the extractor are tampered with by an adversary, and the goal is to guarantee that the output of the extractor on the original inputs is still close to uniform even conditioned on the output of the extractor on the tampered inputs. To discuss non-malleable extractors, we start by defining tampering functions.

**Definition I.2** (Tampering Function)**.** For any function $f : S \to S$, We say $f$ has no fixed points if $f(s) \neq s$ for all $s \in S$. For any $n > 0$, let $\mathcal{F}_n$ denote the set of all functions $f : \{0,1\}^n \to \{0,1\}^n$. Any subset of $\mathcal{F}_n$ is a family of tampering functions.

It is clear that if the tampering function is the identity function, then non-malleability is impossible. Thus, without loss of generality, for non-malleable extractors we only consider tampering functions with no fixed points (although this can be extended to the more general setting, see for example [28]). See for example [28]). Depending on what the tampering function acts on, there are different models of non-malleable extractors. If the tampering acts on the seed of a seeded extractor, we get the notion of *seeded non-malleable extractors*, introduced by Dodis and Wichs [45]:

**Definition I.3** ( [45])**.** A function snmExt : $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a strong seeded non-malleable extractor for min-entropy $k$ and error $\epsilon$ if the following holds: For any $(n, k)$ source $X$ and tampering function $\mathcal{A} : \{0,1\}^d \to \{0,1\}^d$ with no fixed points, we have

$$|\text{snmExt}(X, U_d) \circ \text{snmExt}(X, \mathcal{A}(U_d)) \circ U_d - U_m \circ \text{snmExt}(X, \mathcal{A}(U_d)) \circ U_d| < \epsilon,$$

where $U_m$ is independent of $U_d$ and $X$.

Alternatively, if the tampering function acts on the inputs to a seedless extractor, then we get the notion of *seedless*

[1]We focus on affine sources over the field $\mathbb{F}_2$. For larger fields there are constructions with better parameters.

*non-malleable extractors*. This was first introduced by Cheraghchi and Guruswami [28] for the model of two independent sources:

**Definition I.4** ( [28])**.** A function $\mathsf{nmExt} : (\{0,1\}^n)^C \to \{0,1\}^m$ is a $(k,\epsilon)$-seedless non-malleable extractor for $C$ independent sources, if it satisfies the following property: Let $X_1, \cdots, X_C$ be $C$ independent $(n,k)$ sources, and $f_1, \cdots, f_C : \{0,1\}^n \to \{0,1\}^n$ be $C$ arbitrary tampering functions such that there exists an $f_i$ with no fixed points, then

$$|\mathsf{nmExt}(X_1, \cdots, X_C) \circ \mathsf{nmExt}(f_1(X_1), \cdots, f_C(X_2)) - U_m \circ \mathsf{nmExt}(f_1(X_1), \cdots, f_C(X_2))| < \epsilon.$$

Chattopadhyay and Li [21] adapted the definition to affine sources and affine tampering, thus leading to affine non-malleable extractors:

**Definition I.5** ( [21])**.** A function $\mathsf{anmExt} : \{0,1\}^n \to \{0,1\}^m$ is a $(k,\epsilon)$ affine non-malleable extractor if for any affine source $X$ with entropy at least $k$ and any affine function $f : \{0,1\}^n \to \{0,1\}^n$ with no fixed point, we have

$$|\mathsf{anmExt}(X) \circ \mathsf{anmExt}(f(X)) - \mathbf{U}_m \circ \mathsf{anmExt}(f(X))| \leq \epsilon.$$

Using the probabilistic method, one can prove the existence of all these non-malleable extractors with excellent parameters. For example, [45] showed that seeded non-malleable extractors exist when $k > 2m + 2\log(1/\varepsilon) + \log d + 6$ and $d > \log(n - k + 1) + 2\log(1/\varepsilon) + 5$. [28] showed that two source non-malleable extractors exist for $(n,k)$ sources when $k \geq m + \frac{3}{2}\log(1/\varepsilon) + O(1)$ and $k \geq \log n + O(1)$. Similarly, it can be also shown that affine non-malleable extractors exist for entropy $k \geq 2m + 2\log(1/\varepsilon) + \log n + O(1)$.

However, constructing explicit non-malleable extractors turns out to be significantly harder than constructing standard extractors, despite considerable effort [18], [19], [21], [33]–[35], [37]–[39], [44], [72], [73], [79], [80]. Previously, the best explicit seeded non-malleable extractors are due to Li [79], [80], which achieve $k \geq C(\log\log n + a\log(1/\epsilon))$, $d = O(\log n) + \log(1/\epsilon)2^{O(a(\log\log(1/\epsilon))^{\frac{1}{a}})}$ and output length $\Omega(k)$, for some constant $C > 1$ and any integer $a \in \mathbb{N}$; or $k \geq C(\log\log n + \log\log(1/\epsilon)\log(1/\epsilon))$ and $d = O(\log n + \log\log(1/\epsilon)\log(1/\epsilon))$ for some constant $C > 1$. For two source non-malleable extractors, the best explicit constructions are due to Li [80] and Chung, Obremski, Aggarwal [31]. The former achieves $k \geq (1-\gamma)n$ with error $2^{-\Omega(n\log\log n/\log n)}$ and output length $\Omega(n)$, for some constant $\gamma \in (0,1)$; while the latter achieves $k_1 \geq (\frac{4}{5} + \gamma)n$ for the first source, $k_2 \geq C\log n$ for the second source, with some constants $C > 1, \gamma \in (0,1)$, error $2^{-\min(k_1,k_2)^{\Omega(1)}}$, and output length $\Omega(\min(k_1,k_2))$. The only known explicit affine non-malleable extractor is given in [21], which achieves entropy $k \geq n - n^{\delta}$ for some constant $\delta \in (0,1)$, error $2^{-n^{\Omega(1)}}$ and output length $n^{\Omega(1)}$.

*c) Privacy amplification with an active adversary.:* The basic problem of *privacy amplification* was introduced by Bennett, Brassard, and Robert [12]. The situation arises where two parties with local (non-shared) uniform random bits aim to convert a shared secret weak random source $X$ into shared secret uniform random bits. This is achieved by a communication protocol, which is watched by an adversary with unlimited computational power. Such protocols are important in various applications such as quantum key distribution. While standard strong seeded extractors provide optimal one-round protocols for a passive adversary (i.e., an adversary who can only see the communications but cannot change them), they fail badly for an active adversary (i.e., an adversary who can arbitrarily change, delete and reorder messages). The main goal for the latter case is to design a protocol that uses as few number of interactions and as few bits of communications as possible, and achieves a shared uniform random string $R$ which is as long as possible. In this context, the difference between $H_\infty(X)$ and the length of the output is defined as the *entropy loss*, together with a security parameter $s$, which ensures that the probability that any active adversary can successfully cause the two parties to output two different strings without being detected is at most $2^{-s}$. On the other hand, the two parties should achieve a shared secret string that is $2^{-s}$-close to uniform, if the adversary remains passive. We refer the reader to [44] for a formal definition.

A long line of work has been devoted to this problem [16], [18], [19], [33]–[35], [37], [39], [43]–[45], [66], [72], [73], [76], [79], [80], [82], [89]. In contrast to a passive adversary, here one round protocol can only exist when the entropy rate of $X$ is bigger than $1/2$, and the protocol has to incur a large entropy loss. For a source $X$ with entropy rate smaller than $1/2$, [45] showed that any protocol needs at least two rounds with entropy loss at least $\Omega(s)$, and communication complexity at least $\Omega(\log n + s)$. Achieving a two-round protocol that asymptotically match these parameters for all possible security parameters $s$ is thus the ultimate goal (note that $s$ can be at most $\Omega(k)$ where $k = H_\infty(\mathbf{X})$). Previously, the best known protocol is due to Li [80], which achieves two rounds with entropy loss $O(\log\log n + s)$, with communication complexity $O(\log n) + s2^{O(a(\log s)^{\frac{1}{a}})}$ for any constant integer $a \geq 2$ and $s$ up to $\Omega(k)$; or communication complexity $O(\log n + s\log^2 s)$ for $s$ up to $\Omega(k/\log\log k)$.

*d) Non-malleable codes.:* Non-malleable codes, introduced by Dziembowski, Pietrzak and Wichs [50], are a generalization of standard error correcting codes to handle much larger classes of tampering. Informally, such a code is defined with respect to a specific family of tampering functions $\mathcal{F}$. The code consists of a randomized encoding function $E$ and a deterministic decoding function $D$, such that on any modified codeword $f(E(x))$ obtained from some function $f \in \mathcal{F}$ and some message $x$, the decoded message $x' = D(f(E(x)))$ is either the original message $x$, or $\epsilon$-close to a completely unrelated message. [50] shows that non-malleable codes have applications in tamper-resilient cryptography, and

most notably, they can provide security guarantees even if the adversary can completely overwrite the codeword.

Even with this relaxation, it can be seen that no non-malleable codes can exist if $\mathcal{F}$ is completely unrestricted. However, such codes do exist for many broad families of tampering functions. By now the study of non-malleable codes has grown into a large field with numerous publications, and we only survey some of the most related previous works here. One of the most natural and well studied families of tampering functions is the so called *split-state* model, where a $k$-bit message $x$ is encoded into $t$ parts of messages $y_1, \cdots, y_t$, each of length $n$, so the rate of the code is $k/(tn)$. The adversary is then allowed to arbitrarily tamper with each $y_i$ independently.

This model arises in many natural applications, for example when the $y_i$'s are stored in different parts of memory. Non-malleable codes in this model are also used in various non-malleable secret sharing schemes [56]. Obviously, the case of $t = 1$ corresponds to unrestricted tampering functions, and it is not possible to construct non-malleable codes. Thus the case of $t = 2$ is the most general and interesting setting. [50] first proved the existence of non-malleable codes in the split-state model, while Cheraghchi and Guruswami [27] showed that the optimal rate of non-malleable codes in the 2-split-state model is $1/2$. Following a long line of research [1]–[5], [18], [24], [49], [58], [67], [79], [80], Li [80] gave the first explicit construction in the 2-split-state model with constant rate and constant error $\epsilon$, while Aggarwal and Obremski [5] improved the error to be negligible $\epsilon = 2^{-k^{\Omega(1)}}$. The current best construction is due to [4], which achieves rate $1/3$ and error $\epsilon = 2^{-k/\log^3 k}$.

In [21], Chattopadhyay and Li studied the model where the tampering function is any arbitrary affine function on the entire codeword (instead of acting on 2 parts of the codeword independently). They give an explicit non-malleable code with rate $k^{-\Omega(1)}$ and error $2^{-k^{\Omega(1)}}$, which remains the best known construction to date.

*e) Hardness against read-once linear branching program.:* Branching programs are natural models to measure the space complexity of computation. A standard branching program is a directed acyclic graph with one source and two sinks (labeled by 1 and 0), where each non-sink node is marked with an index of an input bit and has out-degree 2. One outgoing edge is labeled by 0 and the other is labeled by 1. For any input, the computation of the branching program follows the natural path from the source to one sink, by reading the corresponding bits and going through the corresponding edges, and the input is accepted if the path ends in the sink with label 1. The size of the branching program is defined as the number of its nodes, which roughly corresponds to $2^{O(s)}$ for space $s$ computation.

Unfortunately, proving non-trivial size lower bounds of explicit functions for general branching programs (e.g., those that can separate $\mathsf{P}$ from $\mathsf{LOGSPACE}$) seems beyond the reach of current techniques, hence essentially almost all research has been focusing on restricted models. Among these, the most well studied model is that of *read once branching program*, or ROBP for short. In this model, in any computational path, each bit of the input is read at most once. Non-explicitly, an optimal lower bound of size $\Theta(2^{n-\log n})$ is known [6]. Explicitly, several previous works gave exponential lower bounds [6], [13], [46], [55], [60], [61], [68], [84], [92], [95], [97]. However, the best known lower bound for an explicit function, due to Andreev, Baskakov, Clementi and Rolim [6], is only $2^{n-O(\log^2 n)}$, and the bound of $2^{n-O(\log n)}$ is only known for a function in $\mathsf{DTIME}(2^{O(\log^2 n)}) \cap \mathsf{P/poly}$.

Recently, motivated by strengthening tree-like resolution refutation lower bounds and average case lower bounds for parity decision trees, Gryaznov, Pudlák, and Talebanfard [57] introduced the model of read once linear branching programs (ROLBP for short), where the queries on each computational path are generalized to be linear functions. To enforce the read once property, [57] defined two kinds of ROLBPs: a *strongly* ROLBP requires that at any node, the span of the linear queries on all paths leading to this node has no non-trivial intersection with the span of the linear queries on all paths starting from this node, while a *weakly* ROLBP only requires that the linear query at any node is not in the span of the linear queries on all paths leading to this node. It can be seen that both kinds of ROLBPs are generalizations of standard ROBPs.

[57] gave an explicit function which requires strong ROLBPs of size $\Omega(2^{n/3})$, which was subsequently improved by Chattopadhyay and Liao [23] to $2^{n-\log^{O(1)} n}$.[2]

### A. Our Results

We improve all of the above results, achieving asymptotically optimal constructions in almost all cases (except seedless non-malleable extractors, and the error and output length of seedless extractors). We list our main results according to the order of the areas that appear in the introduction.

*a) Seedless extractors.:* Our results for seedless extractors can be summarized as follows.

**Theorem I.6.** *For every constant $\epsilon > 0$ there exists a constant $c > 1$ and an explicit extractor $\mathsf{TExt} : \{0,1\}^{2n} \to \{0,1\}$ with error $\epsilon$, for the interleaving of two independent $(n, k)$ sources such that $k \geq c \log n$.*

**Theorem I.7.** *For every constant $\epsilon > 0$ there exists a constant $c > 1$ and an explicit extractor $\mathsf{SumsetExt} : \{0,1\}^n \to \{0,1\}$ with error $\epsilon$, for the sum of two independent $(n, k)$ sources such that $k \geq c \log n$, or an affine source on $n$ bits with entropy $k \geq c \log n$.*

**Theorem I.8.** *For every constant $\epsilon > 0$ there exists a constant $c > 1$ such that for every $s > 0$ there exits an explicit extractor $\mathsf{SpExt} : \{0,1\}^n \to \{0,1\}$ with error $\epsilon$, for space-$s$ sources on $n$ bits with min-entropy $k \geq 2s + c \log n$.*

All of the above theorems achieve asymptotically optimal entropy in the corresponding models. In addition, Theorem I.6 immediately gives the following corollary about explicit Ramsey graphs.

[2] In fact, these results also give average-case hardness for strongly ROLBPs.

**Corollary I.9.** *There is a constant $c > 1$ such that for every integer $N$ there exists a (strongly) explicit Ramsey graph on $N$ vertices with no clique or independent set of size $K = \log^c N$.*

*b) Non-malleable extractors.:* Our results for non-malleable extractors are summarized as follows.

**Theorem I.10.** *For any constant $\gamma > 0$ there is a constant $C > 0$ such that for any $0 < \epsilon < 1$ with $k \geq C \log(d/\epsilon)$ and $d = C \log(n/\epsilon)$, there is an explicit strong seeded non-malleable extractor for $(n, k)$ sources with seed length $d$, error $\epsilon$ and output length $\frac{(1-\gamma)k}{2}$.*

This theorem achieves asymptotically optimal parameters in all aspects. In fact, we can also extend it to the stronger notion of $t$-non-malleable seeded extractors. Next we have seedless non-malleable extractors.

**Theorem I.11.** *There exists a constant $C > 1$ such that for any constant $0 < \gamma < 1$ and $k \geq C \log n$, there exists an explicit construction of a $((\frac{2}{3} + \gamma)n, k, 2^{-\Omega(k)})$ two-source non-malleable extractor with output length $\Omega(k)$.*

This theorem improves both constructions in [80] and [31]. Specifically, like in [31], we can also handle the case where the second source only has logarithmic min-entropy, while we improve the entropy rate of the first source from $4/5 + \gamma$ in [31] and $1 - \gamma$ in [80] to $2/3 + \gamma$. Simultaneously, the error is also improved to an optimal $2^{-\Omega(k)}$, from $2^{-k^{\Omega(1)}}$ in [31] and $2^{-\Omega(k \log\log k/\log k)}$ in [80]. We note that for applications in non-malleable codes, we don't really need such small entropy (any linear entropy suffices), but such two source non-malleable extractors have applications in privacy amplification with tamperable memory, see [31] for details.

**Theorem I.12.** *There exists a constant $0 < \gamma < 1$ such that for any $n \in \mathbb{N}$, there exists an explicit construction of a $((1-\gamma)n, 2^{-\Omega(n)})$ affine non-malleable extractor with output length $\Omega(n)$.*

*c) Privacy amplification.:* Combining our optimal seeded non-malleable extractor with the protocol in [45], we get the following theorem.

**Theorem I.13.** *There exists a constant $0 < \alpha < 1$ such that for any $n, k \in \mathbb{N}$, there is an explicit two-round privacy amplification protocol in the presence of an active adversary, that achieves any security parameter $s \leq \alpha k$, entropy loss $O(\log\log n + s)$, and communication complexity $O(\log n + s)$.*

Our two-round protocol achieves asymptotically optimal parameters in all aspects, for security parameter up to $s = \Omega(k)$. The $O(\log\log n)$ term is the best possible if using the two-round protocol in [45]. This follows from the use of a message authentication code (MAC) that authenticates the seed of a strong seeded extractor with security parameter $s$, which has at least $\Omega(\log n)$ bits. Thus the MAC requires a key of length at least $\log\log n + s$. See [45] for more details.

*d) Non-malleable codes.:* Using our seedless non-malleable extractors, we also get new constructions of non-malleable codes.

**Theorem I.14.** *For any $n \in \mathbb{N}$ there exists a non-malleable code with efficient encoding and decoding against 2-split-state tampering, which has message length $k$, block length $2n$, rate $k/(2n) = \Omega(1)$ and error $2^{-\Omega(k)}$.*

**Theorem I.15.** *For any $n \in \mathbb{N}$ there exists a non-malleable code with efficient encoding and decoding against affine tampering, which has message length $k$, block length $n$, rate $k/n = \Omega(1)$ and error $2^{-\Omega(k)}$.*

Both theorems are asymptotically optimal. Theorem I.14 achieves a smaller constant rate than the rate $1/3$ construction in [4], but improves the error from $2^{-k/\log^3 k}$ to $2^{-\Omega(k)}$. Theorem I.15 significantly improves the construction in [21], with rate only $k^{-\Omega(1)}$ and error $2^{-k^{\Omega(1)}}$.

*e) Hardness against read once linear branching program.:* Our sumset extractor directly gives a hard function for strongly ROLBPs (in fact with any constant average-case hardness). We have

**Theorem I.16.** *There is an explicit function* SumsetExt $: \{0,1\}^n \to \{0,1\}$ *that requires strongly read once linear branching program of size $2^{n-O(\log n)}$.*

Our result improves the results of $\Omega(2^{n/3})$ in [57] and $2^{n-\log^{O(1)} n}$ in [23]. Clearly, it also gives the first explicit function that requires standard ROBPs of size $2^{n-O(\log n)}$, improving the previously best known result of $2^{n-O(\log^2 n)}$ in [6]. By the $\Theta(2^{n-\log n})$ bound for standard ROBPs [6], our result is optimal up to the constant in $O(.)$. We remark that our affine extractor also directly gives an asymptotically optimal $2^{n-O(\log n)}$ size lower bound for DNF circuits with a bottom layer of parity gates, by the result in [41].

### *B. Overview of the Techniques*

Before explaining our new ideas, we first recall the connections and reductions established in previous works. This allows us to reduce all the problems to a couple of central pseudorandom objects.

*a) Connections between different pseudorandom objects and applications.:* Non-malleable extractors have direct motivations and applications in cryptography. For example, [45] shows that an optimal seeded non-malleable extractor gives an optimal two-round privacy amplification protocol with an active adversary. Similarly, [27] and [21] show that good two-source and affine non-malleable extractors give non-malleable codes against 2-split state tampering and affine tampering. The idea is simple: the encoding function is to uniformly sample a pre-image of the message under the extractor function, and the decoding function is the extractor itself. Reducing the average case error of the extractor to the worst case guarantee of the code blows up the error $\epsilon$ to $2^m\epsilon$ where $m$ is the output length of the extractor. Thus, to achieve a constant rate it is crucial to have an exponentially small error $\epsilon = 2^{-\Omega(n)}$, while it is enough to work for any linear entropy $k = \Omega(n)$. For hardness against strongly ROLBPs, [23] observed that, just

like a standard ROBP, if one conditions on an internal node, then the programs before and after this node correspond to two independent sources. Hence this reduces the question of finding a hard function to the question of constructing a good extractor for the sum of two independent sources.

Yet, previous works also established more surprising, and unexpected connections between non-malleable extractors and standard seedless extractors, which have been the underlying source of most of the recent progress on extractor theory. Specifically, the first such connection was established between seeded non-malleable extractors and two-source (and more generally independent source) extractors by Li [72], [74], [75], where he showed sufficiently good seeded non-malleable extractors imply improved two source extractors. Using techniques from non-malleable extractors, this has led to Li's construction of the first explicit extractor for three independent $(n, k)$ sources with $k \geq \log^{O(1)} n$, output length $\Omega(k)$ and error $2^{-k^{\Omega(1)}}$ [74]. The construction uses two sources to produce a *somewhere random* source with $n^{O(1)}$ rows, such that there exist a large fraction of (almost) uniform rows, and these rows are almost $t$-wise independent for some $t = \log^{O(1)} n$. The third source is then used to extract random bits from this somewhere random source.

Chattopadhyay and Zuckerman [26] further formalized this connection, and brought in another key improvement by applying a *resilient function* directly to the somewhere random source, thus giving the first two source extractor for $k \geq \log^{O(1)} n$ with error $n^{-\Omega(1)}$. Afterwards, a series of works [10], [19], [33], [40], [78] improved the reduction and eventually, [10] establishes that an optimal seeded non-malleable extractor[3] would give a two source extractor for entropy $O(\log n)$. Later, Li [79] further established a connection between two source non-malleable extractors and seeded non-malleable extractors, which roughly says the following: a two source non-malleable extractor for any constant (less than 1) entropy rate with error $2^{-\Omega(n)}$ would give an optimal seeded non-malleable extractor. Again, it is crucial here to have an exponentially small error of $2^{-\Omega(n)}$, while the entropy rate can be any constant less than 1. Finally, these connections have been roughly extended to extractors for the sum of two independent sources in [23].[4] In summary, by the established connections, all the problems can be reduced to constructing explicit two-source and affine non-malleable extractors for any constant (less than 1) entropy rate with error $2^{-\Omega(n)}$.

*b) Our new ideas.:* Most of the above connections have been known for a while, yet the goal of constructing two-source non-malleable extractors with error $2^{-\Omega(n)}$ has been elusive so far. Indeed, more and more sophisticated techniques were developed in [19], [33], [37], [38], [40], [79], [80], only resulting in the construction in [80] which achieves error $2^{-\Omega(n \log\log n/\log n)}$. The bottleneck comes from the fact that all these constructions are based on some kind of *alternating extraction* using an advice string. To get error $\epsilon$ the length of the advice string is provably at least $\log(1/\epsilon)$, while the alternating extraction appears to need at least some growing function $f(\log(1/\epsilon))$ number of steps, where each step needs at least $\log(1/\epsilon)$ entropy. This result in a total entropy of $f(\log(1/\epsilon))\log(1/\epsilon)$. Since the total entropy is $< n$ and $f$ is a growing function, this falls short of achieving error $2^{-\Omega(n)}$.

Luckily, there is one previous work by Chattopadhyay and Zuckerman [24] which does achieve error $2^{-\Omega(n)}$. Their constsruction relies on techinques from additive combinatorics, and does not use alternating extraction. However, their construction (CZExt for short) only gives a non-malleable extractor that requires 10 independent $(n, k)$ sources with $k \geq (1-\gamma)n$ for some constant $\gamma > 0$. In addition, the tampering function has to act independently on each of the 10 sources, thus it is not a prior clear that this can give us anything for two source non-malleable extractors. Nevertheless, this construction is our starting point to provide the last missing link in the complete picture.

Essentially, we show how to get some kind of independence from just one weak source and an arbitrary function tampering with this source. To illustrate the basic idea, it helps to start with the example where $X$ is a uniform random string over $\{0,1\}^n$, while $f : \{0,1\}^n \to \{0,1\}^n$ is any linear tampering function. Let us divide $X$ evenly into $\ell$ blocks $X = X_1 \circ \cdots \circ X_\ell$, where each $X_i$ has $m = n/\ell$ bits. Consider the tampered input $X' = f(X) = X'_1 \circ \cdots \circ X'_\ell$. It is easy to see that there are linear functions $\{f^{ij}\}_{i,j\in[\ell]}$ such that for any $i \in [\ell]$, $X'_i = \sum_{j\in[\ell]} f^{ij}(X_j)$. If for some $i \in [\ell]$ there exists a $j \in [\ell], j \neq i$ such that $H(f^{ij}(X_j)) \geq \delta m$ for any constant $\delta > 0$, then since $X_i$ and $X_j$ are independent, we have $H(X_i \circ X'_i) \geq H(X_i) + H(f^{ij}(X_j)) \geq (1+\delta)m$. This implies that the conditional entropy $H(X_i|X'_i)$ is at least $(1+\delta)m - m = \delta m$. In this case, we can apply an affine extractor for any linear entropy in [15], [71], [96], so that the output on $X_i$ is close to uniform conditioned on the output on $X'_i$. This already achieves some kind of non-malleable extractor.

On the other hand, if for any $i \in [\ell]$ and any $j \in [\ell], j \neq i$, we have $H(f^{ij}(X_j)) < \delta m$, then we can *fix* all $f^{ij}(X_j)$ where $i \neq j$. Note that conditioned on this fixing, the $X_i$'s are still independent, and furthermore the fixing does not cause any $X_i$ to lose much entropy. Specifically, each $X_i$ still has entropy at least $(1 - \ell\delta)m$. Most importantly, with this fixing, each $X'_i$ is now a deterministic function of $X_i$! Thus, as long as $\ell\delta$ is small, we have obtained $\ell$ independent weak sources $\{X_i\}$ with $\ell$ tampering functions acting on each $X_i$ independently. Taking $\ell = 10$ for example, at this point we can apply the function CZExt to the $X_i$'s, and the output will again be close to uniform even conditioned on the output on the $X'_i$'s. Thus, if we combine the outputs in both cases, we get a somewhere random source with $\ell + 1$ rows such that one row is close to uniform conditioned on the corresponding row in the tampered output. We call this a non-malleable somewhere random source. With this object, it is now relatively easy to finish our construction using existing techniques.

[3]More accurately, a seeded non-malleable extractor against multiple tampering.

[4] [23] actually reduces extractors for sumset sources to good *correlation breakers*, which are building blocks in two-source non-malleable extractors. We ignore these technical details here.

In summary, the high level key new idea of our constructions can be roughly stated as the following result of dichotomy, which leads to a "win-win" situation: divide a weak source $X$ with sufficiently high entropy into $\ell$ blocks $X = X_1 \circ \cdots \circ X_\ell$, and consider the tampered version $X' = f(X) = X'_1 \circ \cdots \circ X'_\ell$. Then either (1) (in the case where $f$ "mixes" the $X_i$'s well) there exists an $i \in [\ell]$ such that $X_i | X'_i$ has large entropy, or (2) (in the case where $f$ doesn't mix the $X_i$'s well) $X_1 \circ \cdots \circ X_\ell$ can be viewed as independent sources and $f$ can be viewed as $\ell$ functions $f = g_1 \circ \cdots \circ g_\ell$ where each $g_i$ acts on $X_i$ independently.

However, making this idea formally work requires non-trivial techniques in both the constructions and the analysis. We now explain more technical details below.

*c) Affine non-malleable extractors.:* The previous analysis about a uniform random string $X$ can be relatively easily adapted to a high entropy affine source with slight modifications. Specifically, given an affine source on $n$ bits with entropy $k = (1-\gamma)n$ for some small constant $\gamma > 0$, we now divide it into say $\ell + 1$ blocks $X = X_1 \circ \cdots \circ X_\ell \circ X_{\ell+1}$, where each $X_i$ for $i \in [\ell]$ has $3\gamma n$ bits and $X_{\ell+1}$ has $(1 - 3\gamma\ell)n$ bits. Since $\ell = 10$ is a constant, we can choose a small constant $\gamma$ and make sure the size of $X_{\ell+1}$ is much larger than the $X_i$'s. The plan is to use $X_1 \circ \cdots \circ X_\ell$ to generate the non-malleable somewhere random source, and then use $X_{\ell+1}$ to extract random bits. However, one issue here is that $X_1 \circ \cdots \circ X_\ell$ may be the same as $X'_1 \circ \cdots \circ X'_\ell$, in which case it is impossible to generate the non-malleable somewhere random source. To fix this, as in previous works, we need to first generate a small advice string $\alpha$ from $X$ such that $\alpha \neq \alpha'$ with probability $1 - 2^{\Omega(n)}$, where $\alpha'$ is the advice string generated from $X'$. We also need to keep the entropy of $X$ and the structure of an affine source conditioned on the generation of the advice strings. This turns out to be even trickier than the case of two-source non-malleable extractors, and we end up using two more blocks from $X$ and an improved advice generator for affine tampering based on that in [21]. To explain our main ideas we ignore these technical issues here, and refer the reader to the full version for details.

Now assume that we have already generated the advice string $\alpha$, and $X$ still has entropy $(1 - \gamma)n$. The blocks of $X$ are no longer independent in general, but we show it is a convex combination of independent sources. Specifically, we view $X$ as the uniform random string subject to $\gamma n$ affine constraints. Conditioned on the fixing of the corresponding part of each constraint in each block, all blocks become independent. We can now do the same analysis as before. If for some $i \in [\ell]$ there exists a $j \in [\ell + 1], j \neq i$ such that $H(f^{ij}(X_j))$ is large, then $H(X_i | X'_i)$ is also large. Otherwise, we can fix all the $f^{ij}(X_j)$'s with $i \in [\ell], j \in [\ell + 1]$ and $i \neq j$. Conditioned on this fixing, the $X_i$'s are still independent with high entropy, and now all the $X'_i$'s with $i \in [\ell]$ are deterministic functions of the $X_i$'s. Thus we can apply an affine extractor to each $X_i$ with $i \in [\ell]$ and apply CZExt to $\{X_i \circ \alpha\}_{i \in [\ell]}$ (the concatenation with $\alpha$ ensures no fixed points with high probability). Combining all the outputs, we get a non-malleable somewhere random source $R$ with a constant number of rows, where each row has $\Omega(n)$ bits with error $2^{-\Omega(n)}$.

Note that $R$ and the tampered version $R'$ are deterministic functions of $\{X_i\}_{i \in [\ell]}$ and $\{X'_i\}_{i \in [\ell]}$. As long as $X_{\ell+1}$ has large enough entropy compared to the total size of $\{X_i\}_{i \in [\ell]}$ and $\{X'_i\}_{i \in [\ell]}$, a standard argument shows that there is an affine source $A$ contained in $X_{\ell+1}$ which is independent of $\{X_i\}_{i \in [\ell]}$ and $\{X'_i\}_{i \in [\ell]}$, and one can use linear seeded extractors to do alternating extraction between $R$ and $X_{\ell+1}$ to break the correlations. Indeed we apply an *affine correlation breaker*, such as those developed in [22], [78] to $X_{\ell+1}$ and each row of $R$, using the index of the corresponding row as the advice string, and finally take the XOR of all outputs. We argue that the output is non-malleable as follows. Without loss of generality assume that the first row of $R$ (denoted by $R_1$) is close to uniform conditioned on the first row of $R'$ (denoted by $R'_1$). We first fix $R'_1$ and all the outputs produced in the affine correlation breaker with $X'_{\ell+1}$ and $R'_1$. By using linear seeded extractors appropriately and keeping the output length to be small, we can ensure that (1) the affine structure of the sources is preserved, (2) $A$ still has high entropy and is independent of $\{X_i\}_{i \in [\ell]}$ and $\{X'_i\}_{i \in [\ell]}$, and (3) $R_1$ is still close to uniform. Now the affine correlation breaker guarantees that the output from $(X_{\ell+1}, R_1)$ is close to uniform given all the other outputs from $(X_{\ell+1}, R)$ and $(X'_{\ell+1}, R')$. Therefore once we take the XOR of the outputs, the string produced from $X$ is close to uniform conditioned on the string produced from $X'$. The key point is that $R$ only has a constant number of rows, thus the index of each row only has a constant number of bits, and $R_1$ and $X_{\ell+1}$ has $\Omega(n)$ entropy. Hence, we can achieve error $2^{-\Omega(n)}$ with output length $\Omega(n)$.

*d) Two-source non-malleable extractors.:* The case of two-source non-malleable extractors is more complicated, as here we don't have the nice structure of affine sources. Again, we ignore the issue of generating advice strings, and assume that we are given an advice string $\alpha \in \{0,1\}^{\Omega(n)}$ such that $\alpha \neq \alpha'$ with probability $1 - 2^{\Omega(n)}$, where $\alpha'$ is the advice string generated from the tampered inputs.

We show how to use a single source and the advice string to generate a *non-malleable somewhere high entropy source*, which is a source $R$ with a constant number of rows, each row with $\Omega(n)$ bits, and there exists a row $i$ such that $H_\infty(R_i | R'_i) \geq \Omega(n)$ (again $R'$ is the tampered version). We call this function a *non-malleable somewhere condenser with advice*. This is similar in spirit to, and can be viewed as the non-malleable analogue of the reduction given in [8], which shows how to turn an independent source extractor into a somewhere condenser, that converts any weak random source with any linear entropy into a constant number of rows such that one row has entropy rate 0.9.

Specifically, given an $(n, k)$ source $X$ with $k \geq (1 - \beta)n$ for some small constant $\beta > 0$, let us again divide $X$ evenly into $\ell = 10$ blocks $X = X_1 \circ \cdots \circ X_\ell$ where each $X_i$ has $m = n/\ell$ bits. The non-malleable somewhere condenser produces a random variable $R$ with $\ell + 1$ rows, where for each $i \in [\ell]$,

$R_i = X_i$, and $R_{\ell+1} = \mathsf{CZExt}(X_1 \circ \alpha, \cdots, \circ X_\ell \circ \alpha)$.

The analysis is more subtle and relies on carefully dividing $X$ into a convex combination of subsources. Let $X' = X'_1 \circ \cdots \circ X'_\ell$ be the tampered input. Without loss of generality assume $X$ is the uniform distribution on a set $S \subseteq \{0,1\}^n$ with size $2^{(1-\beta)n}$. Similar to [8], for each $i \in [\ell]$, we define $H_i$ to be the set which contains heavy elements in the support of $(X_i, X'_i)$, e.g., $H_i = \{(y, y') \in \{0,1\}^{2m} : \Pr[(X_i, X'_i) = (y, y')] \geq 2^{-(1+3\beta)m}\}$. We divide $S$ into two subsets: $S' = \{x \in S : \exists i, (x_i, x'_i) \notin H_i\}$ and $S'' = \{x \in S : \forall i, (x_i, x'_i) \in H_i\} = S \setminus S'$. If either $S'$ or $S''$ is small, e.g., has size at most $2^{(1-\beta)n-\beta m}$, then we can safely ignore it since it only has probability mass at most $2^{-\beta m}$. Otherwise we consider $S'$ and $S''$ separately, since $X$ is just a convex combination of the uniform distributions over $S'$ and $S''$.

$S'$ is relatively easy to handle. Given that $|S'| \geq 2^{(1-\beta)n-\beta m}$, if we divide $S'$ into disjoint subsets by grouping all $x \in S'$ with the same smallest index $i$ such that $(x_i, x'_i) \notin H_i$ together, then on average each subset has size roughly $2^{(1-\beta)n-\beta m}/\ell$. Since all elements in the subset are light elements, the uniform distribution over the subset has min-entropy at least $(1+3\beta)m - \beta m - \log \ell > (1+\beta)m$. This means that if we consider the subsource corresponding to the uniform distribution over each subset, then roughly $H_\infty(X_i | X'_i) \geq \beta m = \Omega(n)$.

Taking care of $S''$ is much trickier. In this case, we want to argue that somehow, $X_1, \cdots, X_\ell$ can be viewed as independent sources and the tampering function $f$ can be viewed as $f = g_1 \circ \cdots \circ g_\ell$ where each $g_i$ acts on $X_i$ independently. Note that in this case, for any $x \in S''$ and any $i \in [\ell]$, we have $(x_i, x'_i) \in H_i$. Our first step is to remove those elements $x \in S''$ such that there exists an $i \in [\ell]$ and too many $y' \in \{0,1\}^m$ (say $> 2^{\beta n + 6\beta m}$ such $y'$'s) where $(x_i, y') \in H_i$. Intuitively, these are the strings where the tampering function $f$ mixes too much entropy from the blocks $\{X_j, j \neq i\}$ into $X'_i$, and thus are bad for our purpose. By definition of $H_i$, for any $i$ we have $|H_i| \leq 2^{(1+3\beta)m}$. Hence the number of such $x$'s cannot be too large, and is at most $\ell 2^{(1+3\beta)m}/2^{\beta n + 6\beta m} \cdot 2^{(\ell-1)m} < 2^{(1-\beta)n - 2\beta m}$. Thus, removing these strings only cause $X$ to lose probability mass at most $2^{-2\beta m}$.

Let $S^*$ be the subset of $S''$ after removing the bad strings. It is clear that $S^*$ still has a large size, i.e., $|S^*| \geq (1 - 2^{-\beta m}) 2^{(1-\beta)n-\beta m} > 2^{n - 2\ell\beta m}$. We now consider $X^*$, the uniform distribution over $S^*$, and $X'^* = f(X^*)$. Let $S_i$ be the support of $X^*_i$. The large size of $S^*$ guarantees that each $S_i$ also has large size, in fact $|S_i| \geq 2^{(1-2\ell\beta)m}$. We now consider the sources $(Y_1, Y_2, \cdots, Y_\ell)$ where each $Y_i$ is the independent uniform distribution over $S_i$. To construct the functions $g_1, \cdots, g_\ell$, for any $y \in S_i$ we define the set $W^y_i = \{y' \in \{0,1\}^m : y \circ y' \in H_i\}$. Since we have removed the bad $x$'s, we now have $|W^y_i| \leq 2^{\beta n + 6\beta m}$ for any $i$ and any $y \in S_i$. We now consider a *random* function $g = (g^1, g^2, \cdots, g^\ell)$ where for any $i \in [\ell]$ and any $y \in S_i$, let $g^i(y)$ be a random element independently uniformly chosen from $W^y_i$. For all other $y \in \{0,1\}^m$ let $g^i(y) = 0^m$.

With the random functions, for any $x \in S^*$ we have $\Pr[(x, x') = (x, g(x))] \geq (2^{-\ell(\beta n + 6\beta m)}) \geq 2^{-7\ell\beta n}$ by the independence of the $g^i$'s. Now by linearity of expectation, there exists a subset $V \subseteq S^*$ with $|V| \geq 2^{-7\ell\beta n}|S^*| \geq 2^{-O(\ell\beta n)}\Pi_{i\in[\ell]}|S_i|$ such that for any $x \in V$, $(x, x') = (x, g(x))$. We can now remove the set $V$ from $S^*$ and repeat the above process. As long as there are at least $2^{-\beta n}|S^*|$ strings left, the same argument will give us a new set $V \subseteq S^*$ with $|V| \geq 2^{-O(\ell\beta n)}\Pi_{i\in[\ell]}|S_i|$ and a new function $g = (g^1, g^2, \cdots, g^\ell)$ such that for any $x \in V$, $(x, x') = (x, g(x))$. Repeat this process until there are less than $2^{-\beta n}|S^*|$ strings left, and we have divided $S^*$ into large disjoint subsets $\{V_q \subseteq \{0,1\}^n, q \in \mathcal{Q}\}$ with $\ell$-split state tampering functions $\{g_q : (\{0,1\}^m)^\ell \to (\{0,1\}^m)^\ell, q \in \mathcal{Q}\}$, and a small subset left with less than $2^{-\beta n}|S^*|$ strings.

Observe that $X^*$ is $2^{-\beta n}$-close to a convex combination of the uniform distributions on $\{V_q, q \in \mathcal{Q}\}$, while each subset $V_q$ has large density in the set $\Pi_{i\in[\ell]}S_i$. Since each $S_i$ itself is large, with an appropriate choice of parameters, we can ensure that for any $q \in \mathcal{Q}$, $\mathsf{CZExt}(Y_1 \circ \alpha, Y_2 \circ \alpha, \cdots, Y_\ell \circ \alpha)$ is close to uniform conditioned on $\mathsf{CZExt}(g_q(Y_1) \circ \alpha', g_q(Y_2) \circ \alpha', \cdots, g_q(Y_\ell) \circ \alpha')$. We then show that conditioned on the event $(Y_1, Y_2, \cdots, Y_\ell) \in V_q$, $\mathsf{CZExt}(Y_1 \circ \alpha, Y_2 \circ \alpha, \cdots, Y_\ell \circ \alpha)$ is close to having min-entropy $\Omega(n)$ conditioned on $\mathsf{CZExt}(g_q(Y_1) \circ \alpha', g_q(Y_2) \circ \alpha', \cdots, g_q(Y_\ell) \circ \alpha')$. This takes care of $S''$.

Ignoring the error (which is $2^{-\Omega(n)}$) and the issue of convex combination of subsources, we have now obtained a non-malleable somewhere condenser. The rest of the construction and analysis is relatively straightforward. In the actual construction, we will divide $X$ into more blocks, for example $X = X_1 \circ \cdots \circ X_\ell \circ X_{\ell+1}$ where each $X_i$ has $\Omega(n)$ bits, but $X_{\ell+1}$ has much larger size compared to the previous blocks. We use $(X_1, \cdots, X_\ell)$ to obtain the non-malleable somewhere high entropy source with a constant number of rows. Then, using sum-product theorem based condensers in [8], [87], [98], we can boost the conditional min-entropy rate from $\Omega(1)$ to 0.9, while only increasing the number of rows by a constant factor. At this point we apply an extractor by Raz [87] to each row and the second source $Y$, which effectively converts the non-malleable somewhere high entropy source into a non-malleable somewhere random source. Fix $(X_1, \cdots, X_\ell)$ and $(X'_1, \cdots, X'_\ell)$, we argue that $X$ and $Y$ are still independent, and $X_{\ell+1}$ has enough entropy left. We can now use the non-malleable somewhere random source and a standard correlation breaker to extract uniform random bits from $X_{\ell+1}$, thus achieving a two-source non-malleable extractor by a similar argument as that of the affine non-malleable extractor. Again, the key point is that the somewhere random source only has a constant number of rows, and each row and $X_{\ell+1}$ has $\Omega(n)$ entropy. Hence, we can achieve error $2^{-\Omega(n)}$ with output length $\Omega(n)$.

The above gives a two-source non-malleable extractor for entropy rate $1 - \beta$ with some small constant $\beta > 0$. We can decrease the entropy of the first source to $k_1 \geq (2/3 + \gamma)n$ and the entropy of the second source to $k_2 \geq O(\log n)$ by first

taking a slice of the first source with size $n/3$, then applying the sum-product theorem based condensers in [8], [87], [98], Raz's extractor [87] to the second source, and a strong seeded extractor (e.g., those in [59]) to the first source to boost the entropy rate. This will result in a constant number of rows in both sources such that there exists one row where both sources have very high entropy rate. We can then apply the advice generator, our new two-source non-malleable extractor for entropy rate $1-\beta$, and finally the correlation breaker and taking the XOR of the outputs.

*e) Efficiently sampling the pre-image.:* For applications in non-malleable codes, we need to design efficient algorithms to sample uniformly from the pre-image of any output of our seedless non-malleable extractors. Thus we appropriately modify our extractors, roughly following the same approach as in [79]. However, to achieve error $2^{-\Omega(n)}$, we can no longer use a Reed-Solomon code in the advice generator, since this only achieves error $2^{-\Omega(n/\log n)}$. Instead, we use an asymptotically good linear binary code whose dual code is also asymptotically good. This implies that for some constant $\eta > 0$, any $\eta$ fraction of columns in the generator matrix are linearly independent.

## II. Conclusion and Open Problems

Our results partially finish several long lines of research projects, which are contributed by numerous researchers and publications. The connections discovered in these projects are amazingly broad. Indeed the techniques that culminated in our main results span areas like pseudorandomness, additive combinatorics, Fourier analysis, cryptography, coding theory and so on.

There are still interesting and important open problems left. For example, one natural open question is to improve the output length and error of the seedless extractors. Currently for asymptotically optimal entropy, our constructions can only output 1 bit (or a constant number of bits by the techniques in [78]) with constant error, while it is desirable to achieve negligible, or exponentially small error in cryptographic applications. Interestingly, improving the error may also lead to an improvement in output length by the techniques in [78]. As observed in previous works, one possible approach is to design $t$-non-malleable extractors with better dependence on $t$, which appears to be a challenging problem. One could also ask if we can construct explicit two-source extractors with entropy $\log n + O(1)$, which would give optimal Ramsey graphs. For non-malleable codes it would be interesting to improve the rates of our codes to optimal. Finally, it is always interesting to find other applications of the pseudorandom objects studied in this paper.

## Acknowledgment

We thank Songtao Mao for pointing out an inaccuracy in an earlier version, and Venkat Guruswami for pointing us to the construction of explicit binary linear codes such that both the code and its dual are asymptotically good in [91].

## References

[1] D. Aggarwal, Y. Dodis, T. Kazana, and M. Obremski. Non-malleable reductions and applications. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, 2015.

[2] Divesh Aggarwal. Affine-evasive sets modulo a prime. Technical Report 2014/328, Cryptology ePrint Archive, 2014.

[3] Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, 2014.

[4] Divesh Aggarwal, Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, Maciej Obremski, and Sruthi Sekar. Rate one-third non-malleable codes. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1364–1377. ACM, 2022.

[5] Divesh Aggarwal and Maciej Obremski. A constant rate non-malleable code in the split-state model. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 1285–1294. IEEE, 2020.

[6] Alexander E. Andreev, Juri L. Baskakov, Andrea E. F. Clementi, and José D. P. Rolim. Small pseudo-random sets yield hard functions: New tight explict lower bounds for branching programs. In Jirí Wiedermann, Peter van Emde Boas, and Mogens Nielsen, editors, *Automata, Languages and Programming, 26th International Colloquium, ICALP'99, Prague, Czech Republic, July 11-15, 1999, Proceedings*, volume 1644 of *Lecture Notes in Computer Science*, pages 179–189. Springer, 1999.

[7] Boaz Barak, R. Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 384–393, 2004.

[8] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 1–10, 2005.

[9] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2 source dispersers for $n^{o(1)}$ entropy and Ramsey graphs beating the Frankl-Wilson construction. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[10] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Explicit two-source extractors for near-logarithmic min-entropy. Technical Report TR16-088, ECCC, 2016.

[11] Eli Ben-Sasson and Swastik Kopparty. Affine dispersers from subspace polynomials. *SIAM J. Comput.*, 41(4):880–914, 2012.

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

[13] Beate Bollig and Ingo Wegener. A very simple function that requires exponential size read-once branching programs. *Inf. Process. Lett.*, 66(2):53–57, 1998.

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

[15] Jean Bourgain. On the construction of affine-source extractors. *Geometric and Functional Analysis*, 1:33–57, 2007.

[16] 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.

[17] Eshan Chattopadhyay, Jesse Goodman, and Jyun-Jie Liao. Affine extractors for almost logarithmic entropy. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 622–633. IEEE, 2021.

[18] Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, 2016.

[19] Eshan Chattopadhyay and Xin Li. Explicit non-malleable extractors, multi-source extractors and almost optimal privacy amplification protocols. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[20] Eshan Chattopadhyay and Xin Li. Extractors for sumset sources. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC, Cambridge, MA, USA, June 18-21, 2016*, pages 299–311. ACM, 2016.

[21] Eshan Chattopadhyay and Xin Li. Non-malleable codes and extractors for small-depth circuits, and affine functions. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 1171–1184. ACM, 2017.

[22] Eshan Chattopadhyay and Jyun-Jie Liao. Extractors for sum of two sources. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1584–1597. ACM, 2022.

[23] Eshan Chattopadhyay and Jyun-Jie Liao. Hardness against linear branching programs and more. Technical report, Electron. Colloquium Comput. Complex., 2022.

[24] Eshan Chattopadhyay and David Zuckerman. Non-malleable codes against constant split-state tampering. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science*, pages 306–315, 2014.

[25] Eshan Chattopadhyay and David Zuckerman. New Extractors for Interleaved Sources. In Ran Raz, editor, *31st Conference on Computational Complexity (CCC 2016)*, volume 50 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7:1–7:28, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

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

[27] Mahdi Cheraghchi and Venkatesan Guruswami. Capacity of non-malleable codes. In *ITCS*, pages 155–168, 2014.

[28] Mahdi Cheraghchi and Venkatesan Guruswami. Non-malleable coding against bit-wise and split-state tampering. In *TCC*, pages 440–464, 2014.

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

[30] Benny Chor, Oded Goldreich, Johan Hastad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem of t-resilient functions (preliminary version). In *26th Annual Symposium on Foundations of Computer Science, Portland, Oregon, USA, 21-23 October 1985*, pages 396–407, 1985.

[31] Eldon Chung, Maciej Obremski, and Divesh Aggarwal. Extractors: Low entropy requirements colliding with non-malleability. Technical report, arXiv, 2021.

[32] Gil 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.

[33] Gil 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.

[34] Gil Cohen. Non-malleable extractors - new tools and improved constructions. In *Proceedings of the 31st Annual IEEE Conference on Computational Complexity*, 2016.

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

[36] Gil Cohen. Two-source dispersers for polylogarithmic entropy and improved ramsey graphs. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 278–284. ACM, 2016.

[37] Gil 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.

[38] Gil Cohen. Towards optimal two-source extractors and ramsey graphs. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 1157–1170. ACM, 2017.

[39] Gil Cohen, Ran Raz, and Gil Segev. Non-malleable extractors with short seeds and applications to privacy amplification. *SIAM Journal on Computing*, 43(2):450–476, 2014.

[40] Gil Cohen and Leonard Schulman. Extractors for near logarithmic min-entropy. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[41] Gil Cohen and Igor Shinkar. The complexity of DNF of parities. In Madhu Sudan, editor, *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, Cambridge, MA, USA, January 14-16, 2016*, pages 47–58. ACM, 2016.

[42] Evgeny Demenkov and Alexander Kulikov. An elementary proof of 3n-o(n) lower bound on the circuit complexity of affine dispersers. In *Proceedings of the 36th international conference on Mathematical foundations of computer science*, pages 256–265, 2011.

[43] 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.

[44] Yevgeniy Dodis, Xin Li, Trevor D. Wooley, and David Zuckerman. Privacy amplification and non-malleable extractors via character sums. *SIAM Journal on Computing*, 43(2):800–830, 2014.

[45] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pages 601–610, 2009.

[46] Paul E. Dunne. Lower bounds on the complexity of 1-time only branching programs. In Lothar Budach, editor, *Fundamentals of Computation Theory, FCT '85, Cottbus, GDR, September 9-13, 1985*, volume 199 of *Lecture Notes in Computer Science*, pages 90–99. Springer, 1985.

[47] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu 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*, 2009.

[48] Zeev Dvir and Avi Wigderson. Kakeya sets, new mergers and old extractors. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, 2008.

[49] Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In *CRYPTO (2)*, pages 239–257, 2013.

[50] Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. In *ICS*, pages 434–452, 2010.

[51] P. Erdős. Some remarks on the theory of graphs. *Bulletin of the American Mathematics Society*, 53:292–294, 1947.

[52] Magnus Gausdal Find, Alexander Golovnev, Edward A. Hirsch, and Alexander S. Kulikov. A better-than-3n lower bound for the circuit complexity of an explicit function. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 89–98, 2016.

[53] Ariel Gabizon and Ran Raz. Deterministic extractors for affine sources over large fields. *Combinatorica*, 28(4):415–440, 2008.

[54] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM J. Comput.*, 36(4):1072–1094, 2006.

[55] Anna Gál. A simple function that requires exponential size read-once branching programs. *Inf. Process. Lett.*, 62(1):13–16, 1997.

[56] Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing*, 2018.

[57] Svyatoslav Gryaznov, Pavel Pudlák, and Navid Talebanfard. Linear Branching Programs and Directional Affine Extractors. In *37th Computational Complexity Conference (CCC 2022)*, volume 234, pages 4:1–4:16, 2022.

[58] Divya Gupta, Hemanta K. Maji, and Mingyuan Wang. Constant-rate non-malleable codes in the split-state model. Technical Report Report 2017/1048, Cryptology ePrint Archive, 2018.

[59] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. *Journal of the ACM*, 56(4):1–34, 2009.

[60] Stasys Jukna. Entropy of contact circuits and lower bounds on their complexity. *Theor. Comput. Sci.*, 57:113–129, 1988.

[61] Valentine Kabanets. Almost k-wise independence and hard boolean functions. *Theor. Comput. Sci.*, 297(1-3):281–295, 2003.

[62] Yael Kalai, Xin Li, and Anup Rao. 2-source extractors under computational assumptions and cryptography with defective randomness. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 617–628, 2009.

[63] Yael Tauman Kalai, Xin Li, Anup Rao, and David Zuckerman. Network extractor protocols. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 654–663, 2008.

[64] Jesse Kamp, Anup Rao, Salil P. Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. *Journal of Computer and System Sciences*, 77:191–220, 2011.

[65] Jesse Kamp and David Zuckerman. Deterministic Extractors for Bit-Fixing Sources and Exposure-Resilient Cryptography. *Siam Journal on Computing*, 36:1231–1247, 2007.

[66] 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.

[67] Bhavana Kanukurthi, Lakshmibhavana Obbattu, and Sruthi Sekar. Four-state non-malleable codes with explicit constant rate. In *Fifteenth IACR Theory of Cryptography Conference*, 2017.

[68] Matthias Krause, Christoph Meinel, and Stephan Waack. Separating the eraser turing machine classes l_e, nl_e, co-nl_e and p_e. *Theor. Comput. Sci.*, 86(2):267–275, 1991.

[69] Mark Lewko. An explicit two-source extractor with min-entropy rate near $4/9$. *Mathematika*, 65(4):950–957, 2019.

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

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

[72] Xin 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.

[73] Xin 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.

[74] Xin 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.

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

[76] Xin Li. Non-malleable condensers for arbitrary min-entropy, and almost optimal protocols for privacy amplification. In *12th IACR Theory of Cryptography Conference*, pages 502–531. Springer-Verlag, 2015. LNCS 9014.

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

[78] Xin 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.

[79] Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, 2017.

[80] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. In Amir Shpilka, editor, *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA*, volume 137 of *LIPIcs*, pages 28:1–28:49. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[81] C. J. Lu, Omer Reingold, Salil Vadhan, and Avi Wigderson. Extractors: Optimal up to constant factors. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 602–611, 2003.

[82] Ueli M. Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In *Advances in Cryptology — CRYPTO '97, 17th Annual International Cryptology Conference, Proceedings*, 1997.

[83] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[84] Stephen Ponzio. A lower bound for integer multiplication with read-once branching programs. *SIAM Journal on Computing*, 28(3):798–815, 1998.

[85] Anup Rao. Extractors for a constant number of polynomially small min-entropy independent sources. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[86] Anup Rao. Extractors for low-weight affine sources. In *Proc. of the 24th CCC*, 2009.

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

[88] Ran Raz and Amir Yehudayoff. Multilinear formulas, maximal-partition discrepancy and mixed-sources extractors. *Journal of Computer and System Sciences*, 77:167–190, 2011.

[89] Renato Renner and Stefan 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.

[90] Ronen Shaltiel. Dispersers for affine sources with sub-polynomial entropy. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, 2011.

[91] Amir Shpilka. Constructions of low-degree and error-correcting $\varepsilon$-biased generators. *Comput. Complex.*, 18(4):495–525, dec 2009.

[92] Janos Simon and Mario Szegedy. A new lower bound theorem for read-only-once branching programs and its applications. In *Advances In Computational Complexity Theory*, 1992.

[93] Luca Trevisan and Salil P. Vadhan. Extracting Randomness from Samplable Distributions. In *IEEE Symposium on Foundations of Computer Science*, pages 32–42, 2000.

[94] Emanuele Viola. Extractors for circuit sources. *SIAM J. Comput.*, 43(2):655–672, 2014.

[95] Ingo Wegener. On the complexity of branching programs and decision trees for clique functions. *J. ACM*, 35(2):461–471, 1988.

[96] Amir Yehudayoff. Affine extractors over prime fields. *Combinatorica*, 31(2):245–256, 2011.

[97] Stanislav Zák. An exponential lower bound for one-time-only branching programs. In Michal Chytil and Václav Koubek, editors, *Mathematical Foundations of Computer Science 1984, Praha, Czechoslovakia, September 3-7, 1984, Proceedings*, volume 176 of *Lecture Notes in Computer Science*, pages 562–566. Springer, 1984.

[98] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Theory of Computing*, pages 103–128, 2007.