Generalized Strong Extractors and Deterministic Privacy Amplification[*]

Robert König and Ueli Maurer

Department of Computer Science,
Swiss Federal Institute of Technology (ETH), Zurich,
CH-8092 Zurich, Switzerland
{rkoenig, maurer}@inf.ethz.ch

Abstract. Extracting essentially uniform randomness from a somewhat random source X is a crucial operation in various applications, in particular in cryptography where an adversary usually possesses some partial information about X. In this paper we formalize and study the most general form of extracting randomness in such a cryptographic setting. Our notion of strong extractors captures in particular the case where the catalyst randomness is neither uniform nor independent of the actual extractor input. This is for example important for privacy amplification, where a uniform cryptographic key is generated by Alice and Bob sharing some partially secret information X by exchanging a catalyst R over an insecure channel accessible to an adversary Eve. Here the authentication information for R creates, from Eve's viewpoint, a dependence between X and R. We provide explicit constructions for this setting based on strong blenders. In addition, we give strong deterministic randomness extractors for lists of random variables, where only an unknown subset of the variables is required to have some amount of min-entropy.

1 Introduction

1.1 Extracting Uniform Randomness

Extracting essentially uniform randomness from somewhat random information is an important operation arising in different settings, ranging from the derandomization of probabilistic algorithms, the design of pseudo-random generators, to the generation of information-theoretically secure cryptographic keys.

One can distinguish different variations of this problem, depending on whether the randomness extraction is deterministic or makes use of some catalyst randomness, and whether or not the generated random string must be protected from an adversary with side information, including the catalyst (cryptographic vs. non-cryptographic case).

Non-cryptographic randomness extraction has been studied extensively. A deterministic randomness extraction function $f : \Omega \to \Omega'$ is characterized by

[*] This work was partially supported by the Swiss National Science Foundation, project No. 200020-103847/1.

N.P. Smart (Ed.): Cryptography and Coding 2005, LNCS 3796, pp. 322–339, 2005.

the set $\mathcal{S}$ of random variables (often called a source) X for which it generates an essentially uniform output (e.g., has distance at most ε from the uniform distribution). Such a function is called an $(\mathcal{S}, \varepsilon)$-extractor[1] [Dod00]. The question of the existence of such extractors and the problem of finding explicit constructions has been considered for a large number of sources and remains an important research topic [TV00]. Examples include various kinds of "streaming" sources (e.g., [vN51, Eli72, Blu84, SV86, Vaz87b, Vaz87c]), which produce a sequence of symbols (as for example Markov sources), families consisting of pairs or tuples of independent weak random variables (e.g., [CG88, DO03, DEOR04]), families generated by samplers (e.g., [TV00]), and various kinds of bit-fixing and symbol-fixing sources (e.g., [CGH$^+$85, CW89, KZ03, GRS04]).

The term "extractor" is generally used for the probabilistic non-cryptographic case. In this case, the extractor takes as a second input an independent uniform random string R, which can be seen as a catalyst. The source X from which randomness is extracted is usually characterized by a lower bound on the min-entropy. A (k, ε)-extractor [NZ96] extracts ε-close to uniform randomness under the sole condition that the min-entropy of X is at least k.

Such a (k, ε)-extractor is called *strong* if the output is ε-close to uniform even when R is taken as part of the output. This is useful in a setting where R is communicated over an (authenticated) channel accessible to an adversary who should still be completely ignorant about the generated string. This setting, usually referred to as *privacy amplification*, is discussed in the following section.

Note that the concept of an $(\mathcal{S}, \varepsilon)$-extractor is a strict generalization of the concept of a (k, ε)-extractor if one views the catalyst as part of the input to the (then deterministic) extractor. In the same sense, the $(\mathcal{S}, \varepsilon)$-strong extractors defined in this paper are a strict generalization of (k, ε)-strong extractors. The output of an $(\mathcal{S}, \varepsilon)$-extractor is required to be close to uniform even given some additional piece of information, which does not necessarily have to be part of the input, but is characterized by the family $\mathcal{S}$.

1.2 Privacy Amplification

Classical privacy amplification, introduced by Bennett, Brassard, and Robert [BBR88] (and further analysed in [BBCM95]), refers to the following setting. Two parties Alice and Bob are connected by an authenticated but otherwise insecure communication channel, and they share a random variable X about which Eve has partial information, modeled by a random variable Y known to her.[2] The random variable X could for instance be the result of a quantum cryptography protocol or some other protocol.

Alice and Bob's goal is to generate an almost uniform random string S about which Eve has essentially no information, i.e., which is essentially uniform from

[1] Note that the term extractor usually refers to the probabilistic case, which is generally attributed to Nisan and Zuckerman [NZ96], see below.

[2] The setting is described by the joint probability distribution P_{XY} or, more precisely, by a set of such distributions. Actually, in the literature X is usually assumed to be a uniformly distributed bitstring, but this restriction is not necessary.

her point of view and can thus be used as a cryptographic key. This is achieved by Alice choosing a random string R, sending it to Bob (and hence also to Eve), and Alice and Bob applying a strong extractor (with catalyst randomness R) to obtain S. This works if the min-entropy of X, when given $Y = y$, is sufficiently high, with overwhelming probability over the values y that Y can take on. As mentioned above, the extractor must be strong since S must be uniform even when conditioned on R.

1.3 Contributions of This Paper

This privacy amplification setting can be unsatisfactory for two different reasons. First, in a practical setting, where the goal is to make as conservative and realistic assumptions as possible, one might worry that the catalyst randomness generated by one of the parties is neither uniform nor fully independent of X. Therefore, a natural question to ask is whether privacy amplification is possible without catalyst randomness, i.e., by a deterministic function. This problem is formalized in Section 3, where our new notion of strong extractors is introduced. We also provide a definition of strong condensers, which only guarantee some amount of min-entropy in the output and show how these concepts are related to each other.

A non-uniform and dependent catalyst can be seen as a special case of the above when viewed as part of the input to the (then deterministic) privacy amplification with two input random variables. In Section 4 we show that the amount of extractable uniform randomness is determined essentially by the difference of the amount of min-entropy and the degree of dependence. These results give rise to new sources allowing for conventional deterministic randomness extraction, in particular for *dependent* pairs of weak random variables, thus relaxing the independence condition needed in the constructions of [CG88, DO03, DEOR04]. As an important example, the type of dependence considered includes the case of outputs generated by a (hidden) Markov model, thus generalizing for example [Blu84].

In Section 5 we present strong extractors (or, equivalently, deterministic privacy amplification) for a setting where Alice and Bob share a list of random variables, some unknown subset of which contains sufficient min-entropy, and where the adversary also knows some unknown subset of them. Note that the problem of constructing such extractors was recently considered by Lee, Lu, Tsai and Tzeng [LLTT05] along with a different cryptographic application in the context of key agreement for group communication. One of our constructions is very similar to theirs, though our analysis is different. We stress that the problem considered here is different from the problem of constructing extractors for several independent sources (each having a specific amount of min-entropy). Concerning the latter problem, there has recently been a significant breakthrough by Barak et al. [BIW04].

A second generalization of standard privacy amplification is to get rid of the need for an authenticated communication channel between Alice and Bob.[3] In this case, the shared random variable X must also be used to authenticate the catalyst R, in addition to being the input to the extraction procedure. This creates a dependence, from the adversary's viewpoint, between X and R, thus requiring the use of our generalized strong extractors. This setting has been considered before [MW97, DO03, RW03]. Our results lead to a more general and modular treatment with simpler proofs, but this application is not discussed here.

1.4 Outline

Section 2 introduces some basic concepts used throughout the paper. We then present our general definition of strong extractors and strong condensers in Section 3.1, and show how these primitives and some of their basic properties are related to privacy amplification. In Section 3.2, we discuss how our definition of a strong extractor generalizes various known definitions of randomness extractors. In Section 3.3 we establish a relation between strong extractors and strong condensers. In Section 4, we show how to construct $(\mathcal{S}, \varepsilon)$-strong extractors for a non-trivial family $\mathcal{S}$ which consists of dependent pairs of random variables. Finally, in Section 5, we show how to construct strong extractors for tuples of independent weak random variables with certain properties.

2 Preliminaries

For $n \in \mathbb{N}$ we denote by $[n]$ the set $\{1, \ldots, n\}$. If $x = (x_1 \cdots x_n) \in \{0,1\}^n$ is a bitstring and $S \subset [n]$ a set of indices, we write $x|_S$ for the concatenation of the bits x_i with $i \in S$.

We will denote by $\mathcal{P}(\Omega)$ the set of probability distributions[4] on an alphabet Ω. Moreover, $\mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2) \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ will be the set of independent distributions on $\Omega_1 \times \Omega_2$. If $X_1, X_2 \in \mathcal{P}(\Omega)$, we write $P_{X_1} \equiv P_{X_2}$ if $P_{X_1}(x) = P_{X_2}(x)$ for all $x \in \Omega$. For $(X_1, X_2) \in \mathcal{P}(\Omega_1 \times \Omega_2)$, the distribution $X_1 \times X_2 \in \mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2)$ is defined by $P_{X_1 \times X_2}(x_1, x_2) := P_{X_1}(x_1) P_{X_2}(x_2)$ for all $(x_1, x_2) \in \Omega_1 \times \Omega_2$.

The statistical distance between two distributions P and P' over the same alphabet Ω is defined as $d(P, P') := \frac{1}{2} \sum_{z \in \Omega} |P(z) - P'(z)|$. Note that the statistical distance satisfies

$$d(P_1 \times Q, P_2 \times Q) = d(P_1, P_2) \tag{1}$$

and is strongly convex, i.e.,

$$d\big(\sum_i \lambda_i P_i, \sum_i \lambda_i Q_i\big) \leq \sum_i \lambda_i d(P_i, Q_i) \qquad \text{if } \lambda_i \geq 0 \text{ and } \sum_i \lambda_i = 1. \tag{2}$$

[3] Actually, in most realistic settings the channel is completely insecure and authenticity must be guaranteed by use of a pre-distributed short secret key.

[4] The terms random variable and probability distribution will be used interchangeably.

Let $U_\Omega \in \mathcal{P}(\Omega)$ denote a random variable with uniform distribution on Ω. A random variable $Z \in \mathcal{P}(\Omega)$ is ε-close to uniform if $d(Z, U_\Omega) \leq \varepsilon$.

For a set $\mathcal{S} \subset \mathcal{P}(\Omega)$ of probability distributions, let

$$\mathfrak{B}^\varepsilon(\mathcal{S}) := \{X \in \mathcal{P}(\Omega) \mid \exists Y \in \mathcal{S} : d(X,Y) \leq \varepsilon\}$$

denote the distributions which are ε-close to some distribution in $\mathcal{S}$. We write $\overline{\mathcal{S}}$ for the convex hull of $\mathcal{S}$, i.e., the set of distributions which can be written as a convex combination of distributions in $\mathcal{S}$.

For $(X,Y) \in \mathcal{P}(\Omega_1 \times \Omega_2)$, we define the min-entropy of X and the conditional min-entropy of X given Y as follows[5]:

$$H_\infty(X) := -\log_2(\max_x P_X(x)) \qquad H_\infty(X|Y) := \min_y H_\infty(X|Y=y).$$

We call a random variable X for which only a lower bound on its min-entropy is known (i.e., $H_\infty(X) \geq k$ for some k) a *weak* random variable. Finally, we use[6]

$$H_0(X) := \log_2(|\operatorname{supp}(X)|)$$

to measure the size of the support $\operatorname{supp}(X) := \{x \in \Omega_1 \mid P_X(x) > 0\}$ of X.

We will use the following property of the statistical distance, which we state as a lemma.

Lemma 1. *Let $(S,Y) \in \mathcal{P}(\Omega_1 \times \Omega_2)$ be an arbitrary pair of random variables. Then there exists[7] a random variable S' which is uniformly distributed on Ω_1, independent of Y, and satisfies $\mathbb{P}[S = S'] \geq 1 - d\big((S,Y), U_{\Omega_1} \times Y\big)$.*

Proof. Let S_y be a random variable with distribution $P_{S_y} \equiv P_{S|Y=y}$ and let $d_y := d(S_y, U_{\Omega_1})$. We use the following well-known fact.

For an arbitrary random variable $T \in \mathcal{P}(\Omega)$, there exists a random variable T' defined by a channel[8] $P_{T'|T}$ with the property that T' is uniformly distributed on Ω and $\mathbb{P}[T = T'] = 1 - d(T, U_\Omega)$. Applying this to S_y, we conclude that there exists a random variable S'_y defined by a channel $P_{S'_y|S_y}$ such that

$$\mathbb{P}[S'_y = S_y] = 1 - d_y \qquad \text{and } P_{S'_y} = P_{U_{\Omega_1}}.$$

Let us define S' by the conditional distributions $P_{S'|Y=y,S=s} := P_{S'_y|S_y=s}$ for all $(s,y) \in \Omega_1 \times \Omega_2$. Then we obtain for all $(s',y) \in \Omega_1 \times \Omega_2$

$$P_{S'|Y=y}(s') = \sum_{s \in \Omega_1} P_{S|Y=y}(s) P_{S'|Y=y,S=s}(s') = \sum_{s \in \Omega_1} P_{S_y}(s) P_{S'_y|S_y=s}(s') = P_{S'_y}(s')$$

[5] $H_\infty(X|Y=y)$ is to be understood as the min-entropy of the conditional distribution $P_{X|Y=y}$.

[6] For $0 < \alpha < \infty$, $\alpha \neq 1$, the *Rényi entropy of order* α is defined as $H_\alpha(X) := \frac{1}{1-\alpha} \log_2 \left(\sum_x P_X(x)^\alpha\right)$. The quantities $H_0(X)$ and $H_\infty(X)$ are obtained by taking the limits $\alpha \to 0$ and $\alpha \to \infty$, respectively.

[7] "exists" is to be interpreted as follows: One can define a new random experiment with random variables S, S', and Y such that P_{SY} is equal in both experiments and such that S' satisfies the stated conditions.

[8] This means that T and T' are jointly distributed according to $P_{TT'}(t,t') := P_T(t) P_{T'|T=t}(t')$.

which implies that S' is indeed uniform and independent of Y. Moreover, we have $\mathbb{P}[S = S'|Y = y] = \sum_{s\in\Omega_1} P_{S|Y=y}(s)P_{S'|Y=y,S=s}(s) = \sum_{s\in\Omega_1} P_{S_y}(s)P_{S'_y|S_y=s}(s)$, hence $\mathbb{P}[S = S'|Y = y] = \mathbb{P}[S_y = S'_y]$ and $\mathbb{P}[S = S'|Y = y] = 1 - d_y$. But $\mathop{\mathbb{E}}_{y\leftarrow Y}[d_y] = d\big((S, Y), U_{\Omega_1} \times Y\big)$. The statement now follows from

$$\mathop{\mathbb{E}}_{y\leftarrow Y}\big[\mathbb{P}[S = S'|Y = y]\big] = \mathbb{P}[S = S'].$$

3 Strong Extraction for General Families of Random Variables

3.1 Basic Definitions and the Relation to Privacy Amplification

In the general setting of privacy amplification described in the introduction, the two parties (possibly after communicating first) have a shared random string X, whereas the adversary holds some information about X which is summarized by a random variable Y. It is important to note that Y does not necessarily have to be a part of X, but may depend in some other more intricate way on X. As an example, if Alice and Bob used X to authenticate some message M using a MAC, then Eve might learn $Y = (M, MAC_X(M))$. As a consequence, we may usually only assume that the pair (X, Y) has some specific structure (depending on the particular setting), i.e., it is contained in some family $\mathcal{S}$ of distributions. The question is then what Alice and Bob have to do in order to *deterministically* extract a key S from X which is uniform from the point of view of the adversary.

According to Lemma 1, if for the extracted key S, the quantity $d\big((S, Y), U_{\Omega'} \times Y\big)$ is small, then S is with high probability identical to a perfectly uniformly distributed "ideal" key which is independent of the part Y known to the adversary. This motivates the following general definition.

Definition 1. *Let $\mathcal{S} \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ be a set of probability distributions on $\Omega_1 \times \Omega_2$. A function* Ext : $\Omega_1 \to \Omega'$ *is an* $(\mathcal{S}, \varepsilon)$-strong extractor *if for every pair $(X, Y) \in \mathcal{S}$,*

$$d\big((\mathrm{Ext}(X), Y), U_{\Omega'} \times Y\big) \leq \varepsilon.$$

Using this new terminology, Alice and Bob simply have to apply an appropriate $(\mathcal{S}, \varepsilon)$-strong extractor in order to obtain the desired result. The following lemma describes some intuitive properties of strong extractors which follow directly from the definition and properties of the statistical distance.

Lemma 2. *An $(\mathcal{S}, \varepsilon)$-strong extractor is*

(i). *an $(\overline{\mathcal{S}}, \varepsilon)$-strong extractor.*
(ii). *a $(\mathfrak{B}^\delta(\mathcal{S}), \varepsilon + \delta)$-strong extractor for every $\delta \geq 0$.*
(iii). *an $(\mathcal{S}', \varepsilon)$-strong extractor for the family of distributions*

$$\mathcal{S}' := \big\{(X, (Y, Z)) \bigm| P_{(X,Y)|Z=z} \in \mathcal{S} \textit{ for all } z \in \mathrm{supp}(Z)\big\} \ .$$

(iv). *an $(\mathcal{S}', \varepsilon)$-strong extractor for the family of distributions*

$$\mathcal{S}' := \big\{(X, Z) \bigm| (X, Y) \in \mathcal{S}, X \leftrightarrow Y \leftrightarrow Z\big\},$$

where the notation $X \leftrightarrow Y \leftrightarrow Z$ means that X, Y and Z form a Markov chain.

Property (i) expresses the obvious fact that an $(\mathcal{S}, \varepsilon)$-strong extractor also works on any convex combination of distributions in $\mathcal{S}$. Property (ii) implies that in the context of privacy amplification, Alice and Bob can obtain an almost perfect secret key even if the initial situation is only close to a situation for which the extractor is appropriate. Property (iii) states that any additional piece of information Z does not help the adversary if conditioned on every value that Z can take on, the extracted key is close to uniform from the adversary's view. Finally, Property (iv) asserts that the extracted key still looks uniform to the adversary even if he processes his piece of information Y to obtain some different random variable Z.

As a weakening of the concept of extractors, it is natural to consider also the concept of condensers (see e.g., [RSW00]). In the privacy amplification setting, this corresponds to a situation where Alice and Bob would like to obtain a key which has a large amount of min-entropy from the point of view of the adversary. This may be used for example in an authentication protocol. We are thus led to the following analogous definition.

Definition 2. *Let $\mathcal{S} \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ be a set of probability distributions on $\Omega_1 \times \Omega_2$. A function* Cond : $\Omega_1 \to \Omega'$ *is an* $(\mathcal{S}, k, \varepsilon)$-strong condenser *if for every $(X, Y) \in \mathcal{S}$, there exists a random variable S such that*

$$d\big((\mathrm{Cond}(X), Y), (S, Y)\big) \leq \varepsilon \qquad \textit{and} \quad H_\infty(S|Y) \geq k.$$

3.2 Relation to Known Definitions

In this section, we show that known definitions of extractors are in fact special instances of $(\mathcal{S}, \varepsilon)$-strong extractors. Let us begin with the following general notion of deterministic[9] extractors, first introduced by Dodis.

Definition 3 ([Dod00]). *Let $\mathcal{S} \subset \mathcal{P}(\Omega)$ be a set of probability distributions on Ω. A function* Ext : $\Omega \to \Omega'$ *is an* $(\mathcal{S}, \varepsilon)$-extractor *if for every $X \in \mathcal{S}$, $d(\mathrm{Ext}(X), U_{\Omega'}) \leq \varepsilon$.*

Obviously, such an extractor corresponds to an $(\mathcal{S}', \varepsilon)$-extractor for the family $\mathcal{S}' := \{(X, \perp) \mid X \in \mathcal{S}\}$ where $\perp$ denotes a constant random variable. Note that an $(\mathcal{S}, \varepsilon)$-strong extractor is an $(\mathcal{S}, \varepsilon)$-extractor according to Definition 3, a fact which follows from Property (iv) of Lemma 2.

Our definition also generalizes the concept of strong (k, ε)-extractors, which are defined as follows.

[9] In this paper, we generally use the term *deterministic* to refer to procedures which (contrary to probabilistic ones) do not require a seed consisting of truly random bits. Note, however, that a probabilistic extractor can be seen as a deterministic one in the sense of Definition 3.

Definition 4 ([NZ96]). *A strong (k,ε)-extractor is a function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *such that for every $X \in \mathcal{P}(\{0,1\}^n)$ with $H_\infty(X) \geq k$,*

$$d\big((\mathrm{Ext}(X,R),R),U_{\{0,1\}^m} \times R\big) \leq \varepsilon.$$

A strong (k,ε)-extractor is an $(\mathcal{S}',\varepsilon)$-extractor for the family of distributions $\mathcal{S}' \subset \mathcal{P}\big((\{0,1\}^n \times \{0,1\}^d) \times \{0,1\}^d\big)$ given by

$$\mathcal{S}' := \left\{\big((X,R),R\big) \;\middle|\; H_\infty(X) \geq k,\ R \text{ independent of } X \text{ and } P_R \equiv P_{U_{\{0,1\}^d}}\right\}.$$

Similarly, our concept generalizes the so-called strong blenders introduced by Dodis and Oliveira in [DO03]. To describe this type of strong extractors, it is convenient to introduce the following families of distributions, which we also use in Sections 4 and 5.

Definition 5 ([CG88]). *The set*$\mathbf{CG}(\Omega_1,)[k_1,\Omega_2]k_2$ *of so-called* Chor-Goldreich-sources *is the set of all pairs $(X_1,X_2) \in \mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2)$ of independent random variables such that $H_\infty(X_1) \geq k_1$ and $H_\infty(X_2) \geq k_2$.*

The set $\mathbf{CG}(\Omega_1,\Omega_2)[k]$ is the set of all pairs of independent random variables $(X_1,X_2) \in \mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2)$ satisfying $H_\infty(X_1X_2) \geq k$. Furthermore, we define $\mathbf{CG}(\Omega)[k] := \mathbf{CG}(\Omega,\Omega)[k]$.

To simplify the notation, we will sometimes refer to the set $\{0,1\}^n$ simply by n in these two definitions. For example, we will write $\mathbf{CG}(n_1,n_2)[k_1,k_2]$ instead of $\mathbf{CG}(\{0,1\}^{n_1},\{0,1\}^{n_2})[k_1,k_2]$.

Definition 6 ([DO03]). *A* (k_1,k_2,ε)-strong blender *is a function* $\mathrm{Ble} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ *such that*

$$d\big((\mathrm{Ble}(X,Y),Y),U_{\{0,1\}^m} \times Y\big) \leq \varepsilon$$

for all pairs $(X,Y) \in \mathbf{CG}(n_1,n_2)[k_1,k_2]$.

With our new notion of strong extraction, a (k_1,k_2,ε)-strong blender is an $(\mathcal{S},\varepsilon)$-strong extractor for the special family of distributions

$$\mathcal{S} := \big\{\big((X,Y),Y\big) \;\big|\; (X,Y) \in \mathbf{CG}(n_1,n_2)[k_1,k_2]\big\}. \tag{3}$$

We will reconsider strong blenders in Section 4.1. Note that in the definition of the family $\mathcal{S}$, the random variables X and Y are independent. In Section 4.2, we show how to construct strong extractors even in the case where X and Y depend on each other.

As already mentioned, our definition of strong extractors also applies to the more general situation where the "public" information Y given to the adversary is not simply a part of X. In particular, this is the case when it is unavoidable to leak certain additional information about X, for instance if we would like to provide some error tolerance with respect to X. A general solution to this problem is accurately modeled by the concept of fuzzy extractors introduced by Dodis, Reyzin and Smith (see [DRS04] for details). It is easy to see that our definition of strong extractors also generalizes these fuzzy extractors.

3.3 Strong Condensers from Strong Extractors

Intuitively, in the setting of privacy amplification, if Alice and Bob derive a secret key S by applying a strong extractor, this key will still have a high amount of min-entropy from the point of view of the adversary even if the adversary is given a (short) additional piece of information. This means that an $(\mathcal{S}, \varepsilon)$-strong extractor is in fact a strong condenser for a different family $\mathcal{S}'$, which models the situation where the adversary gets additional information. This is formally expressed by Lemma 4.

The proof relies on the following technical result, which appears in a more general form in [MW97] and is implicitly used in [NZ96]. For an arbitrary pair $(X, Z) \in \mathcal{P}(\Omega_1 \times \Omega_2)$ of random variables and $\delta > 0$,

$$\mathop{\mathbb{P}}_{z \leftarrow Z}\big[H_\infty(X|Z = z) \geq H_\infty(X) - H_0(Z) - \log_2 \tfrac{1}{\delta}\big] \geq 1 - \delta \ . \tag{4}$$

We reformulate this statement in a way which is more useful for our purpose.

Lemma 3. *Let $(S, Y, Z) \in \mathcal{P}(\Omega_1 \times \Omega_2 \times \Omega_3)$ be arbitrary random variables and let $\delta > 0$. Then there exists a random variable S' defined by a channel $P_{S'|YZ}$ such that*

$$H_\infty(S'|(Y, Z)) \geq H_\infty(S|Y) - H_0(Z) - \log_2 \tfrac{1}{\delta} \qquad \textit{and} \tag{5}$$

$$d\big((S, Y, Z), (S', Y, Z)\big) \leq \delta. \tag{6}$$

Proof. Let us define the set

$$\Gamma_\delta := \{(y, z) \in \Omega_2 \times \Omega_3 \mid H_\infty(S|Y = y, Z = z) \geq H_\infty(S|Y) - H_0(Z) - \log_2 \tfrac{1}{\delta}\}.$$

Then by identity (4),

$$\mathop{\mathbb{P}}_{z \leftarrow Z|Y=y}\big[(y, z) \in \Gamma_\delta\big] \geq 1 - \delta \ . \tag{7}$$

We define S' by

$$P_{S'|Y=y,Z=z} := \begin{cases} P_{S|Y=y,Z=z} & \text{if } (y, z) \in \Gamma_\delta \\ P_{U_{\Omega_1}} & \text{otherwise.} \end{cases}$$

Statement (5) is now a consequence of the definition of Γ_δ.

As the quantity $d(P_{S|Y=y,Z=z}, P_{S'|Y=y,Z=z})$ is at most 1 if $(y, z) \notin \Gamma_\delta$ and 0 otherwise, we conclude, using (7), that

$$\sum_{z \in \Omega_3} P_{Z|Y=y}(z) d(P_{S|Y=y,Z=z}, P_{S'|Y=y,Z=z}) \leq \mathop{\mathbb{P}}_{z \leftarrow Z|Y=y}[(y, z) \notin \Gamma_\delta] \leq \delta \ .$$

Statement (6) then follows from

$$d\big((S, Y, Z), (S', Y, Z)\big) = \mathop{\mathbb{E}}_{y \leftarrow Y}\big[\sum_{z \in \Omega_3} P_{Z|Y=y}(z) d(P_{S|Y=y,Z=z}, P_{S'|Y=y,Z=z})\big] \ .$$

Lemma 3 allows us to derive the main result of this section.

Lemma 4. *Let* Ext : $\Omega_1 \to \{0,1\}^{n_0}$ *be an* $(\mathcal{S},\varepsilon)$*-strong extractor and let* $\delta > 0$*. Then* Ext *is an* $(\mathcal{S}',k,\varepsilon+\delta)$*-strong condenser for the family*

$$\mathcal{S}' := \left\{ \big(X,(Y,Z)\big) \;\middle|\; (X,Y) \in \mathcal{S} \text{ and } H_0(Z) \leq n_0 - k - \log_2 \tfrac{1}{\delta} \right\} .$$

Proof. Let $S := \mathrm{Ext}(X)$. Then by Lemma 1 there is a random variable S' which is uniformly distributed and independent of Y such that $\mathbb{P}[S = S'] \geq 1-\varepsilon$. In particular, we have $H_\infty(S'|Y) = n_0$. Therefore, by Lemma 3, there is a random variable S'' such that

$$H_\infty(S''|(Y,Z)) \geq n_0 - H_0(Z) - \log_2 \tfrac{1}{\delta} \geq k$$

and

$$d\big((S',(Y,Z)),(S'',(Y,Z))\big) \leq \delta \ .$$

The statement now follows from the triangle inequality of the statistical distance and the fact that

$$d\big((S,(Y,Z)),(S',(Y,Z))\big) \leq \varepsilon$$

which holds because $\mathbb{P}[S = S'] \geq 1-\varepsilon$.

This result is implicitly used for example in a protocol by Renner and Wolf [RW03] for privacy amplification over a non-authenticated channel. Without elaborating this any further, we point out that our generalized concepts of condensers and extractors allow to simplify existing security proofs such as the one given in [RW03].

4 Strong Extraction with a Weak and Dependent Catalyst

An important special case of our generalized notion of (deterministic) strong extractors is when the input can be seen as consisting of two parts, an actual input X_1 and a non-uniform and dependent catalyst X_2 which is also part of the output. In this section we introduce a dependence measure for such pairs (X_1,X_2) of random variables and show that the amount of uniform randomness extractable from (X_1,X_2) is determined essentially by the difference of the min-entropies of X_1 and X_2 and the level of dependence. In Section 4.1 we reformulate the definition of strong blenders and in Section 4.2 we state our main result concerning strong extraction from dependent variables. The dependence measure we consider is defined as follows.

Definition 7. *The set of* m-independent distributions $\mathcal{I}_m(\Omega_1,\Omega_2) \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ *is the set of all pairs* (X_1,X_2) *of random variables on* $\Omega_1 \times \Omega_2$ *which can be written as a convex combination of* m *independent distributions, i.e.,*

$$\mathcal{I}_m(\Omega_1,\Omega_2) := \left\{ \sum_{i\in[m]} \lambda_i P_i \times Q_i \;\middle|\; \forall i \in [m] : \lambda_i \geq 0, P_i \in \mathcal{P}(\Omega_1), Q_i \in \mathcal{P}(\Omega_2) \right\}$$

The dependence index *of a pair of random variables* $(X_1, X_2) \in \mathcal{P}(\Omega_1 \times \Omega_2)$ *is defined as the quantity*

$$dep(X_1, X_2) := \log_2 \left(\min\{ m \in \mathbb{N} \mid (X_1, X_2) \in \mathcal{I}_m(\Omega_1, \Omega_2) \} \right) .$$

Obviously, $\mathcal{I}_m(\Omega_1, \Omega_2) \subset \mathcal{I}_{m'}(\Omega_1, \Omega_2)$ for $m < m'$ and $dep(X_1, X_2) = 0$ if and only if X_1 and X_2 are independent. Note that an m-independent distribution is for example obtained by observing the output of a (hidden) Markov source with at most m states at subsequent time steps.

4.1 Strong Blenders

Strong blenders can be used to perform privacy amplification when Alice and Bob have a pair (X, Y) of independent weak random variables and the adversary is given Y (compare e.g., [DO03]). To model this situation using strong extractors, it is convenient to use a "copying" operator[10] **cc** which transforms a pair of random variables (X, Y) into a pair $((X, Y), Y)$. This models the fact that the adversary is given Y.

Note that this operator has the following simple property which can be verified by direct calculation. If $\sum_i \mu_i = 1$ with $\mu_i \geq 0$ and $P_i \in \mathcal{P}(\Omega_1 \times \Omega_2)$ for all i, then

$$\mathbf{cc}\Big(\sum_i \mu_i P_i\Big) = \sum_i \mu_i \, \mathbf{cc}(P_i) \tag{8}$$

With this definition, the family of distributions $\mathbf{cc}(\mathbf{CG}(n_1, n_2)[k_1, k_2])$ is exactly the family given in equation (3). In other words, a (k_1, k_2, ε)-strong blender according to Definition 6 is a $(\mathbf{cc}(\mathbf{CG}(n_1, n_2)[k_1, k_2]), \varepsilon)$-strong extractor. In the sequel, we will use the terms strong blender and $(\mathbf{cc}(\mathbf{CG}(n_1, n_2)[k_1, k_2]), \varepsilon)$-strong extractor interchangeably, depending on whether or not we would like to refer to the parameters explicitly.

Recently, new results concerning extraction from independent weak sources were obtained by Barak et al. ([BIW04, BKS$^+$05]) and Raz[11] [Raz05]. We use these extractors in Section 4.2.

4.2 m-Independence and Strong Extraction

The following lemma states that every pair (X_1, X_2) of random variables is close to a convex combination of independent random variables having some specific amount of min-entropy which depends on $dep(X_1, X_2)$. An analogous statement holds for the pair $((X_1, X_2), X_2)$.

[10] Formally, the copying operator $\mathbf{cc} : \mathcal{P}(\Omega_1 \times \Omega_2) \to \mathcal{P}((\Omega_1 \times \Omega_2) \times \Omega_2)$ is defined as follows. If $P_{X_1 X_2} \in \mathcal{P}(\Omega_1 \times \Omega_2)$ and $P := \mathbf{cc}(P_{X_1 X_2})$, then $P((x_1, x_2), x_3) := P_{X_1 X_2}(x_1, x_2) \cdot \delta_{x_2, x_3}$ for all $x_i \in \Omega_i$, $i = 1, \ldots, 3$, where δ_{x_2, x_3} denotes the Kronecker-delta, which equals 1 if $x_2 = x_3$ and 0 otherwise.

[11] In particular, Raz [Raz05] presents $(\mathbf{CG}(n, n)[k_1, k_2], \varepsilon)$-extractors for parameters $k_1 = (\frac{1}{2} + \delta)n$ and $k_2 = \Theta(\log n)$ where $\delta > 0$ is an arbitrarily small constant.

Lemma 5. *Let* $(X_1, X_2) \in \mathcal{P}(\Omega_1 \times \Omega_2)$ *and* $\delta_1, \delta_2 > 0$ *be arbitrary and define*

$$k_i = H_\infty(X_i) - dep(X_1, X_2) - \log_2 \tfrac{1}{\delta_i} \qquad \textit{for } i = 1, 2.$$

Then we have

$$\begin{aligned} (X_1, X_2) &\in \mathfrak{B}^{\delta_1+\delta_2}(\overline{\mathbf{CG}(\Omega_1, \Omega_2)[k_1, k_2]}) \qquad \textit{and} \\ \mathbf{cc}(X_1, X_2) &\in \mathfrak{B}^{\delta_1+\delta_2}(\overline{\mathbf{cc}(\mathbf{CG}(\Omega_1, \Omega_2)[k_1, k_2])})\ . \end{aligned}$$

Proof. Let $m := 2^{dep(X_1, X_2)}$. Then there is a distribution $(X_1', X_2', Z) \in \mathcal{P}(\Omega_1 \times \Omega_2 \times [m])$ such that $P_{X_1X_2} \equiv P_{X_1'X_2'} \equiv \sum_{z \in [m]} P_Z(z) P_{X_1'|Z=z} P_{X_2'|Z=z}$. For $i = 1, 2$, applying identity (4) to the pair (X_i', Z) shows that there are two subsets $\mathcal{A}_1, \mathcal{A}_2 \subseteq [m]$ and distributions $\{P_j^i\}_{j \in [m]} \subset \mathcal{P}(\Omega_i)$ for $i = 1, 2$ such that $P_{X_1X_2}$ has the form $P_{X_1X_2} \equiv \sum_{j \in [m]} \lambda_j P_j^1 \times P_j^2$, where for every $i = 1, 2$, $\sum_{j \in \mathcal{A}_i} \lambda_j \geq 1 - \delta_i$ as well as $H_\infty(P_j^i) \geq k_i$ for all $j \in \mathcal{A}_i$. In particular, we may write

$$P_{X_1X_2} \equiv \sum_{j \in \mathcal{A}_1 \cap \mathcal{A}_2} \lambda_j Q_j + \big(1 - \sum_{j \in \mathcal{A}_1 \cap \mathcal{A}_2} \lambda_j\big) Q \tag{9}$$

for some distributions $\{Q_j\}_{j \in \mathcal{A}_1 \cap \mathcal{A}_2} \in \mathbf{CG}(\Omega_1, \Omega_2)[k_1, k_2]$ and a distribution $Q \in \mathcal{P}(\Omega_1 \times \Omega_2)$, where $(1 - \sum_{j \in \mathcal{A}_1 \cap \mathcal{A}_2} \lambda_j) \leq \delta_1 + \delta_2$. By the strong convexity (2) of the statistical distance, the first statement follows. The second statement follows similarly by application of $\mathbf{cc}$ to both sides of (9), using (8) and (2).

Using Lemma 5, Lemma 2 and an explicit construction by Raz [Raz05], we immediately obtain[12] a strong extractor with a weak and dependent catalyst. Theorem 1 gives the exact parameters. Intuitively, it expresses the fact that the amount of required min-entropy rises with the amount of dependence there is. In the setting of privacy amplification, this result states that there is an (explicit) deterministic function which Alice and Bob can use to extract a secret key S in a situation where they initially share a pair of weak random variables (X_1, X_2), and where X_2 is known to the adversary. We emphasize that contrary to the setting analysed in [DO03], X_2 need not be completely independent of X_1.

Theorem 1. *For any parameters* n, m, κ_1, κ_2 *and* $0 < \delta < \frac{1}{2}$ *satisfying*

$$\begin{aligned} \kappa_1 &\geq \big(\frac{1}{2} + \delta\big) \cdot n + 4 \log_2 n \\ \kappa_2 &\geq 5 \log_2(n - \kappa_1) \\ m &\leq \delta \cdot \min\{\frac{n}{8}, \frac{\kappa_1}{40}\} - 1, \end{aligned}$$

where n *is sufficiently large, there exists an (explicit)* $(\mathcal{S}^m_{\kappa_1,\kappa_2}, \varepsilon)$*-strong extractor* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *for the family of distributions*

$$\mathcal{S}^m_{\kappa_1,\kappa_2} := \Big\{\big((X_1, X_2), X_2\big) \mid H_\infty(X_i) - dep(X_1, X_2) \geq \kappa_i + \frac{3}{2}m \textit{ for } i = 1, 2\}.$$

[12] Note that our techniques also apply, e.g., to the constructions provided by Barak et al. ([BIW04, BKS+05]).

with error $\varepsilon := 3 \cdot 2^{-\frac{3}{2}m}$*. Moreover, this extractor is also strong in the first input, i.e., it is an* $(\widehat{\mathcal{S}^m_{\kappa_1,\kappa_2}}, \varepsilon)$*-strong extractor for the family of distributions*

$$\widehat{\mathcal{S}^m_{\kappa_1,\kappa_2}} := \big\{ \big((X_1, X_2), X_1\big) \mid \big((X_1, X_2), X_2\big) \in \mathcal{S}^m_{\kappa_1,\kappa_2} \big\}.$$

5 Strong Extractors for the Family $\mathcal{T}^N_\Omega(k)$

In this section, we consider the following family of random variables, which is somewhat related to symbol-fixing sources (see [CGH$^+$85, CW89, KZ03, GRS04]) since the "positions" having "good" randomness are unknown.

Definition 8. *For an* N*-tuple of random variables* $(X_1, \ldots, X_N) \in \mathcal{P}(\Omega^N)$ *and a subset* $A \subset [N]$*, let* $X|_A$ *denote the concatenation of those random variables* X_i *with* $i \in A$*. The family* $\mathcal{T}^N_\Omega(k)$ *is the set of all pairs of the form* $\big((X_1, \ldots, X_N), X|_A\big)$ *where* $(X_1, \ldots, X_N) \in \mathcal{P}(\Omega)^n$ *are independent random variables and* $A \subset [N]$ *is such that there exists two distinct indices* $i, j \in [N]$ *with* $j \notin A$ *and* $H_\infty(X_i X_j) \geq k$*.*

In the privacy amplification setting, this corresponds to a situation where Alice and Bob have a sequence of independent random variables and the adversary obtains a subsequence. The only thing guaranteed is that two of these random variables (say X_i and X_j) have joint min-entropy at least[13] k and at least one of the two (say X_j) is unknown to the adversary. Note that extractors for this family have been used for other applications than privacy amplification [LLTT05].

We give two new constructions for strong extractors for the family $\mathcal{T}^n_\Omega(k)$. The first construction is based on special group-theoretic strong blenders. It is presented in Section 5.1. The second construction is very similar to a recent construction due to Lee, Lu, Tsai and Tzeng [LLTT05] and related to the construction of strong blenders presented in [DEOR04]. Our proof proceeds along the lines of similar derivations in [DO03] and [DEOR04].

5.1 Group-Theoretic Extractors for the Family $\mathcal{T}^N_\Omega(k)$

Theorem 2 below shows how a $(\mathcal{T}^N_\Omega(k), \varepsilon)$-strong extractor can be constructed in a generic way, using the following simple observation. We omit the trivial proof.

Lemma 6. *Let* $(\mathcal{G}, +)$ *be a group and let* $X_1, X_2 \in \mathcal{P}(\mathcal{G})$ *be two independent random variables defined on* $\mathcal{G}$*. Then* $H_\infty(X_1 + X_2) \geq \max\{H_\infty(X_1), H_\infty(X_2)\}$*.*

Intuitively, this lemma says that taking a random step according to X_2 on the Cayley graph defined by the group $\mathcal{G}$, starting from a random position defined

[13] Note that usually, we have $k > \log_2 |\Omega|$, implying that both X_i and X_j must have some amount of min-entropy individually. As pointed out in the introduction, this problem is different than the extraction problem considered, e.g., by Barak et al. [BIW04, BKS$^+$05].

by X_1, we end up with an element that is at least as random as the variable which contains more randomness. This observation allows us to give a generic construction of a $(\mathcal{T}_{\Omega}^{N}(k), \varepsilon)$-strong extractor, solving this problem in a (at least conceptually) similar manner as the way Kamp and Zuckerman [KZ03] treat the problem of randomness extraction from symbol-fixing sources.

Theorem 2. *Let $(\mathcal{G}, +)$ be an abelian group and let $\mathrm{Ext} : \mathcal{G} \times \mathcal{G} \to \Omega$ be a $(\mathbf{cc}(\mathbf{CG}(\mathcal{G})[k]), \varepsilon)$-strong extractor of the form $\mathrm{Ext}(x_1, x_2) := f(x_1 + x_2)$. Then the function $F(x_1, \ldots, x_N) := f(\sum_{i \in [N]} x_i)$ is a $(\mathcal{T}_{\mathcal{G}}^{N}(k), \varepsilon)$-strong extractor.*

Proof. Let $\big((X_1, \ldots, X_N), X|_A\big) \in \mathcal{T}_{\mathcal{G}}^{N}(k)$. Suppose $i, j \in [N]$ are such that $i \in A$, $j \notin A$ and $H_\infty(X_i X_j) \geq k$. Then $F(X_1, \ldots, X_N) = \mathrm{Ext}\,(X', Y')$ where $X' := \sum_{\ell \in [N] \setminus A} X_\ell$ is independent of $Y' := \sum_{\ell \in A} X_\ell$ and $H_\infty(X') + H_\infty(Y') \geq k$ by assumption and Lemma 6. Because $(X_1, \ldots, X_N) \leftrightarrow X|_A \leftrightarrow Y'$ is a Markov chain, the statement follows from Lemma (iv). The case where $i \notin A$ is treated similarly.

Using the following function $\mathrm{Ext} : \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_{n_0}$ (originally proposed in [CG88] and later shown to be a strong blender in [DO03])

$$\mathrm{Ext}(x_1, x_2) := \begin{cases} \log_g(x_1 + x_2 \mod p) \mod n_0 & \text{if } x_1 + x_2 \neq 0 \mod p \\ 0 & \text{otherwise,} \end{cases}$$

where $p > 2$ is a prime, g a generator of $\mathbb{Z}_p^*$ and n_0 a divisor of $p - 1$, Theorem 2 immediately gives a $\big(\mathcal{T}_{\mathbb{Z}_p}^{n}(\log_2 p + 2\log_2(\frac{1}{\varepsilon}) + 2\log_2 n_0), \varepsilon\big)$-strong extractor for every $\varepsilon \geq \frac{2}{p}$. Note that for appropriately chosen parameters p, g and n_0, the resulting extractor is efficiently computable (see [CG88] for details).

5.2 More Extractors for the Family $\mathcal{T}_{\Omega}^{N}(k)$

It is easy to see that any $(\mathbf{cc}(\mathbf{CG}(n)[k]), \varepsilon)$-strong extractor which is symmetric in its arguments is a $(\mathcal{T}_{\{0,1\}^n}^{2}(k), \varepsilon)$-strong extractor. This, combined with the result by [DO03] that the inner product is a strong blender gives a very simple $(\mathcal{T}_{\{0,1\}^n}^{2}(k), \varepsilon)$-strong extractor.

Lemma 7 ([DO03]). *The inner product modulo 2, denoted $\langle \cdot, \cdot \rangle : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, is a $(\mathcal{T}_{\{0,1\}^n}^{2}(k), \varepsilon)$-strong extractor, where $\log \frac{1}{\varepsilon} = \frac{k-n}{2} + 1$.*

A slight extension of this statement allows us to construct a $(\mathcal{T}_{\{0,1\}^n}^{2}(k), \varepsilon)$-strong extractor which extracts just a single bit. The construction given here is more general than necessary (the parameters a, b, c could be omitted), but allows to prove Lemma 9 more easily.

Lemma 8. *Let M be an invertible $n \times n$ matrix over $GF(2)$ and let $a, b \in \{0,1\}^n$ and $c \in \{0,1\}$ be arbitrary. Define the function $\mathrm{Ext}_M^{a,b,c}(x_1, x_2) := \langle x_1, M x_2 \rangle + \langle a, x_1 \rangle + \langle b, x_2 \rangle + c$ where addition is modulo 2. Then the function $\mathrm{Ext}_M^{a,b,c}$ is a $(\mathcal{T}_{\{0,1\}^n}^{2}(k), \varepsilon)$-extractor, where $\log(\frac{1}{\varepsilon}) = \frac{k-n}{2} + 1$.*

In the following proofs, we use the *non-uniformity* $\delta(Z) := d(Z, U_\Omega)$ to denote the distance of a distribution $Z \in \mathcal{P}(\Omega)$ from the uniform distribution on Ω.

Proof. We have to prove that

$$\mathop{\mathbb{E}}_{x_2 \leftarrow X_2}[\delta(\mathrm{Ext}_M^{a,b,c}(X_1, x_2))] \leq \varepsilon \qquad \text{and} \qquad \mathop{\mathbb{E}}_{x_1 \leftarrow X_1}[\delta(\mathrm{Ext}_M^{a,b,c}(x_1, X_2))] \leq \varepsilon \tag{10}$$

for all pairs $(X_1, X_2) \in \mathbf{CG}(k)[n]$ with k as specified. Since the operation of adding a constant is a bijection, we have for every fixed $x_2 \in \{0,1\}^n$

$$\delta(\mathrm{Ext}_M^{a,b,c}(X_1, x_2)) = \delta(\mathrm{Ext}_M^{a,b,0}(X_1, x_2)) = \delta(\mathrm{Ext}_M^{a,0,0}(X_1, x_2)) \ .$$

Hence, as $\mathrm{Ext}_M^{a,0,0}(x_1, x_2) = \langle x_1, Mx_2 \rangle + \langle a, x_1 \rangle = \langle x_1, Mx_2 + a \rangle$, we obtain

$$\mathop{\mathbb{E}}_{x_2 \leftarrow X_2}[\delta(\mathrm{Ext}_M^{a,b,c}(X_1, x_2))] = \mathop{\mathbb{E}}_{x_2 \leftarrow X_2}[\delta(\langle X_1, Mx_2 + a \rangle)] \ .$$

Define the random variable X_2' by $X_2' := MX_2 + a$. Since the mapping $x_2 \mapsto Mx_2 + a$ is a bijection, we have

$$\mathop{\mathbb{E}}_{x_2 \leftarrow X_2}[\delta(\langle X_1, Mx_2 + a \rangle)] = \mathop{\mathbb{E}}_{x_2' \leftarrow X_2'}[\delta(\langle X_1, x_2' \rangle)] \ .$$

This combined with the fact that $H_\infty(X_2') = H_\infty(X_2)$ and Lemma 7 proves the first inequality in (10). The second inequality then follows from the first with the identity $\mathrm{Ext}_M^{a,b,c}(x_1, x_2) = \mathrm{Ext}_{M^T}^{b,a,c}(x_2, x_1)$ and the fact that the transpose M^T is invertible if M is invertible.

Lemma 8 allows us to obtain a $(\mathcal{T}_{\{0,1\}^n}^N(k), \varepsilon)$-strong extractor for $N > 2$ which again extracts only a single bit.

Lemma 9. *Let M be an invertible $n \times n$-matrix over $GF(2)$ and let* $\mathrm{Ext}_M : (\{0,1\}^n)^N \to \{0,1\}$ *be the function* $\mathrm{Ext}_M(x_1, \ldots, x_N) := \sum_{s<t} \langle x_s, Mx_t \rangle$, *where Mx_t is the matrix-vector multiplication over $GF(2)$ and addition is modulo 2. Then* Ext_M *is a $(\mathcal{T}_{\{0,1\}^n}^N(k), \varepsilon)$-strong extractor, where* $\log(\frac{1}{\varepsilon}) = \frac{k-n}{2} + 1$.

Proof. Suppose that $\big((X_1, \ldots, X_N), X|_A\big) \in \mathcal{T}_{\{0,1\}^n}^N(k)$ and let $i, j \in [N]$ be such that $i \in A$, $j \notin A$ and $H_\infty(X_i X_j) \geq k$. Without loss of generality, we may assume that $i < j$, since $\big((X_{\pi(1)}, \ldots, X_{\pi(N)}), X|_A\big) \in \mathcal{T}_{\{0,1\}^n}^N(k)$ for every permutation $\pi \in S_N$. A straightforward calculation shows (compare [LLTT05]) that we can write $\mathrm{Ext}_M(x_1, \ldots, x_N) = \mathrm{Ext}_M^{a,b,c}(x_i, x_j)$ with a, b, c depending only on the variables x_ℓ with $\ell \notin \{i, j\}$. [14] Hence we have

$$\begin{aligned} &d((\mathrm{Ext}_M(X_1, \ldots, X_N), X|_A), (U_{\{0,1\}}, X|_A)) = \\ &\mathop{\mathbb{E}}_{\tilde{x} \leftarrow X_{A\setminus\{i\}}} \mathop{\mathbb{E}}_{x_i \leftarrow X_i}[\delta(\mathrm{Ext}_M^{a(\tilde{x}), b(\tilde{x}), c(\tilde{x})}(x_i, X_j))] \end{aligned}$$

[14] More precisely, a, b and c are given by the expressions

$$\begin{aligned} a &:= M \textstyle\sum_{t>i, t\neq j} x_t + M^T \sum_{s<i} x_s, \\ b &:= M \textstyle\sum_{t>j} x_t + M^T \sum_{s<j, s\neq i} x_s \\ c &:= \quad \textstyle\sum_{s<t \in [m]\setminus\{i,j\}} \langle x_s, Mx_t \rangle \end{aligned}$$

for appropriately defined functions a, b, c from $(\{0,1\}^n)^{|A|-1}$ to $\{0,1\}^n$ and $\{0,1\}$, respectively. The claim then follows from Lemma 8.

To get an extractor which produces several bits, we use the following reformulation of Vaziranis parity lemma [Vaz87a].

Lemma 10. *Let $\mathcal{S} \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ and let $\mathrm{Ext} : \Omega_1 \to \{0,1\}^m$ be such that the function $x \mapsto \langle v, \mathrm{Ext}(x)\rangle$ is an $(\mathcal{S}, \varepsilon)$-strong extractor for every $v \in \{0,1\}^m \backslash \{0^m\}$. Then Ext is an $(\mathcal{S}, 2^m \cdot \varepsilon)$-strong extractor.*

Proof. We use the following so-called parity lemma by Vazirani [Vaz87a]. For every $A \in \mathcal{P}(\{0,1\}^m)$, the non-uniformity $\delta(A)$ of A is bounded as follows: $\delta(A) \leq \sum_{v \in \{0,1\}^m \backslash \{0^m\}} \delta(\langle v, A\rangle)$. It implies that for any pair of random variables $(A, B) \in \mathcal{P}(\{0,1\}^m \times \Omega_2)$,

$$\begin{aligned} d((A,B),(U_{\{0,1\}^m},B)) &= \mathop{\mathbb{E}}_{b \leftarrow B}[\delta(A|_{B=b})] \\ &\leq \sum_{v \in \{0,1\}^m \backslash \{0^m\}} \mathop{\mathbb{E}}_{b \leftarrow B}[\delta(\langle v, A|_{B=b})]\rangle \\ &= \sum_{v \in \{0,1\}^m \backslash \{0^m\}} d((\langle v, A\rangle, B),(U_{\{0,1\}}, B)), \end{aligned}$$

where the linearity of the expectation was used. Applying this to $(A, B) = (\mathrm{Ext}(X), Y)$ with $(X, Y) \in \mathcal{S}$ immediately yields the claim.

Finally, this implies the main result of this section. Note that the efficient construction of suitable matrices $M_1, \ldots, M_m$ in the following theorem is discussed in [DEOR04].

Theorem 3. *Let $M_1, \ldots, M_m$ be $n \times n$-matrices over $GF(2)$ such that for every non-empty subset $I \subset [m]$ the matrix $\sum_{i \in I} M_i$ is invertible. Then the function $\mathrm{Ext} : (\{0,1\}^n)^N \to \{0,1\}^m$ defined by*

$$\mathrm{Ext}(x_1, \ldots, x_n) = \Big(\sum_{s<t} \langle x_s, M_1 x_t\rangle, \ldots, \sum_{s<t} \langle x_s, M_m x_t\rangle\Big)$$

is a $(\mathcal{T}^N_{\{0,1\}^n}(k), \varepsilon)$-strong extractor, where $\log(\frac{1}{\varepsilon}) = \frac{k-n}{2} - m + 1$.

Proof. Let $v \in \{0,1\}^m \backslash \{0^m\}$ and $I(v) := \{i \in [m] \mid v_i = 1\}$. Then

$$\langle v, \mathrm{Ext}(x_1, \ldots, x_n)\rangle = \sum_{s<t} \langle x_s, (\sum_{i \in I(v)} M_i) x_t\rangle = \mathrm{Ext}_{M(v)}(x_1, \ldots, x_n) \ ,$$

where $M(v) := \sum_{i \in I(v)} M_i$ is invertible by assumption and $\mathrm{Ext}_{M(v)}$ is defined as in Lemma 9. Hence, by Lemma 9, the function $(x_1, \ldots, x_n) \mapsto \langle v, \mathrm{Ext}(x_1, \ldots, x_n)\rangle$ is a $(\mathcal{T}^N_{\{0,1\}^n}(k), \varepsilon)$-strong extractor for every $v \in \{0,1\}^m \backslash \{0^m\}$, where $\log(\frac{1}{\varepsilon}) = \frac{k-n}{2} + 1$. Lemma 10 then implies the desired result.

References

[ABH$^+$86] M. Ajtai, L Babai, P. Hajnal, J. Komlos, P. Pudlak, V. Rodl, E. Szemeredi, and G. Turan. Two lower bounds for branching programs. In *ACM Symposium on Theory of Computing*, pages 30–38, 1986.

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

[BBR88] C. Bennett, G. Brassard, and J. Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, 1988.

[BIW04] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness from few independent sources. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2004.

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

[Blu84] M. Blum. Independent unbiased coin flips from a correlated biased source: a finite state markov chain. *IEEE Symposium on the Foundations of Computer Science*, 1984.

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

[CGH$^+$85] B. Chor, O. Goldreich, J. Håstad, J. Freidmann, S. Rudich, and R. Smolensky. The bit extraction problem or t-resilient functions. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 1985.

[CW89] A. Cohen and A. Wigderson. Dispersers, deterministic amplification, and weak random sources (extended abstract). In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 14–19, 1989.

[DEOR04] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz. Improved randomness extraction from two independent sources. *International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, August 2004.

[DO03] Y. Dodis and R. Oliveira. On extracting private randomness over a public channel. *International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 143–154, August 2003.

[Dod00] Y. Dodis. *Exposure-Resilient Cryptography.* PhD thesis, Massachussetts Institute of Technology, August 2000.

[DRS04] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *Advances in Cryptology — EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 523–539, May 2004.

[DSS01] Y. Dodis, A. Sahai, and A. Smith. On perfect and adaptive security in exposure-resilient cryptography. *Lecture Notes in Computer Science, EUROCRYPT '01*, 2045:301–324, 2001.

[Eli72] P. Elias. The efficient construction of an unbiased random sequence. *Annals of Mathematics Statistics*, 43(3):865–870, 1972.

[GRS04] A. Gabizon, R. Raz, and R. Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2004.

[KZ03] J. Kamp and D. Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. In *IEEE Symposium on Foundations of Computer Science*, 2003.

[LLTT05] C.J. Lee, C.J. Lu, S.C. Tsai, and W.G. Tzeng. Extracting randomness from multiple independent sources. *IEEE Transaction on Information Theory*, 51(6):2224–2227, June 2005.

[MW97] U. Maurer and S. Wolf. Privacy amplification secure against active adversaries. In *Advances in Cryptology — CRYPTO '97*, volume 1294 of *Lecture Notes in Computer Science*, pages 307–321, August 1997.

[NZ96] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[Raz05] R. Raz. Extractors with weak random seeds. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 11–20, 2005.

[RSW00] O. Reingold, R. Shaltiel, and A. Wigderson. Extracting randomness via repeated condensing. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 22–31, 2000.

[RW03] R. Renner and S. Wolf. Unconditional authenticity and privacy from an arbitrarily weak secret. In *Advances in Cryptology — CRYPTO 2003*, volume 2729 of *Lecture Notes in Computer Science*, pages 78–95, August 2003.

[Sak96] M. Saks. Randomization and derandomization in space-bounded computation. In *SCT: Annual Conference on Structure in Complexity Theory*, 1996.

[Sha02] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the European Association for Theoretical Computer Science*, 77:67–95, June 2002.

[SV86] M. Santha and U.V. Vazirani. Generating quasi-random sequences from slightly random sources. *Journal of Computer and System Sciences*, 33:75–87, 1986.

[TV00] L. Trevisan and S. P. Vadhan. Extracting randomness from samplable distributions. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 32–42, 2000.

[Vaz87a] U. Vazirani. Strong communcation complexity or generating quasi-random sequences from two communicating semi-random sources. *Combinatorica*, 7(4):375–392, 1987.

[Vaz87b] U. V. Vazirani. Efficiency considerations in using semi-random sources. In *Proceedings of the nineteenth annual ACM conference on Theory of computing*, pages 160–168, 1987.

[Vaz87c] U. V. Vazirani. Strong communication complexity or generating quasirandom sequences from two communicating semirandom sources. *Combinatorica*, 7(4):375–392, 1987.

[vN51] J. von Neumann. Various techniques used in connection with random digits. *Applied Math Series*, 12:36–38, 1951.

[Zuc90] D. Zuckerman. General weak random sources. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 534–543, 1990.

[Zuc91] D. Zuckerman. Simulating BPP using a general weak random source. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 79–89, 1991.