Improving the Hadamard extractor

Jan Bouda\textsuperscript{a}, Matej Pivoluska\textsuperscript{a}, Martin Plesch\textsuperscript{a,b,*}

\textsuperscript{a} Faculty of Informatics, Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic
\textsuperscript{b} Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 11 Bratislava 45, Slovak Republic

Article history:
Received 20 September 2011
Received in revised form 26 July 2012
Accepted 30 July 2012
Communicated by M. Kiwi

Keywords:
Weak randomness
Randomness extractor
Hadamard matrix

Abstract

In this paper we construct a strong randomness extractor with two independent \(\ell\)-bit input distributions with min entropies \(b_X\), \(b_Y\), \(b_X + b_Y > \ell\) (the probability of any particular output is upper bounded by \(2^{-b_X}\) and \(2^{-b_Y}\), respectively). For \(b_X\), \(b_Y \leq \ell - 1\), our extractor produces one bit which is by the factor of \(\sqrt{2}\) closer to the uniform distribution, when compared to the Hadamard extractor. What is more, this distance drops to zero if at least one of the min entropies raises to \(\ell\). This is in sharp contrast to the Hadamard extractor which fails to produce even a single unbiased bit, even if one of the input distributions is uniform. We also extend our construction to produce \(k\) bits of output with a bias that is by the factor of \(\sqrt{3}/2\) smaller than that of the corresponding Hadamard extractor and retains the ability to produce unbiased bits if one of the input distributions is uniform. The strongness property of the extractor is maintained in both cases, however, in the multi-bit scenario the bias is increased by the factor of \(\sqrt{3}/2\).

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

True randomness is a valuable resource. Many tasks in cryptography and computation in general require uniform random sources to work properly. In spite of extensive research scientists have been unsuccessful in building a device that produces true randomness out of first principles and is fully independent of outside environment. This is the main reason why a long line of research has been devoted to design efficient procedures to transform “weak sources of randomness”, which do occur in nature, into random sources that are close to uniform.

The first idea of a randomness extractor is due to von Neumann [32], who showed how to produce unbiased bits from a string of independent coin tosses with unknown fixed bias. Later Blum [5] considered sources generated by finite-state Markov chains. Santha and Vazirani [26], Vazirani [30], Dodis et al. [11], Barak et al. [2], Barak et al. [1], Raz [25], Rao [22], Bourgain [6], Shaltiel [27], Li [19] and Barak et al. [3] considered extraction methods from several independent sources that contain “enough randomness”. Chor et al. [9], Ben-or and Linial [4], Cohen and Wigderson [10], Kamp and Zuckerman [18] and Gabizon et al. [13] studied sources, which are uniform on a subset of bits. Later these sources were generalized to so called affines sources, which were studied by Rao [24], Gabizon and Raz [15], Barak et al. [1], Bourgain [7], Yehudayoff [33] and Li [20]. Trevisan and Vadhan [29] and Kamp et al. [17] found extractors for efficiently samplable distributions and Viola [31] considered extractors for circuit sources.

In our work we will consider \((\ell, b)\)-sources, i.e. sources that output bit strings of length \(\ell\) and the probability of any particular string is upper bounded by \(2^{-b}\). Such sources have been introduced in [8] and are widely considered in the literature. Our method will be used to extract randomness from two independent sources \(X, Y\) of the same length with min entropies \(b_X\), \(b_Y\), respectively.

* Corresponding author at: Faculty of Informatics, Masaryk University, Botanická 68a, 602 00 Brno, Czech Republic. Tel.: +420 549 49 5247.
E-mail addresses: bouda@fi.muni.cz (J. Bouda), mmpivoluska@mali.muni.cz (M. Pivoluska), martin.plesch@savba.sk, plesch@savba.sk (M. Plesch).

0304-3975/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
The basic tool to design extractors of this type is to consider the extraction of a single bit, i.e. a function \( f : \{0, 1\}^\ell \times \{0, 1\}^\ell \mapsto \{0, 1\} \). The function \( f \) takes two samples from sources \( X, Y \) as the input. The goal is to produce a single bit that is as close to uniform as possible. To show that a particular function is an extractor, consider its matrix representation. Each row and column of such a matrix is labeled by a string of length \( \ell \). The element in row labeled by \( x \) and column labeled by \( y \) contains the number \(-1 \) if \((x, y) \) is a Hadamard extractor.

Various authors [26,8,11] have shown that the dot product function \( x \cdot y = (\sum_i x_i \cdot y_i) \mod 2 \) is a good one-bit extractor for two independent sources. The matrix representation of the dot product modulo 2 is a Hadamard matrix, and thus this function is often referred to as a Hadamard extractor. A Hadamard matrix is a matrix such that each two rows (columns) are mutually orthogonal, and it can be shown that the sum of elements of each minor is relatively small. The disadvantage of Hadamard matrices is that the amount of elements 1 is not equal to the amount of elements \(-1\) in the matrix and it turns out that this property introduces unnecessary bias to the output distribution, especially for high min entropies (representing high quality sources of randomness). The Hadamard extractor and its extension to multiple bit output are used in more sophisticated constructions that combine different kinds of extractors to achieve better parameters (see for example [1,14]).

What is more, as shown in [23], the celebrated extractor of Bourgain [6] can be viewed as a two step procedure: firstly encoding the input distributions in order to raise the min-entropy of their subset, and subsequently using the Hadamard extractor to obtain bits that are close to uniform.

Our work is a direct continuation of the line of works [26,8,11]. We design a two-source extractor for one bit which performs better than the Hadamard extractor considering the bias of the output bit from a uniform distribution, preserving its strongness property. This is achieved by modifying the function \( x \cdot y = (\sum_i x_i \cdot y_i) \mod 2 \) so that it sums the first bits \( x_1 \) and \( y_1 \) instead of multiplying them. Using the XOR lemma, we show how the suggested extractor can be utilized to construct a two source extractor with a \( k \) bit output. This construction decreases the distance of the output bits from the uniform distribution compared to the Hadamard construction and still preserves the strongness property of the original extractor, however with a bias increased by a constant factor. Apart from the better quality of the output, its main advantage is the simplicity of its construction. Our extractor can fully substitute the Hadamard extractor in the constructions that are not based on its strongness property, including Bourgain’s extractor. For \((\ell, b)\)-sources with \( b \leq \ell - 1 \) we obtain at least a \( \sqrt{\epsilon} \) smaller bias than the \( k \) bit Hadamard extractor analyzed in [11]. Moreover, the bias approaches zero as one of the min entropies approached \( \ell \).

The paper is organized as follows. In Section 2 we introduce necessary notation and preliminary lemmas. In Section 3 we design a one-bit extractor with two input sources \( X, Y \) and an output distribution over \( \{0, 1\} \). In Section 4, we extend our construction for more output bits.

2. Preliminaries

2.1. Notation

Various models of non-uniform sources have been proposed in the literature so far. Throughout the paper we will use the model proposed by Chor and Goldreich [8], which characterizes a random source \( X \) over \( \{0, 1\}^\ell \) via its min-entropy \( H_\infty \):

\[
H_\infty(X) = \min_{x \in \{0, 1\}^\ell} \left( -\log_2 \left( \Pr(X = x) \right) \right).
\]

A random variable \( X \) is an \((\ell, b)\)-source if \( H_\infty(X) \geq b \). The quality of the output string is quantified in terms of statistical distance. For two random variables \( X, Y \) over \( \{0, 1\}^\ell \), their statistical distance is defined as:

\[
\|X, Y\| = \frac{1}{2} \sum_{x \in \{0, 1\}^\ell} |\Pr[X = x] - \Pr[Y = x]|.
\]  

(1)

Let \( U_\ell \) be a uniformly distributed random variable over \( \{0, 1\}^\ell \). A random variable \( X \) taking values in \( \{0, 1\}^\ell \) is \( \epsilon \)-close to uniform if \( \|X, U_\ell\| \leq \epsilon \). An \((\ell, b)\)-source \( X \) is flat, if there exists \( S \subseteq \{0, 1\}^\ell \), \( |S| = 2^b \) such that for each \( x \in S \): \( \Pr[X = x] = 2^{-b} \) and for every \( x' \not\in S \): \( \Pr[X = x'] = 0 \). Chor and Goldreich [8] showed that to analyze the bias of the output distribution of a function \( F \), we need to consider only flat distributions, since the worst case behavior of functions is obtained by them.

2.2. Extractors

Let us define a few primitives used in extracting randomness from non-uniform sources. We begin with a definition of a (seeded) extractor (see for example [21,28]):

**Definition 1.** Let \( b > 0 \) and \( \epsilon > 0 \). Then a function \( E : \{0, 1\}^\ell \times \{0, 1\}^m \mapsto \{0, 1\}^k \) is a \((b, \epsilon)\)-extractor, if for all \((\ell, b)\)-sources \( X \)

\[
\|E(X, U_m), U_k\| \leq \epsilon.
\]
Theorem 1. Suppose $F$ is represented by $m \in \mathbb{F}$, and for arbitrary $S_X, S_Y \subseteq \{0, 1\}^\ell$, $|S_X| = 2^b_X$, $|S_Y| = 2^b_Y$, $|\sum_{x \in S_X, y \in S_Y} m_{x,y}| \leq N$. Then $F$ is a $(b_X, b_Y, \epsilon)$-two source extractor.

Proof. Let $X, Y$ be independent $(\ell, b_X), (\ell, b_Y)$ distributions that are flat on $S_X, S_Y$, respectively. Elements of $S_X, S_Y$ uniquely determine the $2^b_X \times 2^b_Y$ minor $m_{x,y}|_{x \in S_X,y \in S_Y}$ of $M_F$. As we are choosing from both $S_X$ and $S_Y$ uniformly, each element $m_{x,y}$ of such a minor has the same probability $\frac{1}{|S_X||S_Y|} = \frac{1}{2^{b_X+b_Y}}$ to be produced. Without loss of generality let us assume that at least half of the elements of $m_{x,y}$ is equal to 1 (the sum is non-negative). Utilizing the condition $\sum_{x \in S_X, y \in S_Y} m_{x,y} \leq N$, the probability that the function outputs 0 is bounded by $\frac{N}{2} + \frac{N/2}{2^{b_X+b_Y}}$. Using (1) we can conclude that

$$\|F(X,Y), U_1\| \leq \frac{N}{2^{b_X+b_Y+1}}. \quad \square$$

Lemma 2. Suppose $F$ is represented by $M_F = \{m_{x,y}|_{x,y=0}\}$ and for arbitrary $S_X, S_Y \subseteq \{0, 1\}^\ell$, $|S_X| = 2^b_X$, $|S_Y| = 2^b_Y$, $\sum_{y \in S_Y} |\sum_{x \in S_X} m_{x,y}| \leq N$. Then $F$ is a $(b_X, b_Y, \epsilon)$-two source extractor.

Proof. To show that a function $F$ with matrix representation $M_F$ is a $(b_X, b_Y, \epsilon)$-two source extractor, we need to show that for arbitrary flat $(\ell, b_X), (\ell, b_Y)$-sources $X, Y$

$$\|F(X,Y), (Y, U_1)\| = \frac{1}{|S_Y|} \sum_{y \in S_Y} \|F(X,y), U_1\| \leq \epsilon. \quad (2)$$

The term $\sum_{y \in S_Y} \|F(X,y), U_1\|$ is the bias of the probability distribution on the output of $F$ after the column $y$ has been chosen and it can be computed from the sum $\sum_{x \in S_X} m_{x,y}$. Suppose the column $y$ contains more elements 1 than $\epsilon$.

Then the probability that $F$ will output 0, given the column $y$ was chosen, is $\frac{1}{2} + \frac{1}{2^b_Y} \sum_{x \in S_X} m_{x,y}$. In other words

$$\|F(X,y), U_1\| = \frac{1}{2^b_Y} \sum_{x \in S_X} m_{x,y}.$$
2.4. Hadamard matrix

Hadamard matrix $H = \{h_{ij}\}_{i,j=0}^{2^\ell-1}$ is a square matrix with elements $\pm 1$. Each row (column) of the matrix $H$ is orthogonal to every other row (column) of the matrix. In terms of matrix tensor product, Hadamard matrices $H_\ell$ of size $2^\ell \times 2^\ell$, $\ell \geq 1$ can be constructed by Sylvester method as follows:

$$H_\ell = H_1 \otimes H_2 \otimes \cdots \otimes H_1,$$

where $H_1 = \left( \begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array} \right)$ and the matrix $H_\ell = H_1^{\otimes \ell}$ is obtained from higher-rank tensor by flattening. Let $x = (x_1, \ldots, x_\ell)$, $y = (y_1, \ldots, y_\ell) \in \{0, 1\}^\ell$ and $F(x, y) = x \cdot y = (x_1 \cdot y_1) + (x_2 \cdot y_2) + \cdots + (x_\ell \cdot y_\ell)$, where all the operations are taken modulo 2 (i.e. dot product modulo 2). Then $M_\ell$ is a Hadamard matrix of the form given by the previous construction.

Lindsay’s lemma combined with Lemma 1 shows that Hadamard matrices $H_\ell$ of size $2^\ell \times 2^\ell$ are good two source extractors for $\ell$-bit input distributions and 1-bit output distribution:

**Lemma 3** (Lindsay’s). Let $H = \{h_{ij}\}_{i,j=0}^{2^\ell-1}$ be a $2^\ell \times 2^\ell$ Hadamard matrix and $S_X, S_Y, |S_X| = s, |S_Y| = s$ subsets of $\{0, \ldots, 2^\ell - 1\}$ corresponding to the choices of rows and columns. Then

$$\left| \sum_{j \in S_Y} \sum_{i \in S_X} h_{ij} \right| \leq H(\ell, s, s) = \sqrt{2^{2\ell} s^2}.$$

By Lindsay’s lemma and Lemma 1, the dot product modulo 2 is a $(b_X, b_Y, 2^{\ell-2})$ two source extractor with one bit output.

Dodis and Oliveira [12] presented a stronger form of Lindsay’s lemma, which allows to study the strongness property of extractors by matrix representations:

**Lemma 4.** Let $H = \{h_{ij}\}_{i,j=0}^{2^\ell-1}$ be a $2^\ell \times 2^\ell$ Hadamard matrix, and $S_X, S_Y, |S_X| = s, |S_Y| = s$ subsets of $\{0, \ldots, 2^\ell - 1\}$ corresponding to the choices of rows and columns. Then

$$\sum_{j \in S_Y} \left| \sum_{i \in S_X} h_{ij} \right| \leq H(\ell, s, s).$$

By Lemmas 4 and 2, the dot product modulo 2 is a $(b_X, b_Y, 2^{\ell-2})$ strong two source extractor with one bit output.

3. The use of the first bit

In order to improve the performance of the Hadamard extractor, we are going to analyze functions on a string of bits that are equal to the scalar product up to the operation $\odot : \{0, 1\}^2 \rightarrow \{0, 1\}$ on bits $x_1$ and $y_1$, i.e. functions of the form:

$$F_\odot(x, y) = (x_1 \odot y_1) + (x_2 \cdot y_2) + \cdots + (x_\ell \cdot y_\ell).$$

Our aim is to optimize $\odot$ for the extraction quality. Let $M_\odot$ be a $2 \times 2$ matrix with rows and columns labeled by $\{0, 1\}$ and elements defined by $m_{a,b} = (-1)^{a \odot b}$. It is easy to see that $M_{F_\odot} = \{m_{a,y}\}_{a,y=0}^{2^\ell-1}$ has the form

$$M_{F_\odot} = M_\odot \otimes H_{\ell-1}.$$

Let us denote

$$F_\odot(\ell, s, s) = \max \left\{ \sum_{j \in S_Y} \sum_{x \in S_X} m_{x,y} \right\},$$

and

$$\delta_\odot(\ell, s, s) = \max \left\{ \sum_{j \in S_Y} \sum_{x \in S_X} m_{x,y} \right\},$$

where both maximizations are performed over all $S_X, S_Y$, such that $|S_X| = s, |S_Y| = s$ and size of $M_{F_\odot}$ is equal to $2^\ell \times 2^\ell$. It is easy to see that $F_\odot \leq \delta_\odot$.

The block form of $M_{F_\odot}$ will consist of four Hadamard matrices, each of them being either $H_{\ell-1}$ or $-H_{\ell-1}$. With this knowledge we will be able to use Lindsay’s lemma and its stronger form to find upper bounds on $F_\odot$ and $\delta_\odot$. In the rest of this section we perform an analysis for different cases of function $\odot$. 

The operation $\circ$ outputs 1 on three inputs and 0 on one input, or it outputs 0 on three inputs and 1 on one input.

Notice that each row (column) of the matrix $M_{F_{\circ}}$ is orthogonal to every other row (column), i.e., it is a Hadamard matrix and Lindsay’s lemma and Lemma 4 apply. Thus in this case

$$F_{\circ}(\ell, x, y) \leq\delta_{\circ}(\ell, x, y) \leq H(\ell, x, y)$$

and by Lemma 2, $F_{\circ}$ is a $(b_X, b_Y, 2^\ell - b_X - b_Y)$-strong two source extractor, which implies that it is also a $(b_X, b_Y, 2^\ell - b_X - b_Y - 1)$-two source extractor.

Operation $\circ$ is constant.

That is $M_{F_{\circ}}$ has one of two possible block representations:

$$\begin{pmatrix} H_{\ell-1} & H_{\ell-1} \\ -H_{\ell-1} & -H_{\ell-1} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -H_{\ell-1} & -H_{\ell-1} \\ H_{\ell-1} & H_{\ell-1} \end{pmatrix}.$$  

Since the block representation of $M_{F_{\circ}}$ consists of four identical Hadamard matrices, it is easy to see that

$$F_{\circ}(\ell, x, y) \leq\delta_{\circ}(\ell, x, y) \leq 4H(\ell - 1,\frac{x}{2},\frac{y}{2}) = H(\ell + 1, x, y)$$

and by Lemma 2 $F_{\circ}$ is a $(b_X, b_Y, 2^\ell - b_X - b_Y - 1)$-strong two source extractor.

Operation $\circ$ is balanced and has a diagonal form.

$M_{F_{\circ}}$ has one of the following forms:

$$\begin{pmatrix} H_{\ell-1} & -H_{\ell-1} \\ -H_{\ell-1} & H_{\ell-1} \end{pmatrix}, \quad \begin{pmatrix} -H_{\ell-1} & H_{\ell-1} \\ H_{\ell-1} & -H_{\ell-1} \end{pmatrix}.$$  

Without loss of generality, let us consider the first one of them. The choice of the rows $S_X$ can contain rows of two types. For each selected row $(x_0,\ldots,x_{\ell-1})$ it holds that either the selection contains the opposite row $(-x_0,\ldots,-x_{\ell-1})$ as well, or all other rows in the selection are orthogonal to $(x_0,\ldots,x_{\ell-1})$. In the former case the opposite rows cancel out and their contribution to the sum of the minor elements is 0. It is easy to see that if $S_X$ contains $m$ such opposite row pairs, the choice of rows $S_X, S_Y \setminus \{ s \}$ is equivalent to the choice of rows $S_X, |S_X| = s, X, |S_X| = \frac{s}{2} - 2m$ of only mutually orthogonal rows. Note that there are at most $2^{\ell-1}$ mutually orthogonal rows (this will be crucial for min-entropies higher than $\ell - 1$).

We obtained that the worst-case selection of $s_X$ rows (regardless of the column selection) in the original matrix is not worse than selection of $s_X$ rows in the matrix

$$H'_{\ell-1}, -H'_{\ell-1},$$  

where all rows of $H'_{\ell-1}$ are the corresponding rows of either $H_{\ell-1}$ or $-H_{\ell-1}$. All rows (as well as columns) of $H'_{\ell-1}$ are mutually orthogonal, thus $H'_{\ell-1}$ is a Hadamard matrix.

Let us discuss the worst-case selection of $s_Y$ columns from the matrix (3). Analogously to the row selection, for any column either there exist an opposite column in the selection, or all other rows are orthogonal to it. Again, this is not worse than to select $s_Y$ columns from the matrix $H'_{\ell-1}$, where all columns of $H'_{\ell-1}$ are the corresponding columns of either $H_{\ell-1}$ or $-H_{\ell-1}$.

Finally, $H'_{\ell-1}$ is an $(\ell - 1) \times (\ell - 1)$ Hadamard matrix. We can conclude that for $s_X, s_Y \leq 2^{\ell-1}$

$$F_{\circ}(\ell, x, y) \leq H(\ell - 1, x, y).$$  

For $s_X, s_Y > 2^{\ell-1}, S_X$ and $S_Y$ have to contain a growing fraction of rows and columns with opposite signs and the value of the upper bound drops accordingly, i.e.

$$F_{\circ}(\ell, x, y) \leq H(\ell - 1, x, y).$$

This bound reaches 0 if at least one of $s_X, s_Y$ reaches $2^{\ell}$. In conclusion, by Lemma 1, $F_{\circ}$ is a $(b_X, b_Y, 2^\ell - b_X - b_Y)$-source extractor for $b_X, b_Y \leq \ell - 1$ and the bias reaches 0 if at least one of $b_X, b_Y \leq \ell$.

Let us now discuss the strongness property of this extractor. If the adversaries have access to one of the inputs, say $X$, they can adopt a different strategy. In order to maximize the sum $\sum_{x \in X} \sum_{y \in Y} m_{x,y}$, it is sufficient to pick pairs of rows with opposite signs and high absolute value, obtaining

$$\delta_{\circ}(\ell, x, y) \leq 2H(\ell - 1, x, y) = H(\ell, x, y).$$  

Notice that this is the same value as for the Hadamard extractor. Interestingly, the strategy which maximizes the bias in the strong extraction scenario leads to vanishing of the bias in usual extraction scenario (i.e. the output distribution is completely uniform, but is not fully independent of one of the input sources). This is in a sharp contrast to the standard Hadamard extraction, where the same strategy maximizes both the bias of the output distribution and the correlation to one of the input distributions.
Operation $\odot$ is balanced and has a horizontal or vertical form

The rest of the functions are of the form: $x_1 \odot y_1 = x_1$, $x_1 \odot y_1 = 1 + x_1$, $x_1 \odot y_1 = y_1$ and $x_1 \odot y_1 = 1 + y_1$. Their matrices are:

$$
\begin{pmatrix}
H_{\ell-1} & H_{\ell-1} \\
-H_{\ell-1} & -H_{\ell-1}
\end{pmatrix}, \quad
\begin{pmatrix}
-H_{\ell-1} & -H_{\ell-1} \\
H_{\ell-1} & H_{\ell-1}
\end{pmatrix}, \quad
\begin{pmatrix}
H_{\ell-1} & -H_{\ell-1} \\
H_{\ell-1} & -H_{\ell-1}
\end{pmatrix}, \quad
\begin{pmatrix}
-H_{\ell-1} & H_{\ell-1} \\
-H_{\ell-1} & H_{\ell-1}
\end{pmatrix}.
$$

Due to symmetry it suffices to analyze $x_1 \odot y_1 = x_1$. Notice that just like in the previous case, the bottom half of the rows is equal to the top half up to the sign and it does not help to select the opposite rows. As for the choice of columns, we have

$$
\ell \text{ optimized over two identical Hadamard matrices of size } 2^M 
$$

matrix representation $\pm$ in

4. Extracting multiple bits

Let $A = \{A_1, \ldots, A_k\}$ be a set of $\ell \times \ell$ matrices with elements from $\{0, 1\}$ such that for each nonempty subset $S \subseteq \{0, \ldots, k\}$, the rank of $A_S \equiv \sum_{i \in S} A_i (\text{mod } 2)$ is $\ell$. Dodis et al. [11] showed that such sets of matrices exist for all $k \leq \ell$. Using such set of matrices, they proposed the following extractor:

$$
\text{EXT}_A : [0, 1]^\ell \times [0, 1]^\ell \rightarrow [0, 1]^k (x, y) \mapsto ((A_1 x) \cdot y, \ldots, (A_k x) \cdot y),
$$

where all the operations are taken modulo 2. The proof that the output distribution is close to the uniform is based on XOR lemma (see for example [16]):

**Lemma 5** (XOR Lemma). For every k-bit random variable $X$ defined on $[0, 1]^k$, the distance $\|X - U_k\|$ is upper bounded by

$$
\sqrt{\sum_{a \in [0, 1]^k} \|X \cdot a - U_k\|^2}.
$$

If $X$ outputs a string $x \in \{0, 1\}^k$, $X \cdot a$ outputs the parity of those bits in the string $x$ defined by the positions of ones in $a$. $\text{EXT}_A(X, Y) \cdot a$ takes values in $\{0, 1\}$ and depends on both input distributions $X$ and $Y$. As we have shown before, such random variables can be represented by a matrix with elements $\pm 1$. Dodis et al. [11] showed that for every $a$ the matrix representation $M_a$ of a random variable $\text{EXT}_A(X, Y) \cdot a$ corresponds to a Hadamard matrix if $\{A_1, \ldots, A_k\}$ is a set of
matrices with desired properties. Denote \( i \in a \) iff \( a \in \{0, 1\}^k \) has 1 in the \( i \)th position. Let \( A_a = (\sum_{i \in a} A_i) \mod 2 \). The crucial observation is the following:

\[
\text{EXT}_a(x, y) \cdot a = \sum_{i \in a} \text{EXT}_a(x, y)_i = \sum_{i \in a} (A_i) \cdot y = \left( \sum_{i \in a} A_i \right) x \cdot y
\]

where all the operations are taken modulo 2. As \( \sum_{i \in a} A_i \) has full rank, it represents a bijective function on all \( x \in \{0, 1\}^\ell \). Thus \( M_a \) is equal to \( H_\ell \) (given by the construction introduced in Section 1) up to a permutation of rows, which does not change the Hadamard property. This allows to upper bound each summand of the XOR lemma to show that \( \text{EXT}_a \) is a \( (b_X, b_Y, 2^{\ell+b_X+b_Y-2}) \)-strong two source extractor.

Let us now consider \( F_+(x, y) = (x_1 + y_1) + (x_2 \cdot y_2) + \cdots + (x_\ell \cdot y_\ell) \).

**Theorem 1.** The function

\[
E_A : \{0, 1\}^\ell \times \{0, 1\}^\ell \to \{0, 1\}^k
\]

\[
(x, y) \mapsto (F_+(A_1x, y), \ldots, F_+((A_kx), y)),
\]

where all the operations are modulo 2, is a \( (b_X, b_Y, \frac{\sqrt{2}}{2} 2^{\ell+b_X+b_Y-2}) \) two source extractor of \( k \) bits.

**Proof.** To show this we adopt the proof of Dodis et al. [11]. First let us consider only sources with min entropy \( b_X, b_Y \leq \ell - 1 \). We get

\[
\sum_{i \in a} F_+((A_i), y) = (A_0) + (A_0) \cdot y_2 + \cdots + (A_\ell) \cdot y_\ell = F_+(A_0x, y).
\]

By the previous analysis (see Eq. (4)), the upper bound on the distance from the uniform distribution for odd parities is \( 2^{-\ell-b_X-b_Y-3} \). Similarly, for even parities we get:

\[
\sum_{i \in a} F_+((A_i), y) = (A_0) + (A_0) \cdot y_2 + \cdots + (A_\ell) \cdot y_\ell \overset{\text{def}}{=} F_+(A_0x, y),
\]

where \( x_1 \cdot y_1 = x_1 \). By the previous analysis (see Eq. (6)), the upper bound on the distance from the uniform distribution for even parities is \( 2^{-\ell-b_X-b_Y-2} \). Finally, by the XOR lemma:

\[
\|E(X, Y) - U_k\| = \sqrt{2^{\ell-1}2^{-b_X-b_Y-3} + 2^{\ell-1}2^{-b_X-b_Y-2}}.
\]

This brings us to the final result

\[
\sqrt{\frac{3}{4} 2^{2\ell-b_X-b_Y-2}} = \frac{\sqrt{2}}{2} 2^{\ell+b_X+b_Y-2} \]

We have shown that \( F_+ \) is a \( (b_X, b_Y, \frac{\sqrt{2}}{2} 2^{\ell+b_X+b_Y-2}) \) two source extractor. What is more, for min entropies \( b_X, b_Y > \ell - 1 \) the bias decreases, as discussed in Section 3, namely as \( b_X \) reaches \( \ell \), the bias of both odd and even parities drops to zero, thus the bias of the output drops to zero. If \( b_Y \) drops to zero, the bias of odd parities drops to zero and the resulting bias reaches half of the Hadamard bias.

We can use similar arguments in the case of the strong extraction scenario. According to Section 3, odd parities achieve the same bound as the Hadamard extractor and even parities achieve a \( \sqrt{2} \) bigger bound. Thus, the extractor retains the strongness property, but the price to pay for decreasing the bias in the non-strong extraction scenario is the increase of the bias by the factor of \( \frac{\sqrt{2}}{2} \) in the strong extraction scenario.

**5. Conclusion**

In this paper we have presented an extractor for two \( \ell \)-bit input distributions \( X, Y \) with min-entropies \( b_X, b_Y, b_X+b_Y > \ell \). If this extractor is used to extract one bit of information, the quality of the output distribution is by the factor \( \frac{\sqrt{2}}{2} \) better than that of the Hadamard extractor. The strongness of the extractor remains the same as for the Hadamard extractor, thus, our extractor can outperform the Hadamard one in all applications where it is used.
Using the Hadamard extractor, if one or even both of the input sequences are uniform, the output sequence is still biased by a constant factor. This is due to the fact that the Hadamard matrix has a different number of entries 1 and −1. On the contrary, by reestablishing the balance of the different entries, our extractor achieves uniform distribution on the output bit in case when at least one of the input distributions is uniform. Moreover, if one of the distributions is almost uniform (e.g. \( l - b < < 1 \)), the output distribution is far better than that produced by the Hadamard extractor, especially for small \( l \). With this property the extractor is applicable especially in scenarios where one of the sources is expected to be (almost) uniform, but for security reasons is combined with a different (perhaps weaker, but independent) source.

Our extractor can be also used to extract \( 1 < k \leq \ell \) bits. The bias of the produced sequence of bits is by the factor of \( \sqrt{\frac{3}{2}} \) smaller than the bias achieved by the Hadamard extractor. For \( b_X, b_Y > \ell - 1 \) the bias reaches \( 0 \), as \( b_X \) reaches \( \ell \). Thus, in a scenario where we expect one distribution to be (almost) uniform, using our extractor can substantially improve the quality of the extraction, as compared to the Hadamard one. The strongness of the extractor is maintained in this scenario as well, however the price to pay is the increase of the upper bound on the bias by the factor of \( \sqrt{\frac{3}{2}} \), compared to the Hadamard extractor.

Acknowledgments

JB acknowledges the support of the Czech Science Foundation GAČR project P202/12/G061, MP02 acknowledge the support of the Czech Science Foundation GAČR project P202/12/1142 as well as projects CE SAS QUTE and VEGA 2/0072/12. MPI acknowledges the support of GAČR project G102/09/H042. MPI acknowledges the support of SoMoPro project funded under FP7 (People) Grant Agreement no. 229603 and by South Moravian Region.

References