Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 57 (2014)

Dimension, Pseudorandomness and Extraction of Pseudorandomness*

Manindra Agrawal[†1], Diptarka Chakraborty[‡2], Debarati Das[§3], and Satyadev Nandakumar[¶4]

[1,2,4]Department of Computer Science & Engineering, Indian Institute of Technology Kanpur, Kanpur, India
[3]Charles University in Prague, Computer Science Institute of Charles University, Malostranské námesti 25, 118 00 Praha 1, Czech Republic

September 23, 2015

Abstract

In this paper we propose a quantification of distributions on a set of strings, in terms of how close to pseudorandom a distribution is. The quantification is an adaptation of the theory of dimension of sets of infinite sequences introduced by Lutz. Adapting Hitchcock's work, we also show that the logarithmic loss incurred by a predictor on a distribution is quantitatively equivalent to the notion of *dimension* we define. Roughly, this captures the equivalence between pseudorandomness defined via indistinguishability and via unpredictability. Later we show some natural properties of our notion of dimension. We also do a comparative study among our proposed notion of dimension and two well known notions of computational analogue of entropy, namely HILL-type pseudo min-entropy and next-bit pseudo Shannon entropy.

Further, we apply our quantification to the following problem. If we know that the dimension of a distribution on the set of n-length strings is $s \in (0, 1]$, can we extract out $O(sn)$ pseudorandom bits out of the distribution? We show that to construct such extractor, one need at least $\Omega(\log n)$ bits of pure randomness. However, it is still open to do the same using $O(\log n)$ random bits. We show that deterministic extraction is possible in a special case - analogous to the bit-fixing sources introduced by Chor *et al.*, which we term *nonpseudorandom bit-fixing source*. We adapt the techniques of Gabizon, Raz and Shaltiel to construct a deterministic *pseudorandom extractor* for this source.

By the end, we make a little progress towards P vs. BPP problem by showing that existence of optimal stretching function that stretches $O(\log n)$ input bits to produce n output bits such that output distribution has dimension $s \in (0, 1]$, implies P=BPP.

*Research supported in part by Research-I Foundation and the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 616787.

†manindra@cse.iitk.ac.in
‡diptarka@cse.iitk.ac.in
§debaratix710@gmail.com
¶satyadev@cse.iitk.ac.in

ISSN 1433-8092

1 Introduction

Incorporating randomness in any feasible computation is one of the basic primitives in theoretical computer science. Fortunately, any efficient (polynomial time) randomized algorithm does not require pure random bits. What it actually needs is a source that *looks* random to it and this is where the notion of *pseudorandomness* [BM84, Yao82] comes into picture. Since its introduction, pseudorandomness has been fundamental to the domain of cryptography, complexity theory and computational learning theory. Pseudorandomness is mainly a computational approach to study the nature of randomness, and *computational indistinguishability* [GM84] played a pivotal role in this. Informally, a distribution is said to be pseudorandom if no efficient algorithm can distinguish it from the uniform distribution. Another way of looking at computational indistinguishability is via the notion of *unpredictability* of distributions, due to Yao [Yao82]. Informally, a distribution is *unpredictable* if there is no efficient algorithm that, given a prefix of a string coming from that distribution, can guess the next bit with a significant success probability. This line of research naturally posed the question of constructing algorithms that can generate pseudorandom distributions, known as *pseudorandom generators*. Till now we know such constructions by assuming the existence of *one-way functions*. It is well known that constructibility of an *optimal pseudorandom generator* implies complete derandomization (i.e., P=BPP) and *exponential hardness assumption* on one-way function enables us to do that. However, Nisan and Wigderson [NW94] showed that the existence of an exponential *hard function*, which is a much weaker assumption, is also sufficient for this purpose. The assumption was further weakened in [IW96].

In order to characterize the class of random sources, information theoretic notion of *min-entropy* is normally used. A computational analogue of entropy was introduced by Yao [Yao82] and was based on compression. Håstad, Impagliazzo, Levin and Luby [HILL99] extended the definition of min-entropy in computational settings while giving the construction of a pseudorandom generator from any one-way function. This HILL-type *pseudoentropy* basically extends the definition of pseudorandomness syntactically. Relations among above two types of pseudoentropy was further studied in [BSW03]. A more relaxed notion of pseudoentropy, known as *next-bit Shannon pseudoentropy*, was later introduced by Haitner, Reingold and Vadhan [HRV10] in the context of an efficient construction of a pseudorandom generator from any one-way function. In a follow up work [VZ12], the same notion was alternatively characterized by *KL-hardness*. So far it is not clear which of the above notions is the most appropriate or whether they are at all suitable to characterize distributions in terms of the degree of pseudorandomness in it.

In this paper, we first propose an alternative measure to quantify the amount of pseudorandomness present in a distribution. This measure is motivated by the ideas of *dimension* [Lut03b] and *logarithmic loss unpredictability* [Hit03]. Lutz used the betting functions known as *gales* to characterize the *Hausdroff dimension* of sets of infinite sequences over a finite alphabet. The definition given by Lutz cannot be carried over directly, because here we consider the distributions over finite length strings instead of sets containing infinite length strings. To overcome this difficulty, we allow "non-uniform" gales and introduce a new probabilistic notion of *success* of a gale over a distribution. We use this to define the *dimension* of a distribution. In [Hit03], Hitchcock showed that the definition of dimension given by Lutz is equivalent to logarithmic loss unpredictability. In this paper, we show that this result can be adapted to establish a quantitative equivalence between the notion of logarithmic loss unpredictability of a distribution and our proposed notion of dimension. Roughly, this captures the essence of equivalence between pseudorandomness defined via indistinguishability and via unpredictability [Yao82]. We show some important properties of the notion of dimension of a distribution, which eventually makes this characterization much more powerful and flexible. We also do a comparative study between our notion of dimension and two known notions of pseudoentropy, namely HILL-type pseudo min-entropy and next-bit pseudo Shannon entropy. We show that the class of distributions with high dimension is a strict superset of the class of distributions having high HILL-type pseudo min-entropy. Whereas, there is a much closer relationship between dimension and next-bit pseudo Shannon entropy.

Once we have a quantification of pseudorandomness of a distribution, the next natural question is how to extract the pseudorandom part from a given distribution. The question is similar to the question of constructing *randomness extractors* which is an *efficient* algorithm that converts a realistic source to an *almost* ideal source of randomness. The term *randomness extractor* was first defined by Nisan and Zuckerman [NZ93]. Unfortunately there is no such deterministic algorithm and to extract out almost all the randomness, extra $\Omega(\log n)$ pure random bits are always required [NZ96, RTS00]. There is a long line of research on construction of extractors towards achieving this bound. For a comprehensive treatment on this topic, we refer the reader to excellent surveys by Nisan and Ta-Shma [NT99] and Shaltiel [Sha02]. Finally, the desired bound was achieved up to some constant factor in [LRVW03].

Coming back to the computational analogue, it is natural to study the same question in the domain of pseudorandomness. Given a distribution with dimension s, the problem is to output $O(sn)$ many bits that are pseudorandom. A simple argument can show that deterministic pseudorandom extraction is not possible, but it is not at all clear that how many pure random bits are necessary to serve the purpose. In this paper, we show that we need to actually involve $\Omega(\log n)$ random bits to extract out all the pseudorandomness present in a distribution. However explicit construction of one such extractor with $O(\log n)$ random bits is not known. If it is known that the given distribution has high HILL-type pseudo min-entropy, then any randomness extractor will work [BSW03]. Instead of HILL-type pseudoentropy, even if we have Yao-type pseudo min-entropy, then also some special kind of randomness extractor (namely with a "reconstruction procedure") could serve our purpose [BSW03]. Unfortunately both of these notions of pseudoentropy can be very small for a distribution with very high dimension. Actually the same counterexample will work for both the cases. So it is interesting to come up with an pseudorandom extractor for a class of distributions having high dimension.

As a first step towards this goal, we consider a special kind of source which we call the *nonpseudorandom bit-fixing source*. It is similar to the well studied notion of *bit-fixing random source* introduced by Chor *et al.* [CGH$^+$85], for which we know the construction of a deterministic randomness extractor due to [KZ03] and [GRS05]. In this paper, we show that the same construction yields a deterministic pseudorandom extractor for all nonpseudorandom bit-fixing sources having *polynomial-size support*.

In the concluding section, we make a little progress towards the question of P vs. BPP by showing that in order to prove P=BPP, it is sufficient to construct an algorithm that stretches $O(\log n)$ pure random bits to n bits such that the output distribution has a non-zero dimension (not necessarily pseudorandom). The idea is that using such stretching algorithm, we easily construct a hard function, which eventually gives us the most desired optimal pseudorandom generator.

Notations: In this paper, we consider the binary alphabet $\Sigma = \{0, 1\}$. We denote $Pr_{x \in_R D}[E]$ as $D[E]$, where E is an event and x is drawn randomly according to the distribution D. We use U_m to denote the uniform distribution on Σ^m. Given a string $x \in \Sigma^n$, $x[i]$ denote the i-th bit of x and $x[1, \ldots, i]$ denotes the first i bits of x. Now suppose $x \in \Sigma^n$ and $S = \{s_1, s_2, \ldots, s_k\} \subseteq \{1, 2, \ldots, n\}$, then by x_S, we denote the string $x[s_1]x[s_2] \ldots x[s_k]$.

2 Quantification of Pseudorandomness

In this section, we propose a quantification of pseudorandomness present in a distribution. We adapt the notion introduced by Lutz [Lut03b] of an s-gale to define a variant notion of success of an s-gale against a distribution D on Σ^n. Throughout this paper, we will talk about non-uniform definitions. First, we consider the definition of pseudorandomness.

2.1 Pseudorandomness

We start by defining of the notion of *indistinguishability* which we will use frequently in this paper.

Definition 1 (Indistinguishability)**.** *A distribution D over Σ^n is* (S, ϵ)-indistinguishable *from another distribution D' over Σ^n (for $S \in \mathbb{N}, \epsilon > 0$) if for every circuit C of size at most S,* $|D[C(x) = 1] - D'[C(x) = 1]| \leq \epsilon$.

Now we are ready to introduce the notion of pseudorandomness.

Definition 2 (Pseudorandomness)**.** *For a distribution D over Σ^n and for any $S > n$,*[1] *$\epsilon > 0$,*

1. *(via computational indistinguishability) D is said to be (S, ϵ)-*pseudorandom *if D is (S, ϵ)-indistinguishable from U_n; or equivalently,*
2. *(via unpredictability [Yao82]) D is said to be (S, ϵ)-pseudorandom if $D[C(x_1, \cdots, x_{i-1}) = x_i] \leq \frac{1}{2} + \frac{\epsilon}{n}$ for all circuits C of size at most $2S$ and for all $i \in [n]$.*

2.2 Martingales, s-gales and predictors

Martingales are "fair" betting games which are used extensively in probability theory (see for example, [AL06]). Lutz introduced a generalized notion, that of an s-gale, to characterize Hausdorff dimension [Lut03a] and Athreya *et al.* used a similar notion to characterize packing dimension[AHLM07].

Definition 3. *[Lut03a] Let $s \in [0, \infty)$. An* s-gale *is a function $d : \Sigma^* \to [0, \infty)$ such that $d(\lambda) = 1$ and $d(w) = 2^{-s}[d(w0) + d(w1)], \forall w \in \Sigma^*$. A* martingale *is a 1-gale.*

The following proposition establishes a connection between s-gales and martingales.

Proposition 2.1 ([Lut03a])**.** *A function $d : \Sigma^* \to [0, \infty)$ is an s-gale if and only if the function $d' : \Sigma^* \to [0, \infty)$ defined as $d'(w) = 2^{(1-s)|w|}d(w)$ is a martingale.*

In order to adapt the notion of an s-gale to the study of pseudorandomness, we first relate it to the notion of predictors, which have been extensively used in the literature [VZ12]. Given an initial finite segment of a string, a predictor specifies a probability distribution over Σ for the next symbol in the string.

Definition 4. *A function $\pi : \Sigma^* \times \Sigma \to [0, 1]$ is a* predictor *if for all $w \in \Sigma^*$, $\pi(w, 0) + \pi(w, 1) = 1$.*

Note that the above definition of a predictor is not much different from the type of predictor used in Definition 2. If we have a predictor that given a prefix of a string outputs the next bit, then by invoking that predictor independently polynomially many times we can get an estimate on the probability of occurrence of 0 or 1 as the next bit and using Chernoff bound it can easily be shown that the estimation is correct up to some inverse exponential error. For the detailed equivalence, the reader may refer to [VZ12]. In this paper, we only consider the martingales (or s-gales) and predictors that can be computed using non-uniform circuits and from now onwards we refer them just by martingales (or s-gales) and predictors. And by the size of a martingale (or an s-gale or a predictor), we refer the size of the circuit corresponding to that martingale (or s-gale or predictor).

2.3 Conversion Between s-Gale & Predictor

There is an equivalence between an s-gale and a predictor. An early reference to this is [Cov74]. We follow the construction given in [Hit03].

A predictor π induces an s-gale d_π for each $s \in [0, \infty)$ and is defined as follows: $d_\pi(\lambda) = 1$, $d_\pi(wa) = 2^s d_\pi(w)\pi(w, a)$ for all $w \in \Sigma^*$ and $a \in \Sigma$; equivalently $d_\pi(w) = 2^{s|w|} \prod_{i=1}^{|w|} \pi(w[1 \cdots i-1], w[i])$ for all $w \in \Sigma^*$.

Conversely, an s-gale d with $d(\lambda) = 1$ induces a predictor π_d defined as: if $d(w) \neq 0$, $\pi_d(w, a) = 2^{-s} \frac{d(wa)}{d(w)}$; otherwise, $\pi_d(w, a) = \frac{1}{2}$, for all $w \in \Sigma^*$ and $a \in \Sigma$.

Hitherto, s-gales have been used to study the dimension of sets of infinite sequences - for an extensive bibliography, see [Hita] and [Hitb]. Although in this paper, we consider distributions on finite length strings, the conversion procedure between s-gale and predictor will be exactly same as described above.

[1] Throughout this paper, we consider $S > n$ so that the circuit can at least read the full input; however reader can feel free to take any $S \in \mathbb{N}$.

2.4 Defining Dimension

Definition 5. *An s-gale $d : \Sigma^* \to [0, \infty)$ is said to ϵ*-succeed over a distribution *D on Σ^n if $D[d(w) \geq 2] > \frac{1}{2} + \epsilon$.*

Note that the above definition of win of an s-gale is not arbitrary and reader may refer to the last portion of the proof of Theorem 2 to get some intuition behind this definition. The following lemma states the equivalence between the standard definition of pseudorandomness and the definition using martingale.

Lemma 2.1. *There exists a constant $c' > 0$ such that for every $c > c'$ and for any $n \in \mathbb{N}$, if a distribution D over Σ^n is (S, ϵ)-pseudorandom then there is no martingale of size at most $(S - c)$ that ϵ-succeeds on D. Conversely, if there is no martingale of size at most $3S$ that $\frac{\epsilon}{n}$-succeeds on D, then D is (S, ϵ)-pseudorandom.*

Proof. Assume $d : \Sigma^* \to [0, \infty)$ is a martingale which ϵ-succeeds on D. *i.e.*,

$$D[d(w) \geq 2] > \frac{1}{2} + \epsilon.$$

By the Markov Inequality, $U_n[d(w) \geq 2] \leq \frac{1}{2}$.

Let C_d be a circuit of size S obtained by instantiating d at length n. Now let C be a circuit which outputs 1 if $C_d(w) \geq 2$. Then,

$$|D[C(w) = 1] - U_n[C(w) = 1]| > \epsilon.$$

Thus D is not an $(S + c, \epsilon)$-pseudorandom distribution, for some constant $c > 0$.

Now for the converse direction, assume that D is not an (S, ϵ)-pseudorandom distribution. Then there exists an bit position $i \in [0, n - 1)$ and some circuit C of size at most $2S$ for which

$$D[C(w_1, \cdots, w_{i-1}) = w_i] > \frac{1}{2} + \frac{\epsilon}{n}.$$

Now build a martingale $d : \Sigma^* \to [0, \infty)$ using this circuit C as follows. Let $d(\lambda) = 1$. Now, $\forall j \in [n], j \neq i$, $d(w[0 \ldots j-1]0) = d(w[0 \ldots j-1]1) = d(w[0 \ldots j-1])$, and $d(w[0 \ldots i-1]b) = 2d(w[0 \ldots i-1])$ if $C(w[0 \ldots i-1]) = b$ and $d(w[0 \ldots i-1]\bar{b}) = 0$.

Now it is clear that

$$D[d(w) \geq 2] > \frac{1}{2} + \frac{\epsilon}{n}.$$

□

The next definition gives a complete quantification of distributions in terms of dimension.

Definition 6 (Dimension)**.** *The (S, ϵ)-dimension of a distribution D on Σ^n is defined as*

$$\dim_{S,\epsilon}(D) = \min\{1, \inf\{s \in [0, \infty) \mid \exists s- \text{ gale } d \text{ of size at most } S \text{ which } \epsilon\text{-succeeds on } D\}\}.$$

Informally, if the dimension of a distribution is s, we say that it is *s-nonpseudorandom.*

3 Unpredictability and Dimension

It is customary to measure the performance of a predictor utilizing a *loss function* [Hit04]. The loss function determines the penalty incurred by a predictor for erring in its prediction. Let the next bit be b and the probability induced by the predictor on it is p_b.

Commonly used loss functions include the *absolute loss function*, which penalizes the amount $1 - p_b$; and the *logarithmic loss function*, which penalizes $-\log(p_b)$. The latter, which appears complicated at first glance, is intimately related to the concepts of Shannon Entropy and dimension. In this section, adapting the result of Hitchcock [Hit03], we establish that there is an equivalence between the notion of dimension that we define in the previous section, and the logarithmic loss function defined on a predictor.

Definition 7. *The* logarithmic loss function *on $p \in [0,1]$ is defined to be $loss(p) = -\log p$.*

Using this, we define the running loss that a predictor incurs while it predicts successive bits of a string in Σ^n, as the sum of the losses that the predictor makes on individual bits.

Definition 8. *Let $\pi : \Sigma^* \times \Sigma \to [0,1]$ be a predictor.*

1. *The* cumulative loss *of π on $w \in \Sigma^n$, denoted as $Loss(\pi, w)$, is defined by $Loss(\pi, w) = \sum_{i=1}^{n} loss(\pi(w[1 \dots i-1]),\ w[i])$.*
2. *The* loss rate *of π on $w \in \Sigma^n$ is $LossRate(\pi, w) = \frac{Loss(\pi,w)}{n}$.*
3. *The* ϵ-loss rate *of π over a distribution D on Σ^n is $LossRate_\epsilon(\pi, D) = \inf t + \frac{1}{n}$, where t is any number in $[0,1]$ such that $D[LossRate(\pi, w) \leq t] > \frac{1}{2} + \epsilon$.*

Note that for a fixed $n \in \mathbb{N}$, any distribution on Σ^n has loss rate between $\frac{1}{n}$ and 1. The unpredictability of a distribution is defined as the infimum of the loss rate that any predictor has to incur on the distribution.

Definition 9. *The* (S, ϵ)-unpredictability *of a distribution D on Σ^n is*

$$\text{unpred}_{S,\epsilon}(D) = \min\{1, \inf\{LossRate_\epsilon(\pi, D) \mid \pi \textit{ is a predictor of size at most } S\}\}.$$

With this, we can prove that dimension can equivalently be defined using unpredictability. The proof is motivated from the proof of the equivalence between logarithmic loss unpredictability and dimension [Hit03]

Theorem 1. *For any distribution D on Σ^n, if $\dim_{S,\epsilon}(D) \leq s$, then $\text{unpred}_{S^2,\epsilon}(D) \leq s$. Conversely, if $\text{unpred}_{S,\epsilon}(D) \leq s$, then $\dim_{S^2,\epsilon}(D) \leq s$.*

Proof. First, let D be a distribution on Σ^n of (S, ϵ)-dimension at most $s \in [0,1)$ where $n > \frac{1}{s}$. Assume that s' is a number such that $s < s' \leq 1$. Suppose d is an s'-gale of size at most S that ϵ-succeeds on D. Let $\pi_d : \Sigma^* \times \Sigma \to [0,1]$ be defined by

$$\pi_d(w, b) = \begin{cases} 2^{-s'} \frac{d(wb)}{d(w)}, & \text{if } d(w) \neq 0 \\ \frac{1}{2} & \textit{otherwise.} \end{cases}$$

For any $w \in \Sigma^n$ with $d(w) \geq 2$, we have

$$\begin{aligned} \text{Loss}_{\pi_d}(w) &= -\sum_{i=1}^{n} \log \pi_d(w[1 \dots i-1], w[i]) \\ &= -\log \Pi_{i=1}^{n} \pi_d(w[1 \dots i-1], w[i]) \\ &= s'n - \log d(w) \\ &\leq s'n - 1. \end{aligned}$$

So $\text{LossRate}(\pi_d, w) \leq s' - \frac{1}{n}$. Thus,

$$\begin{aligned} D[\text{LossRate}(\pi_d, w) \leq s' - \frac{1}{n}] &\geq D[d(w) \geq 2] \\ &> \frac{1}{2} + \epsilon. \end{aligned}$$

Note that π_d can be implemented using a circuit of size at most S^2.

Conversely, assume that $\text{unpred}_{S,\epsilon}(D) \leq t \in [0,1)$. Assume that t' is a number satisfying $t < t' \leq 1$. Let π be a predictor of size at most S such that

$$D[\text{LossRate}(\pi, w) \leq t' - \frac{1}{n}] > \frac{1}{2} + \epsilon.$$

If d' is the t'-gale defined by $d_\pi(w) = 2^{t'|w|}\Pi_{i=1}^{|w|}\pi(w[0\dots i-1], w[i])$, then for any $w \in \Sigma^n$ with $\text{LossRate}(\pi, w) \le t' - \frac{1}{n}$, we have the following.

$$\begin{aligned}\log d_\pi(w) &= t'n + \sum_{i=1}^{n} \log \pi(w[1\dots i-1], w[i]) \\ &= t'n - \text{Loss}_\pi(w) \\ &\ge 1.\end{aligned}$$

Hence, $D[d_\pi(w) \ge 2] > \frac{1}{2} + \epsilon$. Moreover, d can be implemented by a circuit of size at most S^2. □

Till this point, we have given all the definitions parameterized by the circuit size S and bias term ϵ. However, we can naturally extend our definitions to *asymptotic* definitions (See Section 6) where we consider S to be any polynomial in n and ϵ to be inverse of any polynomial in n. In that case, we will get exact equivalence between dimension and unpredictability.

4 Properties of Dimension

We now establish a few basic properties of our notion of dimension. We begin by exhibiting a distribution on Σ^n with dimension s, for any $s \in (0, 1]$.

First, we observe that the dimension of any distribution D is the infimum of a non-empty subset of $[0, 1]$ and hence the dimension of a distribution is well-defined.

Since it is clear that any distribution on Σ^n has a dimension, the following theorem establishes the fact that our definition yields a nontrivial quantification of the set of distributions.

Theorem 2. *Let $s \in (0, 1]$. Then for large enough n and any $S > n$, $\epsilon > 0$, there is a distribution D on Σ^n with (S, ϵ)-dimension s.*

Proof. Let us take a distribution $D := U_n$, i.e., uniform distribution on Σ^n. If $s = 1$, then by Lemma 2.1, D is a distribution with the required (S, ϵ)-dimension, for any $S > 0$ and $\epsilon > 0$.

Otherwise, assume that $s \in (0, 1)$. To each string $x \in \Sigma^n$, we append $\lfloor \frac{n}{s} \rfloor - n$ many zeros, and denote the resulting string as x'. Let $D'(x') = D(x)$. For strings $y \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$ which do not terminate in a sequence of $\lfloor \frac{n}{s} \rfloor - n$ many zeros, we set $D'(y) = 0$.

Let $\pi : \Sigma^* \times \Sigma \to [0, 1]$ be the predictor which testifies that the (S^2, ϵ)-unpredictability of $D \le 1$. Define the new predictor $\pi' : \Sigma^* \times \Sigma \to [0, 1]$ by

$$\pi'(x, b) = \begin{cases} \pi(x, b) & \text{if} |x| < n, b = 0, 1 \\ 1 & \text{if} |x| \ge n, b = 0 \\ 0 & \text{otherwise.} \end{cases}$$

For every $w \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$ which is in the support of D' such that $\text{LossRate}(\pi, w[1\dots n]) \le (1 + \epsilon_1 - \frac{1}{n})$, for any $\epsilon_1 > 0$, we have that

$$\text{LossRate}(\pi', w) = \frac{\text{Loss}(\pi, w[1\dots n])}{\lfloor \frac{n}{s} \rfloor} \le \frac{(1 + \epsilon_1 - \frac{1}{n})n}{\lfloor \frac{n}{s} \rfloor} \le (s + \epsilon' - \frac{1}{\lfloor \frac{n}{s} \rfloor}), \quad \text{for some } \epsilon' > 0$$

The last inequality holds for small enough s/n and this testifies that the (S^2, ϵ)-unpredictability (hence the (S^4, ϵ)-dimension) of the distribution D' is at most s.

Now, assume that (S^4, ϵ)-dimension of D' is less than s and for some ϵ_1, $0 < \epsilon_1 < s$, there exists a s'-gale ($s' = s - \epsilon_1$) d of size at most S^4 which ϵ-succeeds on D'. We show that this would imply that D is not uniform. Now consider a string $w \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$, which is in the support of D'. For any $k \in n + 1, \cdots, \lfloor \frac{n}{s} \rfloor$, $d(w[1\dots k]) \le 2^{s'} d(w[1\dots k-1])$ and thus $d(w) \ge 2$ will imply that $d(w[1\dots n]) \ge 2^{-s'(\lfloor \frac{n}{s} \rfloor - n) + 1}$. Now consider the martingale d' corresponding to the s'-gale d. According to [Lut03a], we have $d'(w') = 2^{(1-s')|w'|}d(w')$, for any string $w' \in \Sigma^*$. Thus,

$$
\begin{aligned}
D'[d'(w[1\dots n]) \geq 2] &\geq D'[d(w[1\dots n]) \geq 2^{-s'(\lfloor \frac{n}{s} \rfloor - n)+1}] \\
&\geq D'[d(w) \geq 2] \\
&> \frac{1}{2} + \epsilon.
\end{aligned}
$$

Note that $D'[d'(w[1\dots n]) \geq 2]$ is same as $D[d'(x) \geq 2]$ or in other words $U_n[d'(x) \geq 2]$, which contradicts the fact that by Markov Inequality, $U_n[d'(x) \geq 2] \leq \frac{1}{2}$ and this completes the proof. □

In subsequent sections, we will see how to extract pseudorandom parts from a convex combination of distributions. We will need a weaker version of the following theorem which establishes a relationship between the dimension of a convex combination of distributions in terms of the dimension of its constituent distributions.

Theorem 3. *Let D_1 and D_2 be the distributions on Σ^n and $\delta \in [0, 1]$. Suppose D is the convex combination of D_1 and D_2 defined by $D = \delta D_1 + (1-\delta)D_2$. Then for any $S > n$ and $\epsilon > 0$, $\dim_{S,\epsilon}(D) \geq \min\{\dim_{S,\epsilon}(D_1), \dim_{S,\epsilon}(D_2)\}$.*

Proof. The claim clearly holds when δ is either 0 or 1, so assume that $0 < \delta < 1$. Let $\dim_{S,\epsilon}(D_1) = s_1$, and $\dim_{S,\epsilon}(D_2) = s_2$.

For the contrary, let us assume that, $\dim_{S,\epsilon}(D) < \min\{s_1, s_2\}$. Now consider $s = \min\{s_1, s_2\} - \epsilon_1$, for some $\epsilon_1, 0 < \epsilon_1 < \min\{s_1, s_2\}$. Then there exists an s-gale d of size at most S such that $D[d(w) \geq 2] > \frac{1}{2} + \epsilon$.

Let the string w for which $d(w) \geq 2$ holds be w_i, $1 \leq i \leq k$ and the corresponding probabilities in D be $p(w_i), 1 \leq i \leq k$. Let $q(w_i)$ and $r(w_i)$, $1 \leq i \leq k$, be the corresponding probabilities in D_1 and D_2 respectively. So, $\sum_{i=1}^{k} p(w_i) > \frac{1}{2} + \epsilon$, where $p(w_i) = \delta q(w_i) + (1-\delta)r(w_i), 1 \leq i \leq k$. Now, since $\dim_{S,\epsilon}(D_2) = s_2$, we have that $r(w_1) + \cdots + r(w_k) \leq \frac{1}{2} + \epsilon$. Thus $q(w_1) + \cdots + q(w_k) > \frac{1}{2} + \epsilon$ implying $\dim_{S,\epsilon}(D_1) < s_1$, which is a contradiction. □

If we just concentrate on pseudorandom distributions, then by replacing s-gales with martingales in the proof of the above theorem, we will get the following lemma, which will be used in Section 7.1.

Lemma 4.1. *Let D_1 and D_2 be the (S, ϵ)-pseudorandom distributions on Σ^n for any $S > n$, $\epsilon > 0$ and $\delta \in [0, 1]$. Suppose there exists a distribution D which can be expressed as $D = \delta D_1 + (1-\delta)D_2$, then D is also (S, ϵ)-pseudorandom.*

However, it is easy to see that convex combinations of distributions may have larger dimension than any of its constituents. For example, let us consider a $n \in \mathbb{N}$ and take the distribution U_n. Now take two distributions on Σ^{n+1}, namely, D_1 produced by the 0-dilution (padding each string with a 0 at the end) of U_n and D_2 produced by the 1-dilution (padding each string with a 1 at the end) of U_n. Then $D = 0.5D_1 + 0.5D_2$ is nothing but U_{n+1} and has dimension which exceeds the dimensions of D_1 and D_2 by $\frac{1}{n}$.

Theorem 4. *Let D, D_1 and D_2 be the distributions on Σ^n, and consider $S > n$, $\epsilon > 0$ and $\delta \in [0, 1]$. Suppose further that $\dim_{S,\epsilon}(D_1) = s_1$. Now if $D = (1-\delta)D_1 + \delta D_2$, then $\dim_{S,(\epsilon+\delta)}(D) \geq s_1$.*[2]

Proof. For the contrary, lets assume that, $\dim_{S,(\epsilon+\delta)}(D) < s_1$ and assume $s = s_1 - \epsilon_1$, for some ϵ_1, $0 < \epsilon_1 < s_1$. So there exists an s-gale of size at most S that $(\epsilon + \delta)$-succeeds over D. Thus,

$$
D[d(w) \geq 2] > \frac{1}{2} + (\epsilon + \delta).
$$

[2]Note that bias term in the dimension of D_1 depends on δ.

Let the string w for which $d(w) \geq 2$ holds be w_i, $1 \leq i \leq k$ and the corresponding probabilities in D be $p(w_i), 1 \leq i \leq k$. Let $q(w_i)$ and $r(w_i)$, $1 \leq i \leq k$, be the corresponding probabilities in D_1 and D_2 respectively. So,

$$p(w_1) + \cdots + p(w_k) > \frac{1}{2} + (\epsilon + \delta),$$

where $p(w_i) = (1-\delta)q(w_i) + \delta r(w_i), 1 \leq i \leq k$.

Now, for any $\delta \in [0,1]$ as $r(w_1) + \cdots + r(w_k) \leq 1$,

$$q(w_1) + \cdots + q(w_k) > \frac{1}{2} + \epsilon$$

and thus $\dim_{S,\epsilon}(D_1) < s_1$ which is a contradiction. □

If we follow the proof of Theorem 4 with martingale instead of s-gale, we get the following weaker version of the above theorem, which we will require in the construction of deterministic extractor for a special kind of sources in Section 7.1.

Lemma 4.2. *Let D, D_1 and D_2 be the distributions on Σ^n, and consider $S > n$, $\epsilon > 0$ and $\delta \in [0,1]$. If D_1 is (S,ϵ)-pseudorandom and $D = (1-\delta)D_1 + \delta D_2$, then D is $(S, \epsilon+\delta)$-pseudorandom as well.*

The following theorem shows that in order for a distribution to have dimension less than 1, it is not sufficient to have a few positions where we can successfully predict - it is necessary that these positions occur often.

Theorem 5. *For large enough n and for any $S > n$ and $\epsilon > 0$, there is a distribution D_n on Σ^n such that $\dim_{S,\epsilon}(D_n) = 1$, but is not (S,ϵ)-pseudorandom.*

Proof. Let D_n on Σ^n be defined as follows.

$$D_n(x) = \begin{cases} \frac{1}{2^{n-1}} & \text{if } x[n] = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then D_n is clearly not (S,ϵ)-pseudorandom, for any value of $S > n$ and $\epsilon > 0$. Consider a predictor $\pi : \Sigma^* \times \Sigma \to [0,1]$ defined as follows. For strings w of length i, $i \in [1, n-2]$, set $\pi(w,b) = 0.5$, $b = 0,1$ and $\pi(w,0) = 1$, $\pi(w,1) = 0$ otherwise. Then

$$D_n[\pi(x[\ 1 \ldots n-1\ ], x[n]) = 1] = 1.$$

However, $\dim_{S,\epsilon}(D_n) = 1$. For contradiction sake, assume that $\dim_{S,\epsilon}(D_n) < 1$ and for some ϵ_1, $0 < \epsilon_1 < 1$, there exists a s-gale ($s = 1 - \epsilon_1$) d of size at most S which ϵ-succeeds on D_n. Now consider a string $w \in \Sigma^n$, which is in the support of D_n. Now, $d(w) \leq 2^s d(w[1 \ldots n-1])$ and thus $d(w) \geq 2$ will imply that $d(w[1 \ldots n-1]) \geq 2^{1-s}$. Now consider the martingale d' corresponding to the s-gale d. According to [Lut03a], we have $d'(w') = 2^{(1-s)|w'|}d(w')$, for any string $w' \in \Sigma^*$. Thus,

$$\begin{aligned} D_n[d'(w[1 \ldots n-1]) \geq 2] &\geq D_n[d(w[1 \ldots n-1]) \geq 2^{1-s}] \\ &\geq D_n[d(w) \geq 2] \\ &> \frac{1}{2} + \epsilon. \end{aligned}$$

The first inequality holds for large enough n. Note that $D_n[d'(w[1 \ldots n-1]) \geq 2]$ is same as $U_{n-1}[d'(x) \geq 2]$, where $x \in \Sigma^{n-1}$ is drawn according to the distribution U_{n-1}. However by Markov Inequality, $U_{n-1}[d'(x) \geq 2] \leq \frac{1}{2}$, which is a contradiction and this completes the proof.

□

5 Pseudoentropy and Dimension

In this section we study the relation between our notion of dimension and different variants of computational or pseudo (min/Shannon) entropy. We will use standard notions and notations of information theory (e.g., Shannon entropy, KL divergence) without defining them. Readers can find a few basic notations and propositions of information theory in the following subsection and for more details, we refer the reader to a book by Cover and Thomas [CT06].

5.1 Basics of Information Theory

Definition 10 (Shannon Entropy)**.** *The* Shannon entropy *of a discrete random variable X is defined as*

$$H(X) := -\sum_{x} Pr[X = x] \log Pr[X = x] = -\mathbb{E}_{x\sim X}[\log Pr[X = x]].$$

The joint entropy $H(X, Y)$ *is defined to be* $-\mathbb{E}_{x\sim X,y\sim Y}[\log Pr[X = x, Y = y]]$ *and the* conditional entropy $H(Y \mid X)$ *is defined to be* $\mathbb{E}_{x\sim X}[H(Y \mid X = x)]$.

Proposition 5.1 (Chain Rule for Shannon Entropy)**.**

$$H(X, Y) = H(X) + H(Y \mid X).$$

Definition 11 (KL divergence)**.** *The* Kullback-Leibler distance *or* KL divergence *between two distributions P and Q is defined as*

$$\mathbf{KL}(P\|Q) := \mathbb{E}_{p\sim P} \log \frac{Pr[P = p]}{Pr[Q = p]}.$$

Definition 12 (Conditional KL divergence)**.** *For random variables (P_1, P_2) and (Q_1, Q_2), the* conditional KL divergence *from $(P_2|P_1)$ to $(Q_2|Q_1)$ is defined as*

$$\mathbf{KL}((P_2|P_1)\|(Q_2|Q_1)) = \mathbb{E}_{p_1\sim P_1,p_2\sim P_2}\big[\log \frac{Pr[P_2 = p_2|P_1 = p_1]}{Pr[Q_2 = p_2|Q_1 = p_1]}\big].$$

Just like Shannon entropy, in this case also, we have chain rule stated below.

Proposition 5.2 (Chain Rule for KL divergence)**.**

$$\mathbf{KL}(P_1, P_2\|Q_1, Q_2) = \mathbf{KL}(P_1\|Q_1) + \mathbf{KL}((P_2|P_1)\|(Q_2|Q_1)).$$

5.2 High HILL-type pseudo min-entropy implies high dimension

For a distribution D, *min-entropy* of D is defined as $H_\infty(D) = \min_w\{\log(1/D[w])\}$. We start with the standard definition of computational min-entropy, as given by [HILL99].

Definition 13 (HILL-type pseudo min-entropy)**.** *A distribution D on Σ^n has (S, ϵ)-*HILL-type pseudo min-entropy *(or simply (S, ϵ)-pseudo min-entropy) at least k, denoted as $H_\infty^{HILL,S,\epsilon} \geq k$ if the there exists a distribution D' such that*

1. *$H_\infty(D') \geq k$, and*
2. *D' is (S, ϵ)-indistinguishable from the distribution D.*

Several other definitions of pseudo min-entropy (metric-type, Yao-type or compression type) are there in the literature. We refer the reader to [BSW03] for a comprehensive treatment on different definitions and the connections between them. In the remaining potion of this subsection, we focus only on HILL-type pseudo min-entropy. Now we state the main result of this subsection.

Theorem 6. *There exists a constant $c' > 0$ such that for any $c > c'$, for every distribution D on Σ^n and for any $S > n$, $\epsilon > 0$, if $H_\infty^{HILL,(S+c),\epsilon}(D) \geq sn$, then $\dim_{S,\epsilon}(D) \geq s$*

Proof. The theorem is a consequence of the following claim.

Claim 5.1. *For every distribution X on Σ^n, if $H_\infty(X) = k$ then $\dim_{S,\epsilon}(X) \geq k/n$, for any values of S and $\epsilon > 0$.*

Now observe that if a distribution D is $(S + c, \epsilon)$-indistinguishable from another distribution D', then $\dim_{S,\epsilon}(D) = \dim_{S,\epsilon}(D')$ as otherwise the s-gale which ϵ-succeeds over exactly one of them, acts as a distinguishing circuit. This fact along with Claim 5.1 completes the proof. □

It only remains to establish Claim 5.1.

Proof of Claim 5.1. Let us first take $s = k/n$. Now for the sake of contradiction, let us assume that there exists an s-gale d that ϵ-succeeds over X, i.e., $X[d(w) \geq 2] > \frac{1}{2} + \epsilon$. Now consider the set $S := \{w|d(w) \geq 2\}$. As $H_\infty(X) = k$, $|S| > 2^{sn-1} + 2^{sn}.\epsilon$. By taking the corresponding martingale d' according to the Proposition 2.1, we have that for any $w \in S$, $d'(w) \geq 2^{(1-s)n+1}$ and as a consequence, $U_n[d'(w) \geq 2^{(1-s)n+1}] > 2^{sn-n-1} + 2^{sn-n}.\epsilon$, which contradicts the fact that by Markov inequality, $U_n[d'(w) \geq 2^{(1-s)n+1}] \leq 2^{sn-n-1}$. □

The converse direction of the statement of Theorem 6 is also true if the distribution under consideration is pseudorandom. If the converse is true then we can apply any randomness extractor to get pseudorandom distribution from any distribution having high dimension [BSW03]. However, we should always be careful about the circuit size with respect to which we call the output distribution pseudorandom. Unfortunately, in general the converse is not true.

Counterexample for the converse: Suppose *one-way functions* exist, then it is well-known that we can construct a *pseudorandom generator* $G : \Sigma^l \to \Sigma^m$ such that m is any polynomial in l, say $m = l^3$. For the definitions of one-way function, pseudorandom generator and the construction of pseudorandom generator with polynomial stretch from any one-way function, interested reader may refer to [Gol01, HILL99, VZ12]. Now consider the distribution $D := (G(U_l), U_l)$. For large enough l, using the argument similar to the proof of Theorem 2, it can easily be shown that the distribution D has dimension almost 1 as the distribution on the first m bits are pseudorandom, but pseudo min-entropy is not larger than l.

5.3 Equivalence between dimension and next-bit pseudo Shannon entropy

In the last subsection, we have talked about pseudo min-entropy. In similar fashion, one can also define *pseudo Shannon entropy* and a natural generalization of it is *conditional pseudo Shannon entropy* [HLR07, HRV10, VZ12].

Definition 14 (Conditional pseudo Shannon entropy)**.** *Suppose Y is a random variable jointly distributed with X. Y is said to have (S, ϵ)-*conditional pseudo Shannon entropy *at least k given X if there exists a distribution Z jointly distributed with X such that*

1. *$H(Z|X) \geq k$, and*
2. *(X, Y) and (X, Z) are (S, ϵ)-indistinguishable.*

The following is the variant of pseudoentropy that we are looking for in this subsection and was introduced by Haitner *et al.* [HRV10].

Definition 15 (Next-bit pseudo Shannon entropy)**.** *A random variable $X = (X_1, X_2, \cdots, X_n)$ taking values in Σ^n has (S, ϵ)-*next-bit pseudo Shannon entropy *at least k, denoted as $H^{next,S,\epsilon}(X) \geq k$ if there exist random variables $(Y_1, Y_2, \cdots, Y_n)$ such that*

1. *$\sum_i H(Y_i|X_1, \cdots, X_{i-1}) \geq k$, and*
2. *for all $1 \leq i \leq n$, $(X_1, \cdots, X_{i-1}, X_i)$ and $(X_1, \cdots, X_{i-1}, Y_i)$ are (S, ϵ)-indistinguishable.*

Later, Vadhan and Zheng [VZ12] provided an alternative characterization of conditional pseudo Shannon entropy by showing an equivalence between it and *KL-hardness* (defined below). We use this alternative characterization extensively for our purpose.

Definition 16 (KL-hardness)**.** *Suppose (X,Y) is a $\Sigma^n \times \Sigma$-valued random variable and π be any predictor. Then π is said to be a* δ-KL-predictor *of Y given X if $\mathbf{KL}(X,Y \| X, C_\pi) \leq \delta$ where $C_\pi(y|x) = \pi(x,y)$ for all $x \in \Sigma^n$ and $y \in \Sigma$.*

Moreover, Y is said to be (S,δ)-KL-hard *given X if there is no predictor π of size at most S that is a δ-KL-predictor of Y given X.*

The following theorem provides the equivalence among KL-hardness and conditional pseudo Shannon entropy of a distribution.

Theorem 7 ([VZ12])**.** *For a $\Sigma^n \times \Sigma$-valued random variable (X,Y) and for any $\delta > 0$, $\epsilon > 0$,*

1. *If Y is (S,δ)-KL-hard given X, then for every $\epsilon > 0$, Y has (S',ϵ)-conditional pseudo Shannon entropy at least $H(Y|X) + \delta - \epsilon$, where $S' = S^{\Omega(1)}/poly(n, 1/\epsilon)$.*

2. *Conversely, if Y has (S,ϵ)-conditional pseudo Shannon entropy at least $H(Y|X) + \delta$, then for every $\sigma > 0$, Y is (S',δ')-KL-hard given X, where $S' = \min\{S^{\Omega(1)}/poly\log(1/\sigma), \Omega(\sigma/\epsilon)\}$ and $\delta' = \delta - \sigma$.*

Now we are ready to state the main theorem of this subsection which conveys the fact that the distributions with high dimensions also have high next-bit pseudo Shannon entropy.

Theorem 8. *For any $\epsilon' > 0$, there exists a $n' \in \mathbb{N}$ such that for any $n \geq n'$ and $S > n$, $\epsilon > 0$, for every distribution D on Σ^n, if $\dim_{S,\epsilon}(D) > \frac{2s}{1-2\epsilon} + \epsilon'$, then $H^{next,S',\epsilon}(D) > sn$, where $S' = S^{\Omega(1)}/poly(n)$.*

Proof. For the sake of contradiction, let us assume that $H^{next,S',\epsilon}(D) \leq sn$. Now break D into 1-bit blocks, i.e., $D = (D_1, D_2, \cdots, D_n)$. Let $\dim_{S,\epsilon}(D) > \frac{2s}{1-2\epsilon} + \epsilon'$, for some $\epsilon' > 0$ and thus by Theorem 1, $\text{unpred}_{\sqrt{S},\epsilon}(D) > \frac{2s}{1-2\epsilon} + \epsilon' = t(\text{say})$.

For any predictor $\pi : \Sigma^* \times \Sigma \to [0,1]$, let us define $\pi_i : \Sigma^{i-1} \times \Sigma \to [0,1]$ such that $\pi(x,a) = \pi_i(x,a), \forall_{x \in \Sigma^{i-1}, a \in \Sigma}$ for $1 \leq i \leq n$. Then,

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbf{KL}((D_1, \cdots, D_{i-1}), D_i \| (D_1, \cdots, D_{i-1}), \pi_i) \\
&= \sum_{i=1}^{n} \Big[\sum_{w_i \in \Sigma^i} -\log(\pi_i(w_i[1 \cdots i-1], w_i[i])) Pr[w_i] - H(D_i | D_1 \cdots D_{i-1}) \Big] \\
&= \sum_{i=1}^{n} \Big[\sum_{w_i \in \Sigma^i} loss(\pi(w_i[1 \cdots i-1], w_i[i])) Pr[w_i] - H(D_i | D_1 \cdots D_{i-1}) \Big] \\
&= \sum_{w \in \Sigma^n} \text{Loss}(\pi, w) D(w) - H(D)
\end{aligned}$$

where the last equality follows from the chain rule of Shannon entropy (Proposition 5.1).

Now the definitions of next-bit pseudo Shannon entropy, conditional pseudo Shannon entropy and KL-predictor along with Item 1 of Theorem 7 imply that there exists a predictor π of size at most $\sqrt{S}$ such that $\sum_{w \in \Sigma^n} \text{Loss}(\pi, w) D(w) \leq sn + \epsilon$.

Hence for any $\epsilon_1 > 0$,

$$\begin{aligned}
D[\text{LossRate}(\pi, w) \geq (t - \frac{1}{n} - \epsilon_1)] &= D[\text{Loss}(\pi, w) \geq (t - \frac{1}{n} - \epsilon_1) n] \\
&\leq \frac{\sum_{w \in \Sigma^n} \text{Loss}(\pi, w) D(w)}{(t - \frac{1}{n} - \epsilon_1) n} \quad \text{by Markov inequality} \\
&\leq \frac{s + \frac{\epsilon}{n}}{t - \frac{1}{n} - \epsilon_1}.
\end{aligned}$$

Thus,

$$D[\text{LossRate}(\pi, w) < (t - \frac{1}{n} - \epsilon_1)] \geq 1 - \frac{s + \frac{\epsilon}{n}}{t - \frac{1}{n} - \epsilon_1}.$$

As $\text{unpred}_{\sqrt{S},\epsilon}(D) > \frac{2s}{1-2\epsilon} + \epsilon' = t$, by the definition, it implies that

$$1 - \frac{s + \frac{\epsilon}{n}}{t - \frac{1}{n} - \epsilon_1} \leq \frac{1}{2} + \epsilon.$$

For large enough n, the above inequality gives us $t < \frac{2s}{1-2\epsilon} + \epsilon_2$, for any $\epsilon_2 > \epsilon_1$. So a proper choice of ϵ_1 leads to a contradiction and this completes the proof. □

On the contrary, for the other direction, the following weaker version can easily be proven.

Theorem 9. *For any $\epsilon' > 0$, there exists a $n' \in \mathbb{N}$ such that for any $n \geq n'$ and $S > n$, $\epsilon > 0$, for every distribution D on Σ^n, if $H^{next,S,\epsilon}(D) > sn$, then $\dim_{S',\epsilon}(D) > s - \frac{1}{2} - \epsilon''$, where $S' = \min\{S^{\Omega(1)}/poly\log(1/\epsilon'), \Omega(\epsilon'/\sqrt{\epsilon})\}$ and $\epsilon'' = \epsilon' - \epsilon$.*

Proof. Suppose a distribution D is given such that $H^{next,S,\epsilon}(D) > sn$. Now break D into 1-bit blocks, i.e., $D = (D_1, D_2, \cdots, D_n)$. Let us denote $\dim_{\sqrt{S'},\epsilon}(D) = t$ (term S' is same as that appears in Item 2 of Theorem 7) and thus by Theorem 1, $\text{unpred}_{S',\epsilon}(D) \leq t$. This implies that there exists a predictor π of size at most S' such that for any $\epsilon_1 > 0$,

$$D[\text{LossRate}(\pi, w) \leq (t - \frac{1}{n} + \epsilon_1)] > \frac{1}{2} + \epsilon.$$

Thus we can write the following,

$$\sum_{w \in \Sigma^n} \text{Loss}(\pi, w) D(w) < (t - \frac{1}{n} + \epsilon_1)n + (\frac{1}{2} - \epsilon)n.$$

We have already seen that for any predictor $\pi : \Sigma^* \times \Sigma \to [0, 1]$,

$$\sum_{i=1}^{n} \mathbf{KL}((D_1, \cdots, D_{i-1}), D_i \| (D_1, \cdots, D_{i-1}), \pi_i) = \sum_{w \in \Sigma^n} \text{Loss}(\pi, w) D(w) - H(D).$$

Now the definitions of next-bit pseudo Shannon entropy, conditional pseudo Shannon entropy and KL-predictor along with Item 2 of Theorem 7 imply that for every $\sigma > 0$, $\sum_{w \in \Sigma^n} \text{Loss}(\pi, w) D(w) > sn - \sigma$.

Hence,

$$\begin{aligned} sn - \sigma &< (t - \frac{1}{n} + \epsilon_1)n + (\frac{1}{2} - \epsilon)n \\ \Rightarrow s &< t + \frac{1}{2} + \epsilon_2 - \epsilon \end{aligned}$$

for any $\epsilon_2 > \epsilon_1$, for large enough n and this concludes the proof of the theorem. □

The asymptotic version of the above two theorems will be discussed in the following section.

6 Results for Asymptotic case

In the context of pseudorandomness, it is always natural to consider asymptotic definitions. For a parameter n, a distribution D_n on Σ^n is said to be *pseudorandom* if for every constant $c > 0$ and $c' > 0$, D_n is $(n^c, 1/n^{c'})$-pseudorandom, for sufficiently large n [Gol01]. Sometimes we call a distribution D_n on Σ^n ϵ-*pseudorandom*, for some $\epsilon > 0$ if for every constant $c > 0$, D_n is (n^c, ϵ)-pseudorandom, for sufficiently large n. In a similar fashion, we can define dimension and unpredictability of a distribution.

Definition 17 (Dimension; asymptotic version)**.** *For a parameter n, a distribution D_n on Σ^n is said to have* dimension *(denoted as* $\dim(D_n)$*) s if for every constant $c > 0$ and $c' > 0$, $\dim_{n^c,1/n^{c'}}(D_n) = s$, for sufficiently large n.*

Definition 18 (Unpredictability; asymptotic version)**.** *For a parameter n, a distribution D_n on Σ^n is said to have* unpredictability *(denoted as* $\mathrm{unpred}(D_n)$*) s if for every constant $c > 0$ and $c' > 0$, $\mathrm{unpred}_{n^c,1/n^{c'}}(D_n) = s$, for sufficiently large n.*

Theorem 1 established in Section 3 implies that asymptotically both dimension and unpredictability of a distribution denote the same quantity. We state this equivalence formally in the following corollary of Theorem 1.

Corollary 6.1. *For any distribution D_n on Σ^n, $\dim(D_n) = \mathrm{unpred}(D_n)$.*

All the results of Section 4 can naturally be extended to asymptotic version. If we consider asymptotic version of HILL-type pseudo min entropy, we can say same thing for Theorem 6. The extension is so natural, it is not worth specifying explicitly. However, for next-bit pseudo Shannon entropy, we will provide the equivalence with dimension in case of asymptotic version as non-asymptotic case is slightly subtle. For the asymptotic definitions of conditional pseudo Shannon entropy, next-bit pseudo Shannon entropy (denoted as $H^{next}(X)$) and KL-hardness, we refer the reader to [VZ12]. Now we state the equivalence between conditional pseudo Shannon entropy and KL-hardness form [VZ12].

Theorem 10 ([VZ12])**.** *For a $\Sigma^n \times \Sigma$-valued random variable (X, Y), Y has conditional pseudo Shannon entropy at least $H(Y|X) + \delta$ if and only if Y is δ-KL-hard given X.*

Above theorem helps us to derive the asymptotic version of Theorem 8.

Corollary 6.2. *For any $\epsilon > 0$, there exists a $n' \in \mathbb{N}$ such that for any $n \geq n'$, for every distribution D_n on Σ^n, if $\dim(D_n) > 2s + \epsilon$, then $H^{next}(D_n) > sn$.*

In a similar fashion, we get the following asymptotic version of Theorem 9.

Corollary 6.3. *For any $\epsilon > 0$, there exists a $n' \in \mathbb{N}$ such that for any $n \geq n'$, for every distribution D_n on Σ^n, if $H^{next}(D_n) > sn$, then $\dim(D_n) > s - \frac{1}{2} - \epsilon$.*

7 Pseudorandom Extractors & Lower Bound

We now introduce the notion of *pseudorandom extractor* similar to the notion of randomness extractor. Intuitively, a *randomness extractor* is a function that outputs almost random (statistically close to uniform) bits from weakly random sources, which need not be close to the uniformly random source. Two distributions X and Y on a set Λ are said to be $\epsilon-$*close* (statistically close) if $\max_{S\subseteq\Lambda}\{|Pr[X \in S] - Pr[Y \in S]|\} \leq \epsilon$ or equivalently $\frac{1}{2}\sum\limits_{x\in\Lambda} |Pr[X = x] - Pr[Y = x]| \leq \epsilon$.

Definition 19 (Deterministic Randomness Extractor)**.** *A function $E : \Sigma^n \to \Sigma^m$ is said to be a deterministic ϵ-extractor for a class of distributions $\mathcal{C}$ if for every distribution X on n-bit strings in $\mathcal{C}$, the distribution $E(X)$ is ϵ-close to U_m.*

Likewise, a seeded ϵ-extractor is defined and the only difference is that now it takes a d-bit string chosen according to an uniform distribution, as an extra input. Before going further, we mention that for ease of presentation, now onwards we will only talk about asymptotic versions of the definitions and results derived so far related to pseudorandomness and dimension (See Section 6). We now define the notion of a pseudorandom extractor, the purpose of which is to extract out pseudorandom distribution from a given distribution.

Definition 20 (Pseudorandom Extractor)**.** *A function $E : \Sigma^n \to \Sigma^m$ is said to be a* deterministic pseudorandom extractor *for a class of distributions $\mathcal{C}$ if for every distribution X on n-bit strings in $\mathcal{C}$, $E(X)$ is pseudorandom.*

A function $E : \Sigma^n \times \Sigma^d \to \Sigma^m$ is said to be a seeded pseudorandom extractor *for a class of distributions $\mathcal{C}$ if for every distribution X on n-bit strings in $\mathcal{C}$, $E(X, U_d)$ is pseudorandom.*

In this section, we will concentrate on the class of distributions having dimension at least s. It is clear from the results stated in Section 5.2 that this class of distribution is a strict superset of the class of distributions with HILL-type pseudo min-entropy at least sn, for which any randomness extractor will act as a pseudorandom extractor [BSW03]. Thus it is natural to ask the following.

Question 1. *For any $s \in (0, 1]$, does there exist a deterministic/seeded pseudorandom extractor for the class of distributions on Σ^n having dimension at least s?*

Just like the the case of randomness extraction, one can easily argue that deterministic pseudorandom extraction is not possible[3]. Now the most common question comes next is that what the lower bound on the seed length will be. We answer to this question in the following theorem.

Theorem 11. *Suppose for any $s \in (0, 1]$, $E : \Sigma^n \times \Sigma^d \to \Sigma^m$ be a seeded pseudorandom extractor for the class of distributions on Σ^n having dimension at least s and for some $\delta > 0$, $m = (sn)^\delta$. Then $d = \Omega(\log n)$.*

Proof. For the sake of contradiction, let us assume that $d = o(\log n)$. Now by doing a walk according to the output distribution on an odd-length cycle, we achieve the following claim.

Claim 7.1. *There is a deterministic $\frac{1}{\sqrt[4]{m}}$-extractor $E' : \Sigma^m \to \Sigma^{\frac{\log m}{4}}$ for all pseudorandom distributions on Σ^m.*

We defer the proof of the above claim to Section 7.1.1 where we will prove a much stronger claim in Theorem 13 and the above claim will come as a corollary of that theorem. Now construct the following function $Ext : \Sigma^n \times \Sigma^d \to \Sigma^{c \log n}$ for some constant $c > 0$ such that $Ext(x, y) = E'(E(x, y))$ for all $x \in \Sigma^n, y \in \Sigma^d$. The function Ext is a seeded $\frac{1}{(sn)^{\delta/4}}$-extractor with $d = o(\log n)$, but it is well known due to [RTS00](Theorem 1.9) that any such randomness extractor must satisfy $d = \Omega(\log n)$ and hence we get a contradiction. ☐

However, the question on constructing an *explicit* or polynomial time computable seeded pseudorandom extractor with seed length $O(\log n)$ is still open and next, we formally pose this question.

Question 2. *For any $s \in (0, 1]$, can one construct a seeded pseudorandom extractor $E : \Sigma^n \times \Sigma^d \to \Sigma^m$ in polynomial time, for the class of distributions on Σ^n having dimension at least s such that $m = (sn)^\delta$ for some $\delta > 0$ and $d = O(\log n)$?*

In the next part of this section, we will see a special type of nonpseudorandom source and give an explicit construction of deterministic pseudorandom extractor for that particular type of source. Before proceeding further, we want to mention that it is also very interesting to consider *nonpseudorandom distributions samplable by poly-size circuits* and we will discuss on the existence of extractor for that particular source in Section 7.2.

7.1 Deterministic Pseudorandom Extractor for Nonpseudorandom Bit-fixing Sources

In Section 4 while proving Theorem 2, we have introduced a special type of nonpseudorandom distribution which looks similar to the (n, k)-*bit-fixing source* defined as a distribution X over Σ^n such that there exists a subset $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ where all the bits at the indices of I are independent and uniformly chosen

[3]Suppose $E : \Sigma^n \to \Sigma$ is a deterministic pseudorandom extractor, then there exists $x \in \Sigma$ such that $|E^{-1}(x)| \geq 2^{n-1}$. Thus E is not a pseudorandom extractor for a source D that is a uniform distribution on $E^{-1}(x)$ and by Claim 5.1, $\dim(D) \geq (1-1/n)$.

and rest of the bits are completely fixed. This distribution was introduced by Chor *et al.*[CGH+85]. Now we define an analogous notion for the class of nonpseudorandom distributions, which we term *nonpseudorandom bit-fixing sources.*

Definition 21 (Nonpseudorandom Bit-fixing Source)**.** *Let $s \in (0,1)$. For sufficiently large n and $\epsilon > 0$, a distribution D_n over Σ^n with dimension s is an (n,s,ϵ)-*nonpseudorandom bit-fixing source *if there exists a subset $I = \{i_1, \dots, i_{\lceil sn \rceil}\} \subseteq \{1, \dots, n\}$ such that all the bits at the indices of I come from an ϵ-pseudorandom distribution and rest of the bits are fixed.*

We devote the rest of the section to achieve an affirmative answer to the question of constructing deterministic pseudorandom extractor for the nonpseudorandom bit-fixing sources. For this purpose, we show that a careful analysis of the technique used in the construction of the deterministic randomness extractor for bit-fixing random sources by Gabizon, Raz and Shaltiel [GRS05] will lead us to the desired deterministic pseudorandom extractor.

Theorem 12. *There exists a constant $c > 0$ such that for any $s \in (0,1]$ and for large enough n, $0 < \epsilon < \frac{1}{\sqrt{n}}$, there is an explicit deterministic pseudorandom extractor $E : \Sigma^n \to \Sigma^m$ for all (n,s,ϵ)-nonpseudorandom bit-fixing sources having polynomial-size support, where $m = (sn)^{\Omega(1)}$.*

We first extract $O(\log sn)$ amount of almost random bits and then use the same as seed in the seeded extractor. To use the seeded extractor, we modify the source such that it becomes independent of the random bits extracted. Before going into the exact details of the proof, we first discuss the ingredients required in the proof of the theorem.

7.1.1 Pseudorandom walk and extracting a few random bits

Kamp and Zuckerman [KZ03] use a technique of random walk on odd-length cycles to extract almost random bits from a bit-fixing source. We adapt this to extract $\Omega(\log sn)$ almost random bits from a (n,s,ϵ)-nonpseudorandom bit-fixing source.

Theorem 13. *Let $n \in \mathbb{N}$ and $s \in [0,1]$ such that $k = \lceil sn \rceil > 100$, and let $0 < \epsilon < \frac{1}{\sqrt{n}}$. Then there is a deterministic $\frac{1}{\sqrt[4]{k}}$-extractor $E : \Sigma^n \to \Sigma^{\frac{\log k}{4}}$ for all (n,s,ϵ)-nonpseudorandom bit-fixing sources.*

We prove the above theorem using the property of *pseudorandom walk* together with the fact that the second largest eigenvalue of a l length odd cycle is $\cos(\pi/l)$. Note that a corollary of the above theorem is Claim 7.1 used in the proof of Theorem 11. Before proving the above theorem, we state two lemmas required for the proof. The first is a very special case of a lemma given in [KZ03] which was restated in [GRS05].

Lemma 7.1 ([GRS05])**.** *Let $n \in \mathbb{N}$ and $s \in [0,1]$. Suppose G be an odd length cycle having M vertices and having second largest eigenvalue λ. If we take a walk on G according to the bits from a (n,k)-bit-fixing source, starting from any fixed vertex, then at the end of the n step of the walk, the distribution P on the vertices will be $\frac{1}{2}\lambda^k\sqrt{M}$-close to U_M.*

Now we prove a similar result for (n,s,ϵ)-nonpseudorandom bit-fixing source where $\epsilon < \frac{1}{\sqrt{n}}$ using the property of *pseudorandom walk.* The idea of pseudorandom walk was also used previously in the domain of space bounded computation by Reingold *et al.* [RTV06].

Lemma 7.2. *Let $n \in \mathbb{N}$, $s \in [0,1]$ and $k = \lceil sn \rceil$. Suppose G be an odd length cycle having M vertices and having second largest eigenvalue λ. If we take a walk on G according to the bits from a (n,s,ϵ)-nonpseudorandom bit-fixing source starting from any fixed vertex, then at the end of the n step of the walk, the distribution Q on the vertices will be $\frac{1}{2}(\lambda^k + \sqrt{M}\epsilon)\sqrt{M}$-close to U_M, where M is polynomial in n.*

Proof. Let π be the stationary distribution on the vertices and since we consider an odd length cycle (a 2-regular graph), the stationary distribution is the uniform distribution on M vertices. Suppose we take a n step walk on the graph G starting from any vertex v according to the bits from a (n,k)-bit-fixing source, where $k = \lceil sn \rceil$ and the probability vector on the vertices at the end of the walk is $P = \begin{pmatrix} p_1 & p_2 & \dots & p_M \end{pmatrix}$. Now we take a n step walk on the graph G starting from the same vertex v according to the bits from a (n, s, ϵ)-nonpseudorandom bit-fixing source and the probability vector on the vertices at the end of the walk is $Q = \begin{pmatrix} q_1 & q_2 & \dots & q_M \end{pmatrix}$, where $\forall_{1 \le i \le M}$, $q_i \le p_i + \epsilon$. This bound on q_i can be justified as follows.

Note that the only difference between (n, s, ϵ)-nonpseudorandom bit-fixing source and (n, k)-bit-fixing source is that on the set I, in (n, k)-bit-fixing source, we have U_k instead of ϵ-pseudorandom distribution (say D) on Σ^k. Also observe that actually P and Q are the distributions on vertices at the end of a k step walk, where the walk was started from the vertex v and done according to the bits coming from U_k and D respectively, because a step according to a fixed bit will not change the output distribution and in a (n, k)-bit-fixing source (also in a (n, s, ϵ)-nonpseudorandom bit-fixing source), all the $n - k$ bits are fixed. For a step according to a fixed bit gives rise to a transition matrix that is actually a permutation matrix and thus it leaves the distance from uniform unchanged. Hence, if the bound on $q_i, \forall_{1 \le i \le M}$ is not true then we can use this k step walk on G as a polynomial (polynomial in k) time algorithm to distinguish between U_k and D. Thus,

$$||q - \pi||^2 = \sum_{i=1}^{M} (q_i - \frac{1}{M})^2 \le \sum_{i=1}^{M} (p_i + \epsilon - \frac{1}{M})^2 = ||p - \pi||^2 + M\epsilon^2 \le \lambda^{2k} + M\epsilon^2 \le (\lambda^k + \sqrt{M}\epsilon)^2$$

□

The above lemma together with the fact that the second largest eigenvalue of a l length odd cycle is $\cos(\pi/l)$, implies Theorem 13.

Proof of Theorem 13. Let us take an odd cycle G with $M = \sqrt[4]{k}$ vertices. The second largest eigenvalue of G is $\cos(\frac{\pi}{\sqrt[4]{k}})$. Now take a walk starting from a fixed vertex of G according to the bits from (n, s, ϵ)-nonpseudorandom bit-fixing source and finally output the vertex number of the graph G. Thus we get $\frac{\log k}{4}$ bits. From Lemma 7.2, we reach distance $\frac{1}{2}((\cos(\frac{\pi}{\sqrt[4]{k}}))^k + \sqrt[8]{k}\epsilon)\sqrt[8]{k}$ from uniform.

By the Taylor expansion of the cosine function, for $0 < x < 1$, $\cos(x) < 1 - \frac{x^2}{2} + \frac{x^4}{24}$. Therefore, $\left(\cos\left(\frac{\pi}{\sqrt[4]{k}}\right)\right)^k < (1 - \frac{\pi^2}{4\sqrt{k}})^k < (\exp^{-\frac{\pi^2}{4}})^{\sqrt{k}} < 4^{-\sqrt{k}}$. Hence, $\frac{1}{2}((\cos(\frac{\pi}{\sqrt[4]{k}}))^k + \sqrt[8]{k}\epsilon)\sqrt[8]{k} < \frac{1}{\sqrt[4]{k}}$. Thus we get distribution of $\frac{\log k}{4}$ bit strings which is $\frac{1}{\sqrt[4]{k}}$-close to uniform in statistical distance. □

7.1.2 Sampling and Partitioning with a short seed

Here we restate some of the results on sampling and partitioning used in construction of deterministic extractor for bit-fixing sources from [GRS05]. Let $S \subseteq [n]$ be some subset of size k. Now we consider a process of generating a subset $T \subseteq [n]$ such that $k_{min} \le |S \cap T| \le k_{max}$ and this process is known as *Sampling*.

Definition 22. *A function $Samp : \Sigma^t \to P([n])$ is called a $(n, k, k_{min}, k_{max}, \delta)$-sampler if for any subset $S \subseteq [n]$, where $|S| = k$, $Pr_{w \in_R U_t}[k_{min} \le |Samp(w) \cap S| \le k_{max}] \ge 1 - \delta$*

Now consider a similar process known as *Partitioning*, the task of which is to partition $[n]$ into m distinct subsets $T_1, T_2, \cdots, T_m$ such that for every $1 \le i \le m$, $k_{min} \le |S \cap T_i| \le k_{max}$.
According to [GRS05], the above two processes can be performed using only *a few* random bits.

Lemma 7.3 ([GRS05])**.** *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ and $\frac{1}{2} < e < 1$ such that for any $n \ge 16$ and $k \ge (\log n)^c$, there is an explicit construction of a function $Samp : \Sigma^t \to P([n])$ which is a $(n, k, \frac{k^e}{2}, 3k^e, O(k^{-b}))$-sampler, where $t = \alpha \log k$.*

Lemma 7.4 ([GRS05])**.** *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ and $\frac{1}{2} < e < 1$ such that for any $n \ge 16$ and $k \ge (\log n)^c$, there is an explicit construction that uses only $\alpha \log k$ random bits and partition $[n]$ into $m = O(k^b)$ many subsets $T_1, T_2, \cdots, T_m$ such that for any subset $S \subseteq [n]$, where $|S| = k$, $Pr[\forall 1 \le i \le m, \frac{k^e}{2} \le |T_i \cap S| \le 3k^e] \ge 1 - O(k^{-b})$.*

7.1.3 Generating an independent seed

In this subsection, we see the way of obtaining a short seed from a nonpseudorandom bit-fixing source so that we can use them in a seeded pseudorandom extractor to extract out almost all the pseudorandom part from the source. The main problem of using this short seed in a seeded pseudorandom extractor is that the already obtained seed is dependent on the main distribution. Now we describe that this problem can be removed in the case of nonpseudorandom bit-fixing sources. Even though the result is analogous to [GRS05], the proofs differ in essential details.

Definition 23 (Seed Obtainer)**.** *A function $F : \Sigma^n \to \Sigma^n \times \Sigma^d$ is said to be a (s, s', ρ)-seed obtainer ($s' \leq s$) if for every (n, s, ϵ)-nonpseudorandom bit-fixing source X, the distribution $R = F(X)$ can be written as $R = \eta Q + \sum_a \alpha_a R_a$ ($\eta, \alpha_a > 0$ and $\eta + \sum_a \alpha_a = 1$) such that $\eta \leq \rho$ and for every a, there exists a (n, s', ϵ)-nonpseudorandom bit-fixing source Z_a such that R_a is ρ-close to $Z_a \otimes U_d$.*

From the above definition it is clear that given a seed obtainer and a seeded pseudorandom extractor for nonpseudorandom bit-fixing sources, we can easily construct a deterministic pseudorandom extractor. The following theorem provides us the details of such construction, where the correctness follows from the properties of our proposed notion of dimension described in Section 4.

Theorem 14. *Suppose $F : \Sigma^n \to \Sigma^n \times \Sigma^d$ is a (s, s', ρ)-seed obtainer, where $\rho \leq \frac{1}{(sn)^c}$ for some constant $c > 0$ and $E' : \Sigma^n \times \Sigma^d \to \Sigma^m$ is a seeded pseudorandom extractor for (n, s', ϵ)-nonpseudorandom bit-fixing sources, where $m = (sn)^{\Omega(1)}$. Then the function $E : \Sigma^n \to \Sigma^m$ defined as $E(x) = E'(F(x))$ is a deterministic pseudorandom extractor for (n, s, ϵ)-nonpseudorandom bit-fixing sources.*

Proof. By the definition of the seed obtainer, we can write $E(X) = \eta E'(Q) + \sum_a \alpha_a E'(R_a) = \eta E'(Q) + (1 - \eta)Y$, for some distribution Y. Now by Lemma 4.2, for all a, $E'(R_a)$ is pseudorandom and as a consequence, by Lemma 4.1, Y is pseudorandom as well. Then using Lemma 4.2, we can argue that $E(X)$ is also pseudorandom as $\eta \leq \frac{1}{(sn)^c}$, for some constant $c > 0$. □

Now we give an explicit construction of (s, s', ρ)-seed obtainer, which is crucial in the later part of this paper. To understand the notion of *sampler* used in the following theorem, the readers may refer to the last subsection.

Theorem 15. *For every n, let $Samp : \Sigma^t \to P([n])$ be a $(n, \lceil sn \rceil, \lceil s_1 n \rceil, \lceil s_2 n \rceil, \delta)$-sampler and $E : \Sigma^n \to \Sigma^m$ with $m > t$ be a deterministic ϵ'-extractor for (n, s_1, ϵ)-nonpseudorandom bit-fixing sources, where $\sqrt{\epsilon} \leq \epsilon'$. Then there is an explicit (s, s', ρ)-seed obtainer $F : \Sigma^n \to \Sigma^n \times \Sigma^d$, where $d = m - t$, $s' = s - s_2$, and $\rho = \max\{\epsilon' + \delta, \sqrt{\epsilon'} 2^{t+1}\}$.*

The construction of the seed obtainer is same as that mentioned in [GRS05], however the proof requires slightly different argument.

Proof. The construction of F mentioned in the theorem is as follows: (1) Given $x \in \Sigma^n$, compute $E(x)$. Denote the first t bits of $E(x)$ by $E_1(x)$ and the last $(m - t)$ bits by $E_2(x)$; (2) Compute $Samp(E_1(X))$ and denote it as T; (3) Let $x' = x_{[n] \setminus T}$ and $y = E_2(x)$. If $|x'| < n$, pad it with zeros to get n-bit long string. Now output x', y.
Note that the above construction is same as the construction of seed obtainer given in [GRS05]. However, the proof is not same and more specifically the proof of Claim 7.2 differs from that of the similar claim made in [GRS05]. Here, in the proof we use the properties of pseudorandomness and the fact that the distribution under consideration has polynomial-size support.

Let X be a (n, s, ϵ)-nonpseudorandom bit-fixing source and I be the set of indices at which the bits are not fixed. For a string $a \in \Sigma^t$, T_a denotes $Samp(a)$ and T'_a denotes $[n] \setminus Samp(a)$. Given a string $x \in \Sigma^n$, x_a denotes x_{T_a} and x'_a denotes n-bit string obtained by padding $x_{T'_a}$ with zeros. Let $X' = X'_{E_1(X)}$ and $Y = E_2(X)$. We say that a string $a \in \Sigma^t$ *correctly splits* X if $\lceil s_1 n \rceil \leq |I \cap T_a| \leq \lceil s_2 n \rceil$.

Claim 7.2. *For every $a \in \Sigma^t$ which correctly splits X, $(X'_a, E(X))$ is $\sqrt{\epsilon'}$-close to $(X'_a \otimes U_m)$.*

Proof. Let $|Samp(a)| = l$. Given a string $\sigma \in \Sigma^l$ and a string $\sigma' \in \Sigma^{n-l}$, we define $[\sigma; \sigma']$ as follows: Suppose l indices of T_a are $i_1 < \cdots < i_l$ and the $(n-l)$ indices of T'_a are $i'_1 < \cdots < i'_{n-l}$. The string $[\sigma; \sigma'] \in \Sigma^n$ is defined as:

$$[\sigma; \sigma']_i = \begin{cases} \sigma_j & i \in T_a \text{ and } i_j = i \\ \sigma'_j & i \in T'_a \text{ and } i'_j = i \end{cases}$$

In this notation, we denote $X = [X_a; X'_a]$. Now consider the distribution $(X'_a, E(X)) = (X'_a, E([X_a; X'_a]))$. For every $b \in \Sigma^{n-l}$, we consider the event $\{X'_a = b\}$. As a correctly splits X, there are at least $\lceil s_1 n \rceil$ "good" indices in T_a. Now fix some $b \in \Sigma^{n-l}$ such that $X[X'_a = b] > 0$.

Now we claim that for all subsets $B \subseteq \Sigma^{n-l}$ where $\forall_{b \in B}, X[X'_a = b] > 0$, there exists a $b' \in B$ such that the distribution $([X_a; X'_a] | X'_a = b')$ is $(n, s_1, \sqrt{\epsilon})$-nonpseudorandom bit-fixing source if $\sum_{b \in B} X[X'_a = b] > \sqrt{\epsilon}$.

For the sake of contradiction, let us assume that the above claim is not true. It means that there exists a subset $J \subseteq \Sigma^{n-l}$, where (i) $\forall_{b \in J}, X[X'_a = b] > 0$, (ii) $\sum_{b \in J} X[X'_a = b] > \sqrt{\epsilon}$ and also (iii) for all $b \in J$, the distributions $([X_a; X'_a] | X'_a = b)$ are not $(n, s_1, \sqrt{\epsilon})$-nonpseudorandom bit-fixing sources. Now let us consider only the "good" positions which are $\lceil sn \rceil$ many in X and at least $\lceil s_1 n \rceil$ many in $([X_a; X'_a] | X'_a = b)$. So the above assumption implies that the distribution on those $\lceil s_1 n \rceil$ bits (this part of the string b is denoted as $b_{\lceil s_1 n \rceil}$) in $([X_a; X'_a] | X'_a = b)$ is not $\sqrt{\epsilon}$-pseudorandom, i.e., it has its corresponding distinguishing circuits C_b. If this is the case, then the circuit C (by hard-wiring the good random bits) corresponding to the following algorithm A, will act as a distinguishing circuit for the ϵ-pseudorandom distribution P on $\lceil sn \rceil$ many bits; which is a contradiction. The algorithm A is as follows: on input $y \in \{0,1\}^{\lceil s_1 n \rceil}$, if $y_{\lceil s_1 n \rceil} = b_{\lceil s_1 n \rceil}$ for any $b \in J$, then return $C_b(y_{\lceil s_1 n \rceil})$; otherwise return 0 or 1 uniformly. And thus clearly, $|P[A[y] = 1] - U_{\lceil s_1 n \rceil}[A[y] = 1]| > \epsilon$.

The circuit C is of polynomial size as the support of J is at most polynomial. Note that this is the only place where we use the fact that the distribution under consideration is of polynomial-size support.

So, we can write,

$$\frac{1}{2} \sum_{b,c} |Pr[(X'_a, E(X)) = (b, c)] - Pr_{(X'_a \otimes U_m)}[b, c]|$$

$$= \frac{1}{2} \sum_{b,c} |Pr[X'_a = b] Pr[E(X) = c | X'_a = b] - Pr[X'_a = b] Pr_{U_m}[c]| \leq \sqrt{\epsilon} + \epsilon' \leq \sqrt{\epsilon'}$$

where $\sqrt{\epsilon} \leq \epsilon'$. The first inequality follows from the fact that we can split the sum in two parts one in which $([X_a; X'_a] | X'_a = b)$'s are not $(n, s_1, \sqrt{\epsilon})$-nonpseudorandom bit-fixing sources and another in which $([X_a; X'_a] | X'_a = b)$'s are at least $(n, s_1, \sqrt{\epsilon})$-nonpseudorandom bit-fixing sources. ☐

Next we mention a claim from [GRS05] that makes comment on independence of the pair $(X'_a, E_2(X))$ conditioned on the event $E_1(X) = a$).

Claim 7.3 ([GRS05])**.** *For every fixed $a \in \Sigma^t$ that correctly splits X, the distribution $((X'_a, E_2(X)) \mid E_1(X) = a)$ is $\sqrt{\epsilon'} 2^{t+1}$-close to $(X'_a \otimes U_{m-t})$.*

Note that as a correctly splits X, so X'_a is a $(n, s - s_2, \epsilon)$-nonpseudorandom bit-fixing source.

The rest of the proof follows directly from the proof of correctness of the construction of seed obtainer given in [GRS05] with the following parameters $k = \lceil sn \rceil$, $k_{min} = \lceil s_1 n \rceil$, $k_{max} = \lceil s_2 n \rceil$ and $\epsilon = \sqrt{\epsilon'}$. ☐

7.1.4 A seeded pseudorandom extractor

In this subsection, we discuss about how we can extract $(sn)^{\Omega(1)}$ many pseudorandom bits using $O(\log sn)$ random bits. In the next subsection, we will use this seeded pseudorandom extractor and the techniques discussed in the previous subsections, to construct deterministic extractor. The construction of seeded pseudorandom generator given in the proof of the following theorem is same as that of the seeded randomness extractor given in [GRS05]. However, the analysis is quite different and uses some of the properties of dimension.

Theorem 16. *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ such that for any $n \geq 16$ and $sn \geq (\log n)^c$, there is an explicit function $E : \Sigma^n \times \Sigma^d \to \Sigma^m$ which acts as a seeded pseudorandom extractor for (n, s, ϵ)-nonpseudorandom bit-fixing sources with $d = \alpha \log sn$ and $m = \Omega((sn)^b)$.*

Proof. Let X be a (n, s, ϵ)-nonpseudorandom bit-fixing source and x be a string sampled by X. The description of the extractor $E(x, y)$ is as follows: (1) According to Lemma 7.4 provided in Section 7.1.2, using y as seed, we obtain a partition of $[n]$ into $m = \Omega((sn)^b)$ many sets $T_1, T_2, \cdots, T_m$ with the parameter α, (2) For $1 \leq i \leq m$, compute $z[i] = \oplus_{j \in T_i} x[j]$, (3) Output $z = z[1]z[2]\cdots z[m]$.
Let $I \subseteq [n]$ be the set of indices at which the bits are not fixed and let Z be the distribution of the output strings. We need to show that Z is pseudorandom.
Let A be the event $\{\forall i, |T_i \cap I| \neq 0\}$ and $A^c = \{\exists i, |T_i \cap I| = 0\}$ be the complement event. According to Lemma 7.4, $Pr[A] \geq 1 - O((sn)^{-b})$. Now we can write the output distribution as $Z = Pr[A](Z|A) + Pr[A^c](Z|A^c)$ and hence due to Lemma 4.2, Z is pseudorandom. □

7.1.5 Deterministic pseudorandom extractor

Now it only remains to combine all the components we discussed so far to build the final deterministic pseudorandom extractor mentioned in Theorem 12. We first extract $O(\log sn)$ amount of almost random bits by Theorem 13 and then use the same as seed in the seeded extractor described in Theorem 16. To use the seeded extractor it is required to modify the source such that it becomes independent of the random bits extracted using Theorem 13. For that purpose, we use the technique developed in Section 7.1.3 and this concludes the proof of Theorem 12.

Proof of Theorem 12. Due to Lemma 7.3, we have a $(n, sn, \frac{(sn)^e}{2}, 3(sn)^e, (sn)^{-\Omega(1)})$-sampler $Samp : \Sigma^t \to P([n])$, where $t = \frac{\log sn}{32}$ and $e > \frac{1}{2}$. From Theorem 13, we have a deterministic $\frac{1}{\sqrt[4]{s'n}}$-extractor $E^* : \Sigma^n \to \Sigma^{m'}$ for (n, s', ϵ)-nonpseudorandom bit-fixing sources where $s' = \frac{(sn)^e}{2n}$ and $m' = \frac{\log s'n}{4}$. Now we use Theorem 15 to get (s, s'', ρ)-seed obtainer $F : \Sigma^n \to \Sigma^n \times \Sigma^{m'-t}$ where $s'' = \frac{3(sn)^e}{n}$ and $\rho = \frac{1}{(sn)^p}$, for some constant p. According to Theorem 16, we have a seeded pseudorandom extractor $E' : \Sigma^n \times \Sigma^d \to \Sigma^m$ with $d = \frac{\log sn}{32}$ and $m = (sn - s''n)^{\Omega(1)}$ for $(n, s - s'', \epsilon)$-nonpseudorandom bit-fixing sources. Since $m' = \frac{\log s'n}{4} \geq \frac{\log sn}{16} = t + d$, we use F and E' in Theorem 14 to construct a deterministic pseudorandom extractor $E : \Sigma^n \to \Sigma^m$. For a large enough n, $m = (sn - s''n)^{\Omega(1)} = (sn)^{\Omega(1)}$ and this completes the proof. □

7.2 Discussion on Pseudorandom Extractor for Nonpseudorandom Samplable Distributions

Another interesting special kind of source is samplable distributions studied by Trevisan and Vadhan [TV00]. In a natural way, one can extend the definition of samplable distribution to nonpseudorandom distribution as follows: for any $s \in (0, 1]$, a distribution D_n on Σ^n is said to be s-nonpseudorandom samplable by circuit of size S if there exists a circuit C of size at most S that samples from it and for large enough n, $\dim(D_n) = s$. Observe that the negative results for deterministic randomness extractor in case of samplable distributions will also applicable for deterministic pseudorandom extractor in case of s-nonpseudorandom samplable distribution. By Claim 5.1, for large enough n, if $H_\infty(D_n) \geq n - 1$, then $\dim(D_n) = 1$. Now by applying the argument in [TV00], we get the following.

Theorem 17. *Suppose $\{E_n : \Sigma^n \to \Sigma\}_{n>0}$ is a family of function computable in time $t(n)$ such that for every n, E_n is a deterministic pseudorandom extractor for distributions that are 1-nonpseudorandom samplable by circuit of size $s(n)$. Then for sufficiently large n, there is a language in $DTIME(t(n))$ of circuit complexity at least $\Omega(s(n))$.*

As the existence of deterministic pseudorandom extractor implies separation between deterministic complexity classes and non-uniform circuit classes which is not known so far. So one might have to consider some complexity theoretic assumptions like in [TV00] to construct deterministic pseudorandom extractor. However, we do not think construction with such strong assumption like in [TV00] will be interesting in this case as it is known that certain hardness assumption already leads to a construction of optimal pseudorandom generator (See Section 8). Nevertheless, it is natural to ask the question of constructing explicit extractor using $O(\log n)$ amount of extra randomness. We do not know any such result so far, but in the next section we will see that if some distribution is samplable using very few ($O(\log n)$) random bits, then it is possible to extract out all the pseudorandom bits using extra $O(\log n)$ random bits.

8 Approaching Towards P=BPP

We now show that if there is an exponential time computable algorithm $G : \Sigma^{O(\log n)} \to \Sigma^n$ where the output distribution has dimension s ($s > 0$), then this will imply P=BPP. We refer to this algorithm G as *optimal nonpseudorandom generator*. The proof of this is similar to the proof of Theorem 18 [NW94]. We start with some basic definitions.

A pseudorandom generator against a class of circuits is a function which takes a random seed as input and outputs a sequence of bits which is a pseudorandom distribution.

Definition 24 (Pseudorandom Generators). *A function G is said to be a $l(n)$-pseudorandom generator if*

1. $G = \{G_n\}_{n>0}$ *with* $G_n : \Sigma^{l(n)} \to \Sigma^n$
2. G_n *is computable in* $2^{O(l(n))}$ *time*
3. *For sufficiently large n, $G_n(U_{l(n)})$ is $(n^2, 1/n)$-pseudorandom.*

Definition 25 (optimal Pseudorandom Generators). *A function G is said to be an optimal pseudorandom generator if it is an $O(\log n)$-pseudorandom generator.*

Nisan and Wigderson [NW94] showed that there is a connection between pseudorandom generators and *hard functions* in EXP:

Definition 26 (Hard Function). *A function $f : \Sigma^n \to \Sigma$ is (S, ϵ)-hard if for all circuits of size at most S,*

$$U_n[C(x) = f(x)] \leq \frac{1}{2} + \epsilon.$$

The following theorem shows that under the assumption of existence of hard function in EXP, optimal pseudorandom generator exists [NW94].

Theorem 18 ([NW94]). *There exists an optimal pseudorandom generators if and only if there is a language L in* EXP *and $\exists \delta > 0$ such that L on inputs of length n is $(2^{\delta n}, 1/2^{\delta n})$-hard.*

The proof of the above theorem is constructive and thus we can explicitly convert optimal pseudorandom generators to the hard function and conversely. However this is still a very strong requirement and later Impagliazzo and Wigderson weakened it.

Theorem 19 ([IW96]). *Suppose there is a language L in* EXP *and $\exists \delta > 0$ such that L on inputs of length n cannot be solved by circuits of size at most $2^{\delta n}$. Then there exists a language L' in* EXP *and $\exists \delta' > 0$ such that L' on inputs of length n is $(2^{\delta' n}, 1/2^{\delta' n})$-hard and as a consequence optimal pseudorandom generator exists.*

Now let us state and prove the main result of this section.

Theorem 20. *Consider any $s \in (0, 1]$ and $c > 0$. If there exists an algorithm $G_n : \Sigma^{c \log n} \to \Sigma^n$ computable in $2^{O(\log n)}$ such that for sufficiently large n, $\dim(G_n(U_{c \log n})) \geq s$, then* P=BPP.

Proof. Suppose $X := G_n(U_{c \log n})$. If $\dim(X) = s > 0$, then there must be a subset of indices $S \subseteq \{1, 2, \cdots, n\}$ such that $|S| = \log n$ and for any $i \in S$, loss incurred by any poly-size predictor at i-th bit position is non-zero or in other words, for any poly-size circuit C, $X[C(x_1, \cdots, x_{i-1}) = x_i] < 1$. Otherwise according to Theorem 1 and by the argument used in the proof of Theorem 5, one can show that $\dim(X) = 0$, for large enough n. Suppose S contains first $\log n$ many such indices. Also assume that $S = \{i_1, i_2, \cdots, i_{\log n}\}$ and $i_1 < i_2 < \cdots < i_{\log n}$. Now we define two languages L_0 and L_1 as follows: for $j = 0, 1$, $L_j := \{y \in \Sigma^{\log n - 1} | \exists x \in \Sigma^n$ in the support of G_n and $x_S = jy\}$.

First of all, note that as $i_1 \in S$, none of L_0 and L_1 is a constant function. Now clearly either L_0 or L_1 is the language that satisfies all the conditions of Theorem 19 [IW96]. Otherwise, there exists a predictor circuit of size at most $2^{\delta \log n}$, for some $\delta > 0$, i.e., polynomial in n, by which we can predict $i_{\log n}$-th bit position or loss incurred by that predictor at $i_{\log n}$th bit position will be zero implying $i_{\log n} \notin S$ which is a contradiction. Thus either L_0 or L_1 can be used to construct an *optimal pseudorandom generator* and which eventually implies P=BPP. □

Acknowledgements

The authors thank Somenath Biswas for helpful discussions and comments. The second author also thanks Andrej Bogdanov for suggesting the study of the notion of pseudoentropy.

References

[AHLM07] K. B. Athreya, J. M. Hitchcock, J. H. Lutz, and E. Mayordomo. Effective strong dimension, algorithmic information, and computational complexity. *SIAM Journal on Computing*, 37:671–705, 2007.

[AL06] K. B. Athreya and S. N. Lahiri. *Measure Theory and Probability Theory.* Springer Verlag, 2006.

[BM84] Manuel Blum and Silvio Micali. How to generate cryptographically strong sequences of pseudorandom bits. *SIAM J. Comput.*, 13(4):850–864, November 1984.

[BSW03] Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In Sanjeev Arora, Klaus Jansen, Jos D. P. Rolim, and Amit Sahai, editors, *RANDOM-APPROX*, volume 2764 of *Lecture Notes in Computer Science*, pages 200–215. Springer, 2003.

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Freidmann, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions, 1985.

[Cov74] T. Cover. Universal gambling schemes and the complexity measures of Kolmogorov and Chaitin. Technical Report 12, Stanford University Department of Statistics, October 1974.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing).* Wiley-Interscience, 2006.

[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[Gol01] Oded Goldreich. *The Foundations of Cryptography - Volume 1, Basic Techniques.* Cambridge University Press, 2001.

[GRS05] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed, 2005.

[HILL99] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[Hita] J. M. Hitchcock. Effective Fractal Dimension Bibliography, http://www.cs.uwyo.edu/~jhitchco/bib/dim.shtml (current April, 2011).

[Hitb] J. M. Hitchcock. Resource Bounded Measure - Bibliography, http://www.cs.uwyo.edu/~jhitchco/bib/rbm.shtml (current April, 2011).

[Hit03] J. M. Hitchcock. Fractal dimension and logarithmic loss unpredictability. *Theoretical Computer Science*, 304(1–3):431–441, 2003.

[Hit04] John M. Hitchcock. Fractal dimension and logarithmic loss unpredictability, 2004.

[HLR07] Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. Conditional computational entropy, or toward separating pseudoentropy from compressibility. In *Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007, Proceedings*, pages 169–186, 2007.

[HRV10] Iftach Haitner, Omer Reingold, and Salil P. Vadhan. Efficiency improvements in constructing pseudorandom generators from one-way functions. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 437–446, 2010.

[IW96] Russell Impagliazzo and Avi Wigderson. P=BPP unless E has sub-exponential circuits: Derandomizing the xor lemma (preliminary version). In *In Proceedings of the 29th STOC*, pages 220–229. ACM Press, 1996.

[KZ03] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. In *In Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 92–101, 2003.

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Extractors: optimal up to constant factors. In Lawrence L. Larmore and Michel X. Goemans, editors, *STOC*, pages 602–611. ACM, 2003.

[Lut00] J. H. Lutz. Gales and the constructive dimension of individual sequences. In *Proceedings of the 27th International Colloquium on Automata, Languages, and Programming*, pages 902–913, 2000. Revised as [Lut03a].

[Lut03a] J. H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187:49–79, 2003. Preliminary version appeared as [Lut00].

[Lut03b] Jack H. Lutz. Dimension in complexity classes. *SIAM J. Comput.*, 32(5):1236–1259, 2003.

[NT99] Noam Nisan and Amnon Ta-Shma. Extracting randomness: A survey and new constructions. *J. Comput. Syst. Sci.*, 58(1):148–173, 1999.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[NZ93] Noam Nisan and David Zuckerman. More deterministic simulation in logspace. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, May 16-18, 1993, San Diego, CA, USA*, pages 235–244, 1993.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52:43–52, 1996.

[RTS00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13:2000, 2000.

[RTV06] Omer Reingold, Luca Trevisan, and Salil Vadhan. Pseudorandom walks on regular digraphs and the rl vs. l problem. In *In Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC 06)*, pages 457–466, 2006.

[Sha02] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.

[TV00] Luca Trevisan and Salil P. Vadhan. Extracting randomness from samplable distributions. In *FOCS*, pages 32–42. IEEE Computer Society, 2000.

[VZ12] Salil P. Vadhan and Colin Jia Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 817–836, 2012.

[Yao82] Andrew C. Yao. Theory and application of trapdoor functions. In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, SFCS '82, pages 80–91, Washington, DC, USA, 1982. IEEE Computer Society.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de