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

Measure of Non-pseudorandomness and Deterministic Extraction of Pseudorandomness*

Manindra Agrawal† Diptarka Chakraborty‡ Debarati Das § and Satyadev Nandakumar ¶

Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur,
Kanpur, U.P., PIN-208016, India

Abstract

In this paper, we propose a quantification of distributions on a set of strings, in terms of how close to pseudorandom the distribution is. The quantification is an adaptation of the theory of dimension of sets of infinite sequences first introduced by Lutz [13]. We show that this definition is robust, by considering an alternate, equivalent quantification. It is known that pseudorandomness can be characterized in terms of predictors [20]. Adapting Hitchcock [10], we show that the log-loss function incurred by a predictor on a distribution is quantitatively equivalent to the notion of dimension we define. We show that every distribution on a set of strings of length n has a dimension $s \in [0,1]$, and for every $s \in [0,1]$ there is a distribution with dimension s. We study some natural properties of our notion of dimension.

Further, we propose an application of 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 deterministically extract out *sn pseudorandom* bits out of the distribution? We show that this is possible in a special case - a notion analogous to the bit-fixing sources introduced by Chor *et. al.* [5], which we term a *nonpseudorandom bit-fixing source*. We adapt the techniques of Kamp and Zuckerman [12] and Gabizon, Raz and Shaltiel [7] to establish that in the case of a non-pseudorandom bit-fixing source, we can deterministically extract the pseudorandom part of the source. Further, we show that the existence of optimal nonpseudorandom generator is enough to show P = BPP.

*Research supported by Research-I Foundation
†manindra@cse.iitk.ac.in
‡diptarka@cse.iitk.ac.in
§debarati@cse.iitk.ac.in
¶satyadev@cse.iitk.ac.in

ISSN 1433-8092

1 Introduction

Randomness is a powerful tool in algorithm design. When an algorithm uses randomness, the source of randomness is considered as an ideal process that outputs *independent* and *unbiased* random bits. A fundamental question in Computer Science is to fill the gap between the realistic source of randomness and the ideal one. One way of doing this is to construct an *efficient* algorithm that converts a realistic source to an *almost* ideal source of randomness. Such algorithms are known as *randomness extractors* and were first proposed by Nisan and Zuckerman [16]. It was shown in [17], that there is no such deterministic algorithm. However, if we allow a few ($O(\log n)$) random bits as inputs to such an extractor, then we know the construction due to [19].

Fortunately, many randomized algorithms do not require pure random bits. What we need is a source that *looks* random to those algorithms. Thus leads us to the notion of *pseudorandomness*. Pseudorandom generators have been investigated extensively [1]. We know that if we have an *optimal pseudorandom generator*, then P=BPP. Under reasonable assumptions, we can construct such optimal pseudorandom generators [15], but still we do not know any unconditional construction. Analogously, we could ask for the design of algorithms that convert sources emitting *almost pseudorandom* bits to sources that give pseudorandom bits. The first hindrance in this process is the lack of a quantification that measures the amount of pseudorandomness present in a distribution.

In this paper, we first propose a measure that can quantify the amount of pseudorandomness present in a particular distribution. This measure is motivated by the idea of *dimension* [14] and *unpredictability* [10]. Lutz used the betting function known as *gales* to characterize the *Hausdroff dimension* of sets of infinite sequences over a finite alphabet. The definitions given by Lutz cannot be carried over directly as instead of sets containing infinite length strings, we consider distributions over finite length strings. Thus we we allow our gale to access some amount of pure random bits and introduce a new probabilistic notion of *success* of a gale over a distribution. We use this to define the *dimension* of a pseudorandom distribution. We also show that the definition is robust by showing an equivalent definition in terms of unpredictability, which Yao [20] used to characterize pseudorandomness. In [10], Hitchcock showed that the dimension definition given by Lutz is same as *log-loss unpredictability*. In this paper, we show that this result can be adapted to show that a notion of log-loss unpredictability of a distribution is quantitatively equivalent to our notion of dimension.

Once we have a quantification of the pseudorandomness of a distribution, we consider extractionof its pseudorandom part. Our main objective is to construct a *deterministic pseudorandom extractor*. As a first step in doing so, we consider a special kind of source defined by *nonpseudorandom bit-fixing source* (see Section 5), which is similar to well studied *bit-fixing random source* introduced by Chor *et. al.* [5]. From [12] and [7], we know the construction of deterministic randomness extractor for bit-fixing random source. In this paper, we show that similar construction yields a deterministic pseudorandom extractor for a nonpseudorandom bit-fixing source with *polynomial support*.

The rest of the paper is organized as follows. In the next section, we define a characterization over nonpseudorandom distributions using the notion of dimension and then in section 3, we establish a relationship between dimension and unpredictability of the nonpseudorandom distribution. In section 4, we discuss some useful properties of dimension. Further sections deal with the extractor. In section 5, we introduce the notion of pseudorandom extractor and define a special kind of nonpseudorandom source, named Nonpseudorandom Bit-Fixing Source. We give an explicit construction of deterministic pseudorandom extractor for such a source in section 6. We conclude by showing that the existence of a certain nonpseudorandom generator suffices to conclude P=BPP.

2 Quantification of Nonpseudorandomness

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. In this section, we propose a quantification of pseudorandomness. We adapt the notion introduced by Lutz [14] of an s-gale to define a variant notion of success of an s-gale against a distribution D on Σ^n. First, we consider the definition of pseudorandomness.

2.1 Pseudorandomness

Definition 2.1 (Pseudorandomness)**.** A distribution D over Σ^n is (S, ϵ)-*pseudorandom* (for $S \in \mathbb{N}, \epsilon > 0$) if for every circuit C of size at most S, $|D[C(x) = 1] - U_n[C(x) = 1]| \leq \epsilon$.

Definition 2.2 (Unpredictability implies Pseudorandomness [20])**.** A distribution D over Σ^n is (S, ϵ)-pseudorandom (for some $S > 10n, \epsilon > 0$) 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 $i \in [n]$.

In this paper, we consider only (S, ϵ)-pseudorandom distributions where S is a polynomial and ϵ is an inverse polynomial in in the length of the input strings. In the rest of the paper, where unambiguous, we refer to (S, ϵ)-pseudorandom distributions as ϵ-pseudorandom distributions, or simply as pseudorandom distributions. This assumption is meaningful because of the definition of pseudorandomness given by Blum-Micali in [4].

2.2 Martingales, s-gales and predictors

Martingales are "fair" betting games which are used extensively in probability theory (see for example, [3]). Lutz introduced a generalized notion, that of an s-gale, to characterize Hausdorff dimension [13] and Athreya et al. used a similar notion to characterize packing dimension[2]. We now introduce these notions, adapt them to the quantification of distributions, and establish some of their properties.

Definition 2.3. [13] 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.

In this paper, we consider probabilistic polynomial martingales (and s-gales). These are martingales (or s-gale) that have access to random coin tosses, and can be computed in polynomial time.

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 [1]. Given an initial finite segment of a string, a predictor specifies a probability distribution over Σ for the next symbol in the string.

Definition 2.4. A function $\pi : \Sigma^* \times \Sigma \to [0, 1]$ is a *predictor* if for all $w \in \Sigma^*$, $\pi(w, 0) + \pi(w, 1) = 1$.

2.3 Conversion Between s-Gale & Predictor

There is an equivalence between gales and predictors. An early reference to this is [6]. We follow the construction in [10].

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 is defined as:

$$\pi_d(w, a) = \begin{cases} 2^{-s\frac{d(wa)}{d(w)}} & \text{if } d(w) \neq 0 \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

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 [8] and [9]. 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 except for this minor difference.

2.4 Defining Nonpseudorandomness

Definition 2.5. A martingale $d : \Sigma^* \to [0, \infty)$ is said to *succeed over a distribution D* on Σ^n if $Pr_{w \in_R D, \text{randomness of } d}[d(w) \geq 2] > \frac{1}{2} + \frac{1}{n^c}$, for all constants c.
Let $s \in [0, \infty)$. An s-gale $d : \Sigma^* \to [0, \infty)$ is said to *succeed over a distribution D* on Σ^n if $Pr_{w \in_R D, \text{randomness of } d}[d(w) \geq 2] > \frac{1}{2}2^{-n(1-s)(1+\frac{1}{s})} + \frac{1}{n^c}$, for all constants c if $s > 0$, and if $Pr_{w \in_R D, \text{randomness of } d}[d(w) \geq 2] > \frac{1}{n^c}$ otherwise.

In the above definition, observe that when $s = 1$, the winning criteria for the 1-gale and that of a martingale coincide. Using the above definition of success, we prove an equivalent characterization of pseudorandomness. For notational convenience in the remainder of the paper, define $f(s, n) = 2^{-n(1-s)(1+\frac{1}{s})}$ for $s > 0$, and $f(0, n) = 0$. The following lemma states the equivalence between the standard definition of pseudorandomness and the definition using martingale.

Lemma 2.6. *Let n be a positive integer, S be a polynomial in n and ϵ be an inverse polynomial in n. Then a distribution D over Σ^n is (S, ϵ)-pseudorandom if and only if there is no martingale which succeeds on D.*

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

$$Pr_{w \in_R D, \text{randomness of } d}[d(w) \geq 2] > \frac{1}{2} + \frac{1}{n^c}, \text{ for all constants } c$$

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

Then by an averaging argument there exists a random string r of length polynomial in$n$ such that the deterministic polynomial-time martingale $d' : \Sigma^* \to [0, \infty)$ obtained by hardcoding r into d satisfies

$$D[d'(w) \geq 2] > \frac{1}{2} + \frac{1}{n^c}, \text{ for all constants } c$$

Let $C_{d'}$ be a polynomial-size circuit obtained by instantiating d' at length n. Now let C be a circuit which outputs 1 if $C_{d'}(w) \geq 2$. Then,

$$|D[C(w) = 1] - U_n[C(w) = 1]| > \frac{1}{n^c}, \text{ for all constants } c$$

Thus D is not an (S, ϵ)-pseudorandom distribution.

Now for the converse direction, assume that D is not a (S, ϵ)-pseudorandom distribution. Then there exists an bit position $i \in [0, n-1)$ and some circuit C of polynomial size 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 $d(w[0 \ldots i-1]\overline{b}) = 0$.

Now it is clear that

$$D[d(w) \geq 2] > \frac{1}{2} + \frac{1}{n^c}, \text{ for all constants } c.$$

□

Definition 2.7. (Dimension and Nonpseudorandomness) The dimension of a distribution D on Σ^n is defined as $\dim(D) = \inf\{s \in [0,1] \mid \exists s-$ gale d which succeeds on $D\}$. If the dimension of a distribution is s, we say that it is s-nonpseudorandom.

No martingale can succeed over a pseudorandom (and, *a fortiori* random) distribution. However, note that there are distributions over which some martingale can succeed, but no s-gale can, for $s \in [0, 1)$.

3 Unpredictability and Dimension

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

Commonly used loss functions include the *absolute loss function* which penalizes 1 for every bit incorrectly predicted and 0 for every correct prediction, the *square loss function* which penalizes the predictor the amount $(b - p_b)^2$ when the outcome is b, and the *log loss function*, which penalizes $-\log(p_b)$ when the outcome is 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 [10], we establish that there is an equivalence between the notion of dimension that we define in the previous section, and the log loss function defined on a predictor.

Definition 3.1. The *logarithmic loss function* on a probability $p \in [0, 1]$ 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 3.2. Let $\pi : \Sigma^* \times \Sigma \to [0, 1]$ be a predictor.

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

The unpredictability of a distribution is defined as the infimum of the loss rate that any polynomial-time predictor has to incur on the distribution.

Definition 3.3. The *unpredictability of a distribution* D on Σ^n is
$\mathrm{unpredictability}(D) = \inf\{\mathrm{LossRate}(\pi, D) \mid \pi \text{ is a polynomial-time predictor}\}$.

With this, we can prove that dimension can equivalently be defined using unpredictability and the proof is motivated from the proof of the equivalence between log-loss unpredictability and dimension [10].

Theorem 3.4. *For any distribution D on Σ^n, unpredictability$(D) = \dim(D)$.*

Proof. First, let D be a distribution on Σ^n with dimension $s \in [0,1)$ where $n > \frac{2}{s}$. Assume that s' is a number such that $s < s' - \frac{1}{n} \leq 1$. 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

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

So $\text{LossRate}(\pi_d, w) \leq s' - \frac{1}{n}$. It is easy to see that $D[d(w) \geq 2] = D[w \mid \text{Loss}_{\pi_d}(w) \leq s' - \frac{1}{n}]$, which is greater than

$$2^{-1-n(1-s')(1+\frac{1}{s'})} + \frac{1}{n^c}, \forall c > 0.$$

Conversely, assume that unpredictability(D)=$t \in [0,1)$. Assume that t' is a number satisfying $t < t' \leq 1$. Let π be a predictor such that

$$D[w \in \Sigma^n \mid \text{LossRate}(\pi, w) \leq t' - \frac{1}{n}] > 2^{-1-n(1-t')(1+\frac{1}{t'})} + \frac{1}{n^c}, \forall c > 0.$$

If d' is the t'-gale defined by $d_\pi(w) = 2^{t'|w|}\Pi_{i=1}^{|w|}\pi(w[0\ldots i-1], w[i])$, then we have the following.

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

Hence, $D[w \in \Sigma^n \mid d_\pi(w) \geq 2] > \frac{1}{2}f(t', n) + \frac{1}{n^c}$, for all constants c. □

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 for any $\epsilon > 0$, there is a $1+\epsilon$ gale which succeeds on a given distribution D. Hence the dimension of D is the infimum of a non-empty subset of $[0, 1+\epsilon]$. Thus the dimension of a distribution is well-defined.

Since it is clear that any distribution on Σ^n has a dimension, the following lemma establishes the fact that our definition yields a nontrivial quantification of the set of distributions over Σ^n. First, we recall that a pseudorandom distribution over Σ^n has dimension 1.

Lemma 4.1. *Let $s \in [0,1]$. Then there is a distribution D on Σ^n with dimension s, for large enough n.*

Proof. Consider a pseudorandom distribution D on Σ^n. If $s = 1$, then D is a distribution with the required dimension.

Otherwise, assume that $s \in (0,1)$. To each string $x \in \Sigma^n$, we append $\lfloor \frac{n}{s} \rfloor - n$ many zeroes, 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 zeroes, we set $D'(y) = 0$.

Let $\pi : \Sigma^* \times \Sigma \to [0,1]$ be the predictor which testifies that the unpredictability of $D < 1 + \epsilon$. Define the new predictor $\pi' : \Sigma^* \times \Sigma \to [0,1]$ by

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

For every $w \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$, we have that

$$\text{LossRate}(\pi', w) = \frac{\text{LossRate}(\pi, w[1 \ldots n])}{\lfloor \frac{n}{s} \rfloor} \leq \frac{\text{LossRate}(\pi, w[1 \ldots n])}{\frac{n-s}{s}} = \frac{(1+\epsilon)s}{1 - \frac{s}{n}}$$

when s/n is small enough, testifying that the unpredictability (hence the dimension) of the distribution is at most s.

Now, assume that $\gamma : \Sigma^* \times \Sigma \to [0,1]$ is a predictor which testifies that the unpredictability of D' is $s - \epsilon$. We show that this would imply that $\dim(D) < 1$. Let $w \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$ be a string such that $\text{Loss}(\gamma, w) \leq \lfloor \frac{n}{s} \rfloor (s - \epsilon) < (n - \frac{\epsilon n}{s})$.

Since $\text{Loss}(\gamma, w[n \ldots \lfloor \frac{n}{s} \rfloor]) \geq 0$, we have that $\text{Loss}(\gamma, w[1 \ldots n]) \leq (n - \frac{\epsilon n}{s})$.

Construct a predictor $\gamma' : \Sigma^* \times \Sigma \to [0,1]$ such that $\gamma'(x,b) = \gamma(x,b)$ for all strings x with length at most n and $b = 0, 1$. Then, $\text{Loss}(\gamma', w[1 \ldots n]) < (n - \frac{\epsilon n}{s})$, implying that $\text{LossRate}(\gamma, w[1 \ldots n]) < 1 - \frac{\epsilon}{s} < 1$.

Since this happens for every string $w \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$ with $\text{LossRate}(\gamma, w) < s - \epsilon$, we have that $\text{unpredictability}(D) < 1 - \epsilon/s$, which completes the proof. □

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

Lemma 4.2. *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 defined by $D = \delta D_1 + (1-\delta) D_2$. Then* $\dim(D) \geq \min\{\dim D_1, \dim D_2\}$.

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

For the contrary, lets assume that, $dim(D) < \min\{s_1, s_2\}$ and assume $s = \min\{s_1, s_2\} - \epsilon$, for all $\epsilon, 0 < \epsilon < 1$. Then there exists an s-gale such that

$$Pr_{w \in_R D, \text{randomness of } d}[d(w) \geq 2] > \frac{1}{2} f(s,n) + \frac{1}{n^c}, \text{ for all constants } c.$$

Now define the polynomial sized circuit C_d obtained from d by hard wiring "good" random bits. We can write $D[C_d(w) \geq 2] > \frac{1}{2}f(s,n) + \frac{1}{n^c}$, for all constants c.

Let the string w for which $C_d(w) \geq 2$ 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,

$$\sum_{i=1}^{k} p(w_i) > \frac{1}{2}f(s,n) + \frac{1}{n^c}, \text{ for all constants } c$$

where $p(w_i) = \delta q(w_i) + (1-\delta)r(w_i), 1 \leq i \leq k$.
Now, since $\dim(D_2) \geq s_2$, we have that

$$r(w_1) + \cdots + r(w_k) \leq \frac{1}{2}f(s,n) + \frac{1}{n^c}, \text{ for some constant } c$$

Thus

$$q(w_1) + \cdots + q(w_k) > \frac{1}{2}f(s,n) + \frac{1}{n^c}, \text{ for all constants } c$$

and thus $dim(D_1) < s_1$. □

However, it is easy to see that convex combinations of distributions may have larger dimension than any of its constituents. For example, let us take a pseudorandom distribution P on Σ^n and then 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 P and D_2 produced by the 1-dilution (padding each string with a 1 at the end) of P. Then $D = 0.5D_1 + 0.5D_2$ has dimension which exceeds the dimensions of D_1 and D_2 by $\frac{1}{n}$.

Lemma 4.3. *Let D, D_1 and D_2 be the distributions on Σ^n, and let $\delta = \frac{1}{n^k}$ for some $k > 0$. Suppose further that $dim(D_1) = s_1$. If $D = (1-\delta)D_1 + \delta D_2$, then $dim(D) \geq s_1$.*

Proof. For the contrary, lets assume that, $\dim(D) < s_1$ and assume $s = s_1 - \epsilon$, for some ϵ, $0 < \epsilon < 1$. So there exists an s-gale that wins over D.

$$Pr_{w \in_R D, \text{randomness of } d}[d(w) \geq 2] > \frac{1}{2}f(s,n) + \frac{1}{n^c}, \text{ for all constants } c.$$

Now we view this s-gale as a polynomial sized circuit C_d and by hard wiring "good" random bits, we can write

$$D[C_d(w) \geq 2] > \frac{1}{2}f(s,n) + \frac{1}{n^c}, \text{ for all constants } c$$

Let the string w for which $C_d(w) \geq 2$ 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,

$$p(w_1) + \cdots + p(w_k) > \frac{1}{2}f(s,n) + \frac{1}{n^c}, \text{ for all constants } c,$$

where $p(w_i) = \delta q(w_i) + (1-\delta)r(w_i), 1 \leq i \leq k$.
Now, as $\delta = \frac{1}{n^k}$, for some constant k and $r(w_1) + \cdots + r(w_k) \leq 1$,

$$q(w_1) + \cdots + q(w_k) > \frac{1}{2}f(s,n) + \frac{1}{n^c}, \text{ for all constants } c$$

and thus $dim(D_1) < s_1$ which is a contradiction. □

The following lemma 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.

Lemma 4.4. *For all sufficiently large n, there is a non-pseudorandom distribution D_n on Σ^n such that $\dim(D_n) = 1$.*

Proof. Let D_n on Σ^n be defined as follows. Let $0 < k < n-1$ be an integer.

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

Then D_n is not pseudorandom: consider a (deterministic) polynomial time predictor $\pi : \Sigma^* \times \Sigma \to [0,1]$ defined as follows. For strings w of length i, $i \in [0, n-2] \setminus \{k\}$, set $\pi(w, b) = 0.5$, $b = 0, 1$ and $\pi(w, 0) = 1$, $\pi(w, 1) = 0$ otherwise. Then

$$D_n\{x \in \Sigma^n \mid \pi(x[\ 0 \ldots k-1\ ], x[k]) = 1\} = 1.$$

However, $\dim(D_n) = 1$. To see this, let $\epsilon > 0$ be arbitrary. It suffices to show that $\dim(D_n) > 1 - \epsilon$ for all sufficiently large n. Let n be large enough that $1 - \epsilon + \frac{1}{n} < 1$. Let s be such that $1 - \epsilon < s - \frac{1}{n} < 1$. We show that no $(s - \frac{1}{n})$-gale succeeds on D_n. Let d be an s-gale.

We define an *extender* algorithm $E_d : \Sigma^{<n} \to \Sigma^{\leq n}$, which given a string $w \in \Sigma^{<n}$, outputs a string $w0$ or $w1$ in such a way that $d(w) \leq 2^{-s|w|}$. The algorithm E_d is defined by

$$E_d(w) = \begin{cases} w0 & \text{if } d(w1) > d(w) \text{ or } |w| = k-1 \\ w1 & \text{otherwise.} \end{cases}$$

Then, for every string x in the support of D_n on which $d(x) > 2^{-sn+1} = 2^{-(s-\frac{1}{n})n}$, we have a string $E_d(x)$ in the support of D_n such that $d(E_d(x)) < 2^{-sn}$. It is also clear that E_d restricted to the set of strings x on which $d(x) > 2^{-sn}$ is a 1-1 function. Thus

$$D_n\{x \in \Sigma^n \mid d \text{ succeeds on } x\} \leq \frac{1}{2}.$$

□

5 Non-pseudorandom Bit-fixing Sources and Deterministic Pseudorandom Extractors

We now give an application of our quantification of nonpseudorandom distributions. In the last section, in Lemma 4.1, we 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 $I = \{i_1, \ldots, i_{sn}\} \subseteq \{1, \ldots, n\}$ where all the bits at the indices of I are independent and uniformly chosen and rest of the bits are completely fixed. This distribution was introduced by Chor *et. al.*[5]. Now we define an analogous notion for the class of nonpseudorandom distributions, which we term nonpseudorandom bit-fixing sources.

Definition 5.1 (Nonpseudorandom Bit-fixing Source)**.** Let $n \in \mathbb{N}$, $s \in [0, \infty)$ and ϵ be an inverse polynomial in n. A distribution over Σ^n is an (n, s, ϵ)-*nonpseudorandom bit-fixing source* if there exists a subset $I = \{i_1, \ldots, i_{sn}\} \subseteq \{1, \ldots, n\}$ such that all the bits at the indices of I come from an (S, ϵ)-pseudorandom distribution(where S is polynomial in sn) and rest of the bits are fixed.

Randomness extractors extract out "almost random bits" from an input distribution. Our objective now is to extract pseudorandom bits from a given nonpseudorandom bit-fixing source. For this, we now define the notion of a pseudorandomness extractor.

Definition 5.2 (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 $X \in \mathcal{C}$, $\dim(E(X)) = 1$.
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 $X \in \mathcal{C}$, $\dim(E(X, U_d)) = 1$.

6 Deterministic Pseudorandom Extractor for Non-pseudorandom Bit-fixing Sources

We adapt the technique of Gabizon, Raz and Shaltiel [7]. The main result of this section is the following theorem.

Theorem 6.1. *Let $n \in \mathbb{N}$ and $\epsilon < \frac{1}{\sqrt{n}}$. Then there exists a constant $c > 0$ such that for any large enough n and any s satisfying $sn \geq \log^{(c)} n$, 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 $m = (sn)^{\Omega(1)}$.*

If $G : \Sigma^{O(\log n)} \to \Sigma^n$ is a polynomial time algorithm which outputs the (n, s, ϵ)-nonpseudorandom bit-fixing source, say X, with $\epsilon < \frac{1}{\sqrt{n}}$, then the size of the support of X is polynomial in n (The support of a distribution is the set of strings with non-zero probability).

Corollary 6.2. *Let $n \in \mathbb{N}$ and $\epsilon < \frac{1}{\sqrt{n}}$. Suppose there is a polynomial time algorithm $G : \Sigma^{O(\log n)} \to \Sigma^n$ which outputs an (n, s, ϵ)-nonpseudorandom bit-fixing source. Then there exits a constant $c > 0$ such that, for any large enough n and any s satisfying $sn \geq \log^{(c)} n$, there is a deterministic pseudorandom extractor $E : \Sigma^n \to \Sigma^m$, where $m = (sn)^{\Omega(1)}$.*

We first discuss the ingredients required in the proof of the above theorem.

6.1 Pseudorandom walk and extracting a few random bits

Kamp and Zuckerman [12] 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 $\Omega(\log sn)$ bits from (n, s, ϵ)-nonpseudorandom bit-fixing sources. We first introduce the notion of a randomness extractor, and then outline our construction.

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 Ω are said to be $\epsilon - close$ (statistically close) if $\max_{S \subseteq \Omega}\{|Pr[X \in S] - Pr[Y \in S]|\} \leq \epsilon$ or equivalently $\frac{1}{2} \sum_{x \in \Omega} |Pr[X = x] - Pr[Y = x]| \leq \epsilon$.

Definition 6.3 (Deterministic Extractor)**.** A function $E : \Sigma^n \to \Sigma^m$ is said to be a deterministic ϵ-extractor for a class of distribution C if for every distribution X of n-bit strings in C, the distribution $E(X)$ is ϵ-close to the uniform distribution on m-bit strings.

Theorem 6.4. *Let $s \in [0,1]$ be arbitrary, n be a natural number such that $sn > 100$, and let $0 < \epsilon < \frac{1}{\sqrt{n}}$. Then there is a deterministic $\frac{1}{\sqrt[4]{sn}}$-extractor $E : \Sigma^n \to \Sigma^{\frac{\log sn}{4}}$.*

Before proving the above theorem, we state two lemmas required for the proof. The first is a special case of Lemma 3.3 [12] which was restated as Lemma 4.2 in [7].

Lemma 6.5 (Lemma 4.2 of [7])**.** *Let 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 (n, sn)-bit-fixing source, starting from any fixed vertex, then at the end of the n step of the walk, the distribution D on the vertices will be $\frac{1}{2}\lambda^{sn}\sqrt{M}$-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.* [18].

Lemma 6.6. *Let 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 (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 D on the vertices will be $\frac{1}{2}(\lambda^{sn} + \sqrt{M}\epsilon)\sqrt{M}$-close to U_M, where M is polynomial in n.*

Proof. Let ϵ be a number which satisfies the hypothesis. Suppose we take a n step walk on the graph G starting from any vertex according to the bits from (n, sn)-bit-fixing source and the probability vector on the vertices after the walk is say $p = \begin{pmatrix} p_1 & p_2 & \dots & p_M \end{pmatrix}$.

Now we take a n step walk on the graph G starting from the same vertex according to the bits from (n, s, ϵ)-nonpseudorandom bit-fixing source and the probability vector on the vertices after the walk is say $D = \begin{pmatrix} q_1 & q_2 & \dots & q_M \end{pmatrix}$ where $\forall i,\ q_i \leq p_i + \epsilon$. This can be justified as follows.

If the bound is not true then we can use this walk on G as the polynomial time algorithm to distinguish between uniform distribution on Σ^{sn} and ϵ-pseudorandom distribution on Σ^{sn}. The above bound on q_i is true because the only difference between (n, s, ϵ)-nonpseudorandom bit-fixing source and (n, sn)-bit-fixing source is that on the set I, in (n, sn)-bit-fixing source we have uniform bits instead of ϵ-pseudorandom bits.

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. Thus,

$$\begin{aligned}
||q - \pi||^2 &= \sum_{i=1}^{M} (q_i - \frac{1}{M})^2 \\
&\leq \sum_{i=1}^{M} (p_i + \epsilon - \frac{1}{M})^2 \\
&= \sum_{i=1}^{M} (p_i - \frac{1}{M})^2 + M\epsilon^2 \\
&= ||p - \pi||^2 + M\epsilon^2 \\
&\leq \lambda^{2sn} + M\epsilon^2 \qquad \text{from Lemma 6.5} \\
&\leq (\lambda^{sn} + \sqrt{M}\epsilon)^2
\end{aligned}$$

□

Now using the above lemma, we prove Theorem 6.4.

Proof of Theorem 6.4. Let us take an odd cycle with $M = \sqrt[4]{sn}$ vertices as the graph G. The second largest eigenvalue of G is $\cos(\frac{\pi}{\sqrt[4]{sn}})$. Now take a walk starting from a fixed vertex of G according to the bits from (n, k, ϵ)-nonpseudorandom bit-fixing source and finally output the vertex number of the graph G. Thus we get $\frac{\log sn}{4}$ bits. From Lemma 6.6, we reach distance $\frac{1}{2}((\cos(\frac{\pi}{\sqrt[4]{sn}}))^{sn} + \sqrt[8]{sn\epsilon})\sqrt[8]{sn}$ 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, $\left(\cos\left(\frac{\pi}{\sqrt[4]{sn}}\right)\right)^{sn} < (1 - \frac{\pi^2}{4\sqrt{sn}})^{sn} < (\exp^{-\frac{\pi^2}{4}})^{\sqrt{sn}} < 4^{-\sqrt{sn}}$. Hence, $\frac{1}{2}((\cos(\frac{\pi}{\sqrt[4]{sn}}))^{sn} + \sqrt[8]{sn\epsilon})\sqrt[8]{sn} < \frac{1}{\sqrt[4]{sn}}$. Thus we get distribution of $\frac{\log sn}{4}$ bit strings which is $\frac{1}{\sqrt[4]{sn}}$-close to uniform in statistical distance. □

Note that from the above technique we have an explicit construction of an deterministic extractor $E : \Sigma^n \to \Sigma^{\frac{\log n}{4}}$ for any (S, ϵ)-pseudorandom distribution on n-length strings where $\epsilon \leq n^{-0.5}$ and S is polynomial in n and the output distribution is $n^{-0.25}$-close to uniform.

6.2 Generating an Independent Seed

In this subsection, we see how to obtain a short seed from a nonpseudorandom bit-fixing source so that we can use them in a seeded pseudorandom extractor to extract 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 source. Even though the result is analogous to [7], the proofs differ in essential details.

Definition 6.7 (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.

Theorem 6.8. *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 and $E' : \Sigma^n \times \Sigma^d \to \Sigma^m$ is a seeded pseudorandom extractor for (n, s', ϵ)-nonpseudorandom bit-fixing sources, where $m = s'n$. 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 as $dim(E'(R_a)) = 1$, so by Lemma 4.2, $dim(Y) = 1$ and then by Lemma 4.3, $dim(E(X)) = 1$ as $\eta \leq \frac{1}{(sn)^c}$, for some constant c. □

Now we give an explicit construction of (s, s', ρ)-seed obtainer, which is crucial in the later part of this paper.

Theorem 6.9. *For every n, let $Samp : \Sigma^t \to P([n])$ be a $(n, sn, s_1n, s_2n, \delta)$-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 $\rho = \max\{\epsilon' + \delta, \sqrt{\epsilon'}2^{t+1}\}$.*

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.
Let X be the (n, s, ϵ)-nonpseudorandom bit-fixing source and I be the set of indices at which 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 $Y = E_2(X)$. A string $a \in \Sigma^t$ is said to *correctly split* X if $s_1n \leq |I \cap T_a| \leq s_2n$.

Claim 6.10. *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 s_1n "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 $\sum_{b \in B} X[X'_a = b] > \sqrt{\epsilon}$.

Suppose the above claim is not true, that means there exists a subset $J \subseteq \Sigma^{n-l}$, where $\forall_{b \in J} X[X'_a = b] > 0$, such that $\sum_{b \in J} X[X'_a = b] > \sqrt{\epsilon}$ and 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 lets consider only the "good" positions which are sn in X and at least s_1n in $([X_a; X'_a]|X'_a = b)$. So the above assumption implies that the distribution on those s_1n bits (this part of the string b is denoted as b_{s_1n}) in $([X_a; X'_a]|X'_a = b)$ are not $\sqrt{\epsilon}$-pseudorandom, i.e., they have 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 sn bit ϵ-pseudorandom distribution P which is a contradiction. The algorithm A is as follows: on input $y \in \{0,1\}^{sn}$, if $y_{s_1n} = b_{s_1n}$ for any $b \in S$, then return $C_b(y_{s_1n})$; otherwise return 0 or 1 randomly. And thus clearly, $|P[A[y] = 1] - U_{sn}[A[y] = 1]| > \epsilon$.

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 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. □

Claim 6.11 (Lemma 3.6 of [7]). *For every fixed $a \in \Sigma^t$ that correctly splits X, the distribution $((X'_a, E_2(X))|E_1(X) = a)$ 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 of correctness for the construction of F follows directly from the proof of Theorem 3.3 of [7] with the following parameters $k = sn$, $k_{min} = s_1 n$, $k_{max} = s_2 n$ *and* $\epsilon = \sqrt{\epsilon'}$.

□

6.3 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 [7]. Let $S \subseteq [n]$ be some subset of size k. Now we consider a process of generating a subset $T \subseteq [n]$ such that $k_{min} \leq |S \cap T| \leq k_{max}$ and this process is known as *Sampling*.

Definition 6.12. 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 $|S| = k$, $Pr_{w \in_R U_t}[k_{min} \leq |Samp(w) \cap S| \leq k_{max}] \geq 1 - \delta$

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 $1 \leq i \leq m$, $k_{min} \leq |S \cap T_i| \leq k_{max}$.
According to [7], the above two processes can be performed using only *a few* random bits.

Lemma 6.13 (Lemma 5.2 of [7]). *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ and $\frac{1}{2} < e < 1$ such that for any $n \geq 16$ and $k \geq \log^c n$, 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 6.14 (Lemma 5.3 of [7]). *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ and $\frac{1}{2} < e < 1$ such that for any $n \geq 16$ and $k \geq \log^c n$, 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 $|S| = k$, $Pr[\forall 1 \leq i \leq m, \frac{k^e}{2} \leq |T_i \cap S| \leq 3k^e] \geq 1 - O(k^{-b})$.*

6.4 A Seeded Pseudorandom Extractor

In this subsection, we discuss about how we can extract $(sn)^{\Omega(1)}$ bits of dimension 1 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.

Theorem 6.15. *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ such that for any $n \geq 16$ and $sn \geq \log^c n$, 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 $x = x_1 x_2 \cdots x_n$ be a string sampled by X. The description of the extractor $E(x, y)$ is as follows: (1) According to Lemma 6.14, using the seed y, 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 bits are not fixed and let Z be the distribution of the output strings. We need to show that $dim(Z) = 1$.

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 6.14, $Pr[A] \geq 1 - O((sn)^{-b})$. Now we can write the output distribution Z as $Z = Pr[A](Z|A) + Pr[A^c](Z|A^c)$. Now due to Lemma 4.3, $dim(Z) = 1$.

□

6.5 Deterministic Pseudorandom Extractor

Now we are ready to prove the Theorem 6.1.

Proof of Theorem 6.1. Due to Lemma 6.13, 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 6.4, 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 6.9 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 6.15, 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 6.8 to construct deterministic pseudorandom extractor $E : \Sigma^n \to \Sigma^m$. For large enough n, $m = (sn - s''n)^{\Omega(1)} = (sn)^{\Omega(1)}$ and this completes the proof. □

7 Approaching Towards P=BPP

We now show that if there is a polynomial time 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 7.4 [15]. Before discussing the theorem and its proof, we first mention some required terminologies and theorems.

7.1 Hard Sets & Optimal Pseudorandom Generators

The definitions and theorems discussed in this section are relevant for the Section 7. 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 7.1 (Pseudorandom Generators)**.** A function G is said to be a $l(n)$-pseudorandom generator against a class of circuits $\mathcal{C}$ 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 all n and for all circuits $C \in \mathcal{C}$ having n input bits,

$$|U_n[C(x) = 1] - U_{l(n)}[C(G(y)) = 1]| \leq \frac{1}{n}$$

Here we consider the class of circuits $\mathcal{C}$ which are of size polynomial in n and according to the definition given in [4], we can also write the expression in the property (3) of the above definition as follows

$$|U_n[C(x) = 1] - U_{l(n)}[C(G(y)) = 1]| \leq \frac{1}{n^c} \text{ for some constant } c$$

Definition 7.2 (Optimal Pseudorandom Generators)**.** A function G is said to be an optimal pseudorandom generator against a class of circuits $\mathcal{C}$ if it is an $O(\log(n))$-pseudorandom generator.

Nisan and Wigderson [15] showed that there is a connection between pseudorandom generators and hard-to-approximate sets (or simply hard sets) in EXP:

Definition 7.3. For a set A and a circuit C having n input bits,

$$adv_C(A) = |U_n[C(x) = A(x)] - U_n[C(x) \neq A(x)]|$$

Here we identify A with its characteristic function. For all circuits of size at most $s(n)$, let $adv_{s(n)}(A)$ be the maximum of $adv_C(A)$ where C varies over all circuits of size at most $s(n)$. A set $A \in$ EXP is said to be *hard-to-approximate* by circuits of size $s(n)$ if $adv_{s(n)}(A) \leq \frac{1}{s(n)}$. The following theorem was shown by Nisan and Wigderson [15].

Theorem 7.4 ([15])**.** *There exist optimal pseudorandom generators against class of circuits of size* $2^{\delta l}$ *for some constant* $0 < \delta < 1$ *if and only if there exist sets in* EXP *that are hard-to-approximate by circuits of size* $2^{\epsilon l}$ *for some constant* $0 < \epsilon < 1$.

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

Theorem 7.5 ([11])**.** *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 set in* EXP *that is hard-to-approximate by circuits of size* $2^{\delta' n}$ *for some constant* $0 < \delta' < 1$.

7.2 Derandomization with Optimal Nonpseudorandom Generator

Theorem 7.6. *If there exists a polynomial time algorithm* $G : \Sigma^{O(\log n)} \to \Sigma^n$ *where the output distribution is* X *and* $dim(X) = s$ $(s > 0)$*, then* P=BPP.

Proof. If $dim(X) = s$ $(s > 0)$, then there must be some bit position i such that for all polynomial circuits C, $X[C(x_1, \cdots, x_{i-1}) = x_i] < 1$; otherwise according to Theorem 3.4, $dim(X) = 0$, for large enough n. Now we define a language L as follows: $L = \{y \in \Sigma^n |\ y = y_1 y_2$ such that $|y_1| = i - 1$ and $\exists z$ such that $|z| = (n - i)$ and $y_1 1 z$ is in the range of $G\}$.
Now clearly L is the language that satisfies all the conditions of Theorem 7.5 and thus it can be used to construct an *optimal pseudorandom generator* and which eventually implies P=BPP. □

Acknowledgements

Authors would like to thank Somenath Biswas for some helpful discussions and comments.

References

[1] S. Arora and B. Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009.

[2] 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.

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

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

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

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

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

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

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

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

[11] 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.

[12] 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.

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

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

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

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

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

[18] 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.

[19] Luca Trevisan. Construction of extractors using pseudo-random generators (extended abstract). In *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, STOC '99, pages 141–148, New York, NY, USA, 1999. ACM.

[20] 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