Construction of Extractors Using Pseudo-Random Generators
[Extended Abstract]

Luca Trevisan*

Abstract

We introduce a new approach to construct extractors. Extractors are algorithms that transform a "weakly random" distribution into an almost uniform distribution. Explicit constructions of extractors have a variety of important applications, and tend to be very difficult to achieve.

We demonstrate an unsuspected connection between extractors and pseudorandom generators. In fact, we show that every pseudorandom generator of a certain kind is an extractor.

A pseudo-random generator construction due to Impagliazzo and Wigderson, once reinterpreted via our connection, is already an extractor that beats most known constructions and solves an important open question. We also show that, using the simpler Nisan-Wigderson generator and standard error-correcting codes, one can build even better extractors with the additional advantage that both the construction and the analysis are extremely simple and admit a short self-contained treatment.

1 Introduction

An *extractor* is an algorithm that converts a "weak source of randomness" into an almost uniform distribution by using a small number of additional truly random bits. Extractors have several applications that we briefly survey below.

The natural application of extractors is to allow the simulation of randomized algorithms even in (realistical) settings where only weak sources of randomness are available. This line of research has a long history, that dates back at least to von Neumann's algorithm for generating a sequence of unbiased bits from a source of biased but identically distributed and indipendent bits. More recent work by Santha and Vazirani [SV86] and Vazirani and Vazirani [VV85] considered much weaker sources of randomness (that they call "slightly random" sources) that are still sufficient to allow (non-trivial) simulation of arbitrary randomized algorithms. These results were generalized by Chor and Goldreich [CG88] and Cohen and Wigderson [CW89], and finally by Zuckerman [Zuc90], who introduced the modern definition of weak random source and a construction of extractors (although the term *extractor* was coined later, in [NZ93]). Improved constructions of extractors appeared in [NZ93, SZ94, TS96, Zuc96b]. Neither of these constructions implies an optimal simulation of randomized algorithms. Dispersers are objects similar to, but less powerful than, extractors. Randomized algorithm having one-sided error probability can be simulated by using weak random sources and dispersers. Saks et al. [SSZ98] give a construction of dispersers that implies an optimal simulation of one-sided error randomized algorithms with weak random sources. Andreev et al. [ACRT97] show how to use the dispersers of Saks et al. in order to give optimal simulations of general randomized algorithms using weak random sources. The result of Andreev et al. leaves open the question of whether there exist a construction of extractors that is good enough to imply directly such optimal simulations.

Extractors are also used to derandomize randomized space-bounded computations [NZ93] and for randomness-efficient reduction of error in randomized algorithms (see [Zuc96b, GZ97] and references therein). They yield oblivious samplers (as defined in [BR94]), that have applications to interactive proofs and leader election in anonymous networks (see [Zuc96b] and references therein). They also yield expander graphs, as discovered by Wigderson and Zuckerman [WZ93], that in turn have applications to superconcentrators, sorting in rounds, and routing in optical networks. Constructions of expanders via constuction of extractors and the Wigderson-Zuckerman connection appeared in [NZ93, SZ94, TS96], among others. Extractors can also be used to give simple proofs of certain complexity-theoretic results [GZ97], and to prove certain hardness of approximation results [Zuc96a]. The literature on explicit construction of extractors and dispersers is vast and technically challenging. An excellent and accessible introduction is given by a recent survey by Nisan [Nis96] (see also [NTS98]).

In this paper we show that pseudorandom generator constructions of a certain kind are extractors. Using our connection and some new ideas we describe constructions of extractors that improve or subsume all the previously known constructions and that are exceedingly simpler than previous ones.

1.1 Definitions

We now give the formal definition of an extractor and state some previous results. We first need to define the notions of

*luca@cs.columbia.edu. Computer Science Department, Columbia University.

This paper appears in the Proceedings of STOC'99

min-entropy and *statistical difference*.

We say that (the distribution of) a random variable X of range $\{0,1\}^n$ has min-entropy at least k if for every $x \in \{0,1\}^n$ it holds $\mathbf{Pr}[X = x] \leq 2^{-k}$. If k is an integer, then a canonical example of a distribution having min-entropy k is the uniform distribution over a set $S \subseteq \{0,1\}^n$ of cardinality 2^k. Indeed, it is implicit in [CG88] that if a distribution has min-entropy k then it is a convex combination of distributions each one of whom is uniform over a set of size 2^k. We will consider distributions of min-entropy k as the formalization of the notion of weak sources of randomness containing k "hidden" bits of randomness. The use of min-entropy to measure "hidden randomness" has been advocated by Chor and Goldreich [CG88] and, in full generality, by Zuckerman [Zuc90]. The statistical difference between two random variables X and Y with range $\{0,1\}^n$ is defined as

$$||X - Y|| = \max_{T:\{0,1\}^n \to \{0,1\}} |\,\mathbf{Pr}[T(X) = 1] - \mathbf{Pr}[T(Y) = 1]|$$

and we say that X and Y are ε-close if $||X - Y|| \leq \varepsilon$. For an integer l we denote by U_l a random variable that is uniform over $\{0,1\}^l$.

A function $Ext : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a (k,ε) extractor if for every random variable X of min entropy at least k it holds that $Ext(X, U_t)$ is ε-close to the uniform distribution over $\{0,1\}^m$. A weaker kind of combinatorial construction has also been considered: A function $Disp : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a (k,ε) disperser if for every subset $S \subseteq \{0,1\}^m$ such that $|S| > \varepsilon 2^m$ and for every X of min-enropy k it holds $\mathbf{Pr}[Disp(X) \in S] > 0$.

One would like to have, for every n and k, constructions where t is very small and m is as close to k as possible. There are some limitations towards this goal: One can show that, for a certain range of k, it must be $t \geq \Omega(\log(n/\varepsilon))$, and also it must be $m \leq k + t - \Omega(1/\varepsilon^2)$ (see [RTS97]). It is possible to show (non-constructively) that for every n, k, ε, there is a (k,ε)-extractor $Ext : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ where $t = O(\log n/\varepsilon)$ and $m = k + t - \Theta(1/\varepsilon^2)$. It is an open question to match such bounds with polynomial-time computable functions Ext.

In Table 1 we summarize the best known constructions, for different combination of the parameters, and we state the parameters of (a special case of) our construction.

1.2 Our Result

The extractors constructed in this paper work for any min-entropy $k = n^{\Omega(1)}$, extracts a slightly sub-linear fraction of the original randomness (i.e. the length of the output is $m = k^{1-\gamma}$ for an arbitrarily small γ) and use $O(\log n)$ bits of true randomness. Indeed, a more general result holds, as formalized below.

Theorem 1 (Main) *For every n, m, k, ε we can construct a polynomial-time computable (k,ε)-extractor Ext : $\{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ where*

$$t = O\left(e^{\log k/(\log k/2m)} \frac{(\log n/\varepsilon)^2}{\log(k/2m)}\right).$$

In particular, for fixed constants $\varepsilon > 0$ and $0 < \gamma' < \gamma < 1$ we have for every n an explicit (n^γ, ε)-extractor $Ext : \{0,1\}^n \times \{0,1\}^{O(\log n)} \to \{0,1\}^{n^{\gamma'}}$.

Our construction improves on the construction of Saks, Srinivasan and Zhou [SSZ98] since we construct an extractor rather than a disperser, and improves over the constructions of Ta-Shma [TS96] since the additional randomness is logarithmic instead of slightly super-logarithmic. The best previous construction of extractors using $O(\log n)$ additional randomness was the one of Zuckerman [Zuc96b], that only works when the min-entropy is a constant fraction of the input length, while in our construction every min-entropy of the form n^γ is admissible. Our construction shows an optimal way of using weak random sources to simulate every randomized procedure. In contrast to the result of [ACRT97] we can use a weak random source to generate almost uniformly distributed random bits independently of the purpose for which the random bits are to be used.[1]

Our construction is not yet the best possible, since we lose part of the randomness of the source and because the additional randomness is logarithmic only as long $k = n^{\Omega(1)}$.

1.3 Techniques

This paper contains two main contributions.

The first one is a connection (outlined in Section 2) between pseudorandom generators of a certain kind and extractors. Our connection applies to certain pseudorandom generator constructions that are based on the (conjectured) existence of predicates (decision problems) that can be uniformly computed in time $t(n)$ but cannot be solved by circuits of size much small than $t(n)$. The analysis of such constructions shows that if the predicate is hard, then it is also hard to distinguish the output of the generator from the uniform distribution. This implication is proved by means of a reduction showing how a circuit that is able to distinguish the output of the generator from the uniform distribution can be transformed into a slightly larger circuit that computes the predicate. (Impagliazzo and Wigderson [IW97] present one such construction with very strong parameters.) Our result is that if the (truth table of the) predicate is chosen randomly, according to a distribution with sufficiently high min-entropy, then the output of the generator is statistically close to uniform. This statement is incomparable with standard analyses: we use a stronger assumption (that the predicate is random instead of fixed and hard) and prove a stronger conclusion (that the output is statistically close to, instead of indistinguishable from, the uniform distribution). An immediate application is that a pseudorandom generator construction of this kind *is* an extractor. Our result has a straightforward proof, based on a simple counting argument. The main contribution, indeed, is the *statement* of the result, rather than its *proof*, since it involves a new, more general, way of looking at pseudorandom generator constructions. The Impagliazzo-Wigderson generator, using our connection, is an extractor that beats some previous constructions and that is good enough to imply optimal simulations of randomized algorithms. We stress that even if the Impagliazzo-Wigderson generator is a pseudorandom generator under unproved conjectures, it *provably is* a good extractor (i.e. we do not use any complexity-theoretic assumption in our work).

Our second contribution is a construction that is simpler to describe and analyse (the generator of Impagliazzo

[1] Andreev et al. [ACRT97] show how to produce a sequence of bits that "look random" to a specific algorithm, and their construction works by having oracle access to the algorithm. So it is not possible to generate random bits "offline" before fixing the application where the bits will be used.

Reference	Min entropy k	Output length m	Additional randomness	Type
[Zuc96b]	$k = \Omega(n)$	$m = \Omega(k)$	$O(\log n)$	Extractor
[TS96]	any k	k	$\text{poly} \log n$	Extractor
[TS96]	$k = n^{\Omega(1)}$	$m = k^{\Omega(1)}$	$O(\log n \log \cdots \log n)$	Extractor
[SSZ98]	$k = n^{\Omega(1)}$	$m = k^{\Omega(1)}$	$O(\log n)$	Disperser
[TS98]	any k	$m = k - \text{poly} \log k$	$O(\log n)$	Disperser
This paper	$k = n^{\Omega(1)}$	$m = k^{\Omega(1)}$	$O(\log n)$	Extractor

Table 1: A summary of previous results and our result.

and Wigderson is quite complicated) and that has somewhat better parameters. Our idea is to use a pseudorandom generator construction due to Nisan and Wigderson [NW94], that is considerably simpler than the one of Impagliazzo and Wigderson (indeed the construction of Impagliazzo and Wigderson contains the one of Nisan and Wigderson as one of its several components). The Nisan-Wigderson generator has weaker properties than the Impagliazzo-Wigderson generator, and our ideas outlined in Section 2 would not imply that it is an extractor as well. In Section 3 we show how to use error-correcting codes in order to turn the Nisan-Wigderson generator into a very good extractor. Section 3 contains a completely self-contained treatment of the construction and the analysis.

1.4 Relevance of Our Results

At the very least, our construction improves upon previous ones and solves the question of constructing extractors that use a logarithmic amount of randomness, work for any min-entropy that is polynomially related to the length of the input and have an output that is polynomially related to the amount of entropy. Such a construction has been considered a relevant open qestion (e.g. in [NTS98, Gol99]), even after Andreev et al. [ACRT97] showed that one does not need such extractors in order to develop an optimal simulation of randomized algorithms via weak random sources. Indeed, it was not clear whether the kind of approach introduced by Andreev et al. was *necessary* in order to have optimal simulations, or whether a more traditional approach based on extractors was still possible. Our result clarifies this point.

Perhaps more importantly, our construction is very simple to describe, analyse and understand, in contrast with the most recent previous constructions. Hopefully, our techniques offer more room for improvement than previous, deeply exploited, ones. Raz et al. [RRV99] have already found improvements to our construction (see below). Tight results may come from some combination of our ideas and previous ones.

Our use of results about pseudorandomness in the construction of extractors may come as a surprise: pseudorandom generation deals with (and takes advantage of) a *computational* definition of randomness, while extractors are combinatorial objects used in a framework where *information-theoretic* randomness is being considered. In the past there have been some instances of results about computational randomness inspired by (typically trivial) information-theoretic analogs, e.g. the celebrated Yao's XOR Lemma and various kind of "direct product" results (see e.g. [GNW95]). On the other hand, it seemed "clear" that one could not go the other way, and have information-theoretic applications of computational results. This prejudice might be one reason why the connection discovered in this paper has been missed by the several people who worked on weak random sources and on pseudo-randomness in the past decade (including those who did foundational work in both areas). Perhaps other important results might be proved along similar lines. On the other hand, it might also be that our results are just an isolated exception to the "rule" that computational randomness results are not useful in information theoretic settings. For example, we note that the pseudorandom generator construction of Blum, Micali and Yao [BM84, Yao82] does not yield an extractor using our techniques.

1.5 Later results

Shortly after the development of the results of this paper, Raz, Reingold and Vadhan [RRV99] devised an improvement to our construction. In our construction, if the input has min-entropy k and the output is required to be of length m, then the additional randomness is $O(m^{1/\log(k/2m)}(\log n)^2/\log(k/2m))$. In [RRV99], the dependency is $O((\log n)^2/\log(k/m))$. Raz et al. [RRV99] also show how to recursively compose their construction with itself (along the lines of [WZ93]) and they obtain in this way another construction where $k = m$ and the additional randomness is $O(\log^3 n)$. Constructions of extractors with parameters $k = m$ have applications to the explicit construction of expander graphs [WZ93]. In particular, Raz et al. [RRV99] present constructions of expander graphs and of superconcentrators that improve previous ones by Ta-Shma [TS96]. Raz et al. [RRV99] also improve the dependancy that we have between additional randomness and error paramter ε.

Organization of the Paper

We present in Section 2 our connection between pseudorandom generator constructions and extractors. The main result of Section 2 is that the Impagliazzo-Wigderson generator [IW97] is a good extractor. In Section 3 we describe and analyse a simpler construction based on the Nisan-Wigderson generator [NW94] and on error correcting codes. Section 3 might be read independently of Section 2.

2 The Connection Between Pseudorandom Generators and Extractors

We start by defining the notion of computational indistinguishability, and pseudorandom generators, due to Blum, Goldwasser, Micali and Yao [GM84, BM84, Yao82].

We denote by U_n the uniform distribution over $\{0,1\}^n$. We say that two random variables X and Y with the same range $\{0,1\}^n$ are (S, ε)-indistinguishable if for every

$T : \{0,1\}^n \rightarrow \{0,1\}$ computable by a circuit of size S it holds

$$|\mathbf{Pr}[T(X)=1] - \mathbf{Pr}[T(Y)=1]| \leq \varepsilon$$

One can see the notion of ε-closeness as the "limit" of the notion of (S,ε)-indistinguishability for unbounded S.

Informally, a pseudorandom generator is an algorithm $G : \{0,1\}^t \rightarrow \{0,1\}^m$ where $t << m$ and $G(U_t)$ is (S,ε)-indistinguishable from U_m, with large S and small ε. In complexity theory, one looks for generators, say, $G : \{0,1\}^{O(\log m)} \rightarrow \{0,1\}^m$ where $G(U_{O(\log m)})$ is $(m^2, 1/3)$-indistinguishable from U_m. Such generators (if they were uniformly computable in time poly(m)) would imply deterministic polynomial-time simulations of randomized polynomial-time algorithms.

Given current techniques, all interesting constructions of pseudorandom generators have to rely on complexity-theoretic conjectures. For example the Blum-Micali-Yao [BM84, Yao82] construction (that has different parameters from the ones exemplified above) requires the existence of strong one-way permutations. In a line of work initiated by Nisan and Wigderson [NW94], there have been results showing that the existence of hard-on-average decision problems is sufficient to construct pseudorandom generators. Impagliazzo and Wigderson [IW97] present a construction that only requires the existence of decision problems having high worst-case complexity.[2]

Definition 2 *Let Gen* $: \{0,1\}^t \rightarrow \{0,1\}^m$ *be a generator having access to a predicate* $P : \{0,1\}^l \rightarrow \{0,1\}$*. On input a seed* $s \in \{0,1\}^t$ *and oracle access to* $P : \{0,1\}^l \rightarrow \{0,1\}$ *we denote by* $Gen_P(s)$ *the output of the generator.*

Suppose that whenever, for a certain $P : \{0,1\}^l \rightarrow \{0,1\}$ *and* $T : \{0,1\}^m \rightarrow \{0,1\}$*, we have that*

$$|\mathbf{Pr}[Gen_P(U_t)=1] - \mathbf{Pr}[T(U_m)=1]| > \varepsilon$$

then there exists a circuit of size S *with* T*-gates that computes* P*.*

Then we say that Gen is a (l,t,m,S,ε)*-good pseudorandom generator construction.*

By a "circuit with T-gates" we mean a circuit that can use ordinary fan-in-2 AND and OR gates and fan-in-1 NOT gates, as well as a special gate (of fan-in m) that computes T with unit cost. This is the non-uniform analog of a computation that makes oracle access to T.

Fact 3 *If Gen is a* (l,t,m,S,ε)*-good pseudorandom generator construction, and* $P : \{0,1\}^l \rightarrow \{0,1\}$ *has circuit complexity* $\geq S'$*, then* $Gen_P(U_t)$ *is* $(\varepsilon, S/S')$*-indistinguishable from* U_m*.*

The result of Impagliazzo and Wigderson can be restated as follows.

Theorem 4 ([IW97]) *For every* $\delta > 0$ *there exists a* $\delta' > 0$ *such that for every* l *there is a* poly(m)*-time computable* $(l,t,m,S,1/3)$*-good generator where* $t = O(l)$*,* $m = 2^{\delta' l}$ *and* $S = 2^{\delta l}$*.*

[2] This is an oversimplified account. Both [NW94] and [IW97] require a *non-uniform* kind of hardness, and recent work has concentrated on *uniform* conditions [IW98]. A construction that only needed worst-case non-uniform assumptions was given by Babai et al. [BFNW93], but the parameters were worse than in the later construction of Impagliazzo and Wigderson [IW97].

The following theorem formalizes our connection between pseudorandom generators and extractors. For a string $x \in \{0,1\}^{2^l}$ we denote by $< x >: \{0,1\}^l \rightarrow \{0,1\}$ the predicate whose truth-table is x.

Theorem 5 *Let Gen be* (l,t,m,S,ε)*-good.*

Then $Ext : \{0,1\}^{2^l} \times \{0,1\}^t \rightarrow \{0,1\}^m$ *defined as* $Ext(x,s) = Gen_{<x>}(s)$ *is a* $(k, 2\varepsilon)$*-extractor with* $k = mS\log S + \log(1/\varepsilon)$*.*

Proof: Let X be a random variable with range $\{0,1\}^{2^l}$ and min-entropy $mS\log S + \log 1/\varepsilon$. We want to prove that for every $T : \{0,1\}^m \rightarrow \{0,1\}$ we have $\mathbf{Pr}[T(Ext(X,U_t) = 1)] - \mathbf{Pr}[T(U_m)=1]| \leq 2\varepsilon$. Let $B \subseteq \{0,1\}^{2^l}$ be the set of values v for which

$$\mathbf{Pr}[T(Ext(v,U_t)=1)] - \mathbf{Pr}[T(U_m)=1]| > \varepsilon\ .$$

Given T, for each v in B the predicate $< v >$ is computed by a circuit of size S, and two different circuits correspond to two different predicates, so the number of elements of B is upper bounded by the number of circuits with T-gates of size S, which in turn is at most $2^{mS\log S}$. This means that the probability that $X \in B$ is at most $|B|2^{-k} = \varepsilon$. An averaging argument shows that

$$\mathbf{Pr}[T(Ext(X,U_t)=1)] - \mathbf{Pr}[T(U_m)=1]| \leq 2\varepsilon$$

□

Then, the Impagliazzo-Wigderson generator gives, for every δ a $(n^\delta, 1/3)$ extractor $Ext : \{0,1\}^n \times \{0,1\}^t \rightarrow \{0,1\}^m$ with $m = n^{\delta'}$ and $t = O(\log n)$, where δ' depends only on δ. This is good enough to give optimal simulations of randomized algorithms using weak random sources, and is an improvement over previous results. Indeed, one can get better bounds on δ and δ', and a generalization to every ε by exploiting details of the Impagliazzo-Wigderson generator.

In the next section we will get even better parameters by using a construction of pseudorandom generator due to Nisan and Wigderson [NW94]. Their construction does not satisfy Definition 2. In particular, if a test T exists for which the distinction probability is more than ε, then there exists a circuit of size m^2 with oracle T such that the circuit computes P on at least a fraction $1/2 + \varepsilon/m$ of the inputs (but not necessarily *all* the inputs). The proof of Theorem 5 above does not apply to this kind of generator, since the counting argument breaks down. Specifically, one can say that every element of B is "approximately" computed by one circuit of size S and that there are at most $2^{mS\log S}$ circuits of size S, however the same circuit may approximate a huge number of elements of B, and so we cannot derive a good upper bound on the size of B from an upper bound on the number of circuits of size S. The argument works again if one encodes X with a proper error-correcting code. We prefer not to describe this part of the argument in full generality here, but rather give it in the next section for the special case of the Nisan-Wigderson generator.

3 Main Result

3.1 Preliminaries

In this section we state some known technical results that will be used in the analysis of our extractor. For an integer n we denote by $[n]$ the set $\{1,\ldots,n\}$. We denote by $u_1 \cdot u_2$ the string obtaining by concatenating the strings u_1 and u_2.

Lemma 6 (Error Correcting Codes) *For every n and δ there is an efficient encoding $EC : \{0,1\}^n \rightarrow \{0,1\}^{\bar{n}}$ where $\bar{n} = \mathrm{poly}(n, 1/\delta)$ such that every ball of Hamming radius $(1/2-\delta)\bar{n}$ in $\{0,1\}^{\bar{n}}$ contains at most $1/\delta^2$ codewords. Furthermore $\bar{n}$ can be assumed to be a power of 2.*

Stronger parameters are achievable. In particular the length of the encoding can be $\bar{n} = n\mathrm{poly}(1/\delta)$. However, the stronger bounds would not improve our constructions. Standard codes are good enough to prove Lemma 6. We sketch a proof of the lemma in the Appendix.

Lemma 7 (Design [NW94]) *For every m, a and l there exists an efficiently constructible family of sets $\mathcal{S} = S_1, \ldots, S_m$ such that*

- *$S_i \subseteq [t]$, where $t = O(e^{(\log m)/a} l^2/a)$*
- *$|S_i| = l$*
- *$|S_i \cap S_j| \leq a$.*

The family $\mathcal{S}$ will be called an (m, l, a)-design.

Lemma 7 was proved in [NW94] for the special case of $a = \log m$. The general case follows using the same proof, but a little care is required while doing a certain probabilistic argument (one has to choose the right Chernoff bound).

The following notation will be useful in the next definition: if $S \subseteq [t]$, with $S = \{s_1, \ldots, s_l\}$ (where $s_1 < s_2 < \cdots < s_l$) and $y \in \{0,1\}^t$, then we denote by $y_{|S} \in \{0,1\}^l$ the string $y_{s_1} \cdot y_{s_2} \cdots y_{s_l}$.

Definition 8 (NW Generator [NW94]) *For a function $f : \{0,1\}^l \rightarrow \{0,1\}$ and an (m,l,a)-design $\mathcal{S} = (S_1, \ldots, S_m)$, the Nisan-Wigderson generator $NW_{f,\mathcal{S}} : \{0,1\}^t \rightarrow \{0,1\}^m$ is defined as*

$$NW_{f,\mathcal{S}}(y) = f(y_{|S_1}) \cdots f(y_{|S_m})$$

For two functions $f, g : \{0,1\}^l \rightarrow \{0,1\}$ and a number $0 \leq \rho \leq 1$ we say that g *approximates* f within a factor ρ if f and g agree on at least a fraction ρ of their domain, i.e. $\mathbf{Pr}_x[f(x) = g(x)] \geq \rho$.

The following result is similar to Lemma 2.4 in [NW94].

Lemma 9 (Analysis of the NW Generator) *Let $\mathcal{S}$ be an (m,l,a)-design, and $T : \{0,1\}^m \rightarrow B$. Then there exists a family $\mathcal{G}_T$ of at most 2^{m2^a} functions such that if $f : \{0,1\}^l \rightarrow \{0,1\}$ is a Boolean function for which*

$$|\mathbf{Pr}_{y\in\{0,1\}^d}[T(NW_{f,\mathcal{S}}(y)) = 1] - \mathbf{Pr}_{r\in\{0,1\}^m}[T(r) = 1]| \geq \varepsilon \ .$$

then there exists a function $g : \{0,1\}^l \rightarrow \{0,1\}^m$ $g \in \mathcal{G}_T$ such that either $T(g(\cdot))$ or its complement approximates $f(\cdot)$ within $1/2 - \varepsilon/m$.

The following result will be used in the proof of Lemma 9. It is typically attributed to Yao [Yao82].

Lemma 10 *Let $T : \{0,1\}^m \rightarrow \{0,1\}$, $g : \{0,1\}^{m-1} \rightarrow \{0,1\}$, $f : \{0,1\}^l \rightarrow \{0,1\}$ and $\varepsilon > 0$; if*

$$\left| \mathbf{Pr}_{x\in\{0,1\}^l}[T(g(x), f(x)) = 1] \ - \ \mathbf{Pr}_{x\in\{0,1\}^l, r\in\{0,1\}}[T(g(x), r) = 1] \right| \geq \varepsilon$$

then there exist two bits $b_0, b_1 \in \{0,1\}$ such that the function $b_0 \oplus T(g(x), b_1)$ agrees with $f(x)$ on at least a fraction $1/2+\varepsilon$ of the inputs.

We now prove Lemma 9.

PROOF: [Of Lemma 9] We follow the proof of Lemma 2.4 in [NW94]. The main idea is that if T distinguishes $NW_{f,\mathcal{S}}(\cdot)$ from the uniform distribution, then we can find a bit of the output where this distinction is noticeable, and then we will apply Lemma 10. In order to find the "right bit", we will use the so-called *hybrid argument*. We define $m+1$ distributions $D_0, \ldots, D_m$; D_i is defined as follows: sample a string $v = NW_{f.\mathcal{S}}(y)$ for a random y, and then sample a string $r \in \{0,1\}^m$ according to the uniform distribution, then concatenate the first i bits of v with the last $m-i$ bits of r. By definition, D_0 is distributed as $NW_{f.\mathcal{S}}(y)$ and D_m is the uniform distribution over $\{0,1\}^m$. Using the hypothesis of the Lemma and the triangle inequality we have

$$\begin{aligned}
\varepsilon &\leq |\mathbf{Pr}_y[T(NW_{f.\mathcal{S}}(y)) = 1] - \mathbf{Pr}_r[T(r)]| \\
&= |\mathbf{Pr}[T(D_0) = 1] - \mathbf{Pr}[T(D_m) = 1]| \\
&= |\sum_{i=0}^{m-1}(\mathbf{Pr}[T(D_i) = 1] - \mathbf{Pr}[T(D_{i+1}) = 1])| \\
&\leq \sum_{i=0}^{m-1}|\mathbf{Pr}[T(D_i) = 1] - \mathbf{Pr}[T(D_{i+1}) = 1]|
\end{aligned}$$

In particular, there exists an index i such that

$$|\mathbf{Pr}[T(D_i) = 1] - \mathbf{Pr}[T(D_{i+1}) = 1]| \geq \varepsilon/m \qquad (1)$$

and there exists a bit $b \in \{0,1\}$ such that

$$\mathbf{Pr}[b \oplus T(D_i) = 1] - \mathbf{Pr}[b \oplus T(D_{i+1}) = 1] \geq \varepsilon/m \qquad (2)$$

Now, recall that

$$D_i = f(y_{|S_1}) \cdots f(y_{|S_{i-1}}) r_i r_{i+1} \cdot r_m$$

and

$$D_{i+1} = f(y_{|S_1}) \cdots f(y_{|S_{i-1}}) f(y_{|S_i}) r_{i+1} \cdot r_m \ .$$

We can use an averaging argument to claim that we can fix $r_{i+1}, \ldots, r_m$ to some values $c_{i+1} \cdots c_m$, as well as all the all the bits of y except those in S_i, and still have an expression like (2). So, we have the relation

$$\mathbf{Pr}[b \oplus T(g_1(x) \cdots g_{i-1}(x) r_i c_{i+1} \cdots c_m) = 1] \ -$$

$$\mathbf{Pr}[b \oplus T(g_1(x) \cdots g_{i-1}(x) f(x) c_{i+1} \cdots c_m) = 1] \geq \varepsilon/m$$

where $g_j(x)$ is $f(y_{|S_j})$ and y is the string whose bits in S_i are fixed according to x, and whose other bits had been set non-uniformly. Since, by the property of the sets $S_1, \ldots, S_m$, every set S_j contains at most a elements of S_i, it follows that $g_j(x)$ depends on at most a bits of its input and therefore it is totally specified by at most 2^a bits. We can now apply Lemma 10 and we have that $b_0 \oplus b \oplus T(g_1(x) \cdots g_{i-1}(x) b_1 c_{i+1} \cdots c_m)$ agrees with f on a fraction $1/2 + \varepsilon/m$ of the inputs. The m-tuple $(g_1(x) \cdots g_{i-1}(x) b_1 c_{i+1} \cdots c_m)$ defines a function g that satisfies the statement of the lemma, and that is entirely specified given at most $m2^a$ bits of information. This is why $\mathcal{G}_T$, that is the set of all such functions over all possible f, contains at most 2^{m2^a} functions. □

3.2 Construction

The construction has parameters n, $k \leq n$, $m \leq k/2$ and $\varepsilon > 0$. We assume that $1 + 3\log(1/\varepsilon) + 2\log m \leq m$, which is true if $m \geq 16$ and $\varepsilon \geq 2^{-m/6}$. This simplifies the expressions below, but is not really necessary for the sake of the construction.

Let $EC : \{0,1\}^n \to \{0,1\}^{\bar{n}}$ be as in Lemma 6, with

- $\delta = \varepsilon/m$,

so that

- $\bar{n} = \text{poly}(n, 1/\varepsilon)$,

and define

- $l = \log \bar{n} = O(\log n/\varepsilon)$.

For an element $u \in \{0,1\}^n$, define

- $\bar{u} =< EC(u) >: \{0,1\}^l \to \{0,1\}$.

Let $\mathcal{S} = S_1, \ldots, S_m$ be as in Lemma 7, such that

- $S_i \subseteq [t]$,
- $|S_i| = l$,
- $|S_i \cap S_j| \leq a = \log(k/2m)$, and
- $t = O\left(l^2 e^{(\log m)/\log k/2m} \frac{1}{\log(k/2m)}\right)$.

Then we define $Ext : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ as

$$Ext(u, y) = NW_{\bar{u},\mathcal{S}}(y) = \bar{u}(y_{|S_1}) \cdots \bar{u}(y_{|S_m}) \ .$$

3.3 Analysis

Lemma 11 *For every fixed predicate $T : \{0,1\}^m \to \{0,1\}$, there are at most $2^{1+m2^a} \cdot (m/\varepsilon)^2$ strings $u \in \{0,1\}^n$ such that*

$$\left| \Pr_{y \in \{0,1\}^d}[T(Ext(u,y)) = 1] - \Pr_{r \in \{0,1\}^m}[T(r) = 1] \right| \geq \varepsilon \quad (3)$$

Proof: It follows from the definition of Ext and from Lemma 9 that if u is such that (3) holds, then there exists a function $g : \{0,1\}^l \to \{0,1\}^m$ in $\mathcal{G}_T$ and a bit $b \in \{0,1\}$ such that the function $b \oplus T(g(\cdot))$ approximates $\bar{u}(\cdot)$ within $1/2 - \varepsilon/m = 1/2 - \delta$.

There are at most 2^{m2^a} functions $g \in \mathcal{G}_T$, furthermore, each such function can be within relative distance $1/2 - \varepsilon/m$ from at most $(m/\varepsilon)^2$ functions $\bar{u}(\cdot)$ coming from the error correcting code of Lemma 6.

We conclude that $2(m/\varepsilon)^2 2^{m2^a}$ is an upper bound on the number of strings u for which Expression (3) can occur. □

Theorem 12 *Ext as described above is a $(k, 2\varepsilon)$-extractor.*

Proof: Fix a predicate $T : \{0,1\}^m \to \{0,1\}$. From Lemma 11 we have that the probability that sampling a u from a source of min-entropy k we can have

$$|\Pr_y[T(Ext(u,y)) = 1] - \Pr_r[T(r) = 1]| \geq \varepsilon$$

is at most $2^{1+m2^a} \cdot \frac{m^2}{\varepsilon^2} \cdot 2^{-k}$ which is at most ε by our choice of parameters. A Markov argument shows that

$$|\Pr_{u,y}[T(Ext(u,y)) = 1] - \Pr_r[T(r) = 1]| \leq 2\varepsilon$$

□

4 Final Remarks

The idea of applying results on pseudorandomness to the context of information-theoretic randomness was inspired by previous work of Andreev et al. [ACRT97]. The use of error-correcting codes was inspired by an alternative proof of the results of [IW97] due to Sudan et al. [STV99].

Both the error correcting codes of Lemma 6 and the design of Lemma 7 can be constructed in logarithmic space. The construction of designs in logarithmic space requires a logarithmic amount of randomness, and only succeeds with high probability (see [IW97] and also [AR98, Section 5] for details), but both these limitations are not a problem in our construction, since the randomness can be taken from the seed, and a small error probability only contributes to a slight increase of the final statistical difference from the uniform distribution. Therefore, our extractors can be constructed in logarithmic space, unlike the dispersers of [SSZ98, TS98] and the extractors of [TS96].

Acknowledgments

Oded Goldreich contributed an important idea in a critical phase of this research; he also contributed very valuable suggestions[3] on how to present the results of this paper. I thank Oded Goldreich, Madhu Sudan, Salil Vadhan, Amnon Ta-Shma, and Avi Wigderson for several helpful conversations. This paper would have not been possible without the help of Adam Klivans, Danny Lewin, Salil Vadhan, Yevgeny Dodis, Venkatesan Guruswami, and Amit Sahai in assimilating the ideas of [NW94, BFNW93, Imp95, IW97]. Thanks also to Madhu Sudan for hosting our reading group in the Spring'98 Complexity Seminars at MIT.

This work was mostly done while the author was at MIT, partially supported by a grant of the Italian CNR. Part of this work was also done while the author was at DIMACS, supported by a DIMACS post-doctoral fellowship.

References

[ACRT97] A.E. Andreev, A.E.F. Clementi, J.D.P. Rolim, and L. Trevisan. Weak random sources, hitting sets, and BPP simulations. In *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, pages 264–272, 1997.

[AR98] E. Allender and K. Reinhardt. Isolation, matching, and counting. Technical Report TR98-019, Electronic Colloquium on Computational Complexity, 1998.

[BFNW93] L. Babai, L. Fortnow, N. Nisan, and A. Wigderson. BPP has subexponential time simulations unless EXPTIME has publishable proofs. *Computational Complexity*, 3(4):307–318, 1993.

[BGS98] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCP's and non-approximability – towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998. Preliminary version in *Proc. of FOCS'95*.

[BM84] M. Blum and S. Micali. How to generate cryptographically strong sequences of pseudorandom

[3] Indeed, I did not follow all of them, and this is why the current presentation is not so good.

bits. *SIAM Journal on Computing*, 13(4):850–864, 1984. Preliminary version in *Proc. of FOCS'82*.

[BR94] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science*, pages 276–287, 1994.

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

[CW89] A. Cohen and A. Wigderson. Dispersers, deterministic amplification, and weak random sources. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 14–19, 1989.

[GM84] S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299, 1984. Preliminary Version in *Proc. of STOC'82*.

[GNW95] O. Goldreich, N. Nisan, and A. Wigderson. On Yao's XOR lemma. Technical Report TR95-50, Electronic Colloquium on Computational Complexity, 1995.

[Gol99] O. Goldreich. *Modern Cryptography, Probabilistic Proofs and Pseudorandomness*. Springer-Verlag, 1999.

[GZ97] O. Goldreich and D. Zuckerman. Another proof that $BPP \subseteq PH$ (and more). Technical Report TR97-045, Electronic Colloquium on Computational Complexity, 1997.

[Imp95] R. Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*, pages 538–545, 1995.

[IW97] R. Impagliazzo and A. Wigderson. $P = BPP$ unless E has sub-exponential circuits. In *Proceedings of the 29th ACM Symposium on Theory of Computing*, pages 220–229, 1997.

[IW98] R. Impagliazzo and A. Wigderson. Randomness versus time: De-randomization under a uniform assumption. In *Proceedings of the 39th IEEE Symposium on Foundations of Computer Science*, pages 734–743, 1998.

[MS77] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 1977.

[Nis96] N. Nisan. Extracting randomness: How and why. In *Proceedings of the 11th IEEE Conference on Computational Complexity*, pages 44–58, 1996.

[NTS98] N. Nisan and A. Ta-Shma. Extrating randomness : A survey and new constructions. *Journal of Computer and System Sciences*, 1998. To appear. Preliminary versions in [Nis96, TS96].

[NW94] N. Nisan and A. Wigderson. Hardness vs randomness. *Journal of Computer and System Sciences*, 49:149–167, 1994. Preliminary version in *Proc. of FOCS'88*.

[NZ93] N. Nisan and D. Zuckerman. More deterministic simulation in Logspace. In *Proceedings of the 25th ACM Symposium on Theory of Computing*, pages 235–244, 1993.

[RRV99] R. Raz, O. Reingold, and S. Vadhan. Extracting all the randomness and reducing the error in Trevisan's extractors. In *Proceedings of the 31st ACM Symposium on Theory of Computing*, 1999.

[RTS97] J. Radhakrishnan and Amnon Ta-Shma. Tight bounds for depth-two superconcentrators. In *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, pages 585–594, 1997.

[SSZ98] M. Saks, A. Srinivasan, and S. Zhou. Explicit OR-dispersers with polylogarithmic degree. *Journal of the ACM*, 45(1):123–154, 1998. Preliminary version in *Proc. of STOC'95*.

[STV99] M. Sudan, L. Trevisan, and S. Vadhan. Pseudorandom generators without the XOR lemma. In *Proceedings of the 31st ACM Symposium on Theory of Computing*, 1999.

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

[SZ94] A. Srinivasan and D. Zuckerman. Computing with very weak random sources. In *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science*, pages 264–275, 1994.

[TS96] A. Ta-Shma. On extracting randomness from weak random sources. In *Proceedings of the 28th ACM Symposium on Theory of Computing*, pages 276–285, 1996.

[TS98] A. Ta-Shma. Almost optimal dispersers. In *Proceedings of the 30th ACM Symposium on Theory of Computing*, 1998.

[VV85] U. Vazirani and V. Vazirani. Random polynomial time is equal to slightly random polynomial time. In *Proceedings of the 26th IEEE Symposium on Foundations of Computer Science*, pages 417–428, 1985.

[WZ93] A. Wigderson and D. Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. In *Proceedings of the 25th ACM Symposium on Theory of Computing*, pages 245–251, 1993.

[Yao82] A.C. Yao. Theory and applications of trapdoor functions. In *Proceedings of the 23th IEEE Symposium on Foundations of Computer Science*, pages 80–91, 1982.

[Zuc90] D. Zuckerman. General weak random sources. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 534–543, 1990.

[Zuc96a] D. Zuckerman. On unapproximable versions of *NP*-complete problems. *SIAM Journal on Computing*, 25(6):1293–1304, 1996. Preliminary Version in *Proc. of Structures'93*.

[Zuc96b] D. Zuckerman. Randomness-optimal sampling, extractors and constructive leader election. In *Proceedings of the 28th ACM Symposium on Theory of Computing*, pages 286–295, 1996.

A Appendix

A.1 A Discussion on Lemma 6

It is a standard result that if an error-correcting code has large minimum distance then there can be few codewords in every large ball. In particular, the following bound holds.

Lemma 13 *Suppose C is an error-correcting code with (relative) minimum distance $\geq 1/2-\beta/2$. Then every Hamming ball of (relative) radius $1/2-\sqrt{\beta}$ contains at most $1/3\beta$ codewords.*

A proof can be found e.g. in [BGS98, Lemma A.1]. The following result is well known, even if we do not know of a source where it is clearly stated in this way.

Lemma 14 *For every δ and n there exists an error-correcting code with 2^n codewords of length $\bar{n} = \mathrm{poly}(n, 1/\delta)$ and with minimum distance $(1/2-\delta)\bar{n}$. The code admits a polynomial-time encoding algorithm.*

Several constructions meet this requirement. In particular one can use a Reed-Solomon code concatenated with a Hadamard code. See e.g. [MS77] for a treatment of error correcting codes. Lemma 6 follows from Lemmas 13 and 14.

A.2 A Sketch of the Proof of Lemma 7

The following version of the Chernoff bound will be used.

Lemma 15 *Let $X_1, \ldots, X_n$ be 0/1 mutually independent random variables such that $\mathbf{E}\left[\sum_i X_i\right] = \mu$. Then, for every $\alpha > 1$ it holds*

$$\mathbf{Pr}\left[\sum_i X_i \geq \alpha\mu\right] \leq e^{-((\ln\alpha)+\frac{1}{\alpha}-1)\alpha\mu}$$

This bound is proved in the usual way, but we have not find a standard reference for this particular statement, so we present below a proof (taken from lecture notes by Leighton and Vempala).

Proof: Let $p = \mathbf{Pr}[X_i = 1] = \mathbf{E}[X_i] = \mu/n$. Then

$$\begin{aligned}
\mathbf{Pr}\left[\sum_i X_i \geq \alpha\mu\right] &= \mathbf{Pr}\left[\alpha^{\sum_i X_i} \geq \alpha^{\alpha\mu}\right] \\
&\leq \frac{\mathbf{E}\left[\alpha^{\sum_i X_i}\right]}{\alpha^{\alpha\mu}} \\
&= \frac{\prod_i \mathbf{E}\left[\alpha^{X_i}\right]}{\alpha^{\alpha\mu}} \\
&= \frac{(1-p+p\alpha)^n}{\alpha^{\alpha\mu}} \\
&\leq \frac{e^{(-p+p\alpha)n}}{\alpha^{\alpha\mu}} \\
&= e^{-\mu+\alpha\mu-\alpha\mu\ln\alpha}
\end{aligned}$$

The second step uses Markov, the third uses independence, and the fifth uses the inequality $(1-x) \leq e^{-x}$. □

We can now sketch the proof of Lemma 7 as it was carried on in [NW94].

Proof:[Of Lemma 7] Sequentially choose m subsets of $[t]$ such that any of the chosen subsets intersects the previously chosen ones in less than a points. A probabilistic argument using the above Chernoff bound shows that the algorithm is always able to choose a new subset as long as the total number of sets in no more than m (in the probabilistic argument we will choose a multi-set of elements, so as to be able to use the Chernoff bound, and then we will discard duplicates.) □