Pseudo-Random Generators and Structure of Complete Degrees

Manindra Agrawal
Dept of CSE, IIT Kanpur 208016, India
email: manindra@iitk.ac.in

Abstract

It is shown that if there exist sets in E *that require* $2^{\Omega(n)}$*-sized circuits then sets that are hard for class* P*, and above, under 1-1 reductions are also hard under 1-1, size-increasing reductions. Under the assumption of the hardness of solving RSA or Discrete Log problem, it is shown that sets that are hard for class* NP*, and above, under many-one reductions are also hard under (non-uniform) 1-1, and size-increasing reductions.*

1 Introduction

Pseudo-random generators, although originally defined for specific usages, have turned out to be fundamental objects with applications in many diverse areas. There exist two types of such generators: one computable in exponential-time in their seed length (defined by Nisan and Wigderson [17] for derandomization purposes), and the other computable in polynomial-time in their seed length (defined in [5, 21] for cryptographic purposes). In this paper, we provide yet another application of pseudo-random generators by using both types of generators to prove structural theorems on complete degrees of NP and other classes.

1.1 Background

The structure of complete sets for classes NP, E, NE etc. has received much attention over the years. Polynomial-time many-one reductions (in short, m-reductions) are considered to be the "most appropriate" notion for defining complete sets for these classes. The strongest possible structure of complete sets for a class is *p-isomorphism*: any two sets are reducible to each other via a polynomial-time computable and invertible isomorphism. It has been a long-standing open question whether complete sets for any of the above classes are all p-isomorphic to each other. Berman and Hartmanis [4] showed that two sets are p-isomorphic to each other iff they are reducible to each other via 1-1, size-increasing, and p-invertible m-reductions. They also conjectured (the *isomorphism conjecture*) that all NP-complete sets are p-isomorphic to each other. However, until now, only partial results are known in this direction:

- Berman [3] showed that all sets complete under m-reductions (in short, $\leq^p_m$-complete sets) for E, and deterministic classes above, are also complete under 1-1 and size-increasing reductions (in short, $\leq^p_{1,li}$-complete). His argument also shows that all $\leq^p_m$-complete for deterministic classes that diagonalize over P are also $\leq^p_{li}$-complete.
- Ganesan and Homer [8] showed that all $\leq^p_m$-complete sets for NE, and non-deterministic classes above, are also $\leq^p_1$-complete.

No results are known for NP-complete sets even under the assumption P $\neq$ NP. In fact, it is now widely believed that complete sets for NP, E, NE, etc are *not* all p-isomorphic to each other. The reason being the widely believed existence of 1-1, *one-way* functions—these are polynomial-time computable 1-1 functions that are p-invertible on only a very small fraction of outputs. Joseph and Young [13] (essentially) conjectured that there is no p-invertible reduction of SAT to $f(\text{SAT})$, where f is a 1-1 one-way function. This conjecture was given the name *encrypted complete set* conjecture by Selman [19]. Since both SAT and $f(\text{SAT})$ are NP-complete, the conjecture implies that NP-complete sets are not all p-isomorphic to each other. The same intuition can be used to argue that complete sets for E and NE are also not all p-isomorphic to each other. Adding weight to this conjecture, Kurtz, Mahaney and Royer [14] showed that the conjecture holds (for all classes) relative to a random oracle.

Even if one believes that the isomorphism conjecture is false, the structure of $\leq^p_m$-complete sets for NP (and other standard classes except E) remains unclear. For example, are all NP-complete sets $\leq^p_{1,li}$-complete, or at least $\leq^p_{li}$-complete? As is clear from the abovementioned results, we do not know much about these questions. Even for much bigger non-deterministic classes like NE we do not have a complete answer to these questions. For NE this is a little surprising since for a smaller class E we *do* know that

all $\leq^p_m$-complete sets are $\leq^p_{1,li}$-complete. The reason for this anomaly is that the known method of proving the size-increasing property requires the class to be closed under complement (and this is unlikely to be true for NE). So adding non-determinism to a class *reduces* our ability to to obtain results about its complete degree. A partial result for NE was shown by Tran [20]: all complete sets for NE have an infinite polynomial-time subset (this follows immediately if they are complete under size-increasing reductions).

In fact, there is so far no evidence to believe that all $\leq^p_m$-complete sets for NE are $\leq^p_{1,li}$-complete. The situation for NP is much worse: we have practically no concrete[1] evidence for any kind of structure on NP-complete sets.

1.2 Our Results

We provide, for the first time, strong evidence that $\leq^p_m$-complete sets for NP and other non-deterministic classes are $\leq^p_{1,li}$-complete: under widely believed assumptions, we prove that $\leq^p_m$-complete sets for NP and other classes are $\leq^{p/poly}_{1,li}$-complete ($\leq^{p/poly}_{1,li}$= *non-uniform polynomial-time*, 1-1, and size-increasing). The hypotheses that we use for proving our results are the following:

Hypothesis A: There exists a set in E such that any non-uniform family of circuits computing the set has size $2^{\Omega(n)}$.

Hypothesis B: The *RSA problem* or the *Discrete Log problem* is 2^{n^ϵ}-secure for some $\epsilon > 0$ (see next section for definitions).

Both the above hypotheses are widely believed to be true. Impagliazzo and Wigderson [12] constructed a *true pseudo-random generator* using Hypothesis A. Goldreich et. al. [9] constructed a *one-way permutation* using Hypothesis B. From these one-way permutations, Yao [21] constructed a *cryptographic pseudo-random generator* (see next section for definitions). We will make use of these two generators in our proofs.

We first show that if Hypothesis A holds then, for almost all classes of interest, $\leq^p_1$-complete sets are $\leq^p_{1,li}$-complete:

Theorem 1 *If Hypothesis A holds then for every class $\mathcal{C}$ that is closed under polynomial-time reductions, if A is $\leq^p_1$-hard for $\mathcal{C}$, then A is also $\leq^p_{1,li}$-hard for $\mathcal{C}$.*

The following corollary is immediate:

Corollary 2 *If Hypothesis A holds then:*

- *For any class $\mathcal{C}$ closed under polynomial-time reductions, its $\leq^p_1$-complete degree collapses to $\leq^p_{1,li}$-complete degree.*
- *For class* NE*, its $\leq^p_m$-complete degree collapses to $\leq^p_{1,li}$-complete degree (using the result of [8]).*

The above result does *not* imply anything about $\leq^p_m$-complete degrees of classes below NE, e.g., NP. For some of these classes, we can prove the following:

Theorem 3 *If Hypothesis A holds then for every class $\mathcal{C}$ that can diagonalize over* P*, and is closed under union and non-deterministic polynomial-time reductions, if A is $\leq^p_m$-hard for $\mathcal{C}$, then A is also hard for $\mathcal{C}$ under size-increasing reductions that are 1-1 on $\bar{A}$.*

This result nicely complements the result for deterministic classes that diagonalize over P [3]. Is this result true for NP? Unlikely. Firstly, it is unlikely that NP diagonalizes over P (this would mean that all polynomial time sets can be accepted by an NP machine in time $O(n^k)$ for some fixed $k > 0$). Secondly, there cannot be a relativizable proof of above theorem for NP since relative to an oracle for which P = NP, Hypothesis A is true and $\leq^p_m$-complete degree for NP is clearly not $\leq^p_{li}$-complete.

So for the class NP, we perhaps require a different hypothesis. Using the Hypothesis B, we prove the following for NP:

Theorem 4 *If Hypothesis B holds, then for every class $\mathcal{C}$ closed under non-deterministic polynomial-time reductions, if A is $\leq^p_m$-hard for $\mathcal{C}$, then A is $\leq^{p/poly}_{1,li}$-hard for $\mathcal{C}$.*

We can eliminate non-uniformity for the size-increasing part of the above result using Hypothesis A. By using a hypothesis stronger than A, we can reduce the non-uniformity for the 1-1 part. The stronger hypothesis that we need is:

Hypothesis A*: There exists a set in E such that any non-uniform family of non-deterministic circuits computing the set has size $2^{\Omega(n)}$.

With this hypothesis we get:

Theorem 5 *1. If both Hypotheses A and B hold, then for every class $\mathcal{C}$ closed under non-deterministic polynomial-time reductions, if A is $\leq^p_m$-hard for $\mathcal{C}$, then A is $\leq^p_{1,li}$-hard for $\mathcal{C}$.*

2. If both Hypotheses A* and B hold, then for every class $\mathcal{C}$ closed under non-deterministic polynomial-time reductions, if A is $\leq^p_m$-hard for $\mathcal{C}$, then A is $\leq^{p/log}_{1,li}$-hard for $\mathcal{C}$.

[1] As opposed to the inferred evidence by observing the 'natural' complete sets.

The paper is organized as follows. The next section gives all the definitions that we use. Section 3 gives the proof of Theorem 1 and Section 4 gives proof of Theorem 4. Proofs of Theorems 3 and 5 are given in the Appendix. Section 5 discusses the results and future work.

2 Definitions

2.1 Classes, Reductions, Completeness

We assume the definitions of standard complexity classes [15].

For a class $\mathcal{C}$, ($\leq^p_{li}$-, $\leq^p_1$-, $\leq^p_{1,li}$-, $\leq^p_{1,li,i}$-) $\leq^p_m$-*hard* set is a set that is hard for $\mathcal{C}$ under (respectively size-increasing, 1-1, 1-1 and size-increasing, 1-1 and size-increasing and p-invertible) polynomial-time reductions. Similarly, one defines the various completeness notions. We often refer to $\leq^p_m$-complete sets for a class $\mathcal{C}$ as $\mathcal{C}$-complete sets. The set of all $\leq^p_r$-complete sets (for various types of restrictions r on the reduction) for $\mathcal{C}$ is called $\leq^p_r$-*complete degree* of $\mathcal{C}$.

When a set is hard for $\mathcal{C}$ under non-uniform 1-1, size-increasing polynomial-time reductions, we denote it by $\leq^{p/poly}_{1,li}$-hard. When the length of advice string required by the reduction is $O(\log n)$—instead of poly(n)—on inputs of size n, we denote it by $\leq^{p/log}_{1,li}$-hard.

A *non-deterministic polynomial-time* reduction of set B to A is a (possibly multi-valued) function computable by a non-deterministic polynomial-time TM that, on its guess paths on input x, either aborts or outputs a string with the property that $x \in B$ iff the string output is in A.

A class $\mathcal{C}$ that is closed under polynomial-time reductions *diagonalizes over* P if the set

$$D[t] = \{i \mid \text{ DTM } M_i \text{ accepts } i \text{ within } |i|^{t(i)} + |i| \text{ steps}\},$$

for some monotonically increasing function $t(.)$, is in $\mathcal{C}$ (here M_1, M_2, ... is an enumeration of all deterministic TMs).

2.2 Secure problems

A function $f = \{f_n\}, f_n : \{0,1\}^n \mapsto \{0,1\}^{m(n)}$, is $s(n)$-*secure* if for every $\delta(.)$ such that $\delta(n) < 1$, for every $t(.)$ such that $t(n) \leq \delta(n) \cdot s(n)$, and for every non-uniform circuit family $\{C_n\}$ of size $t(n)$,

$$\Pr_{x \in \{0,1\}^n}[C_n(x) = f_n(x)] \leq \frac{1}{2^{m(n)}} + \delta(n),$$

for sufficiently large n.

This definition is more general than the usual definition in which, instead of a non-uniform family of circuits, a probabilistic algorithm is considered [11]. We would need this more general definition in our proofs.

2.3 RSA and Discrete Log problems

The RSA problem: Given numbers n, e, and y such that

- n is the product of two equal sized primes,
- e is relatively prime to $\phi(n)$, $1 < e < \phi(n)$, and
- y is relatively prime to n, $1 \leq y < n$,

find number x such that $y = x^e (\text{mod } n)$. The security of the RSA public-key encryption algorithm [18] relies on the hardness of this problem.

The Discrete Log problem: Given numbers p, the prime factorization of $\phi(p)$, g, and y such that

- p is prime,
- g is a primitive element of F^*_p, and
- $y \in F^*_p$,

find number x such that $y = g^x (\text{mod } p)$. The security of many cryptographic protocols, including the El Gamal public-key encryption algorithm [7] and Diffie-Hellman key-exchange algorithm [6], is based on the hardness of this problem.[2]

It is widely believed that both the above problems are 2^{n^ϵ}-secure for some $\epsilon > 0$ even under the more general notion of security that we consider.

2.4 One-way permutations

Function p is a $s(n)$-*secure one-way permutation* if

- p is a 1-1, length preserving, and polynomial-time computable function, and
- function p^{-1} is $s(n)$-secure.

In [9], Goldreich, Levin and Nisan showed how to construct permutations from the RSA or Discrete Log problem. These permutations are $s(O(n))$-secure one-way permutations provided the RSA and Discrete Log problem are $s(n)$-secure respectively.

2.5 Cryptographic pseudo-random generators

Function $G = \{G_n\}, G_n : \{0,1\}^n \mapsto \{0,1\}^{m(n)}$ is a $s(n)$-*secure crypto pseudo-random generator* if

- G is computable in polynomial-time in input length,
- $m(n) > n$, and

[2] In the standard version of discrete log problem, the prime factorization of $p-1$ is not given. However, it is believed that the problem is hardest to solve when $p - 1 = 2q$ for a prime q. And when this is the case, prime factorization of $p-1$ can be trivially computed.

- for every $\delta(\cdot)$ such that $\delta(n) < 1$, for every $t(\cdot)$ such that $t(n) \leq \delta(n) \cdot s(n)$, and for for every circuit C of size $t(n)$,

$$|\Pr_{x\in\{0,1\}^{m(n)}}[C(x)=1] - \Pr_{y\in\{0,1\}^{n}}[C(G_n(y))=1]| \leq \delta(n),$$

for sufficiently large n.

In [21, 5] etc. a $\frac{s(n)}{n^{O(1)}}$-secure crypto pseudo-random generator is constructed from a $s(n)$-secure one-way permutation. So assuming that there exist 2^{n^ϵ}-secure one-way permutation, it follows that there exist 2^{n^δ}-secure crypto pseudo-random generators for $0 < \delta < \epsilon$.

2.6 True pseudo-random generators

Function $G = \{G_n\}, G_n : \{0,1\}^\ell \mapsto \{0,1\}^n$ is a *true pseudo-random generator* if

- $\ell = O(\log n)$,
- G is computable in time exponential in input size, and
- for any circuit C of size n,

$$|\Pr_{x\in\{0,1\}^{n}}[C(x)=1] - \Pr_{y\in\{0,1\}^{\ell}}[C(G_n(y))=1]| \leq \frac{1}{n}.$$

In [12], Impagliazzo and Wigderson constructed a true pseudo-random generator—we will refer to it as G^{IW}—using the assumption that there exists a set in E such that any non-uniform family of circuits accepting the set has size $2^{\Omega(n)}$.

3 Proof of Theorem 1

Let A be a $\leq_1^p$-hard set for class $\mathcal{C}$ and $B \in C$. Consider the set $\tilde{B} = B \times \{0,1\}^*$. Since $\mathcal{C}$ is closed under polynomial-time reductions, $\tilde{B} \in C$. Let $\tilde{B} \leq_1^p A$ via function f computable in time $q(n)$ for some polynomial $q(.)$.

Define function g as:

> On input x, $|x| = n$, compute $G_m^{IW}(y)$ for $m = c \cdot q^2(2n+2)+n$ (for a suitable constant $c > 0$) and every string y (size of y is $O(\log m) = O(\log n)$ by definition of G^{IW}). Let these strings be z_1, z_2, ..., z_k ($k = \text{poly}(n)$). For each z_i, let u_i be the first $n+2$ bits of z_i. Compute $f(x, u_i)$ for each i, $1 \leq i \leq k$, and select the index value, say j, for which $|f(x, u_i)|$ is maximum. Output (x, u_j).

Clearly, g is a polynomial-time reduction of B to $\tilde{B}$, and therefore, $h = f \circ g$ is a polynomial-time reduction of B to A. We now prove that h is both 1-1, and size-increasing.

Since both g and f are 1-1, h must be 1-1. Consider any string x, $|x| = n$. Define a circuit C that, on input z, $|z| = m$, works as follows:

> Let $z = uv$ where $|u| = n+2$. C ignores v, and computes $f(x,u)$. If $|f(x,u)| > n$ then it accepts, otherwise rejects.

The size of circuit C is at most $c \cdot q^2(2n+2) + n = m$ since one can simulate time t computation by a circuit of size $c \cdot t^2$ [15] (this determines the value of the constant c). Since f is a 1-1 function, and the set $I_x = \{(x,u) \mid |u| = n+2\}$ has exactly 2^{n+2} strings, at least half of the strings in the set $f(I_x) = \{f(x,u) \mid |u| = n+2\}$ have length at least $n+1$. Therefore, the circuit C would accept at least half the fraction of inputs. Now, by the property of the pseudo-random generator G_m^{IW}, C would accept at least $\frac{1}{2} - \frac{1}{m}$ fraction of strings from the range of G_m^{IW}. This implies that there exists at least one (in fact, *many*) prefix u of some output of G_m^{IW} such that $|f(x,u)| > n$. The definition of function g ensures that $g(x) = (x,u)$ for one such u. Therefore, $|h(x)| > |x|$.

4 Proof of Theorem 4

When f is an m-reduction, we cannot use the above argument. Instead, we use a completely different argument based on Hypothesis B.

Let p be a 2^{n^ϵ}-secure one-way permutation (as constructed in [9] from Hypothesis B). Goldreich and Levin [10] construct the following pseudo-random generator from p:

$$G^{GL}(x,r) = (p(x), r, x \cdot r),$$

where $|x| = |r| = n$ and '$\cdot$' is inner product modulo 2. They show that function G^{GL} is a 2^{n^δ}-secure pseudo-random generator for $0 < \delta < \epsilon$. Notice that G^{GL} is undefined for strings of odd length. As we will require it to be defined everywhere, we extend its definition:

$$G^{GL}(x,rb) = (p(x), rb, x \cdot r),$$

where $|x| = |r| = n$, and $|b| = 1$. This extended function has the same security. Function G^{GL} is clearly a 1-1 function.

The proof is split in three stages. In the first stage, using the generator G^{GL} we show that the set A is hard under reductions with "few collisions." In the next stage, we show, using a pairwise-independent generator, that A is hard under non-uniform size-increasing reductions that are 1-1 on $\{0,1\}^n$. Finally, in third stage, we use a standard padding technique to show that A is $\leq_{1,li}^{p/poly}$-hard.

Stage 1

Let $B \in \mathcal{C}$. Define set

$$\hat{B} = \{G^{GL}(x) \mid x \in B\}.$$

Set $\hat{B}$ is clearly in $\mathcal{C}$ since $\mathcal{C}$ is closed under non-deterministic polynomial-time reductions and G^{GL} is a 1-1 function. Let $\hat{B} \leq_m^p A$ via f. Then, $h = f \circ G^{GL}$ is a reduction of B to A.

We say that function g is *γ-sparsely many-one on $S \subseteq \{0,1\}^n$* if for every $x \in S$, $\mid g^{-1}(g(x)) \cap \{0,1\}^n \mid \leq \frac{2^n}{2^{n^\gamma}}$ (here we use $g^{-1}(z)$ to denote the set of all strings that map to z via g). Say that g is *sparsely many-one on $S \subseteq \{0,1\}^n$* if it is γ-sparsely many-one on S for some $\gamma > 0$.

Lemma 4.1 *For every n, function h is $\frac{\delta}{2}$-sparsely many-one on $B \cap \{0,1\}^n$.*

Proof. Suppose not. Fix an n and $x_0 \in B \cap \{0,1\}^n$ such that

$$\mid h^{-1}(h(x_0)) \cap \{0,1\}^n \mid > \frac{2^n}{2^{n^{\delta/2}}}.$$

Let $S = B \cap \{0,1\}^n$ and $w = h(x_0)$. Since h is a reduction of B to A, $w \in A$.

for every $x \in S$, $h^{-1}(h(x)) \cap \{0,1\}^n \subseteq S$. Let $\mid h^{-1}(h(x_0)) \cap \{0,1\}^n \mid > \frac{2^n}{2^{n^{\delta/2}}}$ for some $x_0 \in S$. Let $T = G^{GL}(h^{-1}(h(x_0)))$. Since G^{GL} is 1-1, $|T| > \frac{2^n}{2^{n^{\delta/2}}}$.

Define a circuit C, that on input y, $|y| = n+1$, accepts iff $f(y) = w$. Since G^{GL} is 1-1, C accepts at least $\frac{2^n}{2^{n^{\delta/2}}}$ strings. Since $w \in A$, all the strings accepted by C are in the range of G^{GL}. So,

$$\mid \Pr_{y \in \{0,1\}^{n+1}}[C(y) = 1] - \Pr_{x \in \{0,1\}^n}[C(G^{GL}(x)) = 1] \mid$$

$$= \frac{1}{2} \cdot \Pr_{x \in \{0,1\}^n}[C(G^{GL}(x)) = 1] > \frac{1}{2^{1+n^{\delta/2}}}.$$

The size of circuit C is a polynomial in n. Therefore, the security of G^{GL} is at most $\text{poly}(n) \cdot 2^{1+n^{\delta/2}} < 2^{n^\delta}$, a contradiction. ■

To obtain a reduction of B to A that is sparsely many-one on entire $\{0,1\}^n$, we need another iteration of the above argument (with a different definition of $\hat{B}$).

Define

$$\hat{B}' = B \cup \{y \mid y \notin \textit{range}(G^{GL})\}.$$

Set $\hat{B}'$ "collects" all the strings that are not in the range of G^{GL}. Observe the following:

Lemma 4.2 $\{y \mid y \notin \textit{range}(G^{GL})\} \in$ NP.

Proof. A non-deterministic TM accepting the above set is defined as follows:

> On input y, let $y = ub$, $|u| = n$ and $|b| = 1$. Let $u = wr$ where $|w| = \lfloor \frac{n}{2} \rfloor$. Guess a v, $|v| = |w|$, such that $p(v) = w$. Compute $G^{GL}(v, r)$ and accepts iff it is not equal to y.

The only point to note here is that there *always* exists a unique v such that $p(v) = w$ since p is a permutation. ■

Therefore, $\hat{B}' \in \mathcal{C}$. Let f' be a reduction of $\hat{B}'$ to A that is sparsely many-one on $\hat{B}' \cap \{0,1\}^n$ for every n. Let $h' = f' \circ G^{GL}$. Clearly, h' is a reduction of B to A.

Lemma 4.3 *For every n, function h' is sparsely many-one on $\{0,1\}^n$.*

Proof. As G^{GL} is a 1-1 reduction of B to $\hat{B}'$ and increases the length by one only, function h' is also sparsely many-one on $B \cap \{0,1\}^n$ for every n. Arguing exactly the same way as before, this time for $\bar{B} \cap \{0,1\}^n$ though, we can show that h' is sparsely many-one on $\bar{B} \cap \{0,1\}^n$ for every n. Therefore, h' is sparsely many-one on $\{0,1\}^n$ for every n. ■

Stage 2

For this stage, we need a *pairwise-independent generator* $F_{n,m}(x,r) : \{0,1\}^n \times \{0,1\}^k \mapsto \{0,1\}^m$—for any two $x_1 \neq x_2$, $|x_1| = |x_2| = n$, $F_{n,m}(x_1, r)$ and $F_{n,m}(x_2, r)$ should be independently and uniformly distributed over $\{0,1\}^m$ when r is uniformly chosen from $\{0,1\}^k$. There are many ways in which such a generator can be defined—we choose the one in which r is a $m \times m$ matrix over GF[2], x is treated as $m \times 1$ vector over GF[2] (by padding $m - n$ trailing zeroes to it, so $m \geq n$), and $F_{n,m}(x,r) = x \cdot r$. Clearly, $F_{n,m}$ is computable in polynomial-time.

We now proceed with our construction. Let $B \in \mathcal{C}$. Define set

$$\tilde{B} = B \times \{0,1\}^*.$$

Clearly, $\tilde{B} \in \mathcal{C}$. Let $\tilde{B} \leq_m^p A$ via f such that f is γ'-sparsely many-one on $\{0,1\}^n$ for every n (as ensured by Stage 1). Define function g as:

$$g(w,r) = (w, F_{n,m}(w,r)),$$

where $m = (n+2)^{2/\gamma'}$. Clearly, g is a reduction of $\tilde{B}$ to itself. Define function $h_n^r(w) = f(g(w,r))$ for $|w| = n$ and $|r| = m$. For any r, function h_n^r is a reduction of B to A for strings of size n. We now show that when r is randomly chosen, at least half of functions h_n^r are size-increasing and 1-1 on $\{0,1\}^n$.

Lemma 4.4 *At least half of functions in the set $\{h_n^r\}$ are size-increasing and 1-1 on $\{0,1\}^n$.*

Proof. We first show that at least $\frac{3}{4}$ fraction of functions are size-increasing. Fix w, $|w| = n$, and consider the set of strings $\{h_n^r(w)\} = f(\{w\} \times \{0,1\}^m)$. At most 2^{n+1} of the strings in this set have size less than or equal to n. Since function f is γ'-sparsely many-one, at most $2^{n+1} \cdot \frac{2^m}{2^{m^{\gamma'}}} = \frac{2^m}{2^{(n+2)^2-n-1}}$ strings in the set $\{w\} \times \{0,1\}^m$ can therefore be mapped by f to strings of size less than or equal to n. As the output of function $g(w,r)$ is uniformly distributed over $\{w\} \times \{0,1\}^m$ when the input r is chosen uniformly (follows since $F_{n,m}(\cdot)$ is uniformly distributed), the probability that $|h_n^r(w)| \leq n$ is at most $\frac{1}{2^{(n+2)^2-n-1}}$. Therefore, the probability that $|h_n^r(w)| \leq n$ for *some* w of length n is at most $\frac{1}{2^{n^2+2n+3}} \leq \frac{1}{4}$. This shows that at least $\frac{3}{4}$ h_n^rs are size-increasing on $\{0,1\}^n$.

Next, we show that at least $\frac{3}{4}$ fraction of functions are 1-1 on $\{0,1\}^n$. The proof is very similar to the above, we now use the pairwise independence of $F_{n,m}$. Fix w_1 and w_2, $|w_1| = |w_2| = n$, $w_1 \neq w_2$. Arguing as above—using the facts that the second components of both $g(w_1,r)$ and $g(w_2,r)$ are independently and uniformly distributed over $\{0,1\}^m$ and that f is γ'-sparsely many-one—one can obtain that the probability that $h_n^r(w_1) = h_n^r(w_2)$ is at most $\frac{1}{2^{m^{\gamma'}}}$. Therefore, the probability that *for some pair* w_1 and w_2, $h_n^r(w_1) = h_n^r(w_2)$ is at most $\frac{2^{n^2}}{2^{m^{\gamma'}}} \leq \frac{1}{2^{4n+4}} \leq \frac{1}{4}$. This completes the proof. ∎

By non-uniformly fixing an appropriate value for r (one such value for each n) so that h_n^r is 1-1 and size-increasing, we obtain a reduction of B to A that is size-increasing and 1-1 on $\{0,1\}^n$ for every n. Notice that this function may *not* be 1-1 everywhere, e.g., there could be two string of *different* lengths that are mapped to the same string by the function.

Stage 3

Finally, we use a standard padding trick to show that A is $\leq_{1,li}^{p/poly}$-hard. Let $B \in \mathcal{C}$, and $\tilde{B} = B \times \{0,1\}^*$. Let $\tilde{B}$ reduce to A via non-uniform f that is 1-1 and size-increasing on $\{0,1\}^n$ for every n. Let $|f(x)| \leq q(|x|)$ for some polynomial $q(\cdot)$.

Define function ℓ with $\ell(j) = q(\ell(j-1))$, $\ell(1) = 1$. Define function g with $g(x) = x01^{\ell(j)-|x|-1}$ where j is the smallest number such that $\ell(j) > |x|$. Clearly, g is a 1-1, size-increasing reduction of B to $\tilde{B}$. So h is a reduction of B to A. We now show,

Lemma 4.5 *Function h is 1-1 and size-increasing.*

Proof. Since f and g are both size-increasing, h is size-increasing. Consider $h(x)$ and $h(x')$ for $x \neq x'$. If $|g(x)| = |g(x')|$ then since f is 1-1 on $\{0,1\}^{|g(x)|}$, $h(x) \neq h(x')$. On the other hand, if $|g(x)| = \ell(j) > |g(x')| = \ell(j')$ then $|h(x')| = |f(g(x'))| \leq q(|g(x')|) = q(\ell(j')) \leq \ell(j'+1) \leq \ell(j) < |h(x)|$. Therefore, h is 1-1. ∎

5 Remarks and Future Work

Until now, results about the structure of m-complete degrees were obtained using diagonalization.[3] Pseudo-random generators, in a sense, provide a strong form of diagonalization. So it is logical (at least on hindsight) that we have been able to obtain stronger results using them.

Also, so far, for proving structure of complete degrees, non-determinism was a drawback instead of a resource. With the help of pseudo-random generators, we have shown how to exploit non-determinism as a resource (we have obtained stronger results for some non-deterministic classes than corresponding deterministic ones).

Many questions remain unanswered. The most important ones are:

1. Can we remove the non-uniformity from Theorems 4 and 5?

2. Can one disprove (or prove!?) the isomorphism conjecture under a similar plausible hypothesis?

3. Can one weaken the hypothesis B to the existence of any one-way function? Notice that the only place we need Hypothesis B is for obtaining sparsely many-one reductions. And this appears to require the stronger hypothesis of one-way permutations.

4. Can one prove that relative to a random oracle, all NP-complete sets are also complete under 1-1 and size-increasing reductions? This would nicely complement the result of [14]. The problem here is that of obtaining one-way permutations relative to a random oracle.

Acknowledgement

The author wishes to thank anonymous referees for suggestions leading to an improved presentation—particularly to a referee who suggested the use of pairwise independent generator in Stage 2 of the proof of Theorem 4 in place of a pseudo-random function generator used in an earlier version of the paper. Also, Harry Buhrman is to be thanked for asking the question that led to this work.

[3] In contrast, for complete degrees under more restricted reductions, e.g., AC^0-reductions, strong structural results have been obtained using other techniques—including, interestingly, special kind of pseudo-random generators [2, 1].

References

[1] M. Agrawal. The first order isomorphism theorem. In *Proceedings of Twenty First FST&TCS*, 2001. to be presented.

[2] M. Agrawal, E. Allender, and S. Rudich. Reductions in circuit complexity: An isomorphism theorem and a gap theorem. *J. Comput. Sys. Sci.*, 57:127–143, 1998.

[3] L. Berman. *Polynomial Reducibilities and Complete Sets*. PhD thesis, Cornell University, 1977.

[4] L. Berman and J. Hartmanis. On isomorphism and density of NP and other complete sets. *SIAM Journal on Computing*, 1:305–322, 1977.

[5] M. Blum and S. Micali. How to generate cryptographically strong sequences of pseudo-random bits. *SIAM Journal on Computing*, 13:850–864, 1984.

[6] W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT-22(6):644–654, 1976.

[7] T. ElGamal. A public-key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, IT-31(4):469–472, 1985.

[8] K. Ganesan and S. Homer. Complete problems and strong polynomial reducibilities. In *Proceedings of the Symposium on Theoretical Aspects of Computer Science*, pages 240–250. Springer Lecture Notes in Computer Science 349, 1988.

[9] O. Goldreich, L. Levin, and N. Nisan. On constructing 1-1 one-way function. Technical Report TR95-029, Electronic Colloquium on Computational Complexity (http://www.eccc.uni-trier.de/eccc), 1995.

[10] O. Goldreich and L. A. Levin. A hardcore predicate for all one-way functions. In *Proceedings of Annual ACM Symposium on the Theory of Computing*, pages 25–32, 1989.

[11] J. Hastad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, pages 221–243, 1998.

[12] R. Impagliazzo and A. Wigderson. P = BPP if E requires exponential circuits: Derandomizing the XOR lemma. In *Proceedings of Annual ACM Symposium on the Theory of Computing*, pages 220–229, 1997.

[13] D. Joseph and P. Young. Some remarks on witness functions for nonpolynomial and noncomplete sets in NP. *Theoretical Computer Science*, 39:225–237, 1985.

[14] S. Kurtz, S. Mahaney, and J. Royer. The isomorphism conjecture fails relative to a random oracle. In *Proceedings of Annual ACM Symposium on the Theory of Computing*, pages 157–166, 1989.

[15] J. V. Leeuwen, editor. *Handbook of Theoretical Computer Science, Volume A*. Elsevier, 1990.

[16] P. B. Milterson and N. V. Vinodchandran. Derandomizing Arthur-Merlin games using hitting sets. In *Proceedings of Annual IEEE Symposium on Foundations of Computer Science*, pages 71–80, 1999.

[17] N. Nisan and A. Wigderson. Hardness vs. randomness. *J. Comput. Sys. Sci.*, 49(2):149–167, 1994.

[18] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of ACM*, 21:120–126, 1978.

[19] A. L. Selman. A survey of one-way functions in complexity theory. *Mathematical Systems Theory*, 25:203–221, 1992.

[20] N. Tran. On p-immunity of nondeterministic complete sets. In *Proceedings of the Structure in Complexity Theory Conference*, pages 262–263, 1995.

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

Appendix

Proof of Theorem 3

When f is an m-reduction, the set $f(I_x)$, as defined in proof of Theorem 1, may not have *any* string of length greater than n. The known techniques for 'converting' m-reductions to 1-1 reductions require the class $\mathcal{C}$ to be at least E. So we cannot use any of these techniques. Instead, using non-determinism and diagonalization over P we obtain a reduction that is 1-1 on the set I_x for $x \in \bar{B}$. Using generator G^{IW}, we then get a reduction that is size-increasing on $\bar{A}$. We then use diagonalization to make it size-increasing everywhere. Finally, we use non-determinism and diagonalization again to make the reduction 1-1 on $\bar{A}$.

Let A be a $\leq^p_m$-hard set for class $\mathcal{C}$ and $B \in C$. Let the diagonal set $D[t] \in \mathcal{C}$. Define set $\hat{B}$ as:

> On input (i, x, u), accept if *either* there exists a $u' > u$, $|u'| = |u|$, such that $M_i(i, x, u) = M_i(i, x, u')$ and M_i on both inputs halts within $(|i| + |x| + |u|)^{t(i,x,u)} + |i| + |x| + |u|$ steps for DTM M_i, *or* $x \in B$.

Define DTM index $j(i, x, u, u')$ coding a TM that, on input $j(i, x, u, u')$ accepts iff $u' > u$, $|u'| = |u|$, $M_i(i, x, u) = M_i(i, x, u')$ and M_i on either input halts within $(|i| + |x| + |u|)^{t(i,x,u)} + |i| + |x| + |u|$ steps. Also define a set $\tilde{B}$ such that $j(i, x, u, u') \in \tilde{B}$ iff $x \in B$.

Clearly, $(i, x, u) \in \hat{B}$ iff for some u': $j(i, x, u, u') \in D[t] \cup \tilde{B}$. Since $\tilde{B} \leq^p_m B$, and class $\mathcal{C}$ is closed under non-deterministic polynomial-time reductions and union, it follows that $\hat{B} \in \mathcal{C}$. Let $\tilde{B} \leq^p_m A$ via function f computable in time $q(n)$ for some polynomial $q(.)$. Also, let DTM M_j compute the function f within $|x|^k + k$ steps on input x.

Lemma 5.1 *For any $x \in \bar{B}$ and for any $\ell > 0$, function f is 1-1 on the set $I^\ell_x = \{(j, x, u) \mid |u| = \ell\}$.*

Proof. Suppose not. Then for some $x \in \bar{B}$ and some ℓ, there are strings u and u' with $|u| = |u'| = \ell$, such that $f(j, x, u) = f(j, x, u')$. Let u and u' be the lexicographically largest two strings with this property and $u' > u$. Then $(j, x, u) \in \hat{B}$ and $(j, x, u') \notin \hat{B}$ by definition of $\hat{B}$. A contradiction. ■

Now defining the function g reducing B to $\hat{B}$ exactly as in proof of Theorem 1—except that instead of pairs (x, u) the function works with triples (j, x, u)—and arguing exactly as before, we can show that the reduction $h = f \circ g$ of B to A is size-increasing whenever $x \in \bar{B}$.

To make the reduction size-increasing everywhere, we use diagonalization. For set B, define set $\hat{B}'$ as:

> On input (i, x), compute $M_i(i, x)$ where DTM M_i halts within $(|i| + |x|)^{t(i,x)} + |i| + |x|$ steps. Reject if $|M_i(i, x)| \leq |x|$. Otherwise, accept iff $x \in B$.

As before, we can show that $\hat{B}' \in \mathcal{C}$. Let f' be a reduction of $\hat{B}'$ to A that is size-increasing on the complement of $\hat{B}'$ (as shown above, we can construct such a reduction). Let $M_{j'}$ be a DTM that computes f' in polynomial-time. Define functions g' and h' as: $g'(x) = (j', x)$ and $h' = f' \circ g'$.

Lemma 5.2 *Function g' is a reduction of B to $\hat{B}'$ and function h' is size-increasing.*

Proof. Suppose $|M_{j'}(j', x)| \leq |x|$ for some x. Then by the definition of set $\hat{B}'$, $(j', x) \notin \hat{B}'$. Since $M_{j'}$ computes reduction f' of $\hat{B}'$ to A, and f' is size-increasing on the complement of $\hat{B}'$, this is impossible. Therefore, $|M_{j'}(j', x)| > |x|$ for every x. The claim follows. ■

Finally, we make use of diagonalization and non-determinism to construct a reduction of B to A that is 1-1 on $\bar{B}$. Define set $\hat{B}''$ as:

> On input (i, x), accept if *either* there exists an x', $x' > x$, $|x'| = |x|$, such that $M_i(i, x) = M_i(i, x')$, and M_i on both inputs halts within $(|i| + |x|)^{t(i,x)} + |i| + |x|$ steps, *or* $x \in B$.

As before, we can show that $\hat{B}'' \in C$. Fix a reduction of $\hat{B}''$ to A that is size-increasing. Let $M_{j''}$ be a DTM computing f'' in polynomial time. As before, one can argue that f'' must be 1-1 on the inputs of the form (j'', x) when $x \in \bar{B} \cap \{0, 1\}^n$ and $g''(x) = (j'', x)$ is a reduction of B to $\hat{B}''$. Therefore, $h'' = f'' \circ g''$ is a size-increasing reduction of B to A that is 1-1 on $\bar{A} \cap \{0, 1\}^n$ for every n. Now using the padding trick described in Stage 4 of proof of Theorem 4, we can obtain another reduction h''' of B to A that is size-increasing and 1-1 on entire $\bar{A}$.

Proof of Theorem 5

This is a minor modification of Stage 2 of proof of Theorem 4. At the end of Stage 2, we have a set of functions $\{h^r_n\}$ such that at least half of them are 1-1 and size-increasing on $\{0, 1\}^n$.

For any give w, define a deterministic circuit C testing if a specified $h^r_n(w)$ is size-increasing on w: on input r, the circuit accepts iff $|h^r_n(w)| > |w|$. This is a polynomial sized circuit and therefore, using the true pseudo-random generator G^{IW} (as in proof of Theorem 1) we can obtain a list of polynomially many r's such that for many of these r's $h^r_n(w)$ is size-increasing. Using this, we can easily define a size-increasing reduction as before.

To reduce non-uniformity from 1-1 part, we use the stronger Hypothesis A*. We first need the definition of a *hitting set generator*:

Definition 5.3 Function $H = \{H_n\}, H_n : \{0,1\}^\ell \mapsto \{0,1\}^n$ is a *hitting set generator* if

- $\ell = O(\log n)$,
- H is computable in time exponential in input size, and
- for any circuit C of size n that accepts at least half the fraction of inputs, there exists a $y \in \{0,1\}^\ell$ such that $C(H_n(y)) = 1$.

Hitting set generators are weaker than pseudo-random generators. However, from the result of [12] it follows that they are, in fact, equivalent. In our proof, we need a stronger version of these generators. Function H is a hitting set generator against *co-nondeterministic circuits* if the circuit C in the above definition is a co-nondeterministic one. It was shown in [16] that if Hypothesis A* is true then there exists a hitting set generator against co-nondeterministic circuits.

We now continue with the proof. Define a co-nondeterministic circuit C testing if a given h_n^r is 1-1: on input r, the circuit checks that for every $w \neq w'$, $h_n^r(w) \neq h_n^r(w')$. The size of the circuit is poly$(|r|)$. Using an appropriate hitting set generator (guaranteed by Hypothesis A*), we can obtain polynomially many r's such that for at least one such r, h_n^r is 1-1. Identifying such a r now requires only $O(\log n)$ many non-uniform bits.