Super-bits, Demi-bits, and $N\tilde{P}/qpoly$-natural Proofs

Steven Rudich
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
rudich@cs.cmu.edu

July 26, 1997

Abstract

We introduce the super-bit conjecture, which allows the development of a theory generalizing the notion of pseudorandomness so as to fool *non-deterministic* statistical tests. This new kind of pseudorandomness rules out the existence of $NP/poly$-natural properties that can work against $P/poly$. This is an important extension of the original theory of $P/poly$-natural proofs [10]. We also introduce the closely related demi-bit conjecture which is more intuitive and is the source of interesting open problems.

1 Introduction

By exploiting the theory of pseudorandom generators [1, 11, 3, 7], Razborov and Rudich [10] give evidence that the proof techniques used in the last sixteen years of non-uniform, non-monotone circuit lower bounds will be unable to solve the barrier problems of complexity theory. They argue that known lower bound arguments are all "natural" which means that they exploit some P-natural combinatorial property (see Section 3.1). Furthermore, they show that if strong enough pseudorandomness is possible, no such property can be used in proving that satisfiability requires super-polynomial size circuits. Subsequently, Razborov [8, 9] showed that proof systems with the constructive interpolation property (see also [6]) cannot prove such lower bounds without implying the existence of the arguments ruled out in [10] under a pseudorandomness assumption. Thus, the existence of pseudorandom generators implies the independence of $SAT \notin P/poly$ from a class of proof systems.

We extend the standard notion of pseudorandomness so as to be secure against *non-deterministic* adversaries. (This is fairly counter-intuitive. A non-deterministic adversary can simply guess the seed of the generator.) Our extension makes the generalization of previous work quite easy. We introduce the super-bit conjecture, which serves as the hardness assumption to found our theory. We show that the super-bit conjecture rules out $NP/poly$ (in fact, $NquasiP/qpoly$) natural proofs that satisfiability requires large circuits. This is a much stronger negative result than the previous [10]. Furthermore, the super-bit conjecture implies the independence of $SAT \notin P/poly$ from any proof system with the *existential* interpolation property.

For concreteness, we propose a super-bit generator based on the non-deterministic hardness of the subset sum problem. Thus all the conjectures in this paper are true if subset sum is suitably hard.

We introduce the demi-bit conjecture, which is more intuitive than the super-bit conjecture. We hope, but we cannot prove, that this will serve as the foundation for our theory. The open problems about demi-hardness are of independent interest.

2 Definitions and Notation

We denote by F_n the set of all Boolean functions in n variables. Most of the time, it will be convenient to think of $f_n \in F_n$ as a binary string of length 2^n, called the *truth-table* of f_n.

A non-deterministic circuit is one which includes gates outputting a single "non-deterministic" bit. The circuit is said to accept iff there is a choice of non-deterministic bits to be output that causes the circuit to accept its input. $N\tilde{P}/qpoly$ is the notation for non-uniform, non-deterministic circuits of quasi-polynomial size.

3 The $N\tilde{P}/qpoly$-natural properties conjecture

3.1 Natural properties in general.

Razborov and Rudich [10], formalize the class of combinatorial properties that arise in published lower bound arguments in boolean circuit complexity. We start by reviewing their definition.

Let Γ and Λ be complexity classes. Call a combinatorial property C_n Γ-*natural* with density δ_n if it contains $C_n^* \subseteq C_n$ with the following two conditions:

Constructivity: The predicate $f_n \overset{?}{\in} C_n^*$ is computable in Γ (recall, C_n^* is a set of truth-tables with 2^n bits);

Largeness: $|C_n^*| \geq \delta_n \cdot |F_n|$.

A combinatorial property C_n is *useful against* Λ if it satisfies:

Usefulness: For any sequence of functions f_n, where the event $f_n \in C_n$ happens infinitely often, $\{f_n\} \notin \Lambda$.

3.2 $NP/poly$-natural properties.

Of interest here will be a choice of specific parameters in the general definition. Let $\Gamma = NP/poly$ and δ_n be any function of growth rate $2^{-O(n)}$.

Call a combinatorial property C_n *$NP/poly$-natural* if it contains $C_n^* \subseteq C_n$ with the following two conditions:

Constructivity: $C_n^* \in NP/poly$.

Largeness: $|C_n^*| \geq 2^{-O(n)} \cdot |F_n|$.

3.3 $N\tilde{P}/qpoly$-natural properties.

Also of interest will be a choice of parameters in the general definition with even more liberal parameters. Let $N\tilde{P}/qpoly$ be the class of languages recognized by non-uniform, quasi-polynomial-size circuit families. (Quasi-polynomial means $n^{\log n^{O(1)}}$.)

Let $\Gamma = N\tilde{P}/qpoly$ and δ_n be any function of growth rate $2^{-n^{O(1)}}$. If C_n is natural with these parameters, we say that it is *$N\tilde{P}/qpoly$-natural.*

3.4 Main conjecture

Conjecture 1 *There are no $N\tilde{P}/qpoly$-natural properties that are useful against $P/poly$.*

A trivial corollary is the weaker statement: *there are no $NP/poly$-natural properties useful against $P/poly$.*

4 The significance of the conjecture.

4.1 A pleasing overall concept.

Conjecture 1 is a much stronger statement than the main negative statement of [10] that there are no $P/poly$-natural proofs. It rules out a wider category of lower bound proof than we have yet encountered.

The concept of an $N\tilde{P}/qpoly$-natural proof is actually simpler than $P/poly$-natural proofs. $N\tilde{P}/qpoly$-natural proofs are any argument that gives a non-negligible fraction of boolean functions a short certificate proving that they lie outside some complexity class.

4.2 Relation to independence results

We say that a propositional proof system (see [5]) T has the *Existential Interpolation Property (EIP)* if there is a polynomial $p(x)$ such that whenever T proves

$A(a_1, \ldots, a_n) \vee B(b_1, \ldots b_m)$ with a proof of length l then T has a $p(l)$ length proof of either A or B. (Notice that the a_i's and b_j's are distinct variables.)

We say that a propositional proof system T has the *Constructive Interpolation Property (CIP)* if it has the EIP and there is a polynomial time procedure that, given a proof of $A \vee B$, will output a length $p(l)$ proof of one or the other.

Statements in complexity theory can be reformulated as families of propositional statements. In order to get rid of all the quantifiers used in a standard mathematical statement, we will allow the propositional equivalent to be exponentially longer than the original. For example, the statement that satisfiability requires circuits of size $\Omega(n^{(\log n)})$ can be handled as follows. S will be a propositional statement of size $2^{O(n)}$ asserting that there is no circuit of size $n^{\log n}$ that can decide satisfiability for all formulas of length n. S has two easy building blocks. (1) A formula SAT which tells us which formulas of length n are satisfiable. SAT is trivial to construct because we allow it to be of length 2^n. (2) A formula CVAL that takes an $O(n^{\log n})$ bit index C into one of the circuits of size $n^{\log n}$ as well as an input X to the circuit and returns true iff $C(X) = 1$. CVAL can be easily constructed with size $O(n^{2\log n})$. Let X_i for $1 \leq i \leq 2^n$ be an enumeration of n-bit strings. $C(X_i)$ will be an abbreviation for CVAL(C, X_i). The free variables of S will be $O(n^{\log n})$ bits which should be thought of as indexing a circuit C. $S(C) \equiv (C(X_1) \neq SAT(X_1)) \vee (C(X_2) \neq SAT(X_2)) \vee \cdots \vee (C(X_n) \neq SAT(X_n))$. S has size $O(2^{2n})$. Let $\langle S_m \rangle$ be the family of propositions where S_m ($m = O(2^{2n})$) asserts that SAT on inputs of length n does not have $n^{\log n}$ size circuits. The brute force way of proving S_m would involve enumerating all the $2^{n^{\log n}}/n^{\log n} = \mathrm{o}(2^{kn})$ circuits of size $n^{\log n}$. The burning question is whether or not standard boolean algebra allows for a polynomial-size proof of $\langle S_m \rangle$. A negative result would imply that S^1_2 can't prove SAT$\notin$P/poly.

Theorem [Razborov [8]]: If a reasonable[1] propositional proof system with CIP proves $\langle S_m \rangle$ with proofs of polynomial-size, then there is a combinatorial property that is $\tilde{P}$-natural against $P/poly$.

Corollary: If 2^{n^ϵ}-hard functions exist, then $\langle S_m \rangle$ does not have polynomial-size proofs in any reasonable propositional proofs systems with CIP.

An easy corollary of Razborov's proof shows:

Theorem 1 *If a reasonable propositional proof system with EIP proves $\langle S_m \rangle$ with proofs of polynomial-size, then there is a combinatorial property that is $N\tilde{P}/qpoly$-natural against $P/poly$.*

Corollary: If Conjecture 1 is true, then $\langle S_m \rangle$ does not have polynomial-size proofs in any reasonable propositional proofs systems with EIP.

Thus, the project of extending the study of $\tilde{P}/qpoly$ natural proofs to $N\tilde{P}/qpoly$-natural corresponds to extending the study of propositional systems with CIP to those with EIP.

[1]Reasonable requires a few simple properties common to all studied propositional proof systems.

4.3 A disjoint pair of NP sets that can't easily be separated.

We say that a set S separates two sets if S contains one of them and $\bar{S}$ (the complement of S) contains the other. In [], Razborov introduces two NP sets that appear to be difficult to separate with a

5 Cryptography against nondeterministic adversaries.

5.1 The super-bit conjecture.

At first glance, it would seem impossible to generalize the notion of a pseudorandom generator to work against nondeterministic adversaries. The reason is simple: a nondeterministic adversary can guess the seed of the generator and verify that the guess is correct.

In this section, we will show how to change the standard definition of pseudorandom generator to handle the case of nondeterministic adversaries. Furthermore, we show that the known generator constructions still work according to this stronger definition.

We start by reviewing the standard definition of pseudorandomness in the nonuniform circuit model.

Definition 1 (Standard Hardness and Pseudorandom Generators): *Let $g_n : \{0,1\}^n \to \{0,1\}^{l(n)}$ be a family in P/poly where $l(n) > n$. The hardness $H(g_n)$ of pseudorandom generator g_n is the minimal S for which there exists a circuit C of size $\leq S$ such that*

$$|Pr_{z\in\{0,1\}^{l(n)}}[C(z)=1] - Pr_{x\in\{0,1\}^n}[C(g_n(x))=1]| \geq \frac{1}{S}$$

Here is the standard hardness conjecture that was shown to imply that *P/poly*-natural properties against *P/poly* do not exist [10].

Conjecture 2 (Standard Conjecture): *There is a pseudorandom generator with hardness 2^{n^ϵ}, for some $\epsilon > 0$.*

Notice that the use of the absolute value does not really matter. This because (especially in the non-uniform model) it is easy to make an adversary circuit C' that determines for a given n if the difference is positive or negative and flips C's output accordingly. Thus, we can not only drop the absolute value, but we can also *order the two probabilities in the difference either of two ways.*

To obtain a stronger definition that uses nondeterministic adversaries, we simply drop the absolute value! Unlike the deterministic case, the order of the two probabilities is *crucial.*

Definition 2 (Non-deterministic Hardness) *Let $g_n : \{0,1\}^n \to \{0,1\}^{l(n)}$ be a family in P/poly where $l(n) > n$. The hardness $H(g_n)$ of pseudorandom generator g_n is the minimal S for which there exists a non-deterministic circuit C of size $\leq S$ such that*

$$Pr_{z\in\{0,1\}^{l(n)}}[C(z) = 1] - Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] \geq \frac{1}{S}$$

If we reversed the order of the probabilities or kept the absolute value, a simple adversary would *always* break the generator: guess the seed x and verify that it is consistent with the observed string.

By placing $Pr_{z\in\{0,1\}^{l(n)}}[c(z) = 1]$ first, we force a nondeterministic circuit to prove that an observed string is probably random, as opposed to proving that the observed string is probably the output of the generator. Non-deterministic circuit classes are not likely to be closed under complement. This asymmetry is what makes the order of the probabilities so important.

Equivalently, we could formulate it in such a way that the two probabilities range over disjoint sets:

$$Pr_{z\in\{0,1\}^{l(n)}, z\notin \text{range}_g}[C(z) = 1] - Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] \geq \frac{1}{S}$$

In other words, there is a function $g \in P/poly$ such that C cannot be true on a significantly larger fraction of non-range elements than the fraction of range elements it is true on.

After consideration, we make the follow conjecture. It could also be called the *non-deterministic hardness* conjecture.

Conjecture 3 (Super-bit Conjecture) *There exists a $g : \{0,1\}^n \to \{0,1\}^{n+1}$ in P/poly with non-deterministic hardness 2^{n^ϵ}, for some $\epsilon > 0$.*

We call it the super-bit conjecture because it says that we can produce a bit of pseudorandomness (n bits in/ $n+1$ bits out) that can fool a super powerful adversary.

We note that a random function oracle $g_n : \{0,1\}^n \to \{0,1\}^{n+1}$ can easily be seen to satisfy the definition of a super-bit. In other words, the conjecture is true relative to a random oracle.

5.2 A super-bit based on subset sum.

We propose that a previously studied generator[4], the subset sum generator, is already the source of a secure super-bit generator. (We are grateful to Moni Naor for this suggestion.)

Let g be a function taking m $m+1$-bit numbers, $a_1, a_2, \ldots, a_m$ and m 1-bit numbers, $b_1, b_2, \ldots, b_m$.

$$\begin{aligned} g(\quad a_1 & , a_2, \ldots, a_n, b_1, b_2, \ldots, b_n) \\ &= (a_1 b_1 + a_2 b_2 + \cdots + a_m b_m \mod 2^{m+1}), a_1, \ldots, a_n, b_1, \ldots, b_n \end{aligned}$$

Notice that g is a function that takes $m(m+1)+m$ bits of input and returns $m(m+1)+m+1$ bits of output.

We conjecture that this function has the super-bit requirement. Intuitively, it is very hard to prove that a given $n+1$-bit sequence is not in the range of g. The length of the shortest proof of non-solvability of subset sum was considered in Furst and Kannan [2]. They were not able to find shorter-than-obvious proofs for subset sum problems on the instance sizes used in the definition of g.

Conjecture 4 (Subset Sum Conjecture) *Let $g : \{0,1\}^n \to \{0,1\}^{n+1}$ be the subset sum function above. g has non-deterministic hardness 2^{n^ϵ}, for some $\epsilon > 0$.*

5.3 Stretching a super-bit.

In the standard theory of pseudorandomness, a single bit of pseudorandomness can be stretched to many bits [1, 11]. In fact, Goldreich, Goldwasser, and Micali[3] showed that it is possible to construct pseudorandom function generators based on any pseudorandom bit generator. In this section, we argue that, using standard constructions, a single super-bit can be stretched to many super-bits, and even used to create pseudorandom function generators with non-deterministic hardness. The proofs are omitted in this abstract. It suffices to peruse the standard proofs [7] and to notice that they never flip the order of the terms in the difference.

Let $g : \{0,1\}^n \to \{0,1\}^{n+1}$ be a function. We stretch g using the standard construction [7]. Let $first(x)$ be the leftmost bit of x, and $rest(x)$ be the remaining $|x|-1$ rightmost bits. $g^0(x) = x, g^1(x) = g(x)$ and, for all $i \geq 1$, $g^{i+1}(x) = first(g^i(x)); g^i(rest(g(x)))$.

Theorem 2 *If g is a pseudorandom generator with non-deterministic hardness S, then $g^{l(n)}$ is a pseudorandom generator with non-deterministic hardness $S/l(n)$.*

We are now in a position to make a pseudorandom function generator. We know that assuming a super-bit generator, we can make a generator $g : \{0,1\}^n \to \{0,1\}^{2n}$ that doubles the length of its input. If x is an even length string, let $left(x)$ be the left half of x, and $right(x)$ be the right half of x. Let $g^0(x) = x$, $g^1(x) = g(x)$ and, for all $i \geq 1$, $g^{i+1}(x) = g^i(left(g(x)))$; $g^i(right(g(x)))$.

Theorem 3 *If g is a pseudorandom generator with non-deterministic hardness S, then $g^{l(n)}$ is a pseudorandom generator with non-deterministic hardness $S/(2^{l(n)})$.*

Furthermore, $g^{l(n)}$ gives us the basis for a pseudorandom function generator. Define $node_x(emptystring) = x$. For a non-empty string y of length less than or equal to $l(n)$, define $node_x(y) = left(g(node_x(rest(y))))$ if $first(y) = 0$, and $right(g(node_x(rest(y))))$ if $first(y) = 1$. Define $f_x(y) = first(node_x(y))$. Notice that, as long as x and y have polynomially related lengths, there is a polynomial-size circuit family to compute $f_x(y)$, given x and y. Define $F(x)$ to be the bit sequence that gives the truth table of f_x.

Corollary to Theorem 3: If g is a pseudorandom generator with non-deterministic hardness S, then F is a pseudorandom generator with output length $2^{l(n)}$ and non-deterministic hardness at least $S/2^{l(n)}$.

This is a trivial corollary, since $F(x)$ presents strictly less information to an adversary than does $g^{l(n)}(x)$. Therefore, we can view $f_x(y)$ as a pseudorandom function family. For each x, f_x can be computed by a polynomial-size circuit family, and for a random x, $F(x)$ looks random to any non-deterministic adversary.

5.4 $N\tilde{P}/qpoly$-natural properties and the super-bit conjecture.

Assume that we can generate a super-bit, then by Theorem 2 we know that we have a pseudorandom generator $g : \{0,1\}^r \to \{0,1\}^{r+1}$ with non-deterministic hardness 2^{n^ϵ}, for some fixed $\epsilon > 0$. Assume that Conjecture 1 is false, i.e., there exists an $N\tilde{P}/qpoly$-natural property C_n that is useful against $P/poly$. We will obtain a contradiction by following reasoning similar to [10] and [8].

Without loss of generality, we can assume that C_n already satisfies the largeness condition (as opposed to some subset C_n^*). Define $N = 2^n$. We can choose an integer k such that for sufficiently large n, the bound on the size of the circuit that computes C_n is less than $N^{\log^{k-2}(N)} = 2^{n^{k-1}}$, and that $|C_n| > 2^{-n^{k-1}} \cdot |F_n| = N^{-\log^{k-2}(N)} \cdot 2^N$.

Let $m = n^{k/\epsilon}$. Choosing $|x| = m$ and $|y| = m^{\epsilon/k} = n$, we can use the construction from Theorem 3 to obtain a pseudorandom function family f_x with non-deterministic hardness $2^{m^\epsilon}/2^n = 2^{n^{k-1}}$ such that $|F(x)| = N$.

Since each function f_x is in $P/poly$, we know that for all x $C_n(F(x)) = 0$. By assumption, $|C_n| > 2^{-n^{k-1}} \cdot |F_n|$. Thus,

$$|Pr_{z\in\{0,1\}^N}[C_n(z) = 1] - Pr_{x\in\{0,1\}^m}[C_n(F(x)) = 1]| > \frac{1}{2^{n^{k-1}}}$$

The circuit C_n that performs this statistical test has size less than $2^{n^{k-1}}$. This contradicts the non-deterministic hardness of f_x.

We conclude:

Theorem 4 *If super-bits exist (Conjecture 3), then Conjecture 1 (no $N\tilde{P}/qpoly$-properties) is true.*

6 The Demi-bit conjecture.

The notion of a super-bit has the advantage of being the generalization of a pseudorandom bit. Unfortunately, Conjecture 3 will be less than intuitive to some people.

We seek here to develop a more intuitive conjecture that might also be sufficient to rule out $N\tilde{P}/qpoly$-natural properties. In fact, we are unable to show that demi-bits can be stretched, as can super-bits. Whether useful in this context or not, demi-bits are of independent interest.

Definition 3 (Demi-Hardness) *Let $g_n : \{0,1\}^n \to \{0,1\}^{l(n)}$ be a family in $P/poly$ where $l(n) > n$. The demi-hardness $H(g_n)$ of pseudorandom generator g_n is the minimal S for which there exists a non-deterministic circuit C of size $\leq S$ such that*

$$Pr_{z\in\{0,1\}^{l(n)}}[C(z) = 1] \geq \frac{1}{S}$$

and

$$Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] \;=\; 0$$

Conjecture 5 (demi-bit Conjecture) *There exists a $b : \{0,1\}^n \to \{0,1\}^{n+1}$ in $P/poly$ with demi-hardness 2^{n^ϵ}, for some $\epsilon > 0$.*

This is more intuitive. Simply, it says that there is function such that most non-range elements have no short certificate that prove that they are not in the range of b. This is similar to formalizing the notion of a $coNP$ predicate that is average case hard for NP.

Open Problem 1 *Does the demi-bit conjecture imply the super-bit conjecture?*

Open Problem 2 *If you have one demi-bit, can you stretch it to 2 demi-bits?*

Open Problem 3 *If you have one demi-bit, can you build a pseudorandom function generator with demi-hardness 2^{n^ϵ}?*

A positive answer to this last problem would solve the next one:

Open Problem 4 *Can you prove that the demi-bit conjecture rules out the existence of $N\tilde{P}/qpoly$-natural properties?*

7 Other Open Problems

Open Problem 5 *On what weaker assumption can one prove the existence of a demi-bit?*

Open Problem 6 *On what weaker assumption can one prove the existence of a super-bit?*

8 Acknowledgements

We would like to thank Moni Naor, Avi Wigderson, Sasha Razborov, Russell Impagliazzo, Oded Goldreich, and Jiří Sgall for useful conversations.

References

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

[2] M. Furst and R. Kannan. Succinct certificates for almost all subset sum problems. *Siam Journal on Computing*, vol 18, 1989, pp. 550-558.

[3] O. Goldreich, S. Goldwasser, and S. Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, 1986.

[4] R. Impagliazzo and M. Naor. Efficient Cryptographic Schemes Provably as Secure as Subset Sum. *Journal of Cryptology*, 9(4):199-216, 1996.

[5] J. Krajíček. *Bounded Arithmetic, Propositional Logic, and Complexity Theory*, Cambridge University Press, 1995.

[6] J. Krajíček. Interpolation theorems, lower bounds for proof systems and independence results for bounded arithmetic. Submitted to *Journal of Symbolic Logic*, 1994.

[7] Michael Luby, *Pseudorandomness and Cryptographic Applications*, Princeton University Press (1996).

[8] A. Razborov. Unprovability of lower bounds on circuit size in certain fragments of Bounded Arithmetic. (*Izvestiya of the RAN*), 59(1):201–222, 1995.

[9] A. Razborov. On provably disjoint **NP**-pairs. Technical Report RS-94-36, Basic Research in Computer Science Center, Aarhus, Denmark, 1994.

[10] A. Razborov and S. Rudich. Natural Proofs. *26th Annual Symposium on Theory of Computing* (1994). To appear in special issue of JCSS.

[11] A. Yao. Theory and Applications of Trapdoor Functions. *Proc.* 23$^{\text{rd}}$ *IEEE Symp. on Foundations of Computer Science*, Chicago, IL (Nov. 1982), 80–91.