Electronic Colloquium on Computational Complexity - Reports Series 1996 - available via:
ECCC
TR96-021
FTP: ftp.eccc.uni-trier.de:/pub/eccc/
WWW: http://www.eccc.uni-trier.de/eccc/
Email: ftpmail@ftp.eccc.uni-trier.de with subject 'help eccc'

NP-hard Sets Are Superterse unless **NP** Is Small

Yongge Wang
Mathematisches Institut
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg
Germany
wang@math.uni-heidelberg.de

Keywords: Computational complexity, **P**-selective set, p-measure.

1 Introduction

One of the important questions in computational complexity theory is whether every **NP** problem is solvable by polynomial time circuits, i.e., $\mathbf{NP} \subseteq ?\mathbf{P}/poly$. Furthermore, it has been asked what the deterministic time complexity of **NP** is if $\mathbf{NP} \subseteq \mathbf{P}/poly$. That is, if **NP** is easy in the nonuniform complexity measure, how easy is **NP** in the uniform complexity measure? Let $\mathbf{P}_T(\mathbf{SPARSE})$ be the class of languages that are polynomial time Turing reducible to some sparse sets. Then it is well known that $\mathbf{P}_T(\mathbf{SPARSE}) = \mathbf{P}/poly$. Hence the above question is equivalent to the following question.

$$\mathbf{NP} \subseteq ?\mathbf{P}_T(\mathbf{SPARSE}).$$

It has been shown by Wilson [18] that this question is oracle dependent. Hence it seems difficult to give an absolute answer to this question at present. In the past, many efforts have been made to consider the question whether **NP** is not included in some subclasses of $\mathbf{P}_T(\mathbf{SPARSE})$. Since $\mathbf{P}_T(\mathbf{SPARSE})$ is the class of languages that are Turing reducible to some sparse sets, one way of obtaining subclasses of $\mathbf{P}_T(\mathbf{SPARSE})$ is to consider some restrictions on the reducibility. For example, Mahaney [10] showed that if all **NP** sets are many-one reducible to some sparse set, then $\mathbf{P} = \mathbf{NP}$. Subsequently this result was improved by Ogihara and Watanabe [12] to the case of truth-table reducibility with constant queries, i.e.,

$$\mathbf{NP} \neq \mathbf{P} \Rightarrow \mathbf{NP} \not\subseteq \mathbf{P}_{btt}(\mathbf{SPARSE}).$$

Other subclasses of $\mathbf{P}_T(\mathbf{SPARSE})$ are obtained by considering $\mathbf{P}$-selective sets introduced by Selman [13]. A set A is $\mathbf{P}$-selective if there exists a polynomial time computable function that selects one of two given input strings such that if any one of the two strings is in A, then also the selected one. Let $\mathbf{SELECT}$ denote the class of $\mathbf{P}$-selective sets. Then we know the following facts:

1. (Selman and Ko (see [14])) $\mathbf{P}_T(\mathbf{SPARSE}) = \mathbf{P}_T(\mathbf{SELECT})$.
2. (Watanabe [17]) $\mathbf{P}_T(\mathbf{SELECT}) \not\subseteq \mathbf{P}_{tt}(\mathbf{SELECT})$.

Regarding our above question, the following results are known:

1. (Selman [13]) If $\mathbf{P} \neq \mathbf{NP}$, then $\mathbf{NP} \not\subseteq \mathbf{P}_m(\mathbf{SELECT})$.
2. (Agrawal and Arvind [1], Beigel, Kummer and Stephan [3], Ogihara [11]) If $\mathbf{P} \neq \mathbf{NP}$, then $\mathbf{NP} \not\subseteq \mathbf{P}_{n^\alpha\text{-}tt}(\mathbf{SELECT})$ for all $\alpha < 1$.

The condition $\alpha < 1$ in the item 2 seems difficult to be removed. In the following, we will remove this condition under a stronger, reasonable hypothesis. We show that

$$\mu_p(\mathbf{NP}) \neq 0 \Rightarrow \mathbf{NP} \not\subseteq \mathbf{P}_{tt}(\mathbf{SELECT}).$$

Many evidences have been presented by Lutz and Mayordomo [9] and Kautz and Miltersen [6] that this stronger hypothesis is reasonable. For example, the following results are known:

1. (Lutz and Mayordomo [8]) If $\mu_p(\mathbf{NP}) \neq 0$, then there exists an $\mathbf{NP}$ search problem which is not reducible to the corresponding decision problem.
2. (Lutz and Mayordomo [8]) If $\mu_p(\mathbf{NP}) \neq 0$, then the "Cook versus Karp-Levin" conjecture holds for $\mathbf{NP}$.
3. (Lutz and Mayordomo [9]) If $\mu_p(\mathbf{NP}) \neq 0$, then for every real number $\alpha < 1$, every $\leq^p_{n^\alpha\text{-}tt}$-hard language for $\mathbf{NP}$ is dense.
4. (Kautz and Miltersen [6]) For a Martin-Löf random language A, $\mu_p^A(\mathbf{NP}^A) \neq 0$.

We also give a partial affirmative answer to a conjecture by Beigel, Kummer and Stephan [3]. They conjectured that every $\leq^p_{tt}$-hard set for $\mathbf{NP}$ is $\mathbf{P}$-superterse unless $\mathbf{P} = \mathbf{NP}$. We will prove that every $\leq^p_{tt}$-hard set for $\mathbf{NP}$ is $\mathbf{P}$-superterse unless $\mathbf{NP}$ has p-measure 0.

We close this section by introducing some notation. N and $Q(Q^+)$ are the set of natural numbers and the set of (nonnegative) rational numbers, respectively.

$\Sigma = \{0,1\}$ is the binary alphabet, Σ^* is the set of (finite) binary strings, and Σ^n is the set of binary strings of length n. The length of a string x is denoted by $|x|$. $<$ is the length-lexicographical ordering on Σ^* and z_n $(n \geq 0)$ is the nth string under this ordering. λ is the empty string. For strings $x, y \in \Sigma^*$, xy is the concatenation of x and y.

A subset of Σ^* is called a language, a problem or simply a set. Capital letters are used to denote subsets of Σ^* and boldface capital letters are used to denote subsets of Σ^∞. The cardinality of a language A is denoted by $\|A\|$. We identify a language A with its characteristic function, i.e., $x \in A$ iff $A(x) = 1$. For a language $A \subseteq \Sigma^*$ and a string $x \in \Sigma^*$, $A\restriction x$ denotes the finite initial segment of A below x, i.e., $A\restriction x = \{y : y < x \ \& \ y \in A\}$, and we identify this initial segment with its characteristic string, i.e., $A\restriction z_n = A(z_0)\cdots A(z_{n-1}) \in \Sigma^*$.

We will use $\mathbf{P}$ and $\mathbf{E}_2$ to denote the complexity classes $DTIME(poly)$ and $DTIME(2^{poly})$, respectively.

2 Resource Bounded Measure and Polynomial Time Membership Comparable Sets

We first introduce a fragment of Lutz's effective measure theory which will be sufficient for our investigation.

Definition 2.1 *A* martingale *is a function $F : \Sigma^* \to R^+$ such that, for all $x \in \Sigma^*$,*

$$F(x) = \frac{F(x1) + F(x0)}{2}.$$

A martingale F succeeds *on a set A if $\limsup_n F(A\restriction z_n) = \infty$. $S^\infty[F]$ denotes the class of sets on which the martingale F succeeds.*

Definition 2.2 *(Lutz [7]) A class $\mathbf{C}$ of sets has p-*measure 0 *$(\mu_p(\mathbf{C}) = 0)$ if there is a polynomial time computable martingale $F : \Sigma^* \to Q^+$ which succeeds on every set in $\mathbf{C}$. The class $\mathbf{C}$ has p-*measure 1 *$(\mu_p(\mathbf{C}) = 1)$ if $\mu_p(\bar{\mathbf{C}}) = 0$ for the complement $\bar{\mathbf{C}} = \{A : A \notin \mathbf{C}\}$ of $\mathbf{C}$.*

We need the following two theorems by Lutz.

Theorem 2.3 *(Lutz [7]) Let $F : N \times \Sigma^* \to Q^+$ be a function such that*

1. *For all $k \in N$ and $x \in \Sigma^*$, $F(k,x)$ is computable in time polynomial in $k + |x|$.*

2. *For each* $k \in N$, $F_k(x) = F(k, x)$ *is a martingale.*

Then $\mathbf{C} = \{A : F_k$ *succeeds on* A *for some* $k \in N\}$ *has p-measure* 0.

Theorem 2.4 *(Lutz [7])* $\mu_p(\mathbf{E}_2) \neq 0$.

Jockusch [5] defined a set A to be *semirecursive* if there is a recursive function f such that for all x and y,

1. $f(x, y) \in \{x, y\}$.
2. If $\{x, y\} \cap A \neq \emptyset$, then $f(x, y) \in A$.

We call the function f a *selector* for A. Selman [13] considered a polynomial time version of semirecursive sets and defined a set A to be **P**-*selective* if A has a polynomial time computable selector. **P**-selective sets have been widely studied, see, e.g., [1, 3, 11].

For a set A, we identify A and its characteristic function. Let f be a selector for A. If f maps a pair (x, y) to y, then we have "$x \in A \to y \in A$", equivalently, "$A(x)A(y) \neq 10$". Thus we can view a selector for A as a function f that maps (x, y) to $z = 01$ or 10. Here we require $z \in \{01, 10\}$. One natural extension is to remove this condition, i.e., let $z \in \Sigma^2$.

Definition 2.5 *(Beigel [3]) A set* A *is* **P**-approximable *if there is a* $k \in N$ *and a polynomial time computable function* $f : \prod_{i=0}^{k-1} \Sigma^* \to \Sigma^k$ *such that for all* $x_0, \cdots, x_{k-1} \in \Sigma^*$, $f(x_0, \cdots, x_{k-1}) \neq A(x_0) \cdots A(x_{k-1})$. *A set* A *is* **P**-superterse *if and only if* A *is not* **P**-*approximable.*

Note that the above definition of **P**-superterseness is a little different from Beigel's [2] original definition. Ogihara [11] further introduced the following notion of polynomial time membership comparability.

Definition 2.6 *(Ogihara [11]) Let* $g : N \to N^+$ *be a monotonic, nondecreasing, polynomial time computable and polynomial bounded function.*

1. *A function* f *is called a* g-membership comparing function *(a* g-mc-function *for short) for* A *if for every* $x_0, \cdots, x_{m-1}$ *with* $m \geq g(\max\{|x_0|, \cdots, |x_{m-1}|\})$,

$$f(x_0, \cdots, x_{m-1}) \in \Sigma^m \text{ and } A(x_0) \cdots A(x_{m-1}) \neq f(x_0, \cdots, x_{m-1}).$$

2. *A set* A *is polynomial time* g-membership comparable *if there exists a polynomial time computable* g-*mc-function for* A.

3. **P**-$mc(g)$ *denotes the class of all polynomial time* g*-membership comparable sets.*

Theorem 2.7 *(Ogihara [11])* $\mathbf{P}_{tt}(\mathbf{SELECT}) \subseteq \mathbf{P}\text{-}mc(\mathbf{LOG})$*, where* $\mathbf{LOG} = \{c \log : c > 0\}$.

Theorem 2.8 *(Ogihara [11])* $\mathbf{P}\text{-}mc(\mathbf{LOG}) \subset \mathbf{P}\text{-}mc(n)$.

The following proposition is obvious.

Proposition 2.9 *1. If* A *is* **P**-*selective, then* A *is* **P**-*approximable.*

2. If A *is* **P**-*approximable, then* $A \in \mathbf{P}\text{-}mc(c)$ *for some constant* $c \in N$*; Moreover,*

$$\mathbf{P}\text{-}appro = \cup_{c \in N} \mathbf{P}\text{-}mc(c),$$

where **P**-*appro is the class of* **P**-*approximable sets.*

The next proposition gives an important property of **P**-approximable sets which we need latter. If A is **P**-approximable then, for strings $x_0, \cdots, x_{s-1} \in \Sigma^*$, we can compute in polynomial time a subset of Σ^s which contains $A(x_0) \cdots A(x_{s-1})$.

Proposition 2.10 *(Beigel [2]) If* A *is* **P**-*approximable via* $k \in N$*, then there is a polynomial time computable function which computes for any* s *strings* $x_0, \cdots x_{s-1}$ *a set of at most*

$$S(s,k) = \binom{s}{0} + \binom{s}{1} + \cdots \binom{s}{k-1}$$

elements from Σ^s *which contains* $A(x_0) \cdots A(x_{s-1})$*. For a fixed* k*,* $S(s,k)$ *is a polynomial in* s *of degree* $k-1$.

Let $\mathbf{P}_{tt}$(**P**-*appro*) be the class of sets which can be $\leq^p_{tt}$-reduced to some **P**-approximable sets. Then we have the following theorem.

Theorem 2.11 $\mathbf{P}_{tt}(\mathbf{P}\text{-}appro) \subseteq \mathbf{P}\text{-}mc(n)$.

Remark: Theorem 2.11 is actually a corollary of Corollary 2.7 in Beigel et al. [3]. For the reason of completeness, we will give the proof here. The idea underlying the following proof is the same as that underlying the proof of Theorem 3.3 in Ogihara [11].

Proof. Let A be a **P**-approximable set via $k \in N$, and let $L \leq^p_{tt} A$ via a machine M. Assume that the number of queries in the reduction $L \leq^p_{tt} A$ is bounded by a polynomial f. Given $n \in N$ and $x_0, \cdots x_{n-1} \in \Sigma^*$

such that $n \geq \max\{|x_0|, \cdots, |x_{n-1}|\}$, for each $i < n$, let Q_i denote the set of queries of M on x_i, and $Q = Q_0 \cup \cdots \cup Q_{n-1}$. Then $\|Q_i\| \leq f(n)$ and, for sufficiently large n,

$$\|Q\|^k \leq (nf(n))^k \leq 2^n.$$

By Lemma 2.10, we can compute, in time polynomial in $\sum_{y \in Q} |y|$, and thus, in time polynomial in n, a set of at most $\|Q\|^k$ elements which contains the characteristic sequence of A on domain Q. So we can compute in time polynomial in $|x|$ a sequence $g(x_0, \cdots, x_{n-1}) \in \Sigma^n$ such that $L(x_0) \cdots L(x_{n-1}) \neq g(x_0, \cdots, x_{n-1})$. I.e., g witnesses that $L \in \mathbf{P}\text{-mc}(n)$. ■

In order to prove our main theorem, we prove a lemma at first.

Lemma 2.12 *Let $1 < n_1, n_2, \cdots$ be a sequence of numbers such that for all i, $n_{i+1} \leq n_i + \log n_i$. Then $\lim_{n\to\infty} \prod_{i=1}^{n}(1 + \frac{1}{n_i}) = \infty$.*

Proof. By a simple induction, it is easy to check that, for almost all i, $n_i \leq i \log i \log\log i$. Hence

$$\lim_{n\to\infty} \prod_{i=1}^{n} \left(1 + \frac{1}{n_i}\right) \geq \lim_{n\to\infty} \prod_{i=1}^{n} \left(1 + \frac{1}{i \log i \log\log i}\right) = \infty.$$

■

Theorem 2.13 **P**-*mc(n) has p-measure 0, i.e., $\mu_p(\mathbf{P}\text{-mc}(n)) = 0$.*

Proof. Let $f_0, f_1, \cdots$ be an enumeration of all polynomial time computable functions.

For each $k \in N$, define a martingale F_k as follows. Let $n_i = i$ for $i \leq 5$ and $n_{i+1} = n_i + [\log n_i]$ for $i > 5$. For $|x| \leq n_5$, let $F_k(x) = 1$. For $x \in \Sigma^{n_{i+1}}$ $(i \geq 5)$, fix the initial segment $y \in \Sigma^{n_i}$ of x and let

$$F_k(x) = \begin{cases} \left(1 + \frac{1}{2^{[\log n_i]} - 1}\right) F_k(y) & \text{if } x \neq y f_k(z_{n_i}, \cdots, z_{n_{i+1}-1}) \\ 0 & \text{if } x = y f_k(z_{n_i}, \cdots, z_{n_{i+1}-1}) \end{cases}$$

And, for $x \in \Sigma^*$ such that $|x| \neq n_i$ $(i \in N)$, let

$$F_k(x) = \frac{F_k(x0) + F_k(x1)}{2}.$$

Now we show that for any set A, if f_k witnesses that A is n-membership comparable, then F_k succeeds on A. Obviously, for $i \geq 5$,

$$F_k(A\restriction z_{n_{i+1}}) = \left(1 + \frac{1}{2^{[\log n_5]} - 1}\right) \cdots \left(1 + \frac{1}{2^{[\log n_i]} - 1}\right) \geq \left(1 + \frac{1}{n_5}\right) \cdots \left(1 + \frac{1}{n_i}\right).$$

By Lemma 2.12, $\limsup_i F_k(A\restriction z_{n_i}) = \infty$, i.e., F_k succeeds on A.

It is straightforward that for all $k \in N$ and $x \in \Sigma^*$, $F_k(x)$ is computable in time polynomial in $k + |x|$. Hence $\mu_p(\mathbf{P}\text{-mc}(n)) = 0$. ■

By combining Theorem 2.4 and Theorem 2.13, we get

Theorem 2.14 $\mathbf{E}_2 \not\subseteq \mathbf{P}_{tt}(\mathbf{P}\text{-}appro)$.

Corollary 2.15 *(Toda [15])* $\mathbf{E}_2 \not\subseteq \mathbf{P}_{tt}(\mathbf{SELECT})$.

Note that Toda proved Corollary 2.15 using a direct diagonalization. The importance of Theorem 2.13 is that it has implications on the structure of **NP**. By combining Theorem 2.13 and Theorem 2.11, we get

Theorem 2.16 *If* **NP** *does not have p-measure* 0*, then no* **P***-approximable set is* $\leq^p_{tt}$*-hard for* **NP***, i.e., every* $\leq^p_{tt}$*-hard set for* **NP** *is* **P***-superterse unless* $\mu_p(\mathbf{NP}) = 0$.

Corollary 2.17 *If* **NP** *does not have p-measure* 0*, then no* **P***-selective set is* $\leq^p_{tt}$*-hard for* **NP**.

Theorem 2.16 gives a partial affirmative answer to the conjecture of Beigel, Kummer and Stephan. Note that our hypothesis $\mu_p(\mathbf{NP}) \neq 0$ is a reasonable scientific hypothesis (see Lutz and Mayordomo [9]). It is worthwhile to mention that, in the above, we used the uniform constructive method initiated by Lutz and Mayordomo [9]. At first, we proved Theorem 2.13, a measure-theoretic result concerning the quantitative structure of $\mathbf{E}_2$, and then get the qualitative separation result: Theorem 2.14. More precisely, the proof of Theorem 2.14 consists of the following two components:

1. Prove that $\mu_p(\mathbf{P}_{tt}(\mathbf{P}\text{-}appro)) = 0$.
2. The measure conservation theorem: Theorem 2.4.

One of the important feature of this method is that it gives an automatic witness for the qualitative separation. For example, in our setting, by Theorem 2.13, for large enough k, every n^k-random language A is not $\leq^p_{tt}$-reducible to any **P**-approximable set.

References

[1] M. Agrawal and V. Arvind. Polynomial time truth-table reductions to **P**-selective sets. In *Proc. 9th Conf. on Structure in Complexity Theory*, pages 24–30. IEEE Computer Society Press, 1994.

[2] R. Beigel. *Query-limited Reducibilities*. PhD thesis, Stanford University, 1987.

[3] R. Beigel, M. Kummer, and F. Stephan. Approximable sets. In *Proc. 9th Conf. on Structure in Complexity Theory*, pages 12–23. IEEE Computer Society Press, 1994.

[4] L. Hemaspaandra and Z. Jiang. **P**-selectivity: Intersections and indices. *Theoretical Computer Science*, 145:371–830, 1995.

[5] C. Jockusch. Semirecursive sets and positive reducibility. *Trans. Amer. Math. Soc.*, 131:420–436, 1968.

[6] S. Kautz and P. Miltersen. Relative to a random oracle, **NP** is not small. In *Proc. 9th Conf. on Structure in Complexity Theory*, pages 162–174. IEEE Computer Society Press, 1994.

[7] J. H. Lutz. Almost everywhere high nonuniform complexity. *J. Comput. System Sci.*, 44:220–258, 1992.

[8] J. H. Lutz and E. Mayordomo. Cook versus Karp-Levin: separating completeness notions if **NP** is not small. In *Proc. 11th STACS*, Lecture Notes in Comput. Sci., 775, pages 415–426. Springer Verlag, 1994.

[9] J. H. Lutz and E. Mayordomo. Measure, stochasticity, and the density of hard languages. *SIAM J. Comput.*, 23:762–779, 1994.

[10] S. R. Mahaney. Sparse complete sets for **NP**: Solution of a conjecture of Berman and Hartmanis. *J. Comput. System Sci.*, 25:130–143, 1982.

[11] M. Ogihara. Polynomial-time membership comparable sets. *SIAM J. Comput.*, 24:1068–1081, 1995.

[12] M. Ogihara and O. Watanabe. On polynomial bounded truth-table reducibility of **NP** sets to sparse sets. *SIAM J. Comput.*, 20:471–483, 1991.

[13] A. Selman. **P**-selective sets, tally languages, and the behavior of polynomial time reducibilities on **NP**. *Math. System Theory*, 13:55–65, 1979.

[14] T. Thierauf, S. Toda, and O. Watanabe. On sets bounded truth-table reducible to **P**-selective sets. In *Proc. 11th STACS*, Lecture Notes in Comput. Sci., 775, pages 427–438. Springer Verlag, 1994.

[15] S. Toda. On polynomial time truth-table reducibilities of intractable sets to **P**-selective sets. *Math. System Theory*, 24:69–82, 1991.

[16] Y. Wang. *Randomness and Complexity*. Ph.D thesis, Universität Heidelberg, 1995. Available on request or from my WWW homepage `http://math.uni-heidelberg.de/logic/wang/wang.html`.

[17] O. Watanabe. reported in [14].

[18] C. B. Wilson. Relativized circuit complexity. *J. Comput. System Sci.*, 31:169–181, 1985.

[19] M. Zimand. On the size of sets with weak membership properties. TR-557, Rochester University, 1994.