KOLMOGOROV COMPLEXITY, RESTRICTED NONDETERMINISM AND GENERALIZED SPECTRA

by

Deborah Joseph and Meera Sitharam

Computer Sciences Technical Report #898

December 1989

KOLMOGOROV COMPLEXITY, RESTRICTED NONDETERMINISM AND GENERALIZED SPECTRA

DEBORAH JOSEPH[1]
University of Wisconsin - Madison

MEERA SITHARAM
University of Wisconsin - Madison

Abstract. This paper uses the technique of generalized spectra and expressibility of complexity classes in logic, developed by Fagin and Immerman, to give alternate characterizations of specific subclasses of NP. These characterizations serve to unify concepts that appear in seemingly different areas of complexity theory; namely, the restricted nondeterminism of Kintala and Fischer and the time bounded Kolmogorov complexity of Daley and Ko. As consequences of these characterizations we show that relatively easy subsets of $NP - P$ can not be pseudorandomly generated, unless $UTIME[t(n)] = DTIME[t(n)]$ for certain exponential functions t. Secondly, we show that no easy subset of the set of all satisfying assignments of satisfiable $g(n)$-easy formulas contains an assignment for each of these formulas, unless $NEXPTIME = EXPTIME$. The latter partially answers a question raised by Hartmanis.

1. INTRODUCTION.

In this paper we use the technique of generalized spectra and expressibility of complexity classes in logic, developed by Fagin and Immerman ([Fa 74] and [Im 82,87]), to give alternate characterizations of subclasses of NP. Our characterizations serve to unify concepts that appear in seemingly different areas of complexity theory, namely restricted nondeterminism of Kintala and Fischer ([KiFi 80], [ADT 89]), and time bounded Kolmogorov complexity of Daley and Ko ([Da 77], [Ko 83]).

In addition, we can use these characterizations to draw conclusions about the Kolmogorov complexity of certain sets. More specifically, our results show that even highly restricted sets in NP do not consist entirely of Kolmogorov-easy strings unless higher complexity classes collapse. This relates to a result of Hartmanis ([HaYe 83]) that sets in P cannot separate SAT from the set of Kolmogorov-easy strings in SAT unless higher complexity classes collapse. Furthermore, since the outputs of pseudorandom generators are Kolmogorov-easy, our results establish the exact complexity of pseudorandomly generated sets. This relates to recent work in [Ya 82], [Al 88], and [NiWi 88] that can

---

[1] This work was supported in part by a National Science Foundation Presidential Young Investigator Award.

Authors' address: Computer Sciences Department, University of Wisconsin, 1210 West Dayton St., Madison, WI 53706, USA.

be viewed as a study of the approximability of pseudorandomly generated sets. On a different level, our results provide a comparison between the power of existential second order explicit definitions and implicit definitions, and thus contribute to the study of expressibility in logic.

Our primary results are the following.

Result 1: We give a logical characterization of Kintala and Fischer's restricted nondeterministic classes, $P_{g(n)}$.

The class $P_{g(n)}$ consists of sets accepted by NP machines that make at most $g(n)$ nondeterministic moves, where n is the length of the input and $g(n)$ is a sublinear function of n. We use the notion of generalized spectra to give an alternate characterization of the classes $P_{g(n)}$. Just as Fagin ([Fa 74]) showed that the sets in NP are exactly the generalized spectra (or, the set of finite models) that satisfy a second order existential ($2^{nd}O\exists_{g(n)}$) sentence, we show that that by restricting the second order quantifier in $2^{nd}O\exists$ sentences we can obtain a class $2^{nd}O\exists_{g(n)}$ that characterizes $P_{g(n)}$.

Kintala and Fischer ([KiFi 80]) introduced these subclasses of NP as a way to study the fine structure of $NP - P$. They constructed oracles that separate the classes $P_{g(n)}$ for functions g that have different growth rates, and argued that the number of nondeterministic guesses is a resource that is independent of the number of steps of an NP computation. Renewed interest in these classes came when Stearns and Hunt ([StHu 86]) related them to the classification of sets in NP based on the (sub) exponential complexities of their deterministic algorithms. More recently, Àlvarez, Díaz, and Torán ([ADT 89]) exhibited natural self-reducible complete problems for the classes $P_{log^j(n)}$ for fixed $j > 1$. In addition, they showed that the classes $P_{log^j(n)}$ have many structural properties similar to those of NP.

Result 2: We give a logical characterization of the uniform subsets of the time bounded Kolmogorov complexity classes, $KT[g(n), n^k]$.

The time bounded Kolmogorov complexity classes, $KT[g(n), n^k]$, were introduced by Daley and Ko ([Da 77], [Ko 83]). Intuitively, the class $KT[g(n), n^k]$ consists of "$g(n)$-easy" strings, i.e, the information content of the string is efficiently retrievable from a compressed string of length $g(n)$, where $g(n)$ is sublinear. We characterize the complexity of uniform subsets of $KT[g(n), n^k]$, by restricting the deterministic checking part of NP computations. In particular, we additionally restrict the first order formulas in $2^{nd}O\exists_{g(n)}$ sentences to be explicit definitions and denote the class of these restricted sentences as $2^{nd}O\exists^E_{g(n)}$.

The above results can be viewed as characterizations of subclasses of NP, obtained by restricting the second order existential quantifier and the first order formula in $2^{nd}O\exists$ sentences. It is worth noting that classes obtained by restricting the second order quantifier have been studied earlier by Fagin ([Fa 74,75]) and Lynch ([Ly 82]). Fagin showed that $2^{nd}O\exists$ sentences, in which the arity of the second order relational variable is bounded by $2k$, characterize sets in $NTIME[n^k]$. Lynch showed that the same restriction on $2^{nd}O\exists$ sentences, in the language of $+$, characterizes $NTIME[n^{2k}]$. Subclasses of NP

obtained by restricting the first order formula in $2^{nd}O\exists$ sentences have also been studied earlier. Papadimitriou and Yannakakis ([PaYa 88]) considered the subclass, SNP, of NP, obtained by restricting the first order formula to be universal, i.e. in the prefix class $\forall$. They showed that a corresponding optimization class $MAXSNP$ is easily approximable. However, none of these restrictions differentiate between the number of nondeterministic moves, and the number of steps in an NP computation.

Our next result relates a question in second order logic to a problem in complexity theory.

Result 3: We show that existential second order implicit definitions have more power than explicit definitions, unless higher complexity classes collapse.

Here we compare $2^{nd}O\exists^{E}_{g(n)}$ with another subclass of NP, the class $2^{nd}O\exists^{Implicit}_{g(n)}$, or $2^{nd}O\exists^{I}_{g(n)}$. This class is obtained by restricting the first order formulas in $2^{nd}O\exists_{g(n)}$ sentences to implicit rather than explicit definitions. Intuitively, a set L in $2^{nd}O\exists^{I}_{g(n)}$ has the property that any string w can witness at most one string x in L. If we additionally require x to be polynomially computable from w, then L is in $2^{nd}O\exists^{E}_{g(n)}$. We give evidence that the strings in L may not be polynomially computable from their short witnesses, and thus they may not be $g(n)$-easy. In particular, we show that every set, in $2^{nd}O\exists^{I}_{g(n)}$ is "strongly equivalent" to a set in $2^{nd}O\exists^{E}_{g(n)}$, i.e, the two sets are equal *and* their witness sets are identical, if and only if, certain higher deterministic complexity classes are equal, to their corresponding unambiguous nondeterministic ($UTIME$) classes.

The question of implicit versus expicit definability over finite structures has been studied earlier by Gurevich ([Gu 84]). He showed that implicit definability in fixpoint logic (first order logic with the fixpoint operator) has more power than explicit definability unless $P = UP \cap coUP$. More recently, Kolaitis [Kol 89] strengthened Gurevich's result by showing that formulas that are implicitly definable in fixpoint logic are in fact implicitly definable by a *pair* of formulas in first order logic.

We note that Grollmann and Selman ([GrSe 84]) have shown that if one-way functions exist, then $P \neq UP$. To explain the similarity between the consequence in Grollmann and Selman's result, and those mentioned in the previous paragraphs, we recall some details of Grollmann and Selman's result. Their proof proceeds by showing that if there are functions that are not polynomially computable, but whose graphs are in P, then $P \neq UP$. We will see that this latter assumption is intuitively similar to assumptions about the equivalence of implicit and explicit definabilities.

From the above results we can draw two interesting conclusions. The first shows that the relationship between implicit and explicit definitions is not merely a question of interest to logicians: it can be used to define the exact complexity of the ranges of pseudorandomly generated sets.

Conclusion 1: Relatively easy sets in $NP-P$ can not be pseudorandomly generated, unless certain deterministic and nondeterministic complexity classes collapse.

To explain further, we will see that the sets in $2^{nd}O\exists^I_{g(n)}$ are highly restricted, and are hence relatively easy sets in $NP-P$, and that sets in $2^{nd}O\exists^E_{g(n)}$ are exactly the ranges of pseudorandom generators. Hence, the assumption that relatively easy sets in $NP-P$ can be pseudorandomly generated is equivalent to the hypothesis of Result 1, and yields the same consequence.

This relates to recent work by Allender in [Al 88] and Nisan and Widgerson in [NiWi 88] on the consequences of the existence of pseudorandom generators. In contrast, their results can be viewed as defining the complexity of sets that can approximate the ranges of pseudorandom generators.

Our second conclusion gives a partial answer to a question posed by Hartmanis in [Ha 83]: Do all $log(n)$-easy satisfiable formulas have at least one $log(n)$-easy satisfying assignment?

Conclusion 2: No easy subset of the set of all satisfying assignments of $g(n)$-easy formulas contains an assignment for each $g(n)$-easy satisfiable formula, unless certain deterministic and nondeterministic complexity classes collapse.

In the next sections we discuss our results in more detail and relate them to other work on the restricted nondeterministic classes, time bounded Kolmogorov complexity, pseudorandom generators and generalized spectra. In a concluding section we present some interesting open problems related to this work.

2. CHARACTERIZING THE CLASSES $P_{g(n)}$.

To give a logical characterization of the classes $P_{g(n)}$ we first require some background and definitions concerning these classes and the use of expressibility in logic as a complexity measure.

In the study of nondeterministic computations one commonly equates the number of nondeterministic moves with the number of steps of the computation. However, Kintala and Fischer ([KiFi 80]) began work aimed at classifying NP machines based on the number of "strict" c-ary nondeterministic moves; i.e., moves for which there are at least $c \geq 2$ choices for the next instantaneous description of the machine. They defined the classes $P_{g(n)}$ as follows.

Definition. [KiFi 80] *For any function $g(n)$, the class $P_{g(n)}$ consists of sets that are accepted by a polynomial time Turing machine that makes at most $g(n)$ c-ary nondeterministic moves on inputs of size n, for some constant c.*

Kintala and Fischer's motivation for introducing the classes $P_{g(n)}$ arose from the observation that most known NP-complete sets can be recognized by machines that make a linear number of nondeterministic moves. This led them to ask which languages can be recognized using sublinear functions, for example $g(n) = log^j n$ or $g(n) = n^{1/j}$,

for fixed $j > 1$, and thereby investigate the fine structure of $NP - P$. A second motivation for studying these classes arises from later work of Stearns and Hunt ([StHu 86]) that classifies sets in NP based on the complexity of deterministic simulations of their NP algorithms. For sublinear functions g the sets in $P_{g(n)}$ have deterministic algorithms that run in subexponential time. A third motivation for studying these classes arises from a more recent work of Àlvarez, Díaz, and Torán. They showed that the structural properties of the classes $P_{log^j n}$ are similar to those of NP, for instance the existence of natural self-reducible complete sets.

Kintala and Fischer posed the questions: Is $P_{log^j n} = P_{log^{j+1} n}$, and is $P_{log^j n}$ closed under complement? They provided evidence that answering these questions will be difficult by proving relativized separation and equivalence results. From this they argued that the number of nondeterministic moves is a resource that is independent of the number of steps in an NP computation.

We will characterize the classes $P_{g(n)}$ by defining the logics that express them. For this we assume the readers' acquaintance with the basic notions of expressibility in first and second order logic, and use the notation of Immerman's survey paper [Im 87]. In our discussion, we deal with finite structures that represent Boolean strings. I.e, we consider structures $\langle A, B^1 \rangle$ where the universe, A, is linearly ordered, and B^1 is a one-place relation symbol. We refer to these as *input* structures of length n, where $|A| = n$. We consider *generalized spectra* (or, sets of input structures) that satisfy sentences in different logics. The *generalized spectrum* of a sentence ϕ is the set

$$\{\langle A, B^1 \rangle : \langle A, B^1 \rangle \models \phi\};$$

i.e., the set of input structures that satisfy ϕ, where the semantic notion of satisfaction is defined in the usual manner.

We deal with sentences in first order logic, FO, a logic with the *least fix point* operator applied to first order formulas, $FO + LFP$, and existential second order logic, $2^{nd}O\exists$. Sentences in the latter logic are of the form: $\exists W^k \phi(W^k)$, where ϕ is a first order formula and W^k is a relational variable of arity k. The following are some of the characterizations of complexity classes given by Immerman and Fagin.

Theorem. [Im 82], [Fa 74]

1. *P is the class of spectra of $FO + LFP$ sentences.*
2. *NP is the class of spectra of $2^{nd}O\exists$ sentences.*

We are now ready to characterize the classes $P_{g(n)}$. Throughout this section, we will assume that $g(n)$ is either $log^j n$ or $n^{1/j}$ for some $j \geq 1$; however, our results extend to other nicely behaved sublinear functions. We begin by defining a restricted second order existential quantifier that will, intuitively, quantify over encodings of short sequences of nondeterministic moves. That is, the quantifiers "$\exists_{g(n)}$" semantically quantify over monadic relations (of arity one) that are defined on a fixed $g(n)$-sized subset of the input structures' universe.

Definition. *Let $\langle A, B^1 \rangle$ be a structure such that $|A| = n$, let ϕ be any formula in $FO + LFP$, and let $g(n)$ be $log^j n$ or $n^{1/j}$, $j > 1$.*

$$\langle A, B^1 \rangle \models \exists_{g(n)} W^1 \phi(W^1) \iff \exists W^1 [\forall x \, [x > g(n) \Rightarrow \neg W^1(x)] \wedge \phi(W^1)].$$

The class of sentences of the form $\exists_{g(n)} W^1 \phi(W^1)$, where ϕ is a formula in $FO + LFP$, is denoted as $2^{nd} O \exists_{g(n)}$.

We now have a characterization of $P_{g(n)}$.

Theorem 1. *Let $g(n)$ be $log^j n$ or $n^{1/j}$, $j > 1$. Then, $P_{g(n)}$ is the class of languages that correspond to the generalized spectra of $2^{nd} O \exists_{g(n)}$ sentences.*

Proof. The containment $2^{nd} O \exists_{g(n)} \subseteq P_{g(n)}$ is clear. To prove that $P_{g(n)} \subseteq 2^{nd} O \exists_{g(n)}$ we consider a set S in $P_{g(n)}$ and a nondeterministic polynomial time Turing machine, N, for S that consists of a deterministic machine, M, an input tape y of length n and a guess tape W of length $g(n)$. The machine M takes the tuple $\langle y, W \rangle$ as input, runs in time n^k for some k and N accepts y, if and only if, M accepts $\langle y, W \rangle$ for some W. Without loss of generality, the inputs y are structures of the form $\langle A, B^1 \rangle$, where $|A| = n$. The inputs to the machine M are structures, $\langle A, B^1, W^1 \rangle$, where W^1 is a monadic relation whose domain is restricted to the smallest $g(n)$ elements of A, and encodes the Boolean string on the guess tape W. Let ϕ be the $FO + LFP$ sentence whose generalized spectrum is the set of structures accepted by M. Thus, N accepts $\langle A, B^1 \rangle$, if and only if, for some relation W^1, $\langle A, B^1, W^1 \rangle \models \phi$. That is, the set of structures accepted by N is the generalized spectrum of the $2^{nd} O \exists_{g(n)}$ sentence $\exists_{g(n)} W^1 \phi(W^1)$. ∎

It is important to note that in the above proof we require ϕ to be a $FO + LFP$ sentence. In contrast, in his proof that NP is $2^{nd} O \exists$, Fagin ([Fa 74]) just required that ϕ be an FO sentence. To elaborate: although sentences of the form $\exists W^k \phi(W^k)$, where $\phi(W^k)$ is a formula in $FO + LFP$, are equivalent to sentences of the form $\exists V^j \psi(V^j)$, where $\psi(V^j)$ is a formula in FO, the arity of the second order variable increases in the rewriting, i.e., $j > k$. This is because the relational variable V^j ranges over the encodings of entire polynomial time computations that check the $FO + LFP$ formula $\phi(W^k)$. Thus, the arity j depends on the deterministic time complexity of the generalized spectrum of ϕ. When we restrict the arity and domain of the second order relational variable, W^1, as in $2^{nd} O \exists_{g(n)}$ sentences, it is not clear that a sentence, $\exists_{g(n)} W^1 \phi(W^1)$, where ϕ is in $FO + LFP$, is equivalent to a sentence $\exists_{g(n)} V^1 \psi(V^1)$, where ψ is in FO.

3. CHARACTERIZING THE CLASSES $KT[g(n), n^k]$.

In this section we introduce the logic of *existential second order explicit definitions,* and use it to characterize the complexity of pseudorandomly generated sets and uniform subsets of the time bounded Kolmogorov complexity classes. We begin with a brief discussion of time bounded Kolmogorov complexity.

The time bounded Kolmogorov complexity classes, $KT[g(n), t(n)]$ were introduced by Daley ([Da 77]), and Ko ([Ko 83]).[2] Intuitively, a string y, of length n, is in the class $KT[g(n), t(n)]$ if it can be generated from a string of length $g(n)$ in $t(n)$ steps.[3] More formally,

Definition. *Let M_u be a universal Turing machine. The class $KT[g(n), t(n)] =_{def}$*

$\{y : |y| = n$ *and* $\exists w$ $[|w| \leq g(n)$ *and* $M_u(w) = y$ *and* M_u *halts in at most* $t(n)$ *steps*$\}$

The notion of time bounded Kolmogorov complexity has applications in the theory of pseudorandom number generators. Pseudorandom generators are polynomially computable functions that typically map short seeds, say of length $n^{1/j}$, to longer pseudorandom strings of length n. Hence, the range, or set of outputs, of a pseudorandom generator is a *uniformly generated* subset of $n^{1/j}$-easy strings.

We can use a version of second order logic, which we will denote $2^{nd}O\exists_{g(n)}^{Explicit}$, to characterize the exact complexity of recognizing pseudorandomly generated sets. $2^{nd}O\exists_{g(n)}^{Explicit}$ is obtained by restricting the syntactic complexity of the *FO+LFP* formula, $\phi(W^1)$, in a $2^{nd}O\exists_{g(n)}$ sentence $\exists_{g(n)}W^1\phi(W^1)$ in the following manner. For any relation W^1 on a universe A we demand that $\phi(W^1)$ *explicitly* define a unique relation B^1 such that $\langle A, B^1, W^1\rangle \models \phi$.

Definition. *A FO+LFP formula $\phi(W^1)$ explicitly defines the relation symbol B^1 if $\phi(W^1)$ is of the form:*

$$\forall x[\sigma(W^1, x) \iff B^1(x)],$$

where $\sigma(W^1, x)$ is a FO+LFP formula that does not contain the relational variable B^1. The logic $2^{nd}O\exists_{g(n)}^{Explicit}$, also denoted $2^{nd}O\exists_{g(n)}^{E}$, consists of sentences of the form $\exists_{g(n)}W^1\phi(W^1)$, where $\phi(W^1)$ explicitly defines the relational variable B^1 of the input structures $\langle A, B^1\rangle$.

The next theorem shows that the logic $2^{nd}O\exists_{g(n)}^{E}$ characterizes uniformly generated subsets of $g(n)$-easy strings; i.e., it characterizes pseudorandomly generated sets.

Theorem 2.

1. *A set L is the spectrum of a $2^{nd}O\exists_{g(n)}^{E}$ sentence if and only if L is the set of outputs of a pseudorandom generator that generates strings of length n from random seeds of length $g(n)$, in time polynomial in n.*
2. *If a set L is the spectrum of a $2^{nd}O\exists_{g(n)}^{E}$ sentence, then $L \subseteq KT[g(n) + c, n^k]$, for some constant k.*

Proof. The proof of (2) follows immediately from (1).

---

[2] These citations refer to the first uses of resource bounded Kolmogorov complexity *within* complexity theory. However, we note that the first known result about time bounded Kolmogorov complexity was proved by Barzdin in [Ba 68].

[3] We will not use Levin's version of this notion, commonly referred to as time limited Kolmogorov complexity ([Le 64]). Interested readers are referred to surveys by Longpré ([Lo 86]), and Li and Vitanyí ([LiVi 88]), for definitions and applications of different variants of Kolmogorov complexity.

The forward direction of (1) will follow from the fact that any set, L, which is characterized by a $2^{nd}O\exists^{E}_{g(n)}$ formula, is an NP set with the following additional properties.

a. Each string $\langle A, B^1\rangle$ of length n in L has a short witness string W^1 of length $g(n)$.
b. Any string W^1 witnesses at most one string in L.
c. Each string $\langle A, B^1\rangle$ in L can be generated from its witness string W^1 in time polynomial n.

Suppose that L is the set of input structures, $\langle A, B^1\rangle$, that satisfy a sentence, $\exists_{g(n)}W^1\phi(W^1)$, where $\phi(W^1)$ is of the form:

$$\forall x[\sigma(W^1, x) \iff B^1(x)]$$

for some $FO{+}LFP$ formula, $\sigma(W^1, x)$, that contains the only the relation symbol W^1. In addition assume that the generalized spectrum of ϕ is in $DTIME[n^k]$ and that Q is a program that checks ϕ given the input structure $\langle A, B^1, W^1\rangle$, where $|A| = n$. Without loss of generality, Q consists of a program Q' of size c that, on input $\langle A, W^1\rangle$ of size $g(n)$, generates a structure $\langle A, \sigma^1\rangle$, followed by a program Q'' that checks if $\langle A, \sigma^1\rangle = \langle A, B^1\rangle$. Q' runs in at most n^k steps, and is the required pseudorandom generator.

To prove the reverse direction of (1), we adapt the proof of Theorem 1: we consider the nondeterministic Turing machine N that accepts L. The machine N consists of an input tape y of length n, a guess tape W of length $g(n)$, and the pseudorandom generator M of the hypothesis that takes W as input, and outputs the string σ of length n, in n^k steps. The machine N accepts y if and only if $\sigma = y$. As in Theorem 1, the inputs y are structures of the form $\langle A, B^1\rangle$, the inputs to M are structures of the form $\langle A, W^1\rangle$, and the outputs of M are structures of the form $\langle A, \sigma^1\rangle$. Thus the set, L accepted by N is the generalized spectrum of a sentence,

$$\exists_{g(n)}W^1\forall x\ [\sigma(W^1, x) \iff B^1(x)],$$

and the set of structures, $\langle A, B^1, W^1\rangle$, that satisfy the sentence $\forall x\ [\sigma(x) \iff B^1(x)]$ is in $DTIME[n^k]$, since given $\langle A, B^1\rangle$ and $\langle A, \sigma^1\rangle$, checking the sentence $\forall x\ [\sigma(x) \iff B^1(x)]$ can be done in n steps. ∎

4. THE CLASSES $2^{nd}O\exists^{Implicit}_{g(n)}$.

Our third result requires that we distinguish between explicit and implicit definitions in existential second order logic. In the previous section, we saw that the spectra of $2^{nd}O\exists^{E}_{g(n)}$ sentences consist of elements that can be polynomially generated from their short witnesses. Here we will see that the spectra of $2^{nd}O\exists^{Implicit}_{g(n)}$ sentences, consist merely of elements with short "exclusive" witnesses, i.e., each string witnesses at most one element in such sets. In this sense, the spectra of $2^{nd}O\exists^{I}_{g(n)}$ sentences are highly restricted NP sets, and can be considered as relatively easy sets in $NP - P$. Below, we formally define the logic $2^{nd}O\exists^{I}_{g(n)}$.

Definition. *A FO+LFP formula, $\phi(W^1)$ implicitly defines the relational variable B^1 if for every relation W^1 and pair of structures $\langle A, B^1, W^1\rangle$ and $\langle A, C^1, W^1\rangle$ that satisfy ϕ, the relations B^1 and C^1 are equivalent. I.e., the structure $\langle A, B^1, C^1, W^1\rangle \models \forall x[B^1(x) \iff C^1(x)]$. The logic $2^{nd}O\exists^{Implicit}_{g(n)}$, (also denoted $2^{nd}O\exists^{I}_{g(n)}$) consists of sentences of the form $\exists_{g(n)}W^1\phi(W^1)$, where $\phi(W^1)$ implicitly defines the relational variable B^1 of the input structures $\langle A, B^1\rangle$.*

Our third result will show that certain *UTIME* classes collapse to the corresponding *DTIME* classes under the assumption that the spectra of $2^{nd}O\exists^{I}_{g(n)}$ are equivalent to those of $2^{nd}O\exists^{E}_{g(n)}$ sentences. However, we need a strong notion of equivalence by which not only are two sets X and Y equivalent but in addition, there is some representation for each set such that the sets of witnesses for X and Y induced by these representations are equivalent. Below we formally define the notion of strong equivalence.

Definition. *Two sets X and Y in NP are strongly equivalent if*

1. *$X = Y$ and*
2. *there are NP machines N_X and N_Y for X and Y such that a string w represents an accepting computation of N_X on $y \in X$ if and only if w also represents an accepting computation of N_Y on y.*

We note that Grollmann and Selman ([GrSe 84]) have shown that if one-way functions exist, then $P \neq UP$. Their proof proceeds by considering the intermediate assumption that there are functions that are not polynomially computable, but whose graphs are in P. They show that under this latter assumption, $P \neq UP$. From the above definition of the logic $2^{nd}O\exists^{I}_{g(n)}$, it is intuitively clear that for spectra of $2^{nd}O\exists^{I}_{g(n)}$ sentences, the relation between the elements and witnesses is the graph of a function. For the spectra of $2^{nd}O\exists^{E}_{g(n)}$ sentences, we additionally know that this function is computable in polynomial time. Hence, it is not surprising that we obtain similar consequences starting from the assumption that $2^{nd}O\exists^{I}_{g(n)}$ sentences are are equivalent to $2^{nd}O\exists^{E}_{g(n)}$ sentences. In fact, as mentioned earlier, similar consequences were obtained by Gurevich ([Gu 84]) and Kolaitis ([Kol 89]) by assuming the equivalence of explicit and implicit definitions in other logics.

Theorem 3. *The following statements are equivalent.*

1. *The spectrum of each $2^{nd}O\exists^{I}_{g(n)}$ sentence is strongly equivalent to the spectrum of a $2^{nd}O\exists^{E}_{g(n)}$ sentence.*
2. *$UTIME[2^{O(n^{1/j})}] = DTIME[2^{O(n^{1/j})}]$, if $g(n) = log^j n$, and $UP = P$, if $g(n) = n^{1/j}$.*

Proof. We will use the intuitive notion of "witness sets." Given a set, L, in $P_{g(n)}$, the witness set of L consists of all strings of length $g(n)$ that witness strings of length at least n in L. We use the fact that the spectra of $2^{nd}O\exists^{I}_{g(n)}$ and $2^{nd}O\exists^{E}_{g(n)}$ sentences have witness sets in in the appropriate *UTIME* and *DTIME* classes. It then follows that the strong equivalence of the former two classes results in the equivalence of the later two classes, and vice versa. We prove the theorem for $g(n) = log^j n$. For $g(n) = n^{1/j}$, the proof is similar.

$(1 \Leftarrow 2)$ Let $L \in 2^{nd}O\exists^{I}_{g(n)}$, where L is the generalized spectrum of the sentence $\exists_{g(n)}W^1\phi(W^1)$, and $\phi(W^1)$ can be checked in $DTIME[n^k]$. Let $\langle A_{g(n)}, W^1\rangle$ be the

substructure of $\langle A, W^1 \rangle$ induced by the smallest $g(n)$ elements of A, where $|A| = n$. Let U be a $UTIME[2^{O(n^{1/j})}]$-acceptor that, on input $\langle A_{g(n)}, W_1, b \rangle$, guesses a unique structure $\langle A, B^1 \rangle$ of size n, checks if $\langle A, B^1, W^1 \rangle \models \phi$ (in n^k steps), and verifies if $B^1(b)$. By hypothesis, there is a $DTIME[2^{O(n^{1/j})}]$-acceptor M that accepts the same set as U. Let $\sigma(W^1, x)$ be the formula that expresses "M accepts $\langle A, W^1, x \rangle$." Clearly, the set of structures $\langle A, B^1, W^1 \rangle$ that satisfy ϕ and those that satisfy $\psi = \forall x[\sigma(x) \iff B^1(x)]$ are identical, and thus L is strongly equivalent to the spectrum of the $2^{nd}O\exists^E_{g(n)}$ sentence: $\exists W^1 \; \forall x \; [\sigma(W^1, x) \iff B^1(x)]$.

$(1 \Rightarrow 2)$ Let U be a $UTIME[2^{O(n^{1/j})}]$-acceptor. Without loss of generality, on input $\langle A_{g(n)}, W^1 \rangle$, U first guesses a unique witness structure $\langle A, B^1 \rangle$, such that $|A| \leq n^k$ for some fixed k, and $\exists x \geq n \; [B^1(x)]$. Then, U checks if $\langle A, W^1, B^1 \rangle \models \phi$ where ϕ is a $FO + LFP$ sentence. Then the witness set of structures, $\langle A, B^1 \rangle$ is the spectrum of the $2^{nd}O\exists^I_{g(n)}$ sentence $\exists_{g(n)} W^1 \phi(W^1)$. By assumption, there is a formula $\psi(W^1) = \forall x[\sigma(W^1, x) \iff B^1(x)]$ such that the set of structures $\langle A, B^1, W^1 \rangle$ that satisfy ψ is identical to those that satisfy ϕ, and $\sigma(W^1, x)$ is a $FO + LFP$ formula. Let M be the deterministic machine that, on input $\langle A_{g(n)}, W^1 \rangle$, generates $\langle A, \sigma^1 \rangle$, and checks if $\exists x \geq n \; [\sigma^1(x)]$, in n^k steps. Clearly, M simulates U, and runs in $2^{O(n^{1/j})}$ steps. ∎

Since all sets in $2^{nd}O\exists^E_{g(n)}$ can be pseudorandomly generated, our result gives evidence that pseudorandom generators cannot generate even relatively easy sets in $NP - P$, namely the spectra of $2^{nd}O\exists^I_{g(n)}$ sentences. This is formalized in the following interesting corollary of Theorem 3.

Corollary 1. *The following statements are equivalent.*

1. *The spectrum of each $2^{nd}O\exists^I_{g(n)}$ sentence is strongly equivalent to the set of outputs of a pseudorandom generator that generates strings of length n from random seeds of length $g(n)$.*
2. *$UTIME[2^{O(n^{1/j})}] = DTIME[2^{O(n^{1/j})}]$, if $g(n) = log^j n$, and $UP = P$, if $g(n) = n^{1/j}$.*

This result relates to recent work by Yao ([Ya 82]), Allender ([Al 88]), and Nisan and Widgerson ([NiWi 88]), on the consequences of the existence of good pseudorandom generators. A pseudorandom generator is considered *good* if efficient algorithms cannot distinguish its outputs from a truly random set of strings. A consequence of the existence of a good pseudorandom generator is that probabilistic algorithms can be efficiently simulated deterministically, by using the range of the generators to mimic coin-tosses. The results in [Ya 82], [Al 88] and [NiWi 88] differ in their exact definition of a good pseudorandom generator. The notion of a statistical test to make the notion of a good pseudorandom generator precise. A *statistical test* for pseudorandom generators is typically a probabilistic polynomial time acceptor, or a (non-uniform) family of polynomial sized circuits, whose inputs are the outputs of the pseudorandom generators. A pseudorandom generator passes a statistical test if the set accepted by the statistical test and the range of the generator differ substantially and are uncorrelated; that is, the statistical test cannot distinguish between the range of the generator and a set of truly random strings. Thus, a pseudorandom generator is considered *$\mathcal{C}$-good,* if it passes all statistical tests of complexity $\mathcal{C}$. In other words, the assumption that $\mathcal{C}$-good pseudorandom generators exist is equivalent to the assumption that pseudorandomly

generated sets cannot be *approximated* by sets of complexity $\mathcal{C}$. Alternatively, uniformly generated subsets of $n^{1/j}$-easy strings cannot be approximated by sets of complexity $\mathcal{C}$. In contrast, we have shown consequences of the assumption that a certain complexity class $\mathcal{C}$, namely the class spectra of $2^{nd}O\exists^I_{g(n)}$ sentences, can be pseudorandomly generated.

As another corollary to Theorem 3, we obtain a comparison of the sets in $P_{g(n)}$ with the spectra of $2^{nd}O\exists^I_{g(n)}$ and $2^{nd}O\exists^E_{g(n)}$ sentences. Clearly, sets in $P_{g(n)}$ can contain arbitrarily Kolmogorov-hard strings, as the number of $g(n)$-easy strings of length n is at most $2^{g(n)}$, while a single witness string of length $g(n)$ may witness the membership of arbitrarily many elements (of length n) of some set in $P_{g(n)}$. However, the following question is more interesting: given a set L in $P_{g(n)}$ does every string w of length $g(n)$ that witnesses some element of L, also witness at least one $g(n)$-easy string of length n? By observing that the set, $SAT\text{-}ASSIGN_{g(n)}$, of satisfying assignments of $g(n)$-easy satisfiable formulas is in $P_{g(n)}$, (and is complete,) the above question can be rephrased as: Does every $g(n)$-easy satisfiable formula have at least one $g(n)$-easy satisfying assignment? Hartmanis ([Ha 83]) posed this question for the case when $g(n) = log(n)$. The following corollary gives a partial answer.

Corollary 2. *Let S be any subset of $SAT\text{-}ASSIGN_{g(n)}$ that contains at least one satisfying assignment of each $g(n)$-easy satisfiable formula. Furthermore, let S be the generalized spectrum of a sentence, $\exists_{g(n)}F^1$ ["F^1 is a $g(n)$-easy satisfiable formula and $\langle A, B^1\rangle$ is a satisfying assignment of F^1" and $\phi(F^1)$]. If S is strongly equivalent to the spectrum of a $2^{nd}O\exists^I_{g(n)}$ sentence, then all sparse sets in $NP - P$ are in UP, and if S is strongly equivalent to the spectrum of a $2^{nd}O\exists^I_{g(n)}$ sentence, then $NEXPTIME = EXPTIME$.*

Proof. We use the following theorem of Hartmanis.

Theorem. [Ha 83] *For $g(n) \geq log(n)$,*

1. *$SAT \cap KT[g(n), n^2]$ is a hard set for all sparse sets in NP.*
2. *The following statements are equivalent.*
   a. *$SAT \cap KT[log(n), n^2] \in P$*
   b. *There are no sparse sets in $NP - P$*
   c. *$NEXPTIME = EXPTIME$.*

If S is strongly equivalent to the set of structures $\langle A, B^1\rangle$ that satisfy the sentence $\exists_{g(n)}F^1\psi(F^1)$, where $\psi(F^1)$ is a $FO{+}LFP$ formula that implicitly defines B^1, then the witness set of S :

$$SAT_{g(n)} = \{\langle A, F^1\rangle : \langle A, F^1\rangle \models \exists B^1\psi(B^1)\}$$

is clearly in UP, and consists exactly of $g(n)$-easy satisfiable formulas. By the above theorem, the set of $g(n)$-easy satisfiable formulas, (for $g(n) \geq log\ n$) is hard for all sparse sets in $NP - P$. Hence, all sparse sets in $NP - P$ are in UP. If the formula $\psi(F^1)$ above explicitly defines B^1, then $SAT_{g(n)}$ is in P, and hence by the above theorem, $NEXPTIME = EXPTIME$. ∎

5. OPEN PROBLEMS.

We have seen that interesting results emerge from a careful study of the relationships between sets in NP and their corresponding witness sets. In particular, we have seen that forcing certain relationships between sets in NP and their witness sets leads to robust complexity classes that have alternate characterizations. We gave evidence that if two such classes, $2^{nd}O\exists^{I}_{g(n)}$ and $2^{nd}O\exists^{E}_{g(n)}$ are strongly equivalent, then higher complexity classes collapse. An interesting question arises in this setting.

1. What are the consequences of an assumption that $2^{nd}O\exists^{I}_{g(n)}$ and $2^{nd}O\exists^{E}_{g(n)}$ are *equivalent*? In particular, does Theorem 3 hold if we weaken the assumption of strong equivalence to just equivalence?

An answer to this question would help to determine the exact complexity of $2^{nd}O\exists^{E}_{g(n)}$ more accurately. The question of determining the *approximate* complexity of $2^{nd}O\exists^{E}_{g(n)}$ is also interesting since this characterizes the complexity of statistical tests that pseudorandom generators can not pass. Hence, the following question arises.

2. Using one of the standard definitions of approximability (for example, that given in [NiWi 88]) can we provide an alternate characterization of the smallest complexity class, C, such that every set in $2^{nd}O\exists^{E}_{g(n)}$ can be approximated by a set in C?

A third question arises by observing that thus far we have only considered the complexity of uniformly generated subsets of the class $KT[g(n), n^k]$.

3. Can we characterize the complexity of other interesting subsets of $KT[g(n), n^k]$, for example, $NP \cap KT[g(n), n^k]$?

Even for $g(n) = log(n)$, an answer to this question would characterize the complexity of the set $SAT \cap KT[log(n), n^k]$, which has many interesting applications following the results of Hartmanis mentioned earlier ([Ha 83], [HaYe 83]).

6. BIBLIOGRAPHY

[Al 88] E.W. Allender, "Some consequences of the existence of pseudorandom generators," *to appear, JCSS.*

[ADT 89] C. Àlvarez, J. Díaz, J. Torán, "Complexity classes with complete problems between P and NP-C," *Fund. Comput. Theory conference, Lecture notes in computer science,* **380**, *Springer-Verlag,* pp. 13-24, 1989.

[Ba 68] Y.M. Barzdin, "Complexity of programs to determine whether natural numbers not greater than n belong to a recursively enumerable set," *Soviet Math. Dokl.* **9**, pp. 1251-1254, 1968.

[Da 77] R.P. Daley, "On the inference of optimal descriptions," *Theoretical Comp. Sci.* **4**, pp. 301-309, 1977.

[Fa 74] R. Fagin, "Generalized first order spectra and polynomial time recognizable sets," *Complexity of computation, AMS,* Providence, pp. 44-73, 1974.

[Fa 75] R. Fagin, "Monadic generalized spectra," *Z. Math. Logic Grundlagen Math.* **21**, pp. 89-96, 1975.

[Gu 84] Y. Gurevich, "Toward logic tailored for computational complexity," *Computation and Proof Theory, Lecture notes in mathematics* **1104**, *Springer-Verlag,* pp. 175-216, 1984.

[Ha 83] J. Hartmanis, "Generalized Kolmogorov complexity and the structure of feasible computations," *IEEE FOCS,* 1983.

[HaYe 83] J. Hartmanis, "Computation times of *NP* sets of different densities," *ICALP, Lecture notes in computer science* **154**, *Springer-Verlag* pp. 319-330, 1983.

[Im 82] N. Immerman, "Upper and lower bounds for first order expressibility," *J. of Computer and System Sciences* **22**, *no.3*, 1982.

[Im 87] N. Immerman, "Expressibility as a complexity measure: results and directions," *Structure in complexity theory conf.*, 1987.

[KiFi 80] C.M.R. Kintala, P. Fischer, "Refining nondeterminism in relativized polynomial time bounded computations," *SIAM J. Comput.* **9**, *no. 1*, 1980.

[Ko 83] K-I. Ko, "Resource bounded program size complexity and pseudorandom sequences," *Dept. of comp. sci., University of Houston*, 1983.

[Kol 89] P.G. Kolaitis, "Implicit definability on finite structures and unambiguous computations," *Manuscript*, 1989.

[Le 73] L.A. Levin, "Universal search problems," *Problems in Information Transmission* **9**, 1973.

[LiVi 88] M. Li, P.M.B. Vitanyí, "Two decades of applied Kolmogorov complexity," *Structure in complexity theory conf.*, 1973.

[Lo 86] L. Longpré, "Resource bounded Kolmogorov complexity, a link between complexity theory and information theory," *Ph.D. Thesis, Dept. of Computer Sciences, Cornell Univ.*, 1986.

[Ly 82] J.F. Lynch, "Complexity classes and theories of finite models," *Math. Systems Theory* **15**, pp.127-144, 1982.

[NiWi 88] N. Nisan, A. Widgerson, "Hardness vs. randomness," *IEEE FOCS*, pp.2-24, 1988.

[PaYa 88] C. Papadimitriou, M. Yannakakis, "Optimization, approximation, and complexity classes," *ACM STOC*, pp.229-234, 1988.

[StHu 86] R.E. Stearns, H.B. Hunt, "On the complexity of the satisfiability problem and the structure of NP," *TR 86-21, SUNY Albany*, 1986.

[Ya 82] A. Yao, "Theory and applications of trapdoor functions," *IEEE FOCS*, pp.80-91, 1982.