BRICS

Basic Research in Computer Science

Generating Hard Tautologies Using Predicate Logic and the Symmetric Group

Søren Riis
Meera Sitharam

BRICS Report Series **RS-98-19**

ISSN 0909-0878 **September 1998**

Copyright © 1998, BRICS, Department of Computer Science University of Aarhus. All rights reserved.

Reproduction of all or part of this work is permitted for educational or research use on condition that this copyright notice is included in any copy.

See back inner page for a list of recent BRICS Report Series publications. Copies may be obtained by contacting:

BRICS
Department of Computer Science
University of Aarhus
Ny Munkegade, building 540
DK–8000 Aarhus C
Denmark

Telephone: +45 8942 3360
Telefax: +45 8942 3255
Internet: BRICS@brics.dk

BRICS publications are in general accessible through the World Wide Web and anonymous FTP through these URLs:

`http://www.brics.dk`
`ftp://ftp.brics.dk`
This document in subdirectory `RS/98/19/`

Generating hard tautologies using predicate logic and the symmetric group

Søren Riis[*†]
Meera Sitharam [‡§]

September 1998

Abstract

We introduce methods to generate uniform families of hard propositional tautologies. The tautologies are essentially generated from a single propositional formula by a natural action of the symmetric group S_n.

The basic idea is that any Second Order Existential sentence Ψ can be systematically translated into a conjunction ϕ of a finite collection of clauses such that the models of size n of an appropriate Skolemization $\tilde{\Psi}$ are in one-to-one correspondence with the satisfying assignments to ϕ_n: the S_n-closure of ϕ, under a natural action of the symmetric group S_n. Each ϕ_n is a CNF and thus has depth at most 2. The size of the ϕ_n's is bounded by a polynomial in n. Under the assumption *NEXPTIME* $\neq$ *co-NEXPTIME*, for any such sequence ϕ_n for which the spectrum $S := \{n : \phi_n \text{ satisfiable}\}$ is *NEXPTIME*-complete, the tautologies $\neg\phi_{n \notin S}$ do not have polynomial length proofs in any propositional proof system.

Our translation method shows that most sequences of tautologies being studied in propositional proof complexity can be systematically generated from Second Order Existential sentences

[1]The International PhD Research School at BRICS, Aarhus, Denmark; Email: smriis@daimi.aau.dk

[2]Part of this work was done while visiting the Fields Institute, Toronto, Canada

[3]CISE Department, University of Florida, Gainesville, FL 32611-6120; Email: sitharam@cise.ufl.edu

[4]Supported in part by NSF Grant CCR 94-09809.

and moreover, many natural mathematical statements can be converted into sequences of propositional tautologies in this manner.

We also discuss algebraic proof complexity issues for such sequences of tautologies. To this end, we show that any Second Order Existential sentence Ψ can be systematically translated into a finite collection of polynomial equations $\bar{Q} = 0$ such that the models of size n of an appropriate skolemization $\tilde{\Psi}$ are in one-to-one correspondence with the solutions to $\bar{Q}_n = 0$: the S_n-closure of $\bar{Q} = 0$, under a natural action of the symmetric group S_n. The degree of $\bar{Q}_n$ is the same as that of $\bar{Q}$, and hence is independent of n, and the number of variables is no more than a polynomial in n. Furthermore, using results in [19] and [20], we briefly describe how, for the corresponding sequences of tautologies ϕ_n, the rich structure of the S_n closed, uniformly generated, algebraic systems $\bar{Q}_n$ has profound consequences on on the algebraic proof complexity of ϕ_n.

I Introduction

Algebraic deductive systems have been used in automatic (First Order) theorem proving for several decades. Roughly speaking, these systems translate the axioms (First Order logic sentences), and the theorem to be proved, into polynomial equations, in such a way that the proof of the theorem can be obtained as a proof of membership in the ideal generated by these polynomial equations. Several algorithms for such equational reasoning have been developed, based on Cylindrical Algebraic Decomposition, Hilbert's Nullstellensatz, Gröbner bases, Wu-Ritt Characteristic sets, etc., and have turned out to be effective, e.g, for geometry theorem proving [6], [9], [12], [21]. Many of these algebraic deductive systems have the advantage of being automatizable in that the time taken to find the proof is strongly related to the length of the proof.

More recently, algebraic deductive systems have been employed in the context of proving lower bounds on the proof complexity of propositional tautologies [4], [5], [7], [8]. Showing the nonexistence of polynomial (in n) length proofs for a sequence of tautologies ϕ_n, is directly linked to the $NP \neq$ co-NP question: the link becomes stronger with the strength of the proof system used.

In this paper, we first give a systematic method of translating an Existential Second Order sentence Ψ into a sequence of depth 2 polynomial size propositional formulae ϕ_n such that ϕ_n is satisfiable (or $\neg\phi_n$ is not

a tautology), if and only if a Ψ has a model of size n.

We give an analogous algebraic version of this: a systematic method of translating an Existential Second Order sentence Ψ into a polynomial system $\bar{Q}$ such that the S_n-closure $\bar{Q}_n$ of $\bar{Q}$ has a solution if and only if Ψ has a model of size n. Since $\bar{Q}_n$ is uniformly generated by S_n-closure, its degree is independent of n, and its number of variables is bounded by a polynomial in n.

More specifically, the system $\bar{Q}$ is based on an appropriate (non-unique) Skolemization $\tilde{\Psi}$ of Ψ, and the models $M \models \tilde{\Psi}$ of size n are directly related to the ideal $I_{\Psi,n}$ generated by the polynomials in $\bar{Q}_n$ In particular, there is a 1-1 correspondence between models of $\tilde{\Psi}$ of size n and points on the (discrete) algebraic variety defined by $I_{\Psi,n}$.

Krajicek has pointed out (personal communication) that Paris and Wilkie applied a translation procedure very similar to ours. This is indeed correct. They introduced this translation in their proof of Theorem 26 (see [14]) in the setting of non-standard models for Bounded Arithmetic. The Paris-Wilkie translation shows how bounded first order formulas can be translated into a sequence of polynomial size bounded depth propositional formulas (see for example [10] for a definition of this). Paris and Wilkie do not consider the case where one is translating formulas involving uninterpreted function symbols. Actually in our case both function symbols as well as relational symbols are uninterpreted. This is crucial for getting the S_n-closure properties we are using. Our idea of introducing uninterpreted functions (via Skolemization) plays a crucial role. Without this idea we would NOT get a sequence of polynomial equations of bounded degree. Neither would we get a translation leading to a finitely generated S_n-closed sequence of satisfiability problems.

Our translation shows that many natural mathematical problems can be converted into questions of propositional satisfiability, and thence to questions about the existence of solutions to polynomial equations. For example, the question of whether there is a nilpotent group of size n is identical to the question of whether there is a model of size n for the Existential Second Order statement "the universe is a nilpotent group". Clearly, many natural problems in algebra, number theory, graph theory, or combinatorics can be phrased in this manner.

Our translation highlights the fact that many tautologies used for showing algebraic proof complexity lower bounds, such as various matching principles, versions of the pigeonhole principle, and primality principle [4], [5], [7], [11] can indeed be obtained from natural Second Order

Existential statements. We illustrate this using examples.

More significantly, we introduce a method to systematically generate *hard* propositional tautologies from Second Order Existential sentences. The method relies on sentences that have a hard spectrum and we show how to generate uniform sequences of tautologies that have no polynomial length proofs provided *NEXPTIME* $\neq$ co-*NEXPTIME*.

In addition, we briefly sketch how this method works for algebraic proof systems such as Nullstellensatz and Polynomial Calculus proof systems (unconditionally, with no additional assumptions from complexity theory). The sequences of propositional formulae ϕ_n obtained using our translation have a rich algebraic structure due to the symmetries in the corresponding polynomial systems. This symmetry permits the representation theory of S_n to be used in analyzing the proof complexity of the corresponding tautologies, in various algebraic deductive systems.

Ajtai was the first to consider and analyse uniform S_n-closed families of linear equations [1], [2], [3]. In [19], [20] we prove various new and rather deep theorems related to algebraic proof complexity. These theorems strongly relies on the fact that our translation gives finitely generated S_n-closed systems of propositional formulas (polynomial equations).

II The Translation

We first develop the required notation and definitions.

Let $\exists_{\mathbf{r}}$ denote the second order existential quantifier for r-ary relations. Let $\exists'_{\mathbf{s}}$ denote the Second Order Existential quantifier for s-ary functions. Let $\exists$ and $\forall$ denote the first order quantifiers.

A *pure atomic formula* is of the form: $R(t_1, \ldots, t_l)$ or $f(t_1, t_2, \ldots, t_l) = s$ or $t = s$ where R is an l-ary relation and f is an l-ary function and $s, t, t_1, \ldots, t_l$ are (not necessarily distinct) variables.

A Second Order Existential sentence Ψ is *strict* (recalling the class *MaxSNP* [13]) if it is of the following form:

$$\exists_{\mathbf{r_1}} R_1 \ldots \exists_{\mathbf{r_u}} R_u \exists'_{\mathbf{s_1}} f_1 \ldots \exists'_{\mathbf{s_v}} f_v \forall z_1, \ldots, z_w \; \eta(\vec{z})$$

where $\eta(\vec{z}) := \eta(z_1, z_2, \ldots, z_w)$ is a Boolean combination of pure atomic formulas. The strongest results are achieved if we assume η is in Conjunctive Normal Form. Let Ψ_n denote the sentence which is achieved from Ψ by restricting all quantifiers to the universe $\{1, 2, \ldots, n\}$.

For an arbitrary Second Order Existential sentence (on prenex normal form)

$$\Psi = \exists_{\mathbf{r_1}} R_1 \ldots \exists_{\mathbf{r_u}} R_u \exists'_{\mathbf{s_1}} f_1 \ldots \exists'_{\mathbf{s_v}} f_v \ \tilde{\eta}$$

where $\tilde{\eta} :\equiv \forall z_1 \exists y_1 \ldots \forall z_w \exists y_w \ \eta(z_1, y_1 \ldots, z_w, y_w)$; a corresponding *strict* version $\tilde{\Psi}$ is obtained by *Skolemization*: we eliminate all first order existential quantifiers by introducing Skolem functions g such that

$$\tilde{\Psi} := \exists_{\mathbf{r_1}} R_1 \ldots \exists_{\mathbf{r_u}} R_u \exists'_{\mathbf{s_1}} f_1 \ldots \exists'_{\mathbf{s_v}} f_v \exists'_{\mathbf{1}} g_1 \ldots \exists'_{\mathbf{w}} g_w \eta'$$

where $\eta' \equiv \eta(z_1, g_1(z_1) \ldots, z_w, g_w(z_1, \ldots, z_w))$. It should be noticed that *different Skolemizations in general might lead to different* $\tilde{\Psi}$.

Next, consider the ring $\mathbb{F}[x_{j,e_j} : e_j \in \{1, \ldots, n\}^{r_j}, 1 \leq j \leq u]$ of polynomials over a field $\mathbb{F}$ and the vector space $\Pi_{n,d}$ spanned by those polynomials of degree at most d. Notice that this vector space is closed under the following action of the symmetric group S_n: $Q(x_{j,e_j}) \in \Pi_{n,d} \Rightarrow \forall \pi \in S_n \pi(Q(x_{j,e_j})) \in \Pi_{n,d}$, where $\pi(Q(x_{j,e_j})) := Q(x_{j,\pi(e_j)})$, and for $e_j := (e_{1,j}, \ldots, e_{r_j,j}) \in \{1, \ldots, n\}^{r_j}$, $\pi(e_j)$ is the natural action defined as: $(\pi(e_{1,j}), \ldots, \pi(e_{r_j,j}))$.

In general, given a set of polynomials $\bar{Q}$ where $\bar{Q} := \{Q_1, \ldots, Q_m\}$, each $Q_i \in \Pi_{l,d}$, for some fixed l, we say that a sequence of polynomial systems $\bar{Q}_n$ is *S_n-closed and uniformly generated from* $\bar{Q}$, if for all n, $\bar{Q} \subseteq \bar{Q}_n$, and $Q_i \in \bar{Q}_n \Rightarrow \pi(Q_i) \in \bar{Q}_n \ \forall \pi \in S_n$.

We are now ready to state our translation result.

Theorem 1 *Let Ψ be a Second Order Existential Sentence and let $\tilde{\Psi}$ be a strict version obtained by a (non-unique) Skolemization. Then there is a sequence of propositional formulae $\phi_{\tilde{\Psi},n}$ and a polynomial system $\bar{Q}_{\tilde{\Psi}} \in \Pi_{l,d}$, for some constants l and d depending only on Ψ, such that the following hold.*

1. *Each $\phi_{\tilde{\Psi},n}$ is a CNF (i.e. conjunctions of disjunctions of literals and negation of literals) such that there is a 1-1 correspondence between the models of $\tilde{\Psi}$ of size n and satisfying truth assignments to $\phi_{\tilde{\Psi},n}$.*

2. *For each n, the conjunction of clauses that form $\phi_{\tilde{\Psi},n}$ is closed under the group action S_n and is generated by closing a fixed, finite set ϕ of clauses under a natural action of the symmetric group S_n. (Some of these clauses are formal clauses containing n literals).*

3. *Let $\bar{Q}_{\tilde{\Psi},n} \in \Pi_{n,d}$ be a S_n-closed sequence of polynomial systems uniformly generated from $\bar{Q}_{\tilde{\Psi}}$ (hence $\bar{Q}_{\tilde{\Psi},n}$ has degree d independent of n, and has a number of variables that is polynomially bounded in n). There is a 1-1 correspondence between the models of $\tilde{\Psi}$ of size n and the solutions to $\bar{Q}_{\tilde{\Psi},n} = 0$.*

4. *Let $\phi_{\tilde{\Psi},n}$ be as above. Then Ψ does not have a model of size n if and only if $\phi_{\tilde{\Psi},n}$ not is satisfiable if and only if $\neg\phi_{\tilde{\Psi},n}$ is a tautology if and only if there is no solution to $\bar{Q}_{\tilde{\Psi},n} = 0$*

Proof. Let $\tilde{\Psi}$ be a strict Second Order Existential sentence in which the quantifier-free first order part η is in CNF. We translate the sequence $\tilde{\Psi}_n$ into a sequence of propositional formulas as follows:

For given n we fix the collection of Boolean variables which consists of:

$q^{(1)}_{i_1,i_2,\ldots,i_{r_1}}, \ldots, q^{(u)}_{i_1,i_2,\ldots,i_{r_u}}$ as well as $p^{(1)}_{i_1,i_2,\ldots,i_{s_1},i_{s_1+1}}, \ldots, p^{(v)}_{i_1,i_2,\ldots,i_{s_v+1}}$ where each $i_j \in \{1,2,\ldots,n\}$.

Translate the sequence $\tilde{\Psi}_n$ into a sequence of propositional formulas as follows:

For each selection $i_1, i_2, \ldots, i_w \in \{1,2,\ldots,n\}$ we let $\gamma := \eta_{i_1,i_2,\ldots,i_w}$ denote the propositional formula which appears by performing the following operations.

(1) Replacing each z_j in η by the number i_j.

(2) Replace each atomic formula $R_l(j_1, j_2, \ldots, j_{r_l})$ in γ by $q^{(l)}_{j_1,j_2,\ldots,j_{r_l}}$ where $j_1, \ldots$ are numbers $\in \{1,2,\ldots,n\}$ after performing the replacement in (1).

(3) Replace each atomic formula $f_l(j_1, j_2, \ldots, j_{s_l}) = j_{s_l+1}$ in γ by $p^{(l)}_{j_1,j_2,\ldots,j_{r_l},j_{s_l+1}}$ where $j_1, \ldots$ are numbers $\in \{1,2,\ldots,n\}$ after performing the replacement in (1).

Thus, removing the universal quantifier on the variables z_i of η gives us a conjunction of a set C_n of $O(n^w)$ CNF formulae of fixed length independent of n. Notice that for any n, the set C_n can be obtained uniformly by the natural action of S_n on the fixed set C_l of clauses, where $l \leq \sum_j^u r_u + \sum_j^v (s_v + 1)$.

To complete the process, we need to assert that the f_m are functions, i.e, that they take exactly one value. To do this, we first supplement the set of fixed length formulae in C_n by additional simple clauses of the form $\neg p^{(m)}_{i_1,i_2,\ldots i_{s_m},j} \vee \neg p^{(m)}_{i_1,i_2,\ldots i_{s_m},j'}$, one for each $j, j' \in \{1,2,\ldots,n\}$ with $j \neq j'$

and for each $i_1, i_2, \ldots \in \{1, 2, \ldots, n\}$. Thus for each n, the resulting collection D_n of fixed length, CNF formulae can still be obtained by S_n-closure of the fixed set D_l of formulae.

Finally to assert that the function f_m takes at least 1 value, for each and for each $i_1, i_2, \ldots \in \{1, 2, \ldots, n\}$, we need to include into the collection D_n *formal* clauses of the type $\vee_j p^{(m)}_{i_1, i_2, \ldots i_{s_m}, j}$. Such clauses vary with n (as does the length), however, in a uniform manner. Again, for each n, the resulting collection E_n of formal CNF formulae can still be obtained by S_n-closure of the fixed set E_l of formal formulae.

Now $\phi_{\tilde{\Psi},n}$ is taken to be the conjunction This formula $\phi_{\tilde{\Psi},n}$ is a conjunction of CNF's and formal disjunctive clauses. in E_n. Thus (letting $n = l$) $\phi_{\tilde{\Psi},l}$ can be reduced to a finite conjunction of formal clauses and the clauses in $\phi_{\tilde{\Psi},n}$ are generated from these formal clauses by an application of the symmetric group S_n. Hence $\phi_{\tilde{\Psi},n}$ is a conjunction of a set of polynomially many clauses. Clearly, from any satisfying assignment for $\phi_{\tilde{\Psi},n}$ we can read-off a model of Ψ_n.

Next we notice that the same idea can be used to produce a system of polynomial equations, one for each clause in $\phi_{\tilde{\Psi},n}$, over any given field $\mathbb{F}$, by replacing each disjunction $x \vee y$ by the polynomial $(1 - x).(1 - y)$, $\neg x$ by the polynomial x and replacing the formal disjunction $\vee_j p^{(m)}_{i_1, i_2, \ldots i_{s_m}, j}$ - arising from functions f^m in $\tilde{\Psi}$ - by a formal sum of the form $\Sigma_j \; p^{(m)}_{i_1, i_2, \ldots, i_{s_m}, j} - 1$. This replacement is not valid for general disjunctions, but works in this particular case due to the presence of the other disjunctions $\neg p^{(m)}_{i_1, i_2, \ldots i_{s_m}, j} \vee \neg p^{(m)}_{i_1, i_2, \ldots i_{s_m}, j'}$, for each distinct pair j, j'. Finally we add for each variable x the equation $x^2 - x = 0$.

The closure under S_n ensures that different solutions to $\bar{Q}_{\tilde{\Psi},n} = 0$ (that are not isomorphic under S_n) correspond to non-isomorphic models of $\tilde{\Psi}$. In fact, the number of non-isomorphic models $M \models \tilde{\Psi}_n$ is inversely related to the size of the ideal $I_{\tilde{\Psi},n}$ generated by the polynomials in $\bar{Q}$. Clearly the ideals $I_{\tilde{\Psi},n}$ are closed under S_n as well. Knowledge of $I_{\tilde{\Psi},n}$ allows us uniquely to determine all models of $\tilde{\Psi}_n$, since there is a 1-1 correspondence between models of $\tilde{\Psi}_n$ and points on the algebraic variety defined by $I_{\psi,n}$.

■

III Some Examples

Example (dense linear ordering)

Fix the language $L = L(R, f)$ where R is a binary relation symbol and f is a binary function symbol. Consider the L-sentence: $\eta \equiv (\forall x, y, z\ (R(x,y) \wedge R(y,z) \Rightarrow R(x,z)) \wedge (\forall x, y(R(x,y) \Rightarrow (\neg R(y,x) \vee x = y))) \wedge (\forall x, y(R(x,y) \vee R(y,x) \vee x = y)) \wedge \forall x, y(R(x,y) \Rightarrow R(x, f(x,y)) \wedge R(f(x,y), y))$.

The sentence is a formalization of "R is a linear ordering whose denseness is witnessed by f". The sentence $\exists R \exists f \eta$ is a strict second order existential sentence.

We show how the translation works on this example. Rewrite the $L(R,f)$-sentence as: $\eta \equiv \exists R \exists f\ \ \forall i, j, k[(\neg R(i,j) \vee \neg R(j,k) \vee R(i,k)) \wedge (\neg R(i,j) \vee \neg R(j,i) \vee i = j) \wedge (R(i,j) \vee R(j,i) \vee i = j) \wedge (\neg R(i,j) \vee (R(i, f(i,j)) \wedge \neg R(j, f(i,j)) \wedge R(f(i,j), j)))]$.

Now introduce variables x_{ij} and y_{ijk} with the intended idea that $x_{ij} = 1 \iff R(i,j)$ and $y_{ijk} = 1 \iff f(i,j) = k$. This gives polynomial equations:
$x_{ij}x_{jk}(1 - x_{ik}) = 0$ for all i, j, k, $x_{ij}x_{ji} = 0$ for all $i \neq j$,
$(1 - x_{ij})(1 - x_{ji}) = 0$ for all $i \neq j$, and $x_{ij}y_{ijl}(1 - x_{il}x_{lj}) = 0$ for all i, j, l.
Besides we add the equations:
$x_{ij}^2 - x_{ij} = 0$ for all i, j, $y_{ijk}^2 - y_{ijk} = 0$ for all i, j, k,
$\sum_l y_{ijl} - 1 = 0$ for all i, j, and $y_{ijk}y_{ijl} = 0$ for all $i, j, k \neq l$.
These equations can be simplified and some equations (like $y_{ijk}^2 - y_{ijk} = 0$) are superfluous because they can be derived from the other equations. A natural simplification gives the following system of equations:
$x_{ij}x_{jk}x_{ki} = 0$ for $i \neq k, j$, $x_{ij} + x_{ji} - 1 = 0$ for all i, j,
$x_{ij}y_{ijl}(1 - x_{il}x_{lj}) = 0$ for i, j, l, $\sum_l y_{ijl} - 1 = 0$ for all i, j,
$x_{ij}^2 - x_{ij} = 0$ for all i, j, and $y_{ijk}y_{ijl} = 0$ for all $i, j, k \neq l$
Finally we can get a fixed set of generating equations. For example, $y_{ijk}y_{ijl} = 0$ for all $i, j, k \neq l$ is generated (under the action of S_n) by the 6 equations for indices in $\{1, 2, 3, 4\}$: $y_{123}y_{124} = 0$, $y_{112}y_{113} = 0$, $y_{111}y_{112} = 0$, $y_{121}y_{122} = 0$, $y_{121}y_{123} = 0$, $y_{122}y_{123} = 0$. ♣

The next example shows how one can treat constants in a translation. More significantly, it shows how a sequence of tautologies, commonly studied in the context of algebraic proof complexity lower bounds [4],[5], [7] can be obtained from a Second Order sentence using this translation.

Example (onto-PHP): Fix the language $L = L(f, c)$. Consider the Second Order Universal sentence in the language L.

$$\Psi \equiv \forall f\ ((\exists x, y f(x) = f(y) \wedge x \neq y) \vee (\exists x f(x) = c)).$$

This statement is a formalization of the principle that there is no onto map $f : U \to U$ which avoids the point $\{c\}$. The negation η is on the form:

$$\eta \equiv \exists f\ ((\forall i, j\ f(i) = f(j) \to i = j) \wedge (\forall i\ f(i) \neq c)).$$

Now introduce variables x_{ij} and y_j with the intention that $x_{ij} = 1 \leftrightarrow f(i) = j$ and $y_j = 1 \leftrightarrow c = j$. This gives the equations:
$x_{ij}x_{ik} = 0$ for all $i, j \neq k$ $\quad x_{ij}y_j = 0$ for all i, j,
$\sum_j x_{ij} - 1 = 0$ for all i , $y_i y_j = 0$ for all i, j and the equation $\sum_j y_j - 1 = 0$.
A different set of equations for the onto-PHP (more specifically onto-PHP_n^{n-1}) was considered in [4],[5],[7]. This system of equations essentially appears by letting $y_n = 1$. Notice that this system is only closed under S_{n-1}. Furthermore, other sequences of tautologies studied in proof complexity, such as various matching principles, the primality principle [4],[5], [7], [11] etc. can also be obtained from Second Order sentences using our translation. ♣

IV Construction of Hard Tautologies

One of the most fundamental questions in the theory of proof complexity is the following: how to obtain families of tautologies which do not have short proofs in a given system of propositional calculus?

First, consider the opposite question. How to obtain easy tautologies? One easy class of tautologies arises from a Second Order Universal sentence Ψ which is a tautology of predicate logic. This happens when Ψ is valid in all models, i.e. not only is it valid in all finite models, but also holds in all infinite models. We call such tautologies *absolute tautologies*. Their importance was noticed in [15], [16], [17] and [18] in the setting of Bounded Arithmetic. The term "absolute" is borrowed from forcing in set theory. Absolute tautologies (in non-standard models) always remain true and are preserved in generic extensions of the universe. The corresponding sequence of propositional tautologies ϕ_n are easy to prove: essentially one can take the finite proof in predicate logic and turn it into a proof in propositional calculus. With the right stock of rules one can achieve this in a constant number of steps.

Now consider tautologies like the pigeonhole principle. This holds for finite models, but fails for infinite models. This already suggests that these tautologies are somewhat harder to prove. For such tautologies

there is no longer a proof in predicate logic so different n's require different treatments. On the other hand tautologies like the pigeonhole principle are still provable in a uniform way.

One way of constructing hard sequences of tautologies is to ensure that different values of n require distinctly different proofs. To do this, we consider Second Order Existential sentences Ψ for which the *spectrum* $S := \{n \in N : \exists M \models \Psi_n\}$ has high complexity. The proof of the following theorem formalizes this observation (the proof is fairly straightforward).

Theorem 2 *Consider a Second Order Existential sentence Ψ for which the spectrum $S := \{n \in N : \exists M \models \Psi_n\}$ is NEXPTIME-complete. (Here we assume that the input is n, coded in binary). For any Skolemization $\tilde{\Psi}$, obtain the sequence of propositional formulae $\phi_{\tilde{\Psi},n}$ as in Theorem 1. Now the sequence of propositional tautologies $\{\neg\phi_{\tilde{\Psi},n}\}_{n \notin S}$ does not have polynomial (in n) length proofs (in any propositional proof system), unless NEXPTIME = co-NEXPTIME.*

Next we briefly sketch a related method of generating unconditionally hard propositional tautologies, based on our translation. This method applies only to algebraic proof systems in particular, the Nullstellensatz system (NS) and the Polynomial Calculus system (PC) studied in the context of proof complexity lower bounds [4], [5], [7]. The idea behind NS and PC is to translate the given propositional tautology ϕ_n into an equivalent system of polynomial equations $\bar{Q}(\bar{x}) = 0$ over some field $\mathbb{F}$ in the standard manner described in the proof of Theorem 1. The task of proving ϕ_n can thus be rephrased as the task of showing that $\bar{Q}(\bar{x}) = 0$ does not have a 0/1-solution over $\mathbb{F}$. According to a weak version of Hilbert's Nullstellensatz, this is equivalent to showing that the ideal generated by $\bar{Q} \cup I$ (where $I := \{x^2 - x, x \text{ variable}\}$) contains the constant polynomial 1. An NS-proof is a list of *witnessing* polynomials $\bar{P}$ such that $\sum P_i Q_i = 1$, where $P_i \in \bar{P}$ and $Q_i \in \bar{Q} \cup I$. The degree of the proof is the maximum degree of the witnessing polynomials $\bar{P}$, and acts as a natural proof complexity measure: a constant degree d NS proof implies that one can find the proof in time $O(n^{O(d)})$, simply by solving a linear system to obtain the coefficients of the witnessing polynomials P_i. Notice that the *number of variables* in the system $\bar{Q}$ is a trivial *upper bound on the degree* of the NS proof since the polynomials $x^2 - x$ for each variable x have been added to $\bar{Q}$. See [8] for a description of PC proofs.

The closure under S_n combined with the uniformity allows us to extract strong lower bounds on the degree as well as on the size of Nullstellensatz proofs as well as Polynomial Calculus proofs. Our analysis relies on the structure of the sequence of linear *spans* U_n of the S_n-closed, sequences $\bar{Q}_n \in \Pi_{n,d}$ generated from a fixed polynomial system $\bar{Q} \in \Pi_{l,d}$ obtained in Theorem 1. These vector spaces that are also closed under the natural action of S_n are so-called $\mathbb{F}S_n$ modules, studied extensively in the representation theory of S_n. In [19], we provide a non-traditional analysis of the structure of uniformly generated sequences of $\mathbb{F}S_n$ modules for fields $\mathbb{F}$ of characteristic 0. The gist of these results is that complete and precise decomposition of these module sequences into irreducible modules (beyond isomorphism types) can be uniformly specified, and as a result there is a single polynomial $p(n)$ that captures the dimension of such module sequences for all but finitely many singular values of n, and moreover, for these singular n, the dimension is bounded above by $p(n)$.

References

[1] Ajtai, M.: The independence of the modulo p counting principles. In Proceedings of the 26th ACM STOC, 402-411 (1994)

[2] Ajtai, M.: On the existence of modulo p cardinal functions, in: Feasible Mathematics II, eds. P. Clote and J. Remmel, Birkhauser. 1-14 (1994)

[3] Ajtai, M.: Symmetric systems of linear equations modulo p, TR94-015 of the Electronic Colloquium on Computational Complexity (1994)

[4] Beame, P., Impagliazzo, R., Krajicek, J., Pitassi, T., Pudlak, P.: Lower bounds on Hilbert's Nullstellensatz and propositional proofs. Proceedings of the London Mathematical Society **73(3)** 1-26 (1996)

[5] Beame, P., Riis, S.: More on the relative strength of counting principles. In: Proceedings of the DIMACS workshop on Feasible Arithmetic and Complexity of Proofs, (1996)

[6] G.E. Collins: Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition, In: Lect. Notes in CS, **33**, 134-183 (1975)

[7] Buss, S., Krajicek, j,. Pitassi, T., Razborov, A., Sergal, J.: Polynomial bound on Nullstellensatz for counting principles. To appear in Computational Complexity (1997)

[8] Clegg, M., Edmonds, J., Impagliazzo, R.: Using the Groebner basis algorithm to find proofs of unsatisfiability. In: Proceedings of the 28th ACM STOC 174-183 (1996)

[9] Chou, S. C., Gao, X. S., and Zhang, J. Z., A method of solving geometric constraints, Wichita State University, Tech. Rep, Dept. of Computer Sci., (1996)

[10] Krajicek, J.:Bounded Arithmetic, propositional logic, and complexity theory, Encyclopedia of Mathematics and Its Applications, Vol. 60, Cambridge University Press (1995)

[11] Krajicek, J.: On the degree of ideal membership proofs from uniform families of polynomials over a finite field (manuscript)

[12] McCune, W., Padmanablan, R.: Automated deduction in equational logic and cubic curves. Lecture Notes in Computer Science, 1095 Springer-Verlag, Berlin (1996)

[13] Papadimitriou, C.: Computational Complexity, Addison-Wesley, (1994)

[14] Paris, J., Wilkie, A.: Counting Problems in Bounded Arithmetic, in: Methods in Mathematical Logic, LNM 1130, 317-340, Springer-Verlag (1985)

[15] Riis, S.: Making infinite structures finite in models of Second Order Bounded Arithmetic. In: Arithmetic, proof theory and computorial complexity, 289-319, Oxford: Oxford University Press 1993

[16] Riis, S.: Independence in Bounded Arithmetic. DPhil dissertation, Oxford University (1993)

[17] Riis, S.: Count(q) does not imply Count(p): Annals of Pure and Applied Logic, 90(1-3):1-56, (1997)

[18] Riis, S.: Count(q) versus the pigeon-hole principle. Archive for Mathematical Logic **36** 157-188 (1997)

[19] Riis, S., Sitharam, M: Permutation modules and their uniformly generated submodules.

[20] Riis, S., Sitharam, M: Manuscript in preparation. Incomplete version appear as technical report TR97-048 of the *Electronic Colloquium on Computational Complexity,* http://www.eccc.uni-trier.de/pub/eccc (1997)

[21] Ruiz, O.E., and Ferreira, P.M.: Algebraic geometry and group theory in geometric constraint satisfaction for computer-aided design and assembly planning, In: IIE Transactions on Design and Manufacturing, **28**, 281-294, (1996)

Recent BRICS Report Series Publications

RS-98-19 Søren Riis and Meera Sitharam. *Generating Hard Tautologies Using Predicate Logic and the Symmetric Group*. September 1998. 13 pp.

RS-98-18 Ulrich Kohlenbach. *Things that can and things that can't be done in PRA*. September 1998. 24 pp.

RS-98-17 Roberto Bruni, José Meseguer, Ugo Montanari, and Vladimiro Sassone. *A Comparison of Petri Net Semantics under the Collective Token Philosophy*. September 1998. 20 pp. To appear in *4th Asian Computing Science Conference*, ASIAN '98 Proceedings, LNCS, 1998.

RS-98-16 Stephen Alstrup, Thore Husfeldt, and Theis Rauhe. *Marked Ancestor Problems*. September 1998.

RS-98-15 Jung-taek Kim, Kwangkeun Yi, and Olivier Danvy. *Assessing the Overhead of ML Exceptions by Selective CPS Transformation*. September 1998. 31 pp. To appear in the proceedings of the *1998 ACM SIGPLAN Workshop on ML*, Baltimore, Maryland, September 26, 1998.

RS-98-14 Sandeep Sen. *The Hardness of Speeding-up Knapsack*. August 1998. 6 pp.

RS-98-13 Olivier Danvy and Morten Rhiger. *Compiling Actions by Partial Evaluation, Revisited*. June 1998. 25 pp.

RS-98-12 Olivier Danvy. *Functional Unparsing*. May 1998. 7 pp. This report supersedes the earlier report BRICS RS-98-5. Extended version of an article to appear in *Journal of Functional Programming*.

RS-98-11 Gudmund Skovbjerg Frandsen, Johan P. Hansen, and Peter Bro Miltersen. *Lower Bounds for Dynamic Algebraic Problems*. May 1998. 30 pp.

RS-98-10 Jakob Pagter and Theis Rauhe. *Optimal Time-Space Trade-Offs for Sorting*. May 1998. 12 pp.

RS-98-9 Zhe Yang. *Encoding Types in ML-like Languages (Preliminary Version)*. April 1998. 32 pp.