An Exponential Lower Bound to the Size of Bounded Depth Frege Proofs of the Pigeonhole Principle

Jan Krajíček* Pavel Pudlák† Alan Woods‡

Abstract

We prove lower bounds of the form $exp\,(n^{\varepsilon_d})$, $\varepsilon_d > 0$, on the length of proofs of an explicit sequence of tautologies, based on the Pigeonhole Principle, in proof systems using formulas of depth d, for any constant d. This is the largest lower bound for the strongest proof system, for which any superpolynomial lower bounds are known.

Introduction

Propositional calculus plays a key role not only in logic, but also in complexity theory. The relation of the complexity of propositional logic to the complexity of computations was studied already in the seminal paper of Cook [10]. If one uses a general concept of a proof system for propositional calculus, then the question whether there is a proof system for propositional logic in which every tautology has a proof of size polynomial in the size of the tautology is equivalent to the well-known open question whether the class of predicates accepted by a non-deterministic polynomial time Turing machine is closed under complementation, i.e. whether $NP = coNP$, cf. [11].

As the task of proving $NP \neq coNP$ appears to be very hard, it seems reasonable to prove superpolynomial lower bounds gradually for stronger and stronger proof systems. For quite a long time there were no nontrivial lower bounds for any proof system, except for a very special system, *regular resolution*, see [20]. The real progress started when Haken [13] proved an exponential lower bound for the (unrestricted) resolution system. This is a rather weak system, but it is very important for practical applications.

In [11] Cook and Reckhow defined certain important proof systems, the most important is the class of the so called *Frege systems*. This concept is intended to capture the idea of the most commonly used systems in logic. A typical Frege system consists of a finite set of axioms and the *Modus Ponens* rule

$$\frac{A,\ \neg A \vee B}{B}$$

*Mathematical Institute AVČR, Žitná 25, Praha 1, 115 67, Czech Republic, e-mail: krajicek@earn.cvut.cz

†Mathematical Institute AVČR, Žitná 25, Praha 1, 115 67, Czech Republic, e-mail: pudlak@earn.cvut.cz

‡Department of Mathematics, University of Western Australia, WA 6009, Australia, e-mail: woods@maths.uwa.edu.au

It seems, however, that these systems are quite strong and proving a superpolynomial lower bound for them may be as hard as solving an important question in computational complexity (eg. proving that an explicit language is not in NC_1).

Thus it was a major advance when Ajtai [1] proved that if in a *Frege system* all formulas are required to have depth bounded by a fixed constant, then there is no polynomial upper bound on the size of such proofs. Here we mean by *depth* the number of alternations of conjunctions, disjunctions and negations. Such systems are called *bounded depth Frege systems.* They are stronger than the resolution system; the resolution system can be thought of as depth 1 Frege system. The first truly exponential lower bound for bounded depth Frege systems was proved by Krajíček [15]. In order to state his result precisely, we have to talk about *refutable sets* instead of (the negations of) tautologies and *refutations* instead of proofs (by contradiction). He proved that for every depth d there exists a sequence of refutable sets of depth d formulas which requires exponential size depth d refutations.

In this paper we show that there exists one sequence of tautologies of depth 4 which requires exponential size proofs for any fixed $d \geq 4$.

All the results mentioned above use essentially the same tautologies, proposed by Cook and Reckhow, (only the sets of formulas used by Krajíček in [15] are rather modified). These are tautologies based on a trivial, but basic, combinatorial principle called the *Pigeonhole Principle.* It is the familiar fact that the range R of injective function f must be at least as large as its domain D. Taking D and R to be sets of cardinalities $n+1$ and n respectively we formalize this principle as the formula $PHP(P,n)$:

$$\begin{array}{c} \exists x_1, x_2 \in D \exists y \in R (x_1 \neq x_2 \wedge P(x_1,y) \wedge P(x_2,y)) \vee \\ \exists x \in D \exists y_1, y_2 \in R (y_1 \neq y_2 \wedge P(x,y_1) \wedge P(x,y_2)) \vee \\ \exists x \in D \forall y \in R \neg P(x,y) \ \vee \exists y \in R \forall x \in D \neg P(x,y). \end{array} \tag{1}$$

This can be encoded as the tautology PHP_n by:

$$\begin{array}{l} \bigvee_{i \neq j \in D,\ k \in R} (p_{ik} \wedge p_{jk}) \vee \bigvee_{i \neq j \in R,\ k \in D} (p_{ki} \wedge p_{kj}) \vee \\ \qquad\qquad \bigvee_{i \in D} \bigwedge_{k \in R} \neg p_{ik} \vee \bigvee_{k \in R} \bigwedge_{i \in D} \neg p_{ik}. \end{array} \tag{2}$$

We have included the condition that the function is onto, which makes the principle weaker, hence our result stronger. Note that the size of PHP_n is polynomial in n.

Now we can state our result explicitly.

Theorem. *Let F be any Frege proof system and d any natural number (greater than the depth of PHP_n as formalized in F). Then every depth d F-proof of PHP_n must have the size at least* $\exp\left(n^{(5+o(1))^{-d}}\right)$.

There have been considered a few other combinatorial principles, but PHP is the easiest to analyze. Unfortunately it does have polynomial size proofs in unrestricted Frege systems, see [8]. In fact, we lack good candidates of hard tautologies for such systems.

Lower bounds on the length of proofs have important applications in *bounded arithmetic* (which are logical theories formalizing computationally feasible reasoning about finite structures). We shall mention these results without proofs. The reader should consult [17, 18, 21] or monographs [7, 12, 16].

Ajtai's original result [1] implies that $PHP(P,n)$ is not provable in $I\Delta_0(P)$. This shows that, in a sense, PHP is stronger than the principle of mathematical induction. Our results strengthens this to the following.

Corollary
The pigeonhole principle $\forall x\ PHP(P,x)$ *is not provable in theory*

$$I\Delta_0(P) + \{\Omega_k | k < \omega\}\ ,$$

where Ω_k *is a* Π^0_2*-axiom saying that function* $\omega_k(x)$ *is total, where* $\omega_0(x) = x^2$ *and* $\omega_{k+1}(x) := 2^{\omega_k(|x|)}$*, and* $|x| = \lceil \log_2(x+1) \rceil$.

In fact, this holds for any function $g(x)$ with Δ_0-graph such that 2^x eventually majorizes any fixed iteration of $g(x)$.

The paper is organized as follows. In the first section we define the notion of a *Frege system* and introduce some notation. In the second section we outline the strategy of the lower bound proof. The third section introduces technical notions, the most important being the notion of *k-complete systems*, and proves their basic properties. The technical heart of the paper is in the fourth section which is devoted to the proof of a probabilistic combinatorial lemma (Lemma D2). The lower bound is derived in the fifth section with the help of the concept of *k-evaluation* introduced there.

The paper is intended for specialists in the field, however it uses only standard techniques from combinatorics and very little logic. Thus it should be accessible also to all mathematicians working in complexity theory and combinatorics.

A similar theorem has been proved independently by *Pitassi, Beame* and *Impagliazzo* [19] and the results have been announced in a joint extended abstract [4]. Our and their proofs are different though have some similarities. We think that the method of the proof is at least as important as the result itself and that the two proofs should be published separately. The main difference is that where [19] use decision trees and critical truth 1-to-1 assignments we use a more general concept of complete systems and all 1-1 assignments.

1 Frege systems

We shall consider the language consisting of disjunction $\vee$ (binary), negation $\neg$, falsehood 0, truth 1 and propositional variables. In general Frege systems may use any complete basis of connectives, but it is well-known, see [11], that they are equivalent up to a polynomial increase of the size of proofs, so we can restrict ourselves to this basis, in which the depth can be defined naturally. Since we study the proofs of PHP_n we need only variables $p_{ij}, i \in D, j \in R$.

A *Frege system* is determined by a finite set of rules of the form

$$\frac{\varphi_1(q_1,\ldots,q_m),\ldots,\varphi_r(q_1,\ldots,q_m)}{\varphi_0(q_1,\ldots,q_m)}$$

where $q_1,\ldots,q_m$ are variables and $\varphi_0,\varphi_1,\ldots,\varphi_r$ are formulas in the language $\vee,\neg,0,1$ and $q_1,\ldots,q_m$. Furthermore the rules must be sound and implicationally complete (cf. [11]). (For our lower bound we need only soundness.) An instance of the rule is obtained by substituting particular formulas $\psi_1,\ldots,\psi_m$ (in the language $\vee,\neg,0,1,p_{ij}$) for $q_1,\ldots,q_m$. A rule with $r=0$ is called an *axiom scheme.*

The size of a formula is the number of occurrences of $\vee$ and $\neg$ in it, *the depth* is the maximum number of alternations of $\vee$ and $\neg$ in the formula. *The size of a proof* is the sum of sizes of formulas in it and *its depth* is the maximum depth of a formula in it. By a theorem of [11] we could restrict ourselves to a particular Frege system, but this would not shorten the proof and would make it less transparent.

We shall use PHP_n in the form

$$\bigvee_{i\neq j\in D,k\in R}\neg(\neg p_{ik}\vee\neg p_{jk})\vee\bigvee_{i\neq j\in R,k\in D}\neg(\neg p_{ki}\vee\neg p_{kj})\vee$$

$$\bigvee_{i\in D}(\neg\bigvee_{k\in R}p_{ik})\vee\bigvee_{k\in R}(\neg\bigvee_{i\in D}p_{ik}),$$

where $\bigvee$ denotes repeated binary $\vee$ (the order of parentheses is not important). Thus the size of PHP_n is $O(n^3)$ and the depth is 4.

2 An outline of the proof

As the proof is rather technical we shall outline the main ideas in this section.

First recall the way of proving unprovability of a formula ψ which is implicitly used, when we use a model of $\neg\psi$. To show that ψ is unprovable we assign the truth values 0 and 1 to all formulas in such a way that:

1. the axioms get value 1,

2. the rules preserve the value 1,

3. ψ gets value 0.

We would like to use this idea, but our ψ is actually *provable,* so we must modify this approach. We shall modify it in two respects. Firstly we shall construct evaluation only for any small (= subexponential) set of formulas, thus we only show that ψ does not have a small proof. Secondly, for technical reasons, we shall not use values 0 and 1, but assign to each φ in question a set S_φ and the value of φ will be a set $H_\varphi\subseteq S_\varphi$, i.e. an element of the boolean algebra of subsets of S_φ. Then 1 is replaced by S_φ itself and 0 by $\emptyset$.

It is clear that the evaluations should be somehow connected with the PHP. Suppose for a moment that $D = R$ in the formula (2) defining PHP_n. Then we can show unprovability of such a formula by taking $S_\varphi = S = \{f|\ f$ 1-to-1 , $f : D \to R\}$ and H_φ will be those f's which define truth assignments to $\{p_{ij}\}$ that satisfy φ. Since PHP_n says that $\{p_{ij}\}$ do not define a 1-to-1 mapping from D onto R, it gets value $\emptyset$, while the axioms get S and the rules clearly preserve the value S.

In our proof $|D| \neq |R|$, therefore we take *partial* 1-to-1 mappings and, in a sense, approximate the above argument. The degree of approximation is determined by a parameter k, the maximal size of a mapping in S_φ. Such an S_φ must have special properties in order to look like all 1-to-1 mappings; we call it *complete*. In order to show that 1 is preserved by rules, we use a *common refinement* of S_φ's occurring in the rule.

The most difficult task is to assign H_φ to φ in such a way that H_φ has the same properties as the one consisting of total 1-to-1 mappings in the case of $|D| = |R|$. The problem is that φ may speak about all variables p_{ij}, while partial 1-to-1 mappings are always undefined for some i's and j's. Here we use a technique from boolean complexity which reduces the depth of formulas by applying a (random) restriction of the domain, the so called *switching lemma*. Here a restriction is a partial 1-to-1 mapping ρ from D to R, which means that we restrict the formula to the i and j's on which ρ is undefined. We think of ρ as a partial specification of the alleged total 1-to-1 mapping. Applying the lemma several times the formula is reduced to such a form that its truth can be decided by using only a subset of all remaining variables. Again the lemma needs essential modifications for our purpose, namely we use partial 1-to-1 mappings instead of 0-1-strings.

3 Complete systems of partial maps

Let n be a natural number and D and R two sets of cardinalities $n + 1$ and n respectively, as in Section 1. Partial maps $h :\subseteq D \to R$ are thought of as sets of pairs and as functions whichever is more convenient. The following definition introduces basic technical notions we shall work with.

Definition A.

(a) M is the set of partial $1-$to-1 maps from D to R,

$$M := \{h :\subseteq D \longrightarrow R \mid h \text{ injective }\}.$$

For $H \subseteq M$, the norm $||H||$ of H is

$$||H|| := \max_{h \in H} |h|.$$

(As h is injective, $|h| = |dom(h)| = |rng(h)|$.)

(b) For $k \leq n$, a subset $S \subseteq M$ is k-*complete* iff it satisfies three conditions:

(i) $\forall \delta, \delta' \in S, \delta \neq \delta' \rightarrow \delta \cup \delta' \notin M$,

(ii) $\forall h \in M, |h| + k \leq n \rightarrow \exists \delta \in S, h \cup \delta \in M$,

(iii) $||S|| \leq k$.

(Note that (ii) with $h = \emptyset$ implies that $S \neq \emptyset$.)

(c) For $H, S \subseteq M$, S is a *refinement* of H, $H \lhd S$ in symbols, iff $\forall \delta \in S, (\forall h \in H, h \cup \delta \notin M) \ \vee \ (\exists h' \in H, h' \subseteq \delta)$.

■

Note that the notation $h \cup \sigma \in M$ is used to express that the two partial 1-to-1 maps h and σ are compatible on the common elements of D and R.

The most important example of a k-complete set is the following. Let T be a labelled tree where

1. the vertices are labelled by elements of D and R;

2. edges of T going out of a vertex v are labelled by pairs (i, j) where i or j is the label of v;

3. if $(i_1, j_1), \ldots, (i_r, j_r)$ are the labels on edges of the path going from the root to v and v is not a leaf, then the out-going edges contain all labels (i, j) disjoint with $\{i_1, j_1, \ldots, i_r, j_r\}$ and containing the label of v (according to the previous condition), thus in particular the label of v must not be in $\{i_1, j_1, \ldots, i_r, j_r\}$;

4. the depth (=the maximal length of a path) is at most k.

Each path from the root to a leaf determines a partial mapping from M. Let S be the set of all such mappings. Then one can check that S is k-complete.

The reader should keep this example in mind, since the whole proof can be carried out using these special complete systems.

Next several lemmas are of rather technical nature but with a simple motivation behind. We shall discuss this motivation first but we shall confine to the situation with the boolean variables x_{ij}, $i, j = 1, \ldots, n$, and the notion of a complete system w.r.t. the set M^{iso} of all total truth assignments to the variables corresponding to isomorphisms of the set $\{1, \ldots, n\}$, i.e. every assignment $\alpha \in M^{iso}$ have the form

$$\alpha(x_{ij}) := \begin{cases} 1 & \text{if } f(i) = j \\ 0 & \text{if } f(i) \neq j \end{cases}$$

for some bijection $f : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$. That is a more natural situation than the situation in the previous definitions, but allows a perfectly general explanation of the intuition.

A complete system w.r.t. M^{iso} is a set S of partial truth assignments to x_{ij} such that:

1. every $\sigma \in S$ has the form:

$$\sigma(x_{ij}) := \begin{cases} 1 & \text{if } g(i) = j \\ 0 & \text{if } (i \in dom(g) \wedge g(i) \neq j) \vee (j \in rng(g) \wedge g^{(-1)}(j) \neq i) \\ \text{undefined} & \text{if } i \notin dom(g) \wedge j \notin rng(g) \end{cases}$$

   for some partial injective map $g :\subseteq \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$.

2. $\forall \sigma \neq \delta \in S,\ \sigma \perp \delta$, where $\sigma \perp \delta$ means that σ, δ are incompatible

3. $\forall \sigma \in M^{iso} \exists \delta \in S,\ \delta \subseteq \sigma$.

Identify the elements $\sigma \in S$ with the subsets

$$[\sigma] := \{\delta \in M^{iso} \mid \sigma \subseteq \delta\} .$$

Then S is a complete system iff the set of the sets $[\sigma]$, $\sigma \in S$, is a partitioning of M^{iso}.

Let $\phi(x_{ij})$ be a formula (in variables x_{ij}). We are interested in the truth values $\phi(\sigma)$ of $\phi(x_{ij})$ only for assignments $\sigma \in M^{iso}$. If S is a complete system such that

$$M^{iso} \cap \phi^{(-1)}(0) = \bigcup_{\sigma \in S_0} [\sigma] \quad \text{and} \quad M^{iso} \cap \phi^{(-1)}(1) = \bigcup_{\sigma \in S_1} [\sigma]$$

where $S_0 \cup S_1 = S$ is a partition of S, then the truth table for ϕ for assignments from M^{iso} can be encoded by a map from S into $\{0, 1\}$: map $\sigma \in S$ to 0 iff $\sigma \in S_0$.

Moreover, the formulas ϕ and $\neg\phi$ are equivalent for all assignments from M^{iso} to particular disjunctive normal forms:

$$\phi \equiv_{iso} \bigvee_{\sigma \in S_1} \sigma \subseteq \overline{x} \quad \text{and} \quad \neg\phi \equiv_{iso} \bigvee_{\sigma \in S_0} \sigma \subseteq \overline{x}$$

where $\sigma \subseteq \overline{x}$ abbreviates the conjunction:

$$\bigwedge_{x_{ij} \in dom(\sigma) \wedge \sigma(x_{ij}) = 1} x_{ij}$$

and where $\equiv_{iso}$ means the equivalence for all truth assignments from M^{iso}. In this view a complete system is just a particular disjunctive normal form of the formula 1, namely such a form in which the disjuncts are mutually incompatible.

Disjunctive normal forms for ϕ and $\neg\phi$ with this property are obtained from any decision tree for ϕ of the form described after Definition A. On the other hand, a complete system S allowing an expression of ϕ and $\neg\phi$ as above yields also a decision tree for ϕ of the depth $\leq \|S\|^2$.

For $\phi = x_{ij}$ the complete system

$$S_{x_{ij}} = \{\ \{x_{ij} \mapsto 1\}\ \} \cup \{\ \{x_{iv} \mapsto 1, x_{uj} \mapsto 1\} \mid u \neq i, v \neq j\}$$

of the norm 2 allows the expression of ϕ and $\neg\phi$ as above, and obviously if we have such a system S_ϕ for ϕ then $S_{\neg\phi} := S_\phi$ works for $\neg\phi$. Hence the only non-trivial case in a construction of a complete system S_ϕ of small norm by induction on the depth of ϕ is when $\phi = \bigvee_i \phi_i$. A system S allows an expression of ϕ and $\neg\phi$ as above iff the sets $M^{iso} \cap \phi^{(-1)}(0)$ and $M^{iso} \cap \phi^{(-1)}(1)$ are unions of some sets $[\sigma]$ determined by S. But having the systems $S_i = S_{\phi_i}$ for ϕ_i allows already an expression

$$\phi^{(-1)}(1) = \bigcup_{\sigma \in H} [\sigma]$$

of ϕ as a union of blocks, where

$$H = \bigcup_i \{\sigma \in S_i \mid [\sigma] \subseteq \phi_i^{(-1)}(1)\}$$

and it has the norm $\|H\| = max_i \|S_i\|$ which is small.

The problem is that H is not necessarily a subset of a complete system (of a small norm). However, if S is a system *refining* H:

$$\forall \sigma \in S,\ (\exists \delta \in H,\ \delta \subseteq \sigma)\ \vee (\forall \delta \in H,\ \delta \perp \sigma)$$

then

$$\phi \equiv \bigvee_{\sigma \in S_1} \sigma \subseteq \overline{x} \text{ and } \neg\phi \equiv \bigvee_{\sigma \in S_0} \sigma \subseteq \overline{x}$$

where

$$S_1 = \{\sigma \in S \mid \exists \delta \in H,\ \delta \subseteq \sigma\}$$

and

$$S_0 = \{\sigma \in S \mid \forall \delta \in H,\ \delta \perp \sigma\}\ .$$

A lemma saying that there is a partial truth assignment ρ corresponding to a partial injective map such that there is a small norm S refining H^ρ thus replaces the standard switching lemma [14] in this situation.

Now assume that we have complete systems $\mathbf{S}_\phi$ assigned to formulas ϕ which allow an expression of ϕ and $\neg\phi$ as above. This allows to define for every ϕ the boolean algebra $exp(\mathbf{S}_\phi)$ of the subsets of $\mathbf{S}_\phi$ and a value $\mathbf{H}_\phi$ in that boolean algebra:

$$\mathbf{H}_\phi := \{\sigma \in \mathbf{S}_\phi \mid [\sigma] \subseteq \phi^{(-1)}(1)\}\ .$$

Think of $\mathbf{H}_\phi$ as of those $\sigma \in \mathbf{S}_\phi$ forcing ϕ true. Hence ϕ is a tautology iff $\mathbf{H}_\phi = \mathbf{S}_\phi$ iff all $\sigma \in \mathbf{S}_\phi$ force ϕ true.

In the particular Definition A of k-complete systems, intuitively, no element $\sigma \in M$ forces PHP_n true. Hence one expects to get an "evaluation" of formulas in which PHP_n will not be true. The notion of k-evaluation defined in Definition F formalizes these ideas. The necessary lemma (Lemma D2) analogous to the switching lemma is proved in section 4.

To help the reader we present main technical concepts and some simple lemmas describing these concepts in the rest of this section and defer the difficult proofs to the next sections.

Lemma A. Suppose $H \lhd S \lhd T$ for some $H, S, T \subseteq M$, S is k-complete and $||T|| + k \leq n$. Then $H \lhd T$.

Proof: We have

$$\forall \tau \in T \exists \delta \in S, \delta \cup \tau \in M$$

by the k-completeness of S and by $||T|| + k \leq n$, and so by $S \lhd T$ it must be:

$$\forall \tau \in T \exists \delta' \in S, \delta' \subseteq \tau.$$

To prove the lemma let $h \in H, \tau \in T$ be such that $h \cup \tau \in M$. Take $\delta' \in S$ s.t. $\delta' \subseteq \tau$. Hence also $h \cup \sigma' \in M$ and thus $h' \subseteq \delta'$ for some $h' \in H$, by $H \lhd S$. We have $h' \subseteq \tau$ as we wanted to establish. ■

Definition B.

For $S, T \subseteq M$, the set $S \times T$, *a common refinement* of S and T, is defined by:

$$S \times T = \{\delta \cup \tau \in M | \ \delta \in S, \tau \in T\},$$

i.e. it is the set of elements of M of the form $\delta \cup \tau, \delta \in S, \tau \in T$. ■

Lemma B.

Let $S, T \subseteq M$ and assume that S is k-complete, T is l-complete, $||S|| + l \leq n$, $||T|| + k \leq n$ and $k + l \leq n$. Then the following hold:

(a) $S \times T$ is $k + l$-complete.

(b) $S \lhd S \times T, \quad T \lhd S \times T$.

Proof:

(a) Assume $\delta \cup \tau, \delta' \cup \tau' \in M$ for $\delta \cup \tau, \delta' \cup \tau'$ two distinct elements of $S \times T$. Then either $\delta \neq \delta'$ or $\tau \neq \tau'$ and hence either $\delta \cup \delta' \notin M$ or $\tau \cup \tau' \notin M$ by the completeness of S and T resp.. In both cases $(\delta \cup \tau) \cup (\delta' \cup \tau') \notin M$ which verifies condition (b)(i) of Definition A.

To verify (b)(ii) let $|h| + k + 1 \leq n$. Then, as $||S|| \leq k$ and $||T|| \leq l$, by the completeness of S, T there are $\delta \in S, \tau \in T$ s.t. $h \cup (\delta \cup \tau) \in M$.

Finally, as obviously $||S \times T|| \leq ||S|| + ||T|| \leq k + l$, condition (b)(iii) holds too.

(b) Let $h \in S$ and $h \cup (\delta \cup \tau) \in M$ for some $\delta \cup \tau \in S \times T$. Then, since S is k-complete, $h = \delta$ and thus $h \subseteq \delta \cup \tau$. Hence $S \lhd S \times T$. Identically follows $T \lhd S \times T$.

■

Definition C.

Let $H, S \subseteq M$. The *projection* of H on $S, S(H)$ in symbols, is:

$$S(H) = \{\delta \in S | \exists h \in H, h \subseteq \delta\}.$$

■

We shall consider the symbol $S(H)$ as a two-place function assigning to a pair $S, H \subseteq M$ a set $S(H) \subseteq M$.

Lemma C1

Let $H, S, T \subseteq M$, let S be k-complete, $||T|| + k \leq n$ and $H \triangleleft S \triangleleft T$. Then $T(S(H)) = T(H)$ and $T(S) = T$.

Proof: To see $T(S(H)) \subseteq T(H)$ let $\tau \in T(S(H))$. Then $h \subseteq \delta \subseteq \tau$ for some $\delta \in S(H), h \in H$. So $\tau \in T(H)$ too.

To establish $T(H) \subseteq T(S(H))$ let $\tau \in T(H)$ and $h \subseteq \tau$ for some $h \in H$. Then, identically as in the proof of Lemma A, for some $\delta \in S, \delta \subseteq \tau$. Hence $h \cup \delta \in M$ and, as $H \triangleleft S, h' \subseteq \delta$ for some $h' \in H$. So $h' \subseteq \delta \subseteq \tau$, i.e. $\tau \in T(S(H))$.

To see $T(S) = T$ take $H = \{\emptyset\}$. ■

Lemma C2

Let $H, S, T \subseteq M, H \triangleleft S \triangleleft T, ||S|| + l \leq n, ||T|| + k \leq n,$ S be k-complete and T be l-complete. Then: $S(H) = S$ iff $T(H) = T$.

Proof:

Assume first $S(H) = S$. By Lemma C1 $T(S) = T$ and also $T(S(H)) = T(H)$. Thus $T(H) = T$.

Now assume $T(H) = T$, and let $\delta \in S$ be given. By the l-completeness of T and by $|\delta| + l \leq ||S|| + l \leq n, \delta \cup \tau \in M$ for some $\tau \in T$. By the assumption $h \subseteq \tau$, some $h \in H$. Hence $\delta \cup h \in M$ too and thus, by $H \triangleleft S, h' \subseteq \delta$ some other $h' \in H$. Therefore $\delta \in S(H)$. ■

Lemma C3.

For any $S, \mathbf{H}_i \in M, i \in I$,

$$S\left(\bigcup_I \mathbf{H}_i\right) = \bigcup_I S(\mathbf{H}_i).$$

■

Lemma C4.

Let $S \subseteq M$ be k-complete and $S_0, S_1 \subseteq S$ two disjoint sets, let $T \subseteq M$. Then: $T(S_0) \cap T(S_1) = \emptyset$.

Proof.

For the sake of contradiction assume $\tau \in T(S_0) \cap T(S_1)$. By Definition C, $\delta_0 \subseteq \tau$ and $\delta_1 \subseteq \tau$ for some $\delta_0 \in S_0, \delta_1 \in S_1$. But then $\delta_0 \cup \delta_1 \in M$ which contradicts k-completeness of S, as necessarily $\delta_0 \neq \delta_1$. ■

Lemma C5.

Let $S, T \subseteq M, S$ be k-complete, $||T|| + k \leq n, S \triangleleft T$ and $S_0 \subseteq S$. Then: $T(S \backslash S_0) = T \backslash T(S_0)$.

Proof:

By Lemma $C1, T(S) = T$. By Lemma $C4, T(S)$ is a disjoint union of $T(S_0)$ and $T(S \backslash S_0)$. Hence $T(S \backslash S_0) = T \backslash T(S_0)$. ■

Now we approach the technical heart of the paper, a space of random maps and a lemma inspired by similar results in boolean complexity.

Definition D.

(a) Let $0 < p < 1$. R_p^n is the probability space of maps $\rho \in M$ which are determined by the following process.

 (i) First form $rng(\rho)$ by putting any $j \in R$ into it with probability $1-p$, i.e. $\Pr (j \in rng(\rho)) = 1 - p$.

 (ii) Then randomly choose a bijection ρ from a random subset $X \subseteq D, |X| = |rng(\rho)|$ to $rng(\rho)$ with uniform distribution.

(b) For $\rho, h \in M, h^\rho$ is undefined if $h \cup \rho \notin M$ and, if $h \cup \rho \in M, dom(h^\rho) = dom(h) \backslash dom(\rho)$ and $h^\rho = h\rho$. Also, $D^\rho = D \backslash dom(\rho), R^\rho = R \backslash rng(\rho)$ and $(n)^\rho = |R^\rho|$.

 For $H \subseteq M, H^\rho = \{h^\rho | h \in H$ and h^ρ is defined $\}$.

Note: "h^ρ undefined" and "$h^\rho = \emptyset$" are different things. ■

In the proof of the theorem we shall be forced to move from a situation with n, D, R, M and some $H, S, T, \ldots \subseteq M$ to a situation with $(n)^\rho$, D^ρ, R^ρ, M^ρ and $H^\rho, S^\rho, T^\rho, \ldots$, by choosing random $\rho \in R_p$, while preserving some properties. That is guaranteed by the next lemma.

Lemma D1.

Let $H, S, K \subseteq M$ and $\rho \in M$ be arbitrary. Then:

(a) $H \triangleleft S$ implies $H^\rho \triangleleft S^\rho$,

(b) S k-complete and $|\rho| + k \leq n$ implies that S^ρ is k-complete,

(c) $K = S(H)$ and $H \triangleleft S$ implies $K^\rho = S^\rho(H^\rho)$.

Proof:

(a) Let $h \in H, \delta \in S$ be such that $h^\rho \cup \delta^\rho \in M^\rho$. Then $h \cup \delta \in M$ and so $h' \subseteq \delta$ for some $h' \in H$. Then $(h')^\rho \subseteq \delta^\rho$.

(b) Let $\delta_1^\rho \cup \delta_2^\rho \in M^\rho$ for some $\delta_1, \delta_2 \in S$. Then $\delta_1 \cup \delta_2 \subseteq M$ so $\delta_1 = \delta_2$, i.e. $\delta_1^\rho = \delta_2^\rho$. To verify the second condition of Definition A(b) let $|h| + k \leq (n)^\rho$ for some $h \in M^\rho \subseteq M$. As $|\rho| = n - (n)^\rho$ we have also $|h \cup \rho| + k \leq n$. Hence by k-completeness of S for some $\delta \in S, (h \cup \rho) \cup \delta \in M$. But then $h \cup \delta^\rho \in M^\rho$ as $h = h^\rho$.

Finally, $||S^\rho|| \leq ||S|| \leq k$.

(c) Let $\kappa^\rho \in K^\rho$, some $\kappa \in K$. Then $\kappa \in S$ and $h \subseteq \kappa$ for some $h \in H$, which gives $\kappa^\rho \in S^\rho$ and $h^\rho \subseteq \kappa^\rho$, i.e. $\kappa^\rho \in S^\rho(H^\rho)$.

Now let $\delta^\rho \in S^\rho, h^\rho \in H^\rho$ s.t. $h^\rho \subseteq \delta^\rho$. Then $h \cup \delta \in M$ and, as $H \triangleleft S, h' \subseteq \delta$ for some other $h' \in H$. So $\delta \in S(H)$, i.e. $\delta \in K$, and therefore $\delta^\rho \in K^\rho$. ■

Lemma D2 (Switching Lemma).
Let $H \subseteq M, ||H|| \leq t \leq s$. Assume that $p < \frac{1}{100}$ and $pn \geq 40s$. Then for random $\rho \in R_p$ the statement:

"there is $2s$-complete $S \subseteq M^\rho$ such that $H^\rho \triangleleft S$ "

holds with probability at least:

$$1 - e(16p^4n^3t)^s - 2^{-cpn} .$$

where $c > 0$ is a constant.

This remains true even if we add the condition $|\rho| \leq n - \frac{1}{2}pn$. For the choice of $p = n^{\varepsilon - 1}$ and $t = s = n^\delta$ such that $0 < \delta < \varepsilon < \frac{1}{5}$, this probability is at least $1 - 2^{-n^\delta}$ for n sufficiently large. ■

The proof of this lemma will occupy the next section.

4 The proof of Lemma D2

The proof is based on an unpublished work of *Woods* which, in turn, builds on earlier work by *Yao* [22], *Cai* [9] and *Håstad* [14], in order to get exponential bounds rather than just the superpolynomial bounds given by the switching lemma of *Ajtai* [1].

For $h \in M$ define:

$$\mu(h) = Pr(\rho = h)$$

and define the *support* of h to be:

$$supp(h) = dom(h) \cup rng(h) .$$

The following lemma follows directly from the definition of R_p^n.

Lemma E1 Let $\emptyset \neq g \in M$, $h \in M$ be such that $supp(h) \cap supp(g) = \emptyset$ and let $X \subseteq M$. Then:

(a) $\mu(h) = \dfrac{p^{n-|h|}(1-p)^{|h|}}{(n+1)_{|h|}}$.

(b) $\mu(h \cup g) = \mu(h)\, \dfrac{p^{-|g|}(1-p)^{|g|}}{(n-|h|+1)_{|g|}}$.

(c) $Pr(\rho \in X) = \sum_{h \in X} \mu(h)$.

Here $(n)_k$ denotes the falling factorial $n(n-1)\ldots(n-k+1)$.

■

Let $m(\rho) = n - |\rho|$ be the number of elements of R on which ρ^{-1} is undefined. The next lemma is an instance of the well-known bounds on the tails of a binomial distribution, such as the *Chernoff inequality*. See [3] or [6].

Lemma E2. There is a constant $c > 0$, such that for $0 < p < \frac{1}{2}$ and n for which $pn > 36$ it holds:

$$Pr\Big(\frac{pn}{2} \le m(\rho) \le \frac{4pn}{3} \Big) \;\ge\; 1 - 2^{-cpn} .$$

■

A property $E(\rho)$ of $\rho \in M$ will be called *positive* if whenever $\sigma \subseteq \rho$, and $E(\sigma)$ holds, $E(\rho)$ holds.

We shall later use one more technical lemma. Before stating it, let us motivate it heuristically. We will be interested in probabilities roughly of the form

$$Pr(\, supp(\rho) \cap U = \emptyset \mid h \cup \rho \in M \,) ,$$

where $supp(h) \subseteq U$ and typically $U = supp(\delta)$ for some $\delta \in M$, so $|U| = 2k$ with $k = |\delta|$. Now most ρ have $m(\rho) \approx pn$. If we restrict our considerations to such maps ρ (so the $Pr(\ )$ notation will temporarily be used rather loosely) then for $x \in D$, $y \in R$, we have

$$Pr(\, x \notin dom(\rho)\,) \approx p \approx Pr(\, y \notin rng(\rho)\,) .$$

Assuming these events are roughly independent, and provided k is significantly smaller than pn, we expect that $Pr(supp(\rho) \cap U = \emptyset) \approx p^{2k}$ (since $|U| = 2k$). Also for $p^2 n \ll 1$, $Pr(\, h \cup \rho \in M\,) \approx Pr(\, h \subseteq \rho\,) \approx n^{-s}$, where $s = |h|$. Therefore we guess

$$Pr(\, supp(\rho) \cap U = \emptyset \mid h \cup \rho \in M\,) \approx p^{2(k-s)}\,(np^2)^s .$$

From the definition of conditional probability as a ratio, conditioning on $supp(\rho) \cap V = \emptyset$ where $V \cap U = \emptyset$ and V has similar properties to U, should not change this. Also it seems plausible that conditioning on a positive condition E (which will typically mean ρ has to be *defined* on certain points) should normally not significantly increase the probability of ρ, ρ^{-1} being *undefined* on U.

For several reasons considerable care is required in making these ideas precise. In particular it seems necessary to be careful how $m(\rho)$ is restricted. To this end we will consider conditional probabilities $Pr'(A)$ defined by

$$Pr'(A) = Pr(\ A \mid m(\rho) \leq \tfrac{4}{3}\, pn\ ).$$

We will adopt the convention that $Pr(A|B) = 0$ if $Pr(B) = 0$.

Lemma E3. Let $U, V \subseteq D \cup R$, $U \cap V = \emptyset$, $k = |U \cap D| = |U \cap R|$, $j = |V \cap D| = |V \cap R|$, and let $h \in M$, $supp(h) \subseteq U$, $s = |h|$.

Suppose $p \leq \frac{1}{100}$, $pn \geq 2$, and let $F(\rho)$ denote the event

$$supp(\rho) \cap V = \emptyset \ \wedge\ E(\rho)\ \wedge\ m(\rho) \geq 10J\ .$$

where E is positive and $J \geq j$. Then

$$Pr'(supp(\rho) \cap U = \emptyset\ \wedge\ m(\rho) \geq 10(J+k) \mid h \cup \rho \in M\ \wedge\ F(\rho)) \leq (2p^2n)^s (2p^2)^{k-s}.$$

Proof: Let $G(\rho)$ hold if and only if $E(\rho) \wedge m(\rho) \leq \frac{4}{3}\, pn$. Clearly the conditional probability in question is either 0 or bounded above by

$$\frac{Pr(\ supp(\rho) \cap (U \cup V) = \emptyset\ \wedge\ G(\rho)\ \wedge\ m(\rho) \geq 10(J+k)\ )}{Pr(\ h \subseteq \rho\ \wedge supp(\rho) \cap V = \emptyset\ \wedge\ G(\rho)\ \wedge\ m(\rho) \geq 10J\ )}\ ,$$

so it suffices to show this ratio is at most $(2p^2n)^s(2p^2)^{k-s}$.

Given ρ there is a unique maximal $\rho' \subseteq \rho$ such that $supp(\rho') \cap U = \emptyset$. Consequently

$$\begin{aligned} Pr(\ h \subseteq \rho\ \ \wedge\ \ & supp(\rho) \cap V = \emptyset\ \wedge\ G(\rho)\ \wedge\ m(\rho) \geq 10J\ ) \\ \geq\ \ & \sum_{\sigma} Pr(\ h \subseteq \rho\ \wedge\ supp(\rho) \cap V = \emptyset\ \wedge\ \rho' = \sigma\ ) \quad (*) \end{aligned}$$

where the sum is over all $\sigma \in M$ satisfying

$$supp(\sigma) \cap (U \cup V) = \emptyset\ \wedge\ G(\sigma)\ \wedge\ m(\sigma) \geq 10(J+k)\ .$$

For certainly if $\rho' = \sigma$ for some such σ, then as G is positive, ρ satisfies

$$G(\rho)\ \wedge\ m(\rho) \geq 10J\ .$$

For each such σ,

$$\begin{aligned} Pr(\ h \subseteq \rho\ \wedge\ supp(\rho) \cap V = \emptyset\ \wedge\ \rho' = \sigma\ )\ \ \geq\ \ & \sum_g Pr(\ h \cup g \subseteq \rho\ \wedge\ \rho' = \sigma\ ) \\ =\ & \sum_g Pr(\ \rho = h \cup g \cup \sigma\ ) \end{aligned}$$

where the sum is over all $g \in M$ with $|g| = 2(k-s)$ having the property that $supp(g) \cap supp(h) = \emptyset$, $supp(g) \cap V = \emptyset$ and $g' = \emptyset$, that is, that if $\langle x, y\rangle \in g$ then either

(i) $x \in U \setminus supp(h)$ and $y \in R \setminus (U \cup V)$, or

(ii) $y \in U \setminus supp(h)$ and $x \in D \setminus (U \cup V)$.

If $m(\sigma) = m$, then there are

$$(m-k-j)_{k-s}(m+1-k-j)_{k-s} \geq (m-2k-j)^{2(k-s)}$$

such maps g. For each of them

$$Pr(\,\rho = h \cup g \cup \sigma\,) = \mu(h \cup g \cup \sigma) = \left(\frac{1-p}{p}\right)^{2(k-s)+s} \frac{\mu(\sigma)}{(m+1)_{2(k-s)+s}}$$

by Lemma E1. Therefore

$$Pr(\,h \subseteq \rho\ \wedge supp(\rho) \cap V = \emptyset \wedge\ \rho' = \sigma\,) \geq \left(\left(\frac{1-p}{p}\right)\frac{m-2k-j}{m+1}\right)^{2(k-s)} \left(\frac{1-p}{p\,(m+1)}\right)^{s} \mu(\sigma)\ .$$

But $m = m(\sigma) \geq 10(k+j)$ so

$$\frac{m-2k-j}{m+1} \geq \frac{10(k+j)-2(k+j)}{10(k+j)+(k+j)} = \frac{8}{11} \geq \frac{1}{\sqrt{2}\,(1-p)}\ .$$

Also as $m = m(\sigma) \leq \frac{4}{3}pn$ by $G(\sigma)$, we see that

$$m+1 \leq 2\,(1-p)\,pn\ .$$

Therefore

$$Pr(\,h \subseteq \rho\ \wedge\ supp(\rho) \cap V = \emptyset\ \wedge\ \rho' = \sigma\,) \geq (2p^2)^{-(k-s)}(2\,p^2 n)^{-s}\mu(\sigma)\ .$$

Summing over σ, the lemma now follows immediately from $(*)$, since

$$\sum_{\sigma} \mu(\sigma) = Pr(\,supp(\rho) \cap (U \cup V) = \emptyset\ \wedge\ G(\rho)\ \wedge\ m(\rho) \geq 10(J+k)\,)\ .$$

$\blacksquare$

We shall divide the proof of Lemma D2 into several steps.

(1) Let $H \subseteq M$ with $||H|| \leq t \leq s$ and let $h^1, h^2, h^3, \ldots$ be any fixed enumeration of elements of H.

(2) Two players I and II will play a game in which they construct a sequence of maps $\delta_0 \subseteq \delta_1 \subseteq \ldots$, all from M, $\delta_0 = \emptyset$.

At a step of the game when $\delta_0 \subseteq \delta_1 \subseteq \ldots \subseteq \delta_l$ has already been constructed, player I plays h_{l+1}: the first h in the enumeration $h^1, h^2, \ldots$ such that $h \cup \delta_l \in M$. Player II then chooses any $\subseteq$-minimal $\delta_{l+1} \supseteq \delta_l$ such that $dom(h_{\ell+1}) \subseteq dom(\delta_{l+1})$ and $rng(h_{\ell+1}) \subseteq rng(\delta_{l+1})$. The game ends iff either I has nothing to play or II choses $\delta_{l+1} \supseteq h_{l+1}$. Note that the strategy of I is fixed while II can use different strategies.

(3) Let $h_1, h_2, \ldots, h_{k+1}$ be the elements of H chosen by player I (by his fixed strategy) in a game determined by $\delta_0 \subseteq \delta_1 \subseteq \ldots \subseteq \delta_{k+1}$. We shall call h_i the *i-th critical map.*

(4) Put S to be $S = \{\delta_{k+1} |$ some $\delta_0 \subseteq \delta_1 \subseteq \ldots \subseteq \delta_{k+1}$ is a sequence constructed in a finished game$\}$.
Claim: S is $||S||$-complete.
Proof of the claim: Assume $\delta \neq \delta'$ are two elements of S. Let $\delta_l \neq \delta'_l$ be the first move of II which was different in the two games. As both δ_l, δ'_l are $\subseteq$-minimal, clearly $\delta_l \cup \delta'_l \notin M$, hence $\delta \cup \delta' \notin M$.

Now let $h \in M$ such that $|h| + ||S|| \leq n$ be given, and let player II follows the strategy: on $dom(h)$ or $rng(h)$ answer according to h, otherwise arbitrarily but consistently with h. As $|h| + ||S|| \leq n$, II can always apply this strategy. Obviously $h \cup \delta \in M$ for any output δ of such a game.

(5) Claim: $H \lhd S$.
Proof of the claim: S is a refinement of H by the definition of the termination of a game: in the first $h \cup \delta \notin M$ for all $h \in H$ and in the second $h \subseteq \delta$ for some $h \in H$.

(6) Unfortunately it is not true in general that $||S|| \leq 2s$ and we have to employ random map ρ from R_p to achieve this. Note that, by Lemma $D1$, both Claims (4) and (5) will remain valid after the application of $\rho \in R_p$.

(7) Now we consider the game played with H^ρ instead of H, thus some $h \in H$ will be discarded (if $\rho \cup h \notin M$), and some $h, g \in H, h \neq g$ may became identical, i.e. $h^\rho = g^\rho$. However we shall still use the order of H induced on H^ρ by taking the first element in each class of the identified mappings. It is useful to think of an element $g \in H^\rho$ as h^ρ where h is the first element of H compatible with ρ. Thus the strategy of the player I can be thought of as: "Play the first $h \in H$ such that $h \cup \delta_\ell \cup \rho \in M$."

(8) Suppose we played a game and obtained critical mappings $h_1^\rho, \ldots, h_{k+1}^\rho$ and $\delta_0 \subseteq \delta_1 \subseteq \ldots \subseteq \delta_{k+1}$. The elements of $h_i^\rho \backslash \delta_{i-1}$ will be called *critical pairs.* We want to bound the probability that there are at least s critical pairs. Therefore we assign a set of parameters for each such game with at least s critical pairs and split the computation according to those parameters.
Let us denote by

$$\delta'_i = \delta_i - \delta_{i-1}, (\delta'_0 = \emptyset);$$

$$\delta^*_i = (dom(h_i) \times rng(h_i)) \cap \delta'_i = (dom(h_i^\rho) \times rng(h_i^\rho)) \cap \delta'_i.$$

Note that this equality holds as $supp(\delta_{k+1}) \cap supp(\rho) = \emptyset$. Let

$$s_i = |h_i^\rho \backslash \delta_{i-1}| \text{ for } 1 \leq i \leq k ,$$

i.e. the number of critical pairs in h_i^ρ and

$$s_{k+1} = s - s_1 \ldots - s_k.$$

Let $d_i = |\delta_i^*|$; note that $|\delta_i'| = 2s_i - d_i$. The sets $T_1, \ldots, T_{k+1} \subseteq \{1, \ldots, t\}$ are defined as follows. Let $i \leq k$ be fixed and suppose

$$h_i = \{e_1, \ldots, e_{t_i}\},$$

then

$$h_i^\rho \backslash \delta_{i-1} = \{e_j\}_{j \in T_i}.$$

Thus h_i and T_i determine the critical pairs of h_i^ρ. For $i = k+1$ we define T_{k+1} in the same way, except that we take only the initial part of $h_{k+1}^\rho \backslash \delta_k$ consisting of s_{k+1} pairs. This is because we want to consider only what happens on the first s critical pairs. Let $\gamma_1, \ldots, \gamma_k$ be the partial one-to-one mappings with domain and range contained in $\{1, \ldots, t\}$ defined as follows. Let $i \leq k$ be fixed and suppose

$$dom(h_i) = \{a_1, \ldots, a_{t_i}\}, a_1 < \ldots < a_{t_i};$$

$$rng(h_i) = \{b_1, \ldots, b_{t_i}\}, b_1 < \ldots < b_{t_i}.$$

Then

$$\delta_i^* = \{(a_l, b_{\gamma_i(l)}) | l \in dom(\gamma_i)\}.$$

Thus h_i and γ_i determine δ_i^* for $i \leq k$.
Finally let $\beta_1, \ldots, \beta_k$ be the mappings from $\{1, \ldots, 2(s_i - d_i)\}$ into $\{1, \ldots, n+1\}$ such that for $i = 1, \ldots, k$ β_i determines $\delta_i' - \delta_i^*$. Namely, let i be fixed, let

$$\{a_1', \ldots, a_{s_i - d_i}'\} = (dom(\delta_i') - dom(\delta_i^*)) \cap dom(h_i),\ a_1' < \ldots < a_{s_i-d_i}';$$

$$\{b_1', \ldots, b_{s_i - d_i}'\} = (rng(\delta_i') - rng(\delta_i^*)) \cap rng(h_i),\ b_1' < \ldots < b_{s_i-d_i}'.$$

Then

$$\delta_i(a_j') = \beta_i(j) \text{ for } j = 1, \ldots, s_i - d_i;$$

$$\delta_i^{-1}(b_j') = \beta_i(s_i - d_i + j) \text{ for } j = 1, \ldots, s_i - d_i.$$

(9) The set of parameters π will consist of

$$k;\ s_1, \ldots, s_{k+1};\ T_1, \ldots, T_{k+1};\ \gamma_1, \ldots, \gamma_k;\ \beta_1, \ldots, \beta_k.$$

We shall estimate the number of possible sets π for fixed $s_1, \ldots, s_{k+1}$, $s_1 + \ldots + s_{k+1} = s$, and $d_1, \ldots, d_k$. The number of $T_1, \ldots, T_{k+1}$ is at most

$$\binom{t_1}{s_1} \cdots \binom{t_{k+1}}{s_{k+1}} \leq \binom{t}{s_1} \cdots \binom{t}{s_{k+1}} \leq t^s.$$

The number of $\gamma_1, \ldots, \gamma_k$ is at most

$$\binom{t_1}{d_1} \cdot t_1^{d_1} \cdot \ldots \cdot \binom{t_k}{d_k} \cdot t_k^{d_k} \leq t^{2d},$$

where $d = d_1 + \ldots + d_k$. The number of $\beta_1, \ldots, \beta_k$ is at most

$$(n+1)^{2(s_1-d_1)} \cdot \ldots \cdot (n+1)^{2(s_k-d_k)} = (n+1)^{2(s'-d)}$$

where $s' = s_1 + \ldots + s_k$. Altogether we get an upper bound

$$t^{s+2d} \cdot (n+1)^{2(s'-d)}$$

(10) Let a set of parameters π be given. We shall estimate the Pr' probability of the event A^π that a game with parameters π occurs and $m(\rho) \geq 20s$. (Recall that $m(\rho) = n - |\rho|$. The technical condition that $m(\rho) \geq 20s$ is included in A^π so that we will ultimately be able to apply lemma E3.)

Let $C_1^\pi(h)$ denote the event that h is the first $h \in H$ consistent with ρ (i.e. $h \cup \rho \in M$). In general, let $C_i^\pi(h)$ be the event that Player I plays h^ρ in round i in a game played according to π till Player I chooses h in the i-th step. (Recall that this implies that there is no $g \in H$ before h with $g^\rho = h^\rho$.) Thus $C_i^\pi(h)$ does not exclude that Player II cannot play according to π already in the i-th move nor does it exclude that $h^\rho = \emptyset$; however it implies that $h \cup \rho \in M$. (π determines $\delta_0, \delta_1, \ldots, \delta_k$ and hence the moves of Player II if H^ρ is known.)

Let $P_i^\pi(h), 1 \leq i \leq k$ be the event that $m(\rho) \geq 10(s_1 + \ldots + s_i)$ and in some game played according to π till the i-th round, h^ρ was played by Player I, the critical pairs of h^ρ are determined from h by the parameter T_i and Player II can play his i-th move according to π. The last condition means that the δ'_i determined by π and h is consistent with δ_{i-1} and $supp(\delta'_i) \cap supp(\rho) = \emptyset$. Let $P_{k+1}^\pi(h)$ be the event that $m(\rho) \geq 20s$ and in some game played according to π till $k+1$-st round h^ρ was played by Player I and there are (at least) s_{k+1} critical pairs in h determined by T_{k+1}. In the sequel we shall abbreviate A^π by $A, C_i^\pi(h_i)$ by C_i and $P_i^\pi(h_i)$ by P_i. We have

$$Pr'(A) \leq \max_{h_1} Pr'(A|C_1);$$

$$Pr'(A|C_1) = Pr'(A \wedge P_1|C_1) = Pr'(A|P_1 \wedge C_1) \cdot Pr'(P_1|C_1);$$

$$Pr'(A|P_1 \wedge C_1) \leq \max_{h_2} Pr'(A|C_2 \wedge P_1 \wedge C_1)$$

$$Pr'(A|C_2 \wedge P_1 \wedge C_1) = Pr'(A|P_2 \wedge C_2 \wedge P_1 \wedge C_1) \cdot Pr'(P_2|C_2 \wedge P_1 \wedge C_1) \text{ etc.}$$

Eventually we get

$$Pr'(A|P_k \wedge C_k \wedge \ldots \wedge P_1 \wedge C_1) \leq$$

$$\leq \max_{h_{k+1}} Pr'(A|C_{k+1} \wedge P_k \wedge C_k \wedge \ldots \wedge P_1 \wedge C_1) =$$

$$= \max_{h_{k+1}} Pr'(A \wedge P_{k+1}|C_{k+1} \wedge P_k \wedge C_k \wedge \ldots \wedge P_1 \wedge C_1) =$$

$$= \max_{h_{k+1}} Pr'(P_{k+1}|C_{k+1} \wedge P_k \wedge C_k \wedge \ldots \wedge P_1 \wedge C_1),$$

since A is implied by $P_{k+1} \wedge C_{k+1} \wedge \ldots \wedge P_1 \wedge C_1$. Thus

$$Pr'(A^\pi) \leq$$

$$\max_{h_1,\ldots,h_{k+1}} \prod_{i=1}^{k+1} Pr'(P_i^\pi(h_i)|C_i^\pi(h_i) \wedge P_{i-1}^\pi(h_{i-1}) \wedge C_{i-1}^\pi(h_{i-1}) \wedge \ldots$$

$$\ldots \wedge P_1^\pi(h_1) \wedge C_1^\pi(h_1))$$

(11) The reason for the above decomposition is that under the condition, $C_i \wedge P_{i-1} \wedge C_{i-1} \wedge \ldots \wedge P_1 \wedge C_1$, the mappings δ_{i-1} and $h_1, \ldots, h_i$ are determined. We shall estimate first the i-th term in the last product for $i \leq k$. If π determines δ_i' to be inconsistent with δ_{i-1}, the probability is zero. Otherwise P_i (under the condition) is equivalent to the condition $supp(\delta_i') \cap supp(\rho) = \emptyset \ \wedge \ m(\rho) \geq 20(s_1 + \ldots + s_i)$. Let us denote by $h_i' = h_i \backslash \delta_{i-1}$. The condition $C_i \wedge P_{i-1} \wedge \ldots \wedge C_1$ implies the condition $h_i' \cup \rho \in M \wedge m(\rho) \geq 20(s_1 + \ldots + s_{i-1})$. Thus we have

$$Pr'(P_i|C_i \wedge P_{i-1} \wedge \ldots \wedge C_1) \leq$$

$$Pr'(\ supp(\delta_i') \cap supp(\rho) = \emptyset \ \wedge \ m(\rho) \geq 20(s_1 + \ldots + s_i) \mid h_i' \cup \rho \in M \ \wedge \ F\ )\ ,$$

where F is

$$m(\rho) \geq 20(s_1 + \ldots + s_{i-1}) \wedge C_i \wedge P_{i-1} \wedge \ldots \wedge C_1\ .$$

Assuming $h_i' \cup \rho \in M \ \wedge \ m(\rho) \geq 20(s_1 + \ldots + s_{i-1})$ the condition F is equivalent to the conjunction of the following three conditions:

$B_1 :\quad g \subseteq \rho;$

$B_2 :\quad supp(\rho) \cap supp(\delta_{i-1}) = \emptyset;$

$B_3 :\quad \forall h \in K(\ \rho \cup h \notin M\ ).$

In B_1, $g \subseteq h_1 \cup \ldots \cup h_{i-1}$ are those pairs which should belong to ρ according to π. In B_3, K consists of those $h \in H$ which for some $j \leq i$ are before h_j, after h_{j-1} (if $j > 1$), and have $h \cup \delta_{j-1} \in M$. Note that now g and δ_{i-1} are fixed, i.e. do not depend on ρ.

As F is equivalent to

$$supp(\rho) \cap supp(\delta_{i-1}) = \emptyset \ \wedge \ B_1 \wedge B_3 \ \wedge \ m(\rho) \geq 20(s_1 + \ldots + s_{i-1})$$

and as $B_1 \wedge B_3$ is positive, the probability

$$Pr'(supp(\delta_i') \cap supp(\rho) = \emptyset \ \wedge \ m(\rho) \geq 20(s_1 + \ldots + s_i) \mid h_i' \cup \rho \in M \wedge F\ )\ .$$

can be estimated using Lemma E3 to be:

$$\leq (2p^2n)^{s_i} \cdot (2p^2)^{(2s_i - d_i - s_i)} = (2p^2n)^{s_i}(2p^2)^{(s_i - d_i)}.$$

(Recall that $|supp(\delta_i')| = 2s_i - d_i$ and $|h_i'| = s_i$. Also, observe that on taking $J = 2(s_1 + \ldots + s_{i-1})$, we have $|\delta_{i-1}| \leq J$, and $m(\rho) \geq 20(s_1 + \ldots + s_i)$ implies $m(\rho) \geq 10(J + 2s_i - d_i)$.)

(12) The last factor in the estimate for $Pr'(A^\pi)$ in (10) can be bounded in the same way by $(2p^2n)^{s_{k+1}}$.

(13) Now we get a bound for $Pr'(A^\pi)$ if $s_1, \ldots, s_{k+1}$ and $d_1, \ldots, d_k$ are fixed:

$$Pr'(A^\pi) \leq \left(\prod_{i=1}^{k} (2p^2n)^{s_i} (2p^2)^{(s_i - d_i)} \right) \cdot (2p^2n)^{s_{k+1}} \leq (2p^2n)^s \cdot (2p^2)^{(s'-d)}.$$

Multiplying it by the estimate from (9) we get an upper bound on the Pr' probability that $m(\rho) \geq 20s$ and there are at least s critical pairs, assuming the game was played with fixed parameters $s_1, \ldots, s_{k+1}$ and $d_1, \ldots, d_k$:

$$t^{s+2d} \cdot (n+1)^{2(s'-d)} \cdot (2p^2n)^s \cdot (2p^2)^{(s'-d)} \leq$$

$$\leq (4p^2nt)^s (pn)^{2s'} \cdot (1 + \frac{1}{n})^{2s'} \cdot \left(\frac{t}{(n+1) \cdot p} \right)^{2d} \leq$$

$$\leq e(4p^4n^3t)^s,$$

(provided that $pn \geq 1$, $2s' \leq n$ and $t \leq (n+1)p$ which follow from our assumptions). The number of all sequences $s_1, \ldots, s_{k+1}$ such that $1 \leq s_1, \ldots, 1 \leq s_{k+1}$, $s_1 + \ldots + s_{k+1} = s$ is 2^{s-1}. (Such decompositions $s_1 + \ldots s_{k+1} = s$ are in a 1-to-1 correspondence with subsets $\{s_1, s_1 + s_2, \ldots, s\}$ of $\{1, \ldots, s\}$ containing the element s.) Notice that we must have $s_i \geq 1$ for $i \leq k$, since otherwise $h_i^\rho \subseteq \delta_{i-1}$ so $\delta_i = \delta_{i-1}$ and the game terminates at round i contrary to the assumption that $s_1 + \ldots s_{k+1} = s$. Similarly we can assume k is chosen such that $s_{k+1} \geq 1$. The number of all sequences $d_1, \ldots, d_k$ such that $0 \leq d_1 \leq s_1, \ldots, 0 \leq d_k \leq s_k$ is $(1 + s_1) \cdot \ldots \cdot (1 + s_k)$ which is maximal when $s_1 = s_2 = \ldots = s_k = 1$ and so is at most 2^s. Thus the Pr' probability that $m(\rho) \geq 20s$ and there are at least s critical pairs is at most

$$e(16p^4n^3t)^s.$$

If δ_{k+1} has $\geq 2s$ elements, then there must be at least s critical pairs. Thus this is also a bound to the probability that $||S|| \geq 2s$.

Recalling the definition of Pr' in terms of Pr, and using the assumption that $20s \leq \frac{1}{2}\, pn$, this implies that

$$Pr(\ ||S|| \geq 2s \ \wedge \ \tfrac{1}{2}\, pn \leq m(\rho) \leq \tfrac{4}{3}\, pn \ ) \ \leq \ e(16p^4n^3t)^s.$$

Application of Lemma E2 then completes the proof of Lemma D2. ∎

5 The proof of the theorem

Recall that we consider only formulas in the language $\vee, \neg, 0, 1, p_{ij}, i \in D, j \in R, |D| = n+1, |R| = n$. Let φ be a disjunction. The *reduced form* of φ will be the expression $\bigvee_{i \in I} \varphi_i$ where each φ_i is either a negated formula or a variable and φ is obtained from $\varphi_i, i \in I$ by applying the binary $\vee$ in a suitable order. Equivalently φ_i can be determined as the maximal subformulas of φ whose depth is less than the depth of φ.

Definition F.

Let Γ be a set of formulas, Γ closed under subformulas. A *k-evaluation* of Γ is a pair of mappings $(\mathbf{H}, \mathbf{S})$:

$$\mathbf{H} : \Gamma \to P(M), \mathbf{S} : \Gamma \to P(M)$$

such that

1. for every $\varphi \in \Gamma, \mathbf{H}_\varphi \subseteq \mathbf{S}_\varphi \subseteq M$ and $\mathbf{S}_\varphi$ is k-complete;
2. $\mathbf{S}_0 = \mathbf{S}_1 = \{\emptyset\}$ and $\mathbf{H}_0 = \emptyset, \mathbf{H}_1 = \mathbf{S}_1$,
   $\mathbf{H}_{p_{ij}} = \{\{(i,j)\}\}$;
   $\mathbf{S}_{p_{ij}} = \{\{(i,j'),(i',j)\} | i' \neq i, j' \neq j\} \cup \{\{(i,j)\}\}$;
3. if $\neg\varphi \in \Gamma$, then $\mathbf{H}_{\neg\varphi} = \mathbf{S}_\varphi \setminus \mathbf{H}_\varphi,\ \mathbf{S}_{\neg\varphi} = \mathbf{S}_\varphi$;
4. if $\varphi \in \Gamma$ and $\bigvee_{i \in I} \varphi_i$ is the reduced form of φ, then

$$\bigcup_{i \in I} \mathbf{H}_{\varphi_i} \triangleleft \mathbf{S}_\varphi \text{ and } \mathbf{H}_\varphi = \mathbf{S}_\varphi \left(\bigcup_{i \in I} \mathbf{H}_{\varphi_i} \right) .$$

We use the letters $\mathbf{H}, \mathbf{S}$ to denote the maps of a k-evaluation as S previously denoted k-complete systems and $\mathbf{S}_\varphi$ is always a k-complete system, and H was used to denote an arbitrary subset of M.

Let $\rho \in M$. We define

$$(p_{ij})^\rho = \begin{array}{ll} 1 & \text{if } \rho(i) = j \\ 0 & \text{if } i \in \mathrm{dom}(\rho), \text{ but } \rho(i) \neq j, \\ 0 & \text{if } j \in \mathrm{rng}(\rho), \text{ but } \rho(i') = j \text{ for } i' \neq i, \\ p_{ij} & \text{otherwise.} \end{array}$$

If φ is a formula, then φ^ρ is obtained by applying ρ to all variables of φ; that is by replacing each p_{ij} by $(p_{ij})^\rho$ but performing no further simplifications. If Γ is a set of formulas, then $\Gamma^\rho = \{\varphi^\rho | \varphi \in \Gamma\}$.

Let $(\mathbf{H}, \mathbf{S})$ be a k-evaluation and $\rho \in M$. We define the restriction $(\mathbf{H}^\rho, \mathbf{S}^\rho)$ of the k-evaluation as follows. For $\varphi \in \Gamma$ define:

1. If φ^ρ is different from all formulas of the form $0, \neg 0, \neg\neg 0, \ldots$ then put

$$\mathbf{S}^\rho_\varphi := \{h^\rho \mid h \cup \rho \in M \wedge h \in \mathbf{S}_\varphi\}$$

   and

$$\mathbf{H}^\rho_\varphi := \{h^\rho \mid h \cup \rho \in M \wedge h \in \mathbf{H}_\varphi\} .$$

2. If φ^ρ is one of the formulas $0, \neg 0, \neg\neg 0, \ldots$ then put:

$$\mathbf{S}^\rho_\varphi := \{\emptyset\}$$

   and

$$\mathbf{H}^\rho_\varphi := \begin{cases} \emptyset & \text{if } \varphi^\rho = 0, \neg\neg 0, \ldots \\ \{\emptyset\} & \text{if } \varphi^\rho = \neg 0, \neg\neg\neg 0, \ldots \end{cases}$$

The exceptional case for formulas collapsed by ρ to $0, \neg 0, \ldots$ is needed as for some atom p_{ij} it might be that $p^\rho_{ij} = 0$ while $S^\rho_{p_{ij}} \neq \{\emptyset\}$, contradicting clause 2. of Definition F.

Lemma F1 Let Γ be a set of formulas, $\rho \in M$, and $|\rho| + k \leq n$. If $(\mathbf{H}, \mathbf{S})$ is a k-evaluation of Γ, then $(\mathbf{H}^\rho, \mathbf{S}^\rho)$ is a k-evaluation of Γ^ρ.

Proof - use Lemma D1(a) and (b), and observe that the set $\{\emptyset\}$ is refined by any other set. ∎

Lemma F2 Let $d \geq 1$, be an integer, $0 < \varepsilon < \frac{1}{5}, 0 < \delta < \varepsilon^{d-1}$ and let Γ be a set of formulas of depth d, Γ closed under subformulas. Suppose that $|\Gamma| < 2^{n^\delta}$ and n is sufficiently large. Then there exists $\rho \in M, |\rho| \leq n - n^{\varepsilon^{d-1}}$ such that there exists a $\leq 2n^\delta$-evaluation of Γ^ρ.

Proof:
Proceed by induction.

1. For $d = 1$ the only formulas are single variables for which we have $\mathbf{H}_{p_{ij}}$ and $\mathbf{S}_{p_{ij}}$ by (2) of the definition. Clearly $\mathbf{S}_{p_{ij}}$ is 2-complete, hence we have a 2-evaluation, $(\rho = \emptyset)$.

2. Suppose that the lemma is true for d and let Γ be a set of formulas of depth $d+1, \Gamma$ closed under subformulas. Let Δ be the formulas of Γ of depth $\leq d$. Let $0 < \varepsilon^d (= \varepsilon^{d+1-1})$ be given. By the induction assumption we have a $\rho' \in M, |\rho'| \leq n - n^{\varepsilon^{d-1}}$ and a $\leq 2n^\delta$-evaluation $(\mathbf{H}', \mathbf{S}')$ of $\Delta^{\rho'}$. Let $m = n - |\rho'|$, thus $m \geq n^{\varepsilon^{d-1}}$. We shall extend ρ' to a suitable ρ. This can be thought of as applying some ρ'' to the restricted universe given by $D^{\rho'}$ and $R^{\rho'}$. By Lemma F1 the restrictions of $\Delta^{\rho'}$ and $\mathbf{S}'$ will be $\leq 2n^\delta$-evaluations of $\Delta^{\rho'\rho''}$ again, thus we only need to choose ρ'' so that we can extend this evaluation to the whole Γ. For negation it is straightforward for any ρ''. For disjunction we apply Lemma D2 with n, D, R replaced by $m, D^{\rho'}, R^{\rho'}$, and $t = s = n^\delta, p = \frac{1}{2}m^{\varepsilon - 1}$. Let $\varphi \in \Gamma, \varphi$ of depth $d+1$, let $\bigvee \varphi_i$ be the reduced form of φ. Note that

$$n^\delta = n^{\varepsilon^{d-1} \cdot \frac{\delta}{\varepsilon^{d-1}}} \leq m^{\frac{\delta}{\varepsilon^{d-1}}} \text{ and } \frac{\delta}{\varepsilon^{d-1}} < \varepsilon.$$

By Lemma D2, if n is sufficiently large, then with probability $\leq 1 - 2^{-n^\delta}$, there is $S \subseteq M^{\rho'\rho''}$ such that

$$\bigcup_i (\mathbf{H}'_{\varphi i})^{\rho''} \triangleleft S \text{ and } |\rho''| \leq m - 2pm = m - m^\varepsilon.$$

If this is the case, we extend $(\mathbf{H}', \mathbf{S}')$ to φ by defining

$$\mathbf{S}_\varphi = S \text{ and } \mathbf{H}_\varphi = S\left(\bigcup_i (\mathbf{H}'_{\varphi i})^{\rho''}\right).$$

Since $|\Gamma| < 2^{n^\delta}$, there is at least one ρ'' with the above properties satisfied all for such $\varphi \in \Gamma$. Then we have also

$$|\rho| = |\rho'\rho''| \leq n - m + m - m^\varepsilon = n - m^\varepsilon \leq n - n^{\varepsilon^d}.$$

■

Lemma F3. For every Frege system F there exists a constant f with the following property. If $(\gamma_1, \ldots, \gamma_t)$ is an F-proof, $(\mathbf{H}, \mathbf{S})$ is a k-evaluation of the set of all subformulas of the proof and $k \leq n/f$, then $\mathbf{H}_{\gamma_i} = \mathbf{S}_{\gamma_i}$ for $i = 1, \ldots, t$.

Proof:
Let F be given. Let f be the maximal number of subformulas in a rule of F plus 1 (axioms are special cases of rules with $r = 0$). Clearly it suffices to prove the following claim and apply induction:

Claim. Let

$$\frac{\varphi_1(\psi_1, \ldots, \psi_m), \ldots, \varphi_r(\psi_1, \ldots, \psi_m)}{\varphi_0(\psi_1, \ldots, \psi_m)}$$

be an instance of a rule of F. Let $(\mathbf{H}, \mathbf{S})$ be an n/f-evaluation of the subformulas of $\varphi_0(\psi_1, \ldots, \psi_m), \ldots, \varphi_r(\psi_1, \ldots, \psi_m)$. Suppose that $\mathbf{H}_\xi = \mathbf{S}_\xi$ for $\xi = \varphi_i(\psi_1, \ldots, \psi_m), i = 1, \ldots, r$. Then $\mathbf{H}_\xi = \mathbf{S}_\xi$ for $\xi = \varphi_0(\psi_1, \ldots, \psi_m)$.

Proof of Claim. Let $(\mathbf{H}, \mathbf{S})$ be given. Let Γ be the set of all formulas of the form $\varphi(\psi_1, \ldots, \psi_m)$, where $\varphi(q_1, \ldots, q_m)$ a subformula of some $\varphi_i(q_1, \ldots, q_m), i = 0, \ldots, r$. Let T be an $\frac{n(f-1)}{f}$-complete system such that $\mathbf{S}_\xi \triangleleft T$ for every $\xi \in \Gamma$. Such a system exists by Lemma B. Note also that $||\mathbf{S}_\xi|| + ||T|| \leq n$ for every $\xi \in \Gamma$.

Suppose that $\neg\xi \in \Gamma$. Then

$$\mathbf{H}_{\neg\xi} = \mathbf{S}_\xi \setminus \mathbf{H}_\xi \ ,$$

hence, by Lemma C5,

$$T(\mathbf{H}_{\neg\xi}) = T \setminus T(\mathbf{H}_\xi) \ .$$

Suppose that $\alpha, \beta, \alpha \vee \beta \in \Gamma$. Let $\bigvee_{i \in A} \gamma_i$ resp. $\bigvee_{i \in B} \gamma_i$ be the reduced forms of α resp. β. Hence, using Lemma C3,

$$\mathbf{H}_{\alpha\vee\beta} = \mathbf{S}_{\alpha\vee\beta}\left(\bigcup_{i\in A} \mathbf{H}_{\gamma_i}\right) \cup \mathbf{S}_{\alpha\vee\beta}\left(\bigcup_{i\in\beta} \mathbf{H}_{\gamma_i}\right).$$

Now using Lemma C3 and using Lemma C1 twice we get

$$T(\mathbf{H}_{\alpha\vee\beta}) = T(\mathbf{S}_{\alpha\vee\beta}\left(\bigcup_{i\in A}\mathbf{H}_{\gamma_i}\right)) \cup T(\mathbf{S}_{\alpha\vee\beta}\left(\bigcup_{i\in B}\mathbf{H}_{\gamma_i}\right)) =$$

$$= T\left(\bigcup_{i\in A}\mathbf{H}_{\gamma_i}\right) \cup T\left(\bigcup_{i\in B}\mathbf{H}_{\gamma_i}\right) =$$

$$= T(\mathbf{S}_{\alpha}\left(\bigcup_{i\in A}\mathbf{H}_{\gamma_i}\right)) \cup T(\mathbf{S}_{\beta}\left(\bigcup_{i\in B}\mathbf{H}_{\gamma_i}\right)) =$$

$$= T(\mathbf{H}_{\alpha}) \cup T(\mathbf{H}_{\beta}).$$

Furthermore we have, by Lemma C1

$$T(\mathbf{H}_\xi) = T(\mathbf{S}_\xi) = T,$$

for $\xi = \varphi_i(\psi_1,\ldots,\psi_m), i = 1,\ldots,r$, since $\mathbf{H}_\xi = \mathbf{S}_\xi$ and $\mathbf{S}_\xi$ is complete.

Thus we have shown that the mapping $\xi \to T(\mathbf{H}_\xi)$ of Γ into the Boolean algebra of subsets of T has the following properties

1. it maps $\neg$ on the operation of the complement and $\vee$ on the operation of the union:
2. it maps the premises of the rule on T.

Since the rule is sound it preserves the value 1_B in any boolean algebra B (as otherwise a suitable ultrafilter would define a 0-1 assignment satisfying the premises of the rule but not its conclusion, contradicting the soundness). The value 1_B in the boolean algebra $B = exp(T)$ of the subsets of T is the set T and hence we must have $T(\mathbf{H}_\xi) = T$ for $\xi = \varphi_0(\psi_1,\ldots,\psi_m)$, whence by definition and Lemma C2,

$$\mathbf{H}_\xi = \mathbf{S}_\xi(\mathbf{H}_\xi) = \mathbf{S}_\xi\ ,$$

which concludes the proof of the claim and, consequently, of the lemma. ■

Recall the discussion before Lemma A to draw an intuition for the next lemma.

Lemma F4.

(i) Let $(\mathbf{H},\mathbf{S})$ be a k-evaluation of the subformulas of PHP_n and suppose $k \leq \frac{n}{2} - 3$. Then $\mathbf{H}_{PHP_n} = \emptyset$ and hence $\mathbf{H}_{PHP_n} \neq \mathbf{S}_{PHP_n}$.

(ii) If $\rho \in M, k \leq \frac{n-|\rho|}{2} - 3$, then part (i) holds for ${PHP_n}^\rho$.

Proof:

(i) PHP_n is a disjunction of formulas $\neg\varphi$ where φ ranges over

$$\neg p_{ik} \vee \neg p_{jk}, i \neq j \in D, k \in R,$$

$$\neg p_{ki} \vee \neg p_{kj}, i \neq j \in R, k \in D,$$

$$\bigvee_{k \in R} p_{ik}, i \in D,$$

$$\bigvee_{i \in D} p_{ik}, k \in R.$$

We shall show that $\mathbf{H}_\varphi = \mathbf{S}_\varphi$ for each such formula, hence $\mathbf{H}_{\neg\varphi} = \emptyset$, thus

$$\mathbf{H}_{PHP_n} = \mathbf{S}_{PHP_n}(\bigcup \mathbf{H}_{\neg\varphi}) = \mathbf{S}_{PHP_n}(\emptyset) = \emptyset \ .$$

(a) Let φ be $\neg p_{ik} \vee \neg p_{jk}$. By definition

$$\mathbf{H}_{\neg p_{ik}} = \{\{(i,k'),(i',k)\} | i' \neq i, k' \neq k\}$$

and

$$\mathbf{H}_{\neg p_{jk}} = \{\{(j,k'),(j',k)\} | j' \neq j, k' \neq k\}.$$

Let T be the 3-complete set

$$\{\{(i,k'),(j,k''),(l,k)\} | i \neq l \neq j, k' \neq k'' \neq k \neq k'\} \cup$$

$$\cup \{\{(i,k),(j,k'\} | k \neq k'\} \cup \{\{(i,k'),(j,k)\} | k \neq k'\}.$$

It can be easily verified that

$$T(\mathbf{H}_{\neg p_{ik}} \cup \mathbf{H}_{\neg p_{jk}}) = T.$$

By Lemma B, we have some $n/2$-complete W which is a common refinement of $\mathbf{S}_\varphi$ and T. Hence, by Lemma C2,

$$W(\mathbf{H}_{\neg p_{ik}} \cup \mathbf{H}_{\neg p_{jk}}) = W.$$

and again by Lemma C2,

$$\mathbf{H}_{\neg p_{ik} \vee \neg p_{jk}} = \mathbf{S}_{\neg p_{ik} \vee \neg p_{jk}}(\mathbf{H}_{\neg p_{ik}} \cup \mathbf{H}_{\neg p_{jk}}) = \mathbf{S}_{\neg p_{ik} \vee \neg p_{jk}}.$$

(b) Let φ be $\bigvee_{k \in R} p_{ik}$. By definition

$$\mathbf{H}_\varphi = \mathbf{S}_\varphi(\{(i,k) | k \in R\}).$$

But, clearly, $\{(i,k) | k \in R\}$ is 1-complete,
hence

$$\mathbf{S}_\varphi(\{(i,k) | k \in R\}) = \mathbf{S}_\varphi.$$

(c) The case of $\neg p_{ki} \vee \neg p_{kj}$; resp. $\bigvee_{i\in D} p_{ik}$ is proved in the same way as (a) resp. (b).

(ii) The generalization is straightforward: if φ contains a variable which is changed to 0 or 1 by ρ, then all the variables in φ are fixed and one can easily check that $\mathbf{H}_{\neg\varphi} = \emptyset$.

■

Now we are ready to prove our main result. Let F be a Frege system, $d \geq 4$ (4 is the depth of PHP_n), $0 < \delta < 5^{-d+1}$ let n be sufficiently large and let $(\gamma_1, \dots, \gamma_t)$ be an F-proof of depth d and size $\leq 2^{n^\delta}$. Let f be the constant associated to F by Lemma F3. Choose an ε such that $\varepsilon < \frac{1}{5}$ and $\delta < \varepsilon^{d-1}$. By Lemma F2, there exists $\rho \in M, |\rho| \leq n - n^{\varepsilon^{d-1}}$ and a $2n^\delta$-evaluation $(\mathbf{H}, \mathbf{S})$ of Γ^ρ, where Γ is the set of subformulas of the proof $(\gamma_1, \dots, \gamma_t)$. Clearly $(\gamma_1^\rho, \dots, \gamma_t^\rho)$ is an F-proof with variables $p_{ij}, i \in D^\rho, j \in R^\rho$ and Γ^ρ is the set of its subformulas. Let $m = (n)^\rho = n - |\rho| \geq n^{\varepsilon^{d-1}}$. Since n is large, we have

$$2n^\delta \leq \frac{m}{f} \text{ and } 2n^\delta \leq \frac{m}{2} - 3.$$

Thus we can apply Lemma F3 (with n replaced by m etc.) and Lemma F4. By the first one we get

$$\mathbf{H}_{\gamma_i^\rho} = \mathbf{S}_{\gamma_i^\rho} \text{ for } i = 1, \dots, t;$$

by the second one

$$\mathbf{H}_{PHP_n^\rho} = \emptyset$$

which is different from $\mathbf{S}_{PHP_n^\rho}$. Thus $(\gamma_1, \dots, \gamma_t)$ cannot be a proof of PHP_n, i.e. proof of PHP_n of depth d must have size $> 2^{n^\delta}$. ■

6 Some open problems

1. The most important problem is to extend superpolynomial lower bounds to stronger systems. Our result shows that some techniques from boolean complexity theory can be adapted for propositional calculus, but the progress in propositional calculus is much slower. It is, perhaps, because the problems involve besides combinatorial difficulties also unusual logical concepts. In boolean complexity exponential lower bounds are known for bounded depth circuits with $\vee$, $\wedge$, $\neg$ and $PARITY$ gates (in general with MOD_p, p prime, gates). It is an open problem to prove a superpolynomial lower bound for Frege systems with bounded depth formulas of this form.

2. Even in the case of bounded depth proofs with formulas in the basis $\vee$, $\wedge$, $\neg$ there are some open problems. *Krajíček* [15] has shown that there is an exponential gap between depth d and $d+1$, but he needed formulas with increasing depth. Is there a constant c such that for every d there is a sequence of tautologies of depth at most c such that these tautologies have polynomial size proofs of depth $d+1$, but only exponential (or just superpolynomial) size proofs of depth d?

3. In our lower bound $exp(n^{\varepsilon_d})$ the constant ε_d decreases exponentially, while in the corresponding boolean case it decreases only linearly. (This is due to the fact that the bound in our switching lemma depends on n.) Is it possible to improve our bound so that ε_d decreases only polynomially?

Acknowledgement

Most of this work was done during the workshop *"Arithmetic: Proof Theory and Complexity"* held in *Prague* June 15 - July 15,1991, a part of a joint project of ČSAV and NSF (NSF ČSAV grant INT-8914569). We thank to participants of the workshop for numerous discussions and, in particular, to *Sam Buss* with whom we discussed this topic already during the *San Diego* workshop in 1990.

We also thank to the referees for helpful comments.

References

[1] *Ajtai, M.*: "The complexity of the pigeonhole principle", 29^{th} Annual Symposium on the Foundations of Computer Science, (1988), pp. 346-355.

[2] *Ajtai, M.*: "$\sum_1^1$-formulae on finite structures", Annals of Pure and Applied Logic 45, (1983), pp. 1-48.

[3] *Alon, N.- Spencer, J.*: "The Probabilistic Method", Wiley, (1991).

[4] *Beame, P.-Impagliazzo, R.-Krajíček, J. -Pitassi, T.-Pudlák, P.-Woods, A.*: "Exponential lower bound for the pigeonhole principle ", Proc. 24^{th} Annual ACM Symp. on Theory of Computing, (1992), pp.200-221.

[5] *Bellantoni, S. - Pitassi, T. - Urquhart, A.*: "Approximation of small depth Frege proofs", SIAM J. of Computing 21(6), (1992), pp.1161-1179.

[6] *Bollobás, B.*: "Random Graphs", Academic Press, 1985.

[7] *Buss, S.R.*: "Bounded Arithmetic", Bibliopolis, 1986.

[8] *Buss, S.R.*: "Polynomial size proofs of the propositional pigeonhole principle", Journal of Symbolic Logic 52, (1987), pp. 916-927.

[9] *Cai, J*: "With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy", Proc. 18^{th} Annual ACM Symp. on Theory of Computing , (1986), pp. 21-29.

[10] *Cook, S.A.*: "The complexity of theorem proving procedures", 3-rd Symp. on Theor. of Comput., 1971, pp.151-158.

[11] *Cook, S.A. - Reckhow, R.A.:* "The relative efficiency of propositional proof systems", Journal of Symbolic Logic 44(1), (1979), pp.36-50.

[12] *Hájek, P., - Pudlák, P.*: "Metamathematics of First Order Arithmetic", Springer-Verlag, 1993.

[13] *Haken, A.*: "The intractability of resolution", Theoretical Computer Science 39, (1985), pp. 297-308.

[14] *Håstad, J.*: "Almost optimal lower bounds for small depth circuits", in: Advances Comp. Research, Vol.5, JAI Press, (1989), pp.143-170.

[15] *Krajíček, J.*: "Lower bounds to the size of constant-depth propositional proofs", Journal of Symbolic Logic, 59(1), (1994), pp.73-86.

[16] *Krajíček, J.*: "Bounded Arithmetic, Propositional Logic and Complexity Theory", Cambridge Univ. Press, – to appear.

[17] *Paris, J. - Wilkie, A.:* "Counting problems in bounded arithmetic", in: Methods in Mathematical Logic, LNM 1130, Springer (1985), pp. 317-340.

[18] *Paris, J. - Wilkie, A. - Woods. A.:* "Provability of the pigeonhole principle and the existence of infinitely many primes", Journal of Symbolic Logic 53(4), (1988), pp. 1235-1244.

[19] *Pitassi, T. -Beame, P. - Impagliazzo, R.:* "Exponential lower bounds for the Pigeonhole Principle", Computational Complexity 3, (1993), pp.97-108.

[20] *Tseitin, G.S.*: "On the complexity of derivations in the propositional calculus", in: Studies in Mathematics and mathematical logic, Part II, ed. A.O. Slisenko, (1968), pp. 115-125.

[21] *Woods, A.*: "Some problems in logic and number theory and their connections", PhD. thesis, Manchester University, (1981).

[22] *Yao, A.*: "Separating the polynomial-time hierarchy by oracles", Proc. 26^{th} Annual IEEE Symp. on Foundations of Computer Science, (1985), pp. 1-10.