

Approximation and Small Depth Frege Proofs

Stephen Bellantoni
Dept. of Computer Science

Toniann Pitassi
Dept. of Computer Science

Alasdair Urquhart
Dept. of Philosophy

University of Toronto
Toronto, Ontario, Canada M5S 1A4

Abstract

Ajtai [Ajt] recently proved that if for some fixed d , every formula in a Frege proof of the propositional pigeonhole principle PHP_n has depth at most d , then the proof size is not less than any polynomial in n . By introducing the notion of an “approximate proof” we demonstrate how to eliminate the non-standard model theory, including the non-constructive use of the compactness theorem, from Ajtai’s lower bound. An approximate proof is one in which each inference is sound on a subset of the possible truth assignments — possibly a different subset for each inference. We also improve the lower bound, giving a specific super-polynomial function ($n^{\Omega(\log^{d+1} n)}$) bounding the proof size from below.

1 Introduction

A Frege proof is a sequence of propositional formulas, each of which is either an axiom instance or follows from previous formulas by one of a fixed set of inference rules. The pigeonhole principle can be expressed by a class of propositional formulas, $\{\text{PHP}_n : n \in N\}$, where PHP_n asserts that there is no 1-1 mapping from a set D_0 of size $n + 1$ to a set D_1 of size n .

Ajtai [Ajt] recently proved that if for some fixed d , every formula in a Frege proof of PHP_n has depth at most d , then the proof size is not less than any polynomial in

n . His proof, while combinatorial in part, is proven for a nonstandard model of Peano Arithmetic; the compactness theorem is then applied to obtain the result for standard values of n .

We demonstrate how to eliminate non-standard model theory from Ajtai’s lower bound by introducing the notion of an “approximate proof”. An approximate proof is one in which each inference is sound on a subset of the possible truth assignments — possibly a different subset for each inference.

Our notion of approximation resembles that of Razborov [Raz] where functions are approximated by introducing small errors at each gate. However, instead of approximating just one formula, we are approximating each formula in a sequence of related formulas. The approximation made for each individual formula changes how the formulas relate to each other: instead of each formula being a sound conclusion from previous formulas, the inference is only an “approximately sound” inference. The use of approximation gives a more direct lower bound proof than was obtained using non-standard model theory.

In this paper we also improve on Ajtai’s result, giving a specific super-polynomial function which bounds the Frege proof size from below. The bound is $n^{\Omega(\log^{d+1} n)}$, where $\log^{d+1} n$ is $d + 3$ iterations of log. Although the possibility of an exponential bound remains open, we give a reason why the proof method cannot be improved to yield an exponential bound.

We also demonstrate that if the Frege proof is of poly-

nomial size, then its depth must be $\Omega(\log^* n)$. This improves the statement which can be inferred from Ajtai's result, namely that polynomial size proofs must have non-constant depth.

Constant-depth lower bounds for PHP_n are related to the power of the systems of bounded arithmetic, $I\Delta_0(f)$, and $S_2(f)$. In particular, a super-polynomial bound for PHP_n implies that $I\Delta_0(f)$ cannot prove the sentence asserting the pigeonhole principle for f , while an exponential lower bound implies that $S_2(f)$ cannot prove the pigeonhole principle for f . See Paris, Wilkie & Woods [PWW], Paris & Wilkie [PW], and Ajtai [Ajt] for discussions of this question.

Lower bounds for propositional proof systems also bear on broader complexity issues. For example, the problem "NP =? co-NP" is equivalent to the problem, "Is there a propositional proof system in which the correctness of a derivation can be checked in polynomial time, and which admits polynomial size proofs of all tautologies?" [CR].

Our lower bound is proved using a particular Frege system over the basis $\{\vee, \neg\}$, but it holds for any Frege system: by a theorem of Cook and Reckhow [CR], all Frege systems are polynomially equivalent; and examining their theorem one finds that the small depth of proofs is preserved in the polynomial length simulation.

The base case of our result is a generalization of an argument originally given by Haken [Ha] (and later abstracted by Urquhart [Urq]) showing that any resolution refutation of PHP_n must contain a large clause.

As in previous results ([FSS], [Ajt2], [H], [Ajt]) involving bounded depth formulas, we proceed by induction on the depth. Applying a random restriction at each depth, we can simplify the formulas enough to reduce the depth, without simplifying the problem too much. However, instead of obtaining a depth $d-1$ proof of the (restricted) pigeonhole principle which is completely sound, we obtain a depth $d-1$ "approximate" proof of the (restricted) pigeonhole principle, which is only approximately sound. This approximation is introduced using a "pseudo complement" similar to Ajtai's.

2 Overview, and Definitions

2.1 Overview

We encode PHP_n using $(n+1)n$ propositional variables, $\{P_{ij} : i \in D_0 \ \& \ j \in D_1\}$, where D_0 and D_1 are disjoint sets such that $|D_0| = n+1$ and $|D_1| = n$. Intuitively, $P_{ij} = 1$ iff i is mapped to j . Since our proof system will be a refutation system, we are concerned with the statement $\neg\text{PHP}_n$, which can be written as the conjunction of the following *pigeonhole clauses*:

$$\bigvee\{P_{ij} : j \in D_1\}, i \in D_0;$$

$$\bigvee\{\neg P_{ik}, \neg P_{jk}\}, i \neq j, i, j \in D_0, k \in D_1.$$

In a refutation, one starts with the negated clauses $\neg\text{PHP}_n$ as axioms and then derives $\bigvee\{\}$, i.e. False. More exact definitions of the formal system are given below.

We obtain the lower bound by induction on the depth, d , of the Frege refutation. Applying a random restriction to the refutation, we can simplify the bottom levels so that each occurrence of negation at depth 3 of each formula is replaced by the 'pseudo complement'. This reduces the depth of each formula to $d-1$, but the resulting sequence of formulas may now only be an approximate refutation.

An approximate refutation is a Frege refutation where each inference is sound with respect to a large subset of all truth assignments. In contrast, an inference in a regular Frege refutation is sound with respect to all truth assignments. Note that our notion of an approximate proof is a local one: each inference can be sound with respect to a different subset of truth assignments, and there may be no single assignment which validates all the inferences. A 'good' approximation for an inference can be obtained if every OR of small ANDs at the bottom levels can be 'covered' by a small set after the restriction is applied. This covering set, which was also used by Ajtai, is analogous to the set of variables remaining after restriction in [FSS]; it is dissimilar to the minterms of Håstad [H].

We repeat the restriction argument $d-2$ times to obtain

an approximate depth 2 Frege refutation of the pigeon-hole principle, *i.e.* a refutation in which each formula is an OR of small ANDs. We then apply one more restriction to obtain a refutation in which each formula in the proof is an OR of small ANDs, covered by a small set. The existence of such a refutation contradicts the base case, which states that any good approximation to a Frege proof of the pigeonhole principle must contain a formula which has no small covering set.

2.2 Definitions

The system H

The lower bound for the pigeonhole clauses will be proven for the Frege refutation system H , described in Figure 1, for unbounded fan-in formulas. This system is a modification of the inference system in Shoenfield [Sh p. 21]. The formulas of H are unordered rooted trees defined inductively by the rules: (1) if γ is a set of variables then $\bigvee\{\wedge\gamma\}$ is a formula; if A is a formula then $\neg A$ is a formula, and if Γ is a finite set of formulas, then $\bigvee\Gamma$ is a formula. Thus the system allows \wedge only at the bottom level, and in fact requires \wedge 's there. This syntactic requirement simplifies the exposition.

In the schemas of Figure 1, A , B , and C represent formulas, and Γ and Δ are finite sets of formulas. NOTE: $A\overset{\circ}{\vee}B$ is the formula $A \vee B$ with the OR's merged together. More formally, $A\overset{\circ}{\vee}B = \bigvee(\text{DISJUNCTS}(A) \cup \text{DISJUNCTS}(B))$; where $\text{DISJUNCTS}(X)$ is the set of disjuncts of X if X is a disjunction and $\text{DISJUNCTS}(X) = \{X\}$ otherwise.

The *size* of a formula is one plus the number of occurrences of \vee and \neg in the formula; the size of a Frege proof is the sum of the sizes of the formulas occurring as lines in the proof. Since each formula consists of ORs of ANDs in the bottom 2 levels, and the rest of gates are ORs and NOTs, the depth of a formula is 2 plus the number of alternations of ORs and NOTs. The depth of a Frege proof is the maximum depth of the formulas in the proof.

If α is a propositional formula in the ordinary sense

Excluded Middle Axiom: $A\overset{\circ}{\vee}\neg A$

Weakening Rule: $\frac{A}{(A\overset{\circ}{\vee}B)}$

Cut Rule: $\frac{(A\overset{\circ}{\vee}B), (\neg A\overset{\circ}{\vee}C)}{(B\overset{\circ}{\vee}C)}$

Merging Rule: $\frac{\bigvee(\{\bigvee\Gamma\} \cup \Delta)}{\bigvee(\Gamma \cup \Delta)}$

Unmerging Rule: $\frac{\bigvee(\Gamma \cup \Delta)}{\bigvee(\{\bigvee\Gamma\} \cup \Delta)}$

Figure 1: Rules of the system H

of, say, [Sh], then we can transform it into a formula α^H of the system H as follows: write it using the basis \neg and \vee ; then replace every propositional variable P_{ij} with $\bigvee\{\wedge\{P_{ij}\}\}$; then merge together any adjacent \vee 's created at heights 2 and 3. For example, the images of the negated $PH P_n$ clauses are:

$$\begin{aligned} & \bigvee\{\wedge\{P_{ij} : j \in D_1\}\}, & i \in D_0; \\ (\neg\bigvee\{\wedge\{P_{ik}\}\}) \vee (\neg\bigvee\{\wedge\{P_{jk}\}\}), & i \neq j \in D_0; \\ & k \in D_1. \end{aligned}$$

Now, given a set of ordinary propositional formulas, say $\{\alpha_i\}$, and given another ordinary formula β , we define an H -proof of β from α over D to be a sequence of formulas such that the final formula is β^H , and each formula is either α_i^H for some i or follows from zero or more preceding formulas using one of the axioms or rules of H (see Figure 1). A *refutation* of α over D is an H -proof of $\bigvee\{\}$ from α .

It is easy to see that the system H is implicationally complete, using for example the fact that the propositional fragment of Shoenfield's system is implicationally complete. If $\{\alpha\}_i \vdash \beta$ in Shoenfield's system then we can obtain an H -proof of β from α by replacing every line γ of the Shoenfield proof with γ^H ; additionally we must insert appropriate combinations of the Merg-

ing and Unmerging rules in H .

The system H is not suited to a direct proof of the lower bound. We will describe a modified version of H , H' , that allows certain unsound inferences to be made. In spite of this unsoundness, we can retain control over the complexity by severely restricting the type of unsound inference which we permit. The new inference system will contain all of the rules of H plus additional rules that allow us to replace $\neg A$ by the pseudo complement of A , when A is of a simple form. In order to describe the pseudo complement, we need some definitions.

Maps; t -disjunctions; covering sets; 1-1 assignments.

First recall that the variables over $D = D_0 \cup D_1$ are $\{P_{ij} : i \in D_0, j \in D_1\}$. A map over D is defined to be a conjunction of the form $\bigwedge \Gamma$, where Γ is a set of variables over D such that distinct variables in Γ have distinct left subscripts and distinct right subscripts. Maps describe bijections between subsets of D_0 and subsets of D_1 . The size of a map $\bigwedge \Gamma$ is $|\Gamma|$; if the size of a map is bounded by t , it is said to be a t -map. An OR of maps is called a *map disjunction*; if all the maps are of size at most t , then it is a t -disjunction.

For a map disjunction G , define $\min(G)$ to be the disjunction obtained by deleting every map from G which implies some other map in G . For example, $P_{11} = \min((P_{11}) \vee (P_{11} \wedge P_{34}))$. In other words, we remove the map C from G if there is some other map $B \subsetneq C$ in G . Of course, G and $\min(G)$ have the same truth value on all assignments.

A formula A is *covered* by a set $V \subseteq D_0 \cup D_1$ if every variable in A has either its left or right subscript in V ; A is *k -coverable* if it is covered by some set V of size k . We write $\text{Cover}(X)$ for the size of the smallest covering set of X .

A map or formula B is *properly covered* by V if it is covered by V and every element of V covers some variable of B ; that is, if V covers B and every vertex in V is hit by an edge of B .

A truth assignment φ over D is any total assignment of $\{0, 1\}$ to the variables over D . An assignment φ is *1-1 on V* if $\{(i, j) : \varphi(P_{ij}) = 1 \ \& \ (i \in V \vee j \in V)\}$ is a bijection (a map) properly covered by V .

Conflicting maps; pseudo-complements.

Two maps $\bigwedge \Gamma$ and $\bigwedge \Delta$ are said to *conflict* if there are variables $P_{ij} \in \Gamma$ and $P_{kl} \in \Delta$ so that either $i = k$ and $j \neq l$, or $j = l$ and $i \neq k$. Notice that there is no map which conflicts with $\bigwedge \{\}$.

If A is a map disjunction such that $\min(A)$ is covered by V , then the *pseudo complement*, $c(A, V, D)$, of A with respect to V on universe D is the following map disjunction:

$$\bigvee \{B : B \text{ is a map over } D \text{ properly covered by } V, \text{ and } B \text{ conflicts with all maps in } A\}.$$

Notice $\min(A)$ doesn't have to be properly covered by V , just covered.

The complement we have defined is not quite the same as the complement defined by Ajtai, for two reasons. First, we require the conflicting maps to be properly covered where Ajtai just requires them to be covered; we need this change make the Distribution Lemma hold (see below). Second, we do not require that A be covered by V , only that $\min(A)$ be covered by V . This simplifies the Conversion Lemma (below), and is a harmless change: a map conflicts with all the maps of $\min(A)$ if and only if it conflicts with all the maps of A .

Making these changes to Ajtai's pseudo-complement does not spoil its key property: $c(A, V, D)$ is equivalent to $\neg A$ with respect to all truth assignments over D which are 1-1 on V . More exactly, we have the following easy lemma.

Lemma 2.1 (Complement Property) *If φ is a truth assignment which is 1-1 on V , then $\varphi(c(A, V, D)) = \varphi(\neg A)$.*

Proof. If $V = \emptyset$ then, since V covers $\min(A)$, either $A = \bigvee \{\}$ or else $\bigwedge \{\}$ is a map in A ; the lemma is

easily seen to hold in these cases. Therefore we assume $V \neq \emptyset$.

If $\varphi(c(A, V, D)) = 1$, then let B be a (nonempty) map in $c(A, V, D)$ set to 1 by φ . Every map C in A conflicts with B at some point in V ; since φ is 1-1 on V and assigns 1 to all variables of B , φ assigns 0 to the conflicting variable in C . Considering all C this implies $\varphi(A) = 0$, so $\varphi(\neg A) = 1$.

In the other direction, suppose $\varphi(A) = 0$. Let B be the (nonempty) map properly covered by V , induced by the assignment φ . That is, $B = \bigwedge \{P_{ij} : \varphi(P_{ij}) = 1 \ \& \ P_{ij} \text{ is covered by } V\}$. Considering any map $C \in A$, we have $\varphi(C) = 0$; therefore there is a variable in C which is set to zero by φ (and covered by V). This variable conflicts with some variable in B , because $\varphi(B) = 1$ and φ is 1-1 on V . We conclude $B \in c(A, V, D)$ and therefore $\varphi(c(A, V, D)) = 1$. ■

The system H' ; approximate refutations; t -soundness

The new proof system, H' , is obtained by adding the following schemes to H , for every map disjunction A and set V covering A :

$$\begin{array}{l} \text{Approximate} \\ \text{Excluded Middle Axiom} \end{array} \quad A \overset{\circ}{\vee} c(A, V, D)$$

$$\begin{array}{l} \text{Approximate} \\ \text{Cut Rule} \end{array} \quad \frac{(A \overset{\circ}{\vee} B), (c(A, V, D) \overset{\circ}{\vee} C)}{(B \overset{\circ}{\vee} C)}$$

Notice that these inferences depend on the fixed set D in the same sense as the PHP clauses depend on D . However, V can vary.

It should also be noted that the approximate complements in different parts of an approximate proof can be defined relative to quite different sets.

Neither the approximate excluded middle axiom nor the approximate cut rule are logically sound; however, by the Complement Property they are sound for the class of assignments which define 1-1 maps on V .

More formally, an inference in an approximate proof is t -sound if there is a set $V \subseteq D_0 \cup D_1$ with $|V| \leq t$ so

that any assignment which defines a 1-1 map on V and makes all premises of the inference true also makes the conclusion true. A sound rule of inference is a 0-sound rule. If $|V| = t$ is large, there are only a small number of truth assignments which are 1-1 on V and hence the inference is not very sound. On the other hand, the smaller $|V|$ is, the closer the inference is to a perfectly sound inference.

Lemma 2.2 (Soundness Fact) *The Approximate Rules are $|V|$ -sound.*

Proof. This follows immediately from the Complement Property. ■

Using this fact, we can slightly strengthen the notion of t -soundness as follows: a proof in H' is *strongly t -sound* if every inference is either 0-sound or is one of the Approximate rules involving $c(A, V, D)$ where $|V| \leq t$. In other words, we strengthen the condition so we know that the particular set V used in taking the pseudo-complement is also the set which witnesses the t -soundness.

We can think of “strongly t -sound” as a *syntactic* condition which is used to guarantee that the *semantic* requirement, “ t -sound”, holds.

Restrictions; miscellany

In choosing random restrictions, we use the same probability space as Ajtai. Each random restriction defines a one-to-one function between a subset of D_0 and a subset of D_1 . Specifically, the probability space $\Omega^{n, \epsilon}$ is the set of all pairs $\rho = \langle r, s \rangle$ where: s is a subset of $D = D_0 \cup D_1$ such that $s_0 = s \cap D_0$ is uniformly chosen with size $n^\epsilon + 1$ and, separately, $s_1 = s \cap D_1$ is uniformly chosen with size n^ϵ ; and r is a uniformly chosen bijection from $D_0 \setminus s$ to $D_1 \setminus s$.

Every $\rho = \langle r, s \rangle$ in $\Omega^{n, \epsilon}$ determines a unique *restriction*, ρ , of the variables P_{ij} , ($i \in D_0$, $j \notin D_1$) as follows.

$$\rho(P_{ij}) = \begin{cases} * & \text{if } i \in s_0 \ \& \ j \in s_1 \\ 1 & \text{if } i \notin s_0 \ \& \ j \notin s_1 \ \& \ r(i) = j \\ 0 & \text{otherwise} \end{cases}$$

Notice that variables assigned $*$ by ρ are variables on s .

We think of restrictions as being performed syntactically on a formula: to apply a restriction, we remove from each map those variables which the restriction sets to 1; and we remove from a map disjunction those maps in which some variable was set to 0. Thus, for example, the identity $(A \dot{\vee} B) \upharpoonright_{\rho} = A \upharpoonright_{\rho} \dot{\vee} B \upharpoonright_{\rho}$ holds. Of course, by the definition of a formula, a given map cannot appear twice in a disjunction. When we want to perform additional simplifications, we explicitly mention the $\min()$ operation.

The notation $\Pr_{\rho}^{n,\epsilon}[A]$ denotes the probability that A occurs when ρ is drawn from $\Omega^{n,\epsilon}$. For a Boolean formula F and an element $\rho \in \Omega_{n,\epsilon}$, F restricted by ρ will be denoted by $F \upharpoonright_{\rho}$. The notation $\log^{[l]} n$ indicates l applications of the base-2 log function (*not* $(\log n)^l$).

Throughout this paper, D_0 is a set of size $n + 1$ and D_1 is a set of size n ; where it is convenient, we shall assume that an ordering is given for each of D_0 and D_1 . Whenever we write a real number where an integer is required, we mean the integer part of the real number (floor). When we assert an inequality involving n , we shall often assume tacitly that n is sufficiently large.

3 Reducing the Depth

In this section we show how a proof of depth d is converted into one of depth $d - 1$, while preserving approximate soundness.

All formulas in the proof will be approximated simultaneously in a bottom-up, level-by-level fashion by repeatedly applying restrictions, replacing each negation at height 3 by an approximating OR, and merging, until we eventually obtain all depth 2 formulas. Note that while the approximation of each gate is quite good, an original depth d formula may eventually be transformed into a very different depth 2 formula. The key point is that our inference rules have the syntactic property that only one gate may be eliminated per inference, and hence our gate-by-gate approximation leads to a

new sequence of formulas which are still approximately sound.

At each stage, the depth is reduced by 1 and some of the inferences are converted from being 0-sound to being t' -sound for some t' . Inferences which were made t -sound in some previous stage will remain at worst t -sound; they will automatically be t' -sound since we will have $t' \geq t$.

In this connection, notice that in the Cut Rule (and in the Approximate Cut Rule), replacing, say, B by an approximating formula B' in both the hypothesis and conclusion does not affect the soundness of the inference. The soundness of the inference is only affected when we approximate the negations at the top level of the formulas; for example when we use the pseudo complement on the negation which is explicitly mentioned in the Cut Rule.

We will need to prove that the conversion process results in a syntactically proper approximate proof; as a step towards this, we show in the following lemma that the pseudo complement is in an appropriate sense invariant under restrictions.

Lemma 3.1 (Distribution Lemma) *Let A be a map disjunction over D such that V covers $\min(A)$; and let $\rho = \langle r, s \rangle \in \Omega^{n,\epsilon}$. Then $c(A, V, D) \upharpoonright_{\rho} = c(A \upharpoonright_{\rho}, V \cap s, D \cap s)$.*

Proof. First we show that any given map B' in $c(A, V, D) \upharpoonright_{\rho}$ is also in $c(A \upharpoonright_{\rho}, V \cap s, D \cap s)$. Let B be a map in $c(A, V, D)$ such that $B \upharpoonright_{\rho} = B' \neq 0$. We wish to show that $B \upharpoonright_{\rho}$ is in $c(A \upharpoonright_{\rho}, V \cap s, D \cap s)$. First of all, because B is properly covered by V and B contains only variables set to 1 or $*$ by ρ , $B \upharpoonright_{\rho}$ is properly covered by $V \cap s$. Secondly, variables in $B \upharpoonright_{\rho}$ are all variables over $D \cap s$ because other variables are assigned values by ρ ; since $B \upharpoonright_{\rho}$ is a map, it is a map over $D \cap s$. Now, if $A \upharpoonright_{\rho} = 0 = \bigvee \{ \}$ then trivially $B \upharpoonright_{\rho}$ is in $c(A \upharpoonright_{\rho}, V \cap s, D \cap s)$. Otherwise, let D' be any map in $A \upharpoonright_{\rho}$, and let D be a map in A such that $D' = D \upharpoonright_{\rho}$. By definition of $c(A, V, D)$, B conflicts with D ; using symmetry, let us assume that P_{ij} is in B , and P_{ik} in

D , where $j \neq k$. If P_{ij} (or P_{ik}) were set to 0 by ρ , a contradiction would hold because then $B \upharpoonright_\rho$ (or, respectively, $D \upharpoonright_\rho$) would be 0 and therefore would not be a map in $c(A, V, D) \upharpoonright_\rho$ (respectively, $A \upharpoonright_\rho$). And if either of the two variables were set to 1, the other would be set to 0; hence $\rho(P_{ij}) = \rho(P_{ik}) = *$. Therefore, B' and D' conflict.

In the other direction, let $B' = \bigwedge \Gamma$ be any given map in $c(A \upharpoonright_\rho, V \cap s, D \cap s)$. Define B by

$$\begin{aligned} \Delta &= \{P_{ij} : P_{ij} \text{ is a variable over } D, \\ &\quad P_{ij} \upharpoonright_\rho = 1, \text{ and } P_{ij} \text{ covered by } V\} \\ B &= \bigwedge (\Gamma \cup \Delta) \end{aligned}$$

Notice that $(\bigwedge \Gamma) \upharpoonright_\rho \neq 0$ because Γ consists of variables over $D \cap s$, and any variable set to 0 by ρ is not a variable over $D \cap s$. Therefore the variables in Γ and Δ don't conflict.

By construction, $B \upharpoonright_\rho = B'$. We wish to show that B is in $c(A, V, D)$, implying $B' \in c(A, V, D) \upharpoonright_\rho$. Firstly, B is a map over D covered by V , by construction. Secondly, V covers B properly, because vertices of V are either in $V \setminus s$ and hit by edges of Δ , or are in $V \cap s$ and hit by edges of B' , using the properness of $V \cap s$ for B' .

Now, if $A = 0 = \bigvee \{\}$ then trivially B conflicts with all maps in A and therefore is in $c(A, V, D)$; else consider any map C in A . We must show that B and C conflict. Notice C is a map over D covered by V , and $C \upharpoonright_\rho$ is a map over $D \cap s$ covered by $V \cap s$.

If $B \upharpoonright_\rho$ conflicts with $C \upharpoonright_\rho$, then B conflicts with C , as desired. Otherwise, $B \upharpoonright_\rho$ doesn't conflict with $C \upharpoonright_\rho$; since $B \upharpoonright_\rho \in c(A \upharpoonright_\rho, V \cap s, D \cap s)$, it must be that $C \upharpoonright_\rho$ is not in $A \upharpoonright_\rho$. Yet C is in A , and because $c(A \upharpoonright_\rho, V \cap s, D \cap s) \neq 0$, $A \upharpoonright_\rho \neq 1$; therefore ρ removes C from A , i.e. $C \upharpoonright_\rho = 0$. This means some variable P_{xy} in C is set to 0.

By the properness of $V \cap s$ for Γ , if P_{xy} were covered by $V \cap s$ then it would either be in Γ (contradicting $B \upharpoonright_\rho \neq 0$) or would conflict with Γ (implying that B conflicts with C , as desired). Since C is covered by V , the only remaining case is that P_{xy} is a variable covered by $V \setminus s$ and set to zero by ρ .

Let us say, using symmetry, that $y \in V \setminus s$. Now let

P_{xy} be a variable set to 1 by ρ , and which therefore conflicts with P_{xy} . Since P_{xy} is in Δ , B conflicts with C . ■

Definition 3.2 An approximate refutation of PHP_n is (d, t) -good if it has depth at most d , map size at most t , and is strongly t -sound. Notice that if a refutation is (d, t) -good, then it is (d, t') -good for all $t' \geq t$.

Below, we will describe a sufficient condition which allows us to convert a (d, t) -good refutation into a $(d-1, t)$ -good refutation. First we describe the conversion mechanism.

Let P be a (d, t) -good refutation over D of PHP_n ($d > 2$), and let ρ be a restriction. P is converted into a depth $d-1$ refutation in four steps.

- (1) Let $G_0 \dots G_m$ be the distinct map disjunctions appearing in formulas of $P \upharpoonright_\rho$. (We only need consider *maximal* map disjunctions, which appear in $P \upharpoonright_\rho$ other than as proper subformulas of map disjunctions.)
Let $W_0 \dots W_m \subseteq D \cap s$ be minimum size covering sets for $\min(G_0 \upharpoonright_\rho) \dots \min(G_m \upharpoonright_\rho)$. In case G is just $\bigvee \{\bigwedge \{P_{jk}\}\}$ for some j, k , then we prefer to cover $\min(G \upharpoonright_\rho)$ with $W = \{k\}$.
- (2) Apply the restriction ρ to each formula of P .
- (3) Replace each occurrence of $\neg G_i \upharpoonright_\rho$ by $c(G_i \upharpoonright_\rho, W_i, D \cap s)$.
- (4) Merge together OR gates appearing at heights 2 and 3 in the new proof.

Lemma 3.3 (Conversion Lemma) *Let P be a (d, t) -good approximate refutation over D of PHP_n ($d > 2$), and let $\rho = \langle \tau, s \rangle \in \Omega^{n, \epsilon}$. If $t' \geq t$ and $\text{Cover}(\min(G \upharpoonright_\rho)) \leq t'$ for every maximal map disjunction G in P , then P converted by ρ is a $(d-1, t')$ -good approximate refutation over $D \cap s$ of PHP_{n^ϵ} .*

Proof. Let $G_0 \dots G_m$ and $W_0 \dots W_m$ be as described above.

We must consider each inference of the original proof and see that after the conversion process, it remains a strongly sound inference in the system H' .

Suppose that the inference is $A\overset{\circ}{V}\neg A$ (Excluded Middle Axiom). If $\text{DEPTH}(A) > 2$ then the conversion results in another Excluded Middle Axiom. This is strongly 0-sound and therefore strongly t' -sound. If $\text{DEPTH}(A) = 2$ (i.e. $A = G_i$ is a map disjunction) then the conversion results in $G_i \uparrow_\rho \overset{\circ}{V}c(G_i \uparrow_\rho, W_i, D \cap s)$, an instance of the Approximate Excluded Middle Axiom. Since $|W_i| \leq t'$ is given, the instance is strongly t' -sound.

Suppose the inference is $(A\overset{\circ}{V}B), (\neg A\overset{\circ}{V}C) \Rightarrow (B\overset{\circ}{V}C)$ (Cut Rule). If $\text{DEPTH}(A) > 2$ then the conversion results in another instance of the Cut Rule. If $\text{DEPTH}(A) = 2$ then $A = G_i$ for some i and the conversion results in $(G_i \overset{\circ}{V}B) \uparrow_\rho, (c(G_i \uparrow_\rho, W_i, D \cap s) \overset{\circ}{V}C \uparrow_\rho, ho) \Rightarrow (B \overset{\circ}{V}C) \uparrow_\rho$; by the definition of \uparrow this is identically $(G_i \uparrow_\rho \overset{\circ}{V}B \uparrow_\rho), (c(G_i \uparrow_\rho, W_i, D \cap s) \overset{\circ}{V}C \uparrow_\rho, ho) \Rightarrow (B \uparrow_\rho \overset{\circ}{V}C \uparrow_\rho)$, a strongly $|W_i| \leq t'$ sound instance of the Approximate Cut Rule over $D \cap s$.

If the inference is an instance of the Weakening Rule, the Merging Rule, or the Unmerging Rule, then the converted inference is an instance of the same rule. (Essentially, this holds because \neg does not appear in these rules).

Suppose the inference is $A\overset{\circ}{V}c(A, V, D)$ for some map disjunction A and some set V covering $\text{min}(A)$ (Approximate Excluded Middle Axiom over D). The converted formula is $A \uparrow_\rho \overset{\circ}{V}c(A \uparrow_\rho, V, D) \uparrow_\rho$, which by the Distribution Lemma is $A \uparrow_\rho \overset{\circ}{V}c(A \uparrow_\rho, V \cap s, D \cap s)$, an instance of Approximate Excluded Middle over $D \cap s$. Since $A\overset{\circ}{V}c(A, V, D)$ was a strongly t -sound instance, we have $|V| \leq t$ and therefore $|V \cap s| \leq t \leq t'$. Therefore $A \uparrow_\rho \overset{\circ}{V}c(A \uparrow_\rho, V \cap s, D \cap s)$ is a strongly t' -sound inference.

Suppose the inference is $(A\overset{\circ}{V}B), (c(A, V, D)\overset{\circ}{V}C) \Rightarrow (B\overset{\circ}{V}C)$ (Approximate Cut Rule over D). Using the Distribution Lemma again, the converted inference is an instance of the Approximate Cut Rule over $D \cap s$. Using reasoning similar to that for the Approximate Excluded Middle Axiom, the inference is strongly $t \leq t'$

sound.

Finally, we analyse the PHP_n clauses as follows.

- $\bigvee\{\bigwedge\{P_{ij} : j \in D_1\}\}$ becomes $\bigvee\{\bigwedge\{\}\}$ if $i \notin s$; this is an instance of the Approximate Excluded Middle over $D \cap s$, with $A = \bigvee\{\}$ and $V = \emptyset$. If $i \in s$, it becomes a PHP_n clause over $D \cap s$.
- For $(\neg \bigvee\{\bigwedge\{P_{ik}\}\}) \vee (\neg \bigvee\{\bigwedge\{P_{jk}\}\})$, recall that we preferred to cover both the disjuncts with $\{k\}$. Therefore both complements will be taken with respect to $\{k\}$ and the result will be $\bigvee\{\bigwedge\{P_{hk} : h \in D_0 \cap s\}\}$. This is an instance of the Approximate Excluded Middle over $D \cap s$, with $A = \bigvee\{\}$ and $V = \{k\}$.

The condition that maps in the converted proof be of size at most t' , is easy, because every map disjunction in the converted proof is t' -coverable. ■

4 The Lower Bound

In this section the lower bound is stated and proven using the following Lemma (see §5), which says that under suitable conditions, applying a random restriction to a map disjunction makes it coverable by a small set.

Lemma 5.1 (Covering Lemma) *Let G be a t -disjunction and let ϵ be a constant such that $0 < \epsilon < 1/16$. If $t = o(\log \log n)$ and $8/\sqrt{\epsilon} \leq k \leq n^{\epsilon^{2t}}$ then (for sufficiently large n),*

$$\Pr_\rho^{n, \epsilon^{2t}}[\text{Cover}(\text{min}(G \uparrow_\rho)) > k] \leq \alpha_k^{n, t},$$

where

$$\alpha_k^{n, t} = \left(\frac{1}{n}\right)^{\frac{1}{11} \epsilon^{2t} k}.$$

It is convenient to introduce some constants and functions. We also indicate some relationships between the

quantities.

$$\begin{aligned}
\epsilon &= 1/25 \\
c &= (2 \log \frac{1}{\epsilon}) > 2 \\
t_d(n) &= \frac{1}{3c} \log^{[d+1]} n \leq \delta_d(n) t_{d-1}(n^{\delta_d(n)}) \\
\delta_d(n) &= \epsilon^{2t_d(n)} = (\log^{[d]} n)^{-1/3} \\
S_d(n) &= n^{t_d(n)/11} \leq S_{d-1}(n^{\delta_d(n)})
\end{aligned}$$

Theorem 4.1 (Lower Bound on Size) *For sufficiently large n , any refutation of PHP_n of depth d must have size at least $S_d(n) = n^{\Omega(\log^{[d+1]} n)}$.*

Proof. Suppose there were such a refutation, P . Since it is a refutation in H it is easily seen that P has map size at most 1, and is strongly 0-sound; therefore P is a $(d, t_d(n))$ -good approximate refutation of PHP_n of size less than $S_d(n)$. Without loss of generality we can assume that P has depth at least 2; now applying the Induction and Base lemmas below gives a contradiction. ■

Theorem 4.2 (Lower Bound on Depth) *For sufficiently large n , any Frege refutation of PHP_n of polynomial size must have depth $\Omega(\log^* n)$.*

Proof. The asymptotics in the lower bound on size hold at least for d up to $O(\log^* n)$.

Supposing the size to be bounded by n^k and setting $n^k = S_d(n)$ gives $\log^{[d+1]} n = k$, i.e. $d = \Omega(\log^* n)$. ■

To prove the following results, we let n be sufficiently large that the required asymptotic relationships hold for all depths up to d , and then proceed by induction on the depth, d . Each time we reduce the depth by applying a restriction to a universe of size n , the size of the resulting universe is $n^{\delta_d(n)}$.

Lemma 4.3 (Induction Lemma) *For n sufficiently large, if $d > 2$ and P is any $(d, t_d(n))$ -good refutation of PHP_n of size less than $S_d(n)$, then there is a restriction $\rho \in \Omega^{n, \delta_d(n)}$ such that: P converted by ρ is a $(d-1, t_{d-1}(n^{\delta_d(n)}))$ -good refutation of $PHP(n^{\delta_d(n)})$ of size less than $S_{d-1}(n^{\delta_d(n)})$.*

Proof. For each map disjunction G in P , the Covering Lemma implies that $\text{Cover}(\min(G \upharpoonright_\rho)) > t_{d-1}(n^{\delta_d(n)})$ with probability at most $\alpha_{t_{d-1}(n^{\delta_d(n)})}^{n, t_d(n)} = 1/S_{d-1}(n^{\delta_d(n)})$. The conditions required by the Covering Lemma, that $t_d(n) = o(\log \log n)$ and that $t_{d-1}(n^{\delta_d(n)}) \leq n^{\delta_d(n)}$, are easily seen to hold for $d \geq 2$.

Since $1/S_{d-1}(n^{\delta_d(n)}) \leq 1/S_d(n)$, and there are fewer than $S_d(n)$ map disjunctions in P , the probability is less than one that some map disjunction has $\text{Cover}(\min(G \upharpoonright_\rho)) > t_{d-1}(n^{\delta_d(n)})$. In particular there is a restriction ρ which makes all the map disjunctions coverable by small sets after taking $\min()$. Since $t_{d-1}(n^{\delta_d(n)}) \geq t_d(n)$, we can apply the Conversion Lemma to show that P converted by ρ is $(d-1, t_{d-1}(n^{\delta_d(n)}))$ -good. The size of P after conversion is still at most $S_d(n)$, which is at most $S_{d-1}(n^{\delta_d(n)})$. ■

It is interesting to notice that depth 2 proofs never contain any of the second type of PHP_n clause $(\neg \bigvee \bigwedge P_{ik} \vee \neg \bigvee \bigwedge P_{jk})$, because these have depth 3. These clauses all get converted into instances of Approximate Excluded Middle, the very first time a restriction is applied.

Lemma 4.4 (Base Lemma) *For sufficiently large n , there is no $(2, t_2(n))$ -good refutation of PHP_n of size less than $S_2(n)$.*

Proof. Suppose there were such a refutation, P . The same calculations as in the Induction Lemma allow us to use the Covering Lemma to show that there is a restriction $\rho \in \Omega^{n, \delta_2(n)}$ such that $\text{Cover}(\min(G \upharpoonright_\rho)) \leq t_1(n^{\delta_2(n)})$ for all maximal map disjunctions G in P . These maximal map disjunctions of P are exactly the formulas of P , because $d = 2$.

Applying ρ to P and replacing each map disjunction X with $\min(X)$ gives $t_2(n)$ -sound refutation P' of $PHP_{n^{\delta_2(n)}}$ such that every formula of P' is $t_1(n^{\delta_2(n)})$ -coverable. Since $t_1(n^{\delta_2(n)}) \leq n^{\delta_2(n)}/12$ for sufficiently large n , the existence of P' contradicts the Criticality Lemma below. We have not shown that P' is a refutation in H' , but that doesn't matter to the Criticality Lemma. ■

The Criticality Lemma, which provides the argument for the base case of the theorem, is a modification of Urquhart's argument [Urq] generalizing the resolution-system lower bound of Haken [Ha].

Definition 4.5 An assignment is *i-critical* if it is 1-1 on $D_0 \setminus \{i\}$ (and therefore is also 1-1 on D_1). For any formula A , the *critical set* $\text{CRIT}(A)$ is defined by

$$\text{CRIT}(A) = \{i : A \upharpoonright_{\rho} = 0 \text{ for some } i\text{-critical } \rho\}.$$

Lemma 4.6 (Criticality Lemma) *There is no $n/12$ -sound approximate refutation of PHP_n in which all formulas are $(n/8 - 1)$ -coverable.*

Proof. Suppose P were such an approximate refutation. Let A be the first formula in P such that $|\text{CRIT}(A)| \geq n/3$. There is such a formula because $|\text{CRIT}(\bigvee \emptyset)| = n + 1$.

Let $\{B_1, \dots, B_k\}$ be the k preceding formulas from which A is derived, with $k \leq 2$ and $k \geq 0$. Since the inference is $n/12$ -sound, there is a set V of size at most $n/12$ such that any assignment which is 1-1 on V and makes $B_1 \dots B_k$ true, also makes A true. Whenever $i \notin V$ and φ is *i-critical*, φ is 1-1 on V and therefore $A \upharpoonright_{\varphi} = 0 \Rightarrow \exists l, B_l \upharpoonright_{\varphi} = 0$. Hence

$$\text{CRIT}(A) \subseteq V \cup \bigcup_{l=1}^k \text{CRIT}(B_l).$$

Since $|\text{CRIT}(B_l)| < n/3$ by the minimality of A , this implies $|\text{CRIT}(A)| < n/12 + kn/3 \leq n/12 + 2n/3 = 3n/4$ and $|D_0 \setminus \text{CRIT}(A)| \geq n/4$.

Now we use these facts, that both $\text{CRIT}(A)$ and $D_0 \setminus \text{CRIT}(A)$ are large, to show that A is not $(n/8 - 1)$ -coverable. Specifically, we find $n/8$ variables in A which have disjoint subscripts.

For any *i-critical* assignment φ and $j \neq i$, let $r(j)$ be such that $\varphi(P_{j,r(j)}) = 1$. Now let $\varphi[j, i]$ be the assignment which agrees with φ except that $\varphi[j, i](P_{j,r(j)}) = 0$ and $\varphi[j, i](P_{i,r(j)}) = 1$. By switching j and i in this way, we get $\varphi[j, i]$ to be *j-critical*; when $j \in D_0 \setminus \text{CRIT}(A)$, this implies that $A \upharpoonright_{\varphi[j, i]} \neq 0$.

For each $i \in \text{CRIT}(A)$, fix an *i-critical* assignment φ such that $A \upharpoonright_{\varphi} = 0$. Consider any $j \in D_0 \setminus \text{CRIT}(A)$. Since $A \upharpoonright_{\varphi} = 0$ but $A \upharpoonright_{\varphi[j, i]} \neq 0$, either $P_{j,r(j)}$ or $P_{i,r(j)}$ occurs in A . Let $\text{VAR}(\varphi, i)$ be the variables so discovered, among all $j \in D_0 \setminus \text{CRIT}(A)$. Since φ defines a 1-1 function and $|D_0 \setminus \text{CRIT}(A)| \geq n/4$, there are at least $n/4$ distinct variables in each $\text{VAR}(\varphi, i)$.

Case 1: For some i , $\text{VAR}(\varphi, i)$ contains at least $n/8$ variables of the form $P_{j,r(j)}$, $j \in D_0 \setminus \text{CRIT}(A)$. These variables have mutually disjoint subscripts, because φ defines a 1-1 function (namely, r).

Case 2: Otherwise. For every $i \in \text{CRIT}(A)$, $\text{VAR}(\varphi, i)$ contains at least $n/8$ variables of the form $P_{i,r(j)}$ for some $j \in D_0 \setminus \text{CRIT}(A)$. There are at least $n/8$ values for i and by considering each in turn we can select a matching of size $n/8$ from the variables in $\bigcup_i \text{VAR}(\varphi, i)$. ■

5 Covering Lemma

In this section we prove the Covering Lemma, which states that if you apply a sufficiently strong restriction to a t -disjunction, then the result is probably k -coverable (for suitable t and k).

The proof is a simplification of Ajtai's T2 [Ajt], in which we extract specific bounds from the combinatorics. The Covering Lemma demonstrates that with high probability, applying a random restriction to a map disjunction results in a formula which can be covered by a small set. It is proved using a combinatorial lemma (5.2), which we derived from Ajtai's Lemma C1.¹

Lemma 5.1 (Covering Lemma) *Let G be a t -disjunction and let ϵ be a constant such that $0 < \epsilon < 1/16$. If $t = o(\log \log n)$ and $8/\sqrt{\epsilon} \leq k \leq n^{\epsilon^{2t}}$ for*

¹Ajtai's lemma C1, appearing in [Ajt] and [Ajt3], contains an error in the statement of (***) and a consequent error in the application of (**). He showed the proof of the corrected (**) in a private communication, which does not comment on the application of (**).

sufficiently large n , then for sufficiently large n ,

$$\Pr_{\rho}^{n, \epsilon^{2t}} [\text{Cover}(\min(G \upharpoonright_{\rho})) > k] \leq \alpha_k^{n, t},$$

where

$$\alpha_k^{n, t} = \left(\frac{1}{n}\right)^{\frac{1}{11} \epsilon^{2t} k}.$$

Proof. Given a fixed but sufficiently large value for n , the proof proceeds by induction on t .

Base case. For the base case ($t = 1$), write G in the form $G = \bigvee_{i \in D_0} \bigvee_{j \in W_i} P_{ij}$ for appropriate sets $W_i \subseteq D_1$. Let $B = \{i \in D_0 : |W_i| \geq n^{1-2\epsilon}\}$.

Taking cases on the size of B , suppose $|B| \geq n^{3\epsilon}$, so that for n sufficiently large, $|B \setminus s| \geq n^{5\epsilon/2}$. A restriction $\rho = \langle r, s \rangle$ can be chosen as follows: first choose $s = s_0 \cup s_1$ where $|s_1| = n^{\epsilon^2}$ and $|s_0| = |s_1| + 1$; let $B_s \subseteq B \setminus s$ be any particular subset of size $n^{5\epsilon/2}$; next, for each $i \in B_s$, in increasing order, choose $r(i)$ uniformly from the remaining elements of D_1 ; then choose the rest of r . Each time $r(i)$ is chosen for $i \in B_s$, the probability is $|(W_i \setminus s) \setminus \{r(k) : k \in B_s \text{ \& } k < i\}|$ out of $|(D_1 \setminus s) \setminus \{r(k) : k \in B_s \text{ \& } k < i\}|$ that $r(i) \in W_i$. Hence, for $i \in B_s$, the probability that $(\bigvee_{j \in W_i} P_{ij}) \upharpoonright_{\rho} = 1$ is at least $(n^{1-2\epsilon} - n^{\epsilon^2} - n^{5\epsilon/2}) / (n - n^{\epsilon^2}) \geq 1 / (2n^{2\epsilon})$. It follows that the probability of this happening for at least one of the $n^{5\epsilon/2}$ possible $i \in B_s$ is at least $1 - (1 - \frac{1}{2n^{2\epsilon}})^{n^{5\epsilon/2}}$. Since $(1 - \frac{1}{x})^x \leq \frac{1}{e}$, this probability is at least $1 - (\frac{1}{e})^{\frac{1}{2} n^{\epsilon/2}}$. Because $t = 1$ and $k \leq n^{\epsilon^2}$, this probability is greater than or equal to $1 - \alpha_k^{n, 1}$, for n sufficiently large. Finally, whenever this happens (i.e. whenever $(\bigvee_{j \in W_i} P_{ij}) \upharpoonright_{\rho} = 1$ for any $i \in B \setminus s$), then $G \upharpoonright_{\rho} = 1$ and therefore $\text{Cover}(\min(G \upharpoonright_{\rho})) = 0$.

On the other hand, suppose that $|B| < n^{3\epsilon}$ and a random $\rho = \langle r, s \rangle$ is chosen from $\Omega^{n, \epsilon^{2t}}$.

Firstly, we show that with high probability $|B \cap s_0| \leq k/2$. Applying Lemma 5.3 below, with the parameters $A' = \{B\}$, $c' = 0$, $t' = k/2$, $\delta' = (1 - 3\epsilon)$, and $\epsilon' = \epsilon^2$, we obtain that the probability of $|B \cap s| > k/2$ is at most $2n^{-(1-3\epsilon-\epsilon^2)k/4}$.

Secondly, we show that with high probability all variables $\{P_{ij} \mid i \notin B\}$ in $G \upharpoonright_{\rho}$ are covered by a subset of

s_1 of size at most $k/2$. We apply Lemma 5.2 to the system

$$\{i \rightarrow W_i : i \in D_0 \setminus B\} \cup \{i \rightarrow \emptyset : i \in B\},$$

with parameters $t' = k/2 > 4/\sqrt{\epsilon}$, $\delta' = 4\epsilon$, and $\epsilon' = \epsilon^2$. Lemma 5.2 implies that with probability at most $n^{-\epsilon^2 k/10}$, the subsystem fails to be $k/2$ -coverable after a restriction from $\Omega^{n, \epsilon^{2t}}$ is chosen and applied.

The variables remaining in G after $\langle r, s \rangle$ is applied are those in the subsystem just described, plus some variables which are covered by $B \cap s$. (Variables in $\{P_{ij} : i \in B \setminus s, j \in W_i\}$ are all set to either 0 or 1.) It follows that with probability at most $2n^{-(1-3\epsilon-\epsilon^2)k/4} + n^{-\epsilon^2 k/10} \leq \alpha_k^{n, 1}$, the function $\min(G \upharpoonright_{\rho})$ is not k -coverable.

Induction step. For all pairs (i, j) , $i \in D_0$, $j \in D_1$, construct the formulas $\phi_{ij} = \bigvee \{\alpha : \alpha \text{ is a map in } G \text{ containing } P_{ij}\}$. Then construct the formulas $\phi'_{ij} = \bigvee \{\alpha' : (\alpha' \wedge P_{ij}) \text{ is a map in } \phi_{ij}\}$, where ϕ'_{ij} is ϕ_{ij} with P_{ij} removed.

The induction step proceeds in three phases: in phase 0 we use the induction hypothesis on each ϕ'_{ij} to obtain sets C_{ij} covering the formulas $\min(\phi_{ij} \upharpoonright_{\rho})$; in phase 1 we apply Lemma 5.2 to the systems $\{j \rightarrow C_{ij}\}$ for each i , to obtain sets C_i ; and in phase 2 we apply Lemma 5.2 once more to the system $\{i \rightarrow C_i\}$. The resulting set covers every $\min(\phi_{ij} \upharpoonright_{\rho})$ and therefore covers $\min(G \upharpoonright_{\rho})$. In phases 0, 1 and 2 we apply successive restrictions $\rho_0 \in \Omega_{n_0, \epsilon^{2(t-1)}}$, $\rho_1 \in \Omega_{n_1, \epsilon}$, and $\rho_2 \in \Omega_{n_2, \epsilon}$ whose composition is the restriction ρ required for the lemma. Here we define $n_0 = n$; $n_1 = n^{\epsilon^{2t-2}}$; and $n_2 = n^{\epsilon^{2t-1}}$ corresponding to the domain size remaining before each of the three phases. The domains themselves we denote by $(D_0^0, D_1^0) = (D_0, D_1)$, (D_0^1, D_1^1) , and (D_0^2, D_1^2) , reserving (D_0^3, D_1^3) for the domain at the end of phase 2. Among all the formulas ϕ_{ij} , the only ones which will ultimately be significant are those for which $i \in D_0^3$ and $j \in D_1^3$, since a cover of these formulas after applying ρ is sufficient to cover G after applying ρ .

Phase 0. Each of the n_0^2 different formulas ϕ'_{ij} is a $t - 1$ disjunction. Applying the induction hypothesis,

we have that a random restriction ρ_0 from $\Omega_{n_0, \epsilon^{2(t-1)}}$ has probability at least $n_0^2 \alpha_t^{n_1 t - 1}$ of making all the formulas l -coverable, for appropriately small l .

Let C'_{ij} be a set of size l covering $\min(\phi'_{ij} \upharpoonright_{\rho_0})$. We need a small set C_{ij} covering $\min(\phi_{ij} \upharpoonright_{\rho})$. Observe that every \wedge -clause of $\min(\phi_{ij} \upharpoonright_{\rho_0})$ is a clause of $\min(\phi'_{ij} \upharpoonright_{\rho_0})$, possibly with P_{ij} added. (Consider cases in which $\rho_0(P_{ij}) = 0, 1$, or $*$. If $\rho_0(P_{ij}) = *$ then P_{ij} appears in every clause of $\phi_{ij} \upharpoonright_{\rho_0}$; therefore clauses eliminated from $\phi'_{ij} \upharpoonright_{\rho_0}$ by the min operator will also be eliminated from $\phi_{ij} \upharpoonright_{\rho_0}$.) Therefore, $\min(\phi_{ij} \upharpoonright_{\rho_0})$ is covered by $C_{ij} = C'_{ij} \cup \{i\}$.

Fixing any δ such that $0 < \delta < 1/4$, and choosing $l = n_1^{(1-\delta)}$, the required induction condition $l \leq n^{\epsilon^{2t}}$ follows trivially. Finally, we can observe that $C_{ij} \subseteq D_0^1 \cup D_0^1$ since the only variables set to $*$ by ρ_0 are those in $\{P_{xy} : x \in D_0^1 \text{ \& } y \in D_1^1\}$.

Phase 1. By the choice of l , Lemma 5.2 can be applied to the n_1 different systems $S_i = \{j \rightarrow C_{ij} : j \in D_1^1\}$, where $i \in D_0^1$. We choose a single random restriction ρ_1 from $\Omega_{n_1, \epsilon}$ and obtain that the covering sets described in 5.2 fail to exist with probability at most $n_1 \beta_\lambda^{n_1}$ for appropriately small λ . Thus with high probability we obtain sets C_i , for $i \in D_0^1$, such that $C_i = \cup_{j \in D_1^1} (C_{ij} \cap (D_0^2 \cup D_1^2))$, and $|C_i| \leq \lambda$.

For each $i \in D_0^1$, C_i covers every variable which is both covered by some C_{ij} ($j \in D_1^1$) and set to $*$ by ρ_1 . Since C_{ij} covers $\min(\phi_{ij} \upharpoonright_{\rho_0})$, this implies that C_i covers $\min(\phi_{ij} \upharpoonright_{\rho_0}) \upharpoonright_{\rho_1}$ for $j \in D_1^1$. Hence C_i covers $\min(\phi_{ij} \upharpoonright_{\rho_0 \rho_1})$ for $i \in D_0^2, j \in D_1^2$. Choose $\lambda = n_2^{(1-\delta)}$.

Phase 2. By the choice of λ , Lemma 5.2 can be applied to the system $S = \{i \rightarrow C_i : i \in D_0^2\}$. After choosing $\rho_2 \in \Omega_{n_2, \epsilon}$, the covering set described by 5.2 fails to exist with probability at most $\beta_k^{n_2}$. Thus we probably obtain a set C such that $C = \cup_{i \in D_0^2} (C_i \cap (D_0^3 \cup D_1^3))$ and $|C| \leq k$. The set C covers every variable which is both covered by some C_i ($i \in D_0^2$) and set to $*$ by ρ_2 . Since C_i covers $\min(\phi_{ij} \upharpoonright_{\rho_0 \rho_1})$, this implies that C covers $\min(\phi_{ij} \upharpoonright_{\rho_0 \rho_1}) \upharpoonright_{\rho_2}$. This gives the desired final result that a set C , of size at most k , covers $\min(\phi_{ij} \upharpoonright_{\rho_0 \rho_1 \rho_2})$ for $i \in D_0^3, j \in D_1^3$.

Analysis. The total probability that we fail to obtain the covering set is at most the sum of the probabilities in the three phases. This amount is $n_0^2 \alpha_t^{n_1 t - 1} + n_1 \beta_\lambda^{n_1} + \beta_k^{n_2}$. Using the constraint on k from the statement of the Lemma, it can be seen that $\beta_k^{n_2}$ is the dominant term in this sum. The amount is:

$$\begin{aligned} & n_0^2 \left(\frac{1}{n_0}\right) \left(\frac{1}{11} \epsilon^{2t-2} n_1^{(1-\delta)}\right) \\ & + n_1 \left(\frac{1}{n_1}\right) \left(\frac{1}{5} \epsilon n_2^{(1-\delta)}\right) \\ & + \left(\frac{1}{n_2}\right) \left(\frac{1}{5} \epsilon k\right) \\ & = \left(\frac{1}{n}\right) \left(\frac{1}{11} \epsilon^{2t-2} n^{(1-\delta) \epsilon^{2t-2}} - 2\right) \\ & + \left(\frac{1}{n}\right) \left(\frac{1}{5} \epsilon^{2t-1} n^{(1-\delta) \epsilon^{2t-1}} - \epsilon^{2t-2}\right) \\ & + \left(\frac{1}{n}\right) \left(\frac{1}{5} \epsilon^{2t} k\right) \\ & \leq 3 \left(\frac{1}{n}\right) \left(\frac{1}{5} \epsilon^{2t} k\right) \\ & \leq \alpha_k^{n_1 t} \end{aligned}$$

The last inequalities are obtained as follows. The condition $t \in o(\log \log n)$ implies that $n^{\epsilon^{2t}}$ is increasing (i.e. $\omega(1)$); therefore the conditions $k \leq n^{\epsilon^{2t}}$ and $(1-\delta)/\epsilon > 1$ imply that

$$k \leq \left(n^{\epsilon^{2t}}\right)^{(1-\delta)/\epsilon} - \frac{1}{\epsilon}$$

and

$$k \leq \frac{6}{11} \left(n^{\epsilon^{2t}}\right)^{(1-\delta)/(\epsilon^2)} - 2.$$

These two inequalities imply that the third exponent above, $(\frac{1}{5} \epsilon^{2t} k)$, is smaller than the other two. ■

The following lemma states that if g is a function taking $x \in D_0 \cup D_1$ to a small subset of $D_0 \cup D_1$ not containing x , then g , when restricted to a random subset of size n^ϵ , will have a small sized range. Recall $n = |D_1| = |D_0| - 1$.

Lemma 5.2 *Let $0 < \delta < 1/3$, $0 \leq \epsilon \leq \delta^2/4$ and let g be a function defined on $D_0 \cup D_1$, such that $g(x) \subseteq D_0 \cup D_1$, $|g(x)| \leq n^{1-\delta}$ and $x \notin g(x)$ for all $x \in D_0 \cup D_1$.*

Then for all $t > 4/\sqrt{\epsilon}$ and for all sufficiently large n we have

$$\Pr_{(r,s)}^{n,\epsilon} \left[\left| \bigcup_{x \in s} g(x) \cap s \right| > t \right] \leq \beta_t^n$$

where

$$\beta_t^n = n^{-t(\frac{\epsilon}{5})}$$

We will need the following lemmas, based on Ajtai's (*) and (**), to prove Lemma 5.2.

Lemma 5.3 *Let A be a set of subsets of $D_0 \cup D_1$ such that $|A| \leq n^c$, and $|X| < n^{1-\delta}$, for all $X \in A$. Then for all $t > 0$ and for all sufficiently large n ,*

$$\Pr_{\rho=\langle r,s \rangle}^{n,\epsilon} [\exists X \in A, |X \cap s| \geq t] \leq 2n^{-\frac{1}{2}(\delta-\epsilon)+c}$$

Lemma 5.4 *Let g' be a function defined on $D_0 \cup D_1$ such that $g'(x) \subseteq D$, $|g'(x)| \leq t$, and $x \notin g'(x)$, for every $x \in D_0 \cup D_1$. Then for all $t > 0$ and for all sufficiently large n ,*

$$\Pr_{(r,s)}^{n,\epsilon} [|\{y : \exists x \in s, y \in g'(x) \cap s\}| > t] \leq 2n^{-\frac{1}{4}(1-4\epsilon)}.$$

6 Limitations of the proof method

The covering lemma states that with high probability any given t -disjunction will be covered by a set of size k , after $O(t)$ restrictions. Because of the large number of restrictions that must be applied for every application of the covering lemma, the map size, t , cannot be too large (otherwise we quickly end up with an assignment to all of the variables). Therefore, one way of improving the bound would be to prove the covering lemma for a *single* restriction. This stronger form of the covering lemma could be stated as: For any $t, \epsilon < 1$, and for any t -disjunction, G , $\Pr_{\rho}^{n,\epsilon} [\text{Cover}(\min(G|_{\rho})) > k] \leq \alpha^k$, for some $\alpha < 1$, where α depends possibly on n and t . Setting $t = k$ approximately equal to $n^{1/d}$, this strengthened covering lemma would yield an exponential lower bound for PHP_n .

Unfortunately, this strengthened version is false for $k > \log n$. This situation is similar to the impossibility of obtaining an exponential lower bound for bounded depth circuits computing the parity function by simply improving the combinatorial lemma in [FSS]. Here we briefly describe a function, due to Russell Impagliazzo, which contradicts this strengthened covering statement, for $t = \log n + 1$.

The multiplexor function is a function on $n + \log n$ bits, $\{x_k\}$, where the first $\log n$ bits are used to index the remaining n bits. The function is "1" iff the value indexed by the first $\log n$ bits is "1". This function can be written as the OR of n minterms, each of size $\log n + 1$.

The counterexample to the strengthened covering lemma is a t -disjunction which encodes the multiplexor function on $\{x_k\}$ using the pigeonhole variables P_{ij} . Because the new function has to be monotone, we will encode negation by using the range elements, D_1 . The "pigeonhole" multiplexor function is a function on variables $\{P_{ij} \mid i \in D_0, j \in D_1\}$ where $|D_0| = \log n + n + 1$, and $|D_1| = \log n + n$. Let T be a fixed subset of D_1 of size $|D_1|/2$. A assignment ρ for $\{P_{ab}\}$ which is 1-1 on D_1 induces an assignment to the $n + \log n$ variables $\{x_i\}$ by: $x_i = 1$ iff $\exists j \in T$ such that $\rho(P_{ij}) = 1$. The value of the pigeonhole multiplexor function is the value of the multiplexor function on these induced values. Note that the modified function can be written as a t -disjunction, for $t = \log n + 1$.

Let $\rho = \langle r, s \rangle$ be a random restriction from $\Omega^{n,\epsilon}$. Intuitively, if all of the $\log n$ index variables in D_0 are included in s , then the restricted function will have large covering sets. The probability of this happening is approximately $(\frac{n^\epsilon}{n})^{\log n}$, which is larger than α^k , for $\alpha < 1$, and $k = n^{1/2}$.

This counterexample shows that an exponential lower bound for PHP_n cannot be obtained by simply improving the Covering Lemma. However, we feel that the covering lemma can be improved to yield a lower bound of $n^{\log^c n}$, for a small constant c , independent of the depth.

7 Conclusions and Open Problems

In this paper we have given a proof-theoretic superpolynomial lower bound for constant depth Frege proofs of the pigeonhole principle. Our approach introduces the notion of using approximations for a sequence of formulas, and shows how to use a proof theoretic approach to eliminate the non-standard model theory which was used by Ajtai. We also improve the lower bound of Ajtai. In addition, this proof more directly explains why bounded depth Frege proofs are weak for proving the pigeonhole principle.

We avoid the non-constructivity of the Compactness Theorem; in fact it appears that our proof can be made feasibly constructive as defined in [CU]. Informally, a feasibly constructive lower bound proof is one which involves only polynomial-time concepts. In contrast, it was shown in [CU] that a superpolynomial lower bound for *extended* Frege systems cannot have a feasibly constructive proof. A formalization of our result as a feasibly constructive proof requires describing exactly how to choose the restrictions, using Spencer's "probabilistic method" for transforming probabilistic algorithms into deterministic ones [Sp, p.31].

An outstanding open question is to prove a truly exponential lower bound for bounded depth Frege proofs. Such a bound would imply that S_2 (a subsystem of Peano arithmetic containing $I\Delta_0$) augmented by a function symbol f , cannot prove the sentence asserting *PHP* for f . (See [PWW], [Bu] for the connection between subsystems of bounded arithmetic and bounded depth Frege proofs.) As it is, current results simply imply that $I\Delta_0(f)$ cannot prove *PHP*(f). It is known [PWW] that $I\Delta_0(f)$ together with the existence of the function $n^{\log n}$ can prove the weak pigeonhole principle for f , i.e. the principle that f is not a bijection between $[2n]$ and $[n]$.

Acknowledgements

We wish to thank Stephen Cook, Russell Impagliazzo, and Alan Woods for many valuable conversations which led to this proof. We thank Alexander Razborov and Judy Goldsmith for their comments on earlier drafts of the paper.

8 References

- [Ajt] M. Ajtai, "The complexity of the pigeonhole principle," forthcoming. Preliminary version, *29th Annual Symposium on the Foundations of Computer Science* (1988), pp. 346-355.
- [Ajt2] M. Ajtai, " Σ_1^1 -Formulae on finite structures," *Annals of Pure and Applied Logic*, Volume 24, 1983, pp. 1-48.
- [Ajt3] M. Ajtai, "First order definability on finite structures," *Annals of Pure and Applied Logic*, Volume 45 (1989), pp. 211-225.
- [Bu] S. Buss, "Polynomial size proofs of the propositional pigeonhole principle," *Journal of Symbolic Logic*, v. 52 (1987), pp. 916-927.
- [CR] S. A. Cook and R. Reckhow, "The relative efficiency of propositional proof systems," *Journal of Symbolic Logic*, Volume 44, Number 1, March, 1979, pp. 36-50.
- [CU] S. A. Cook and A. Urquhart, "Functional Interpretations of Feasibly Constructive Arithmetic," Technical Report 210/88, Department of Computer Science, University of Toronto, 1988.
- [FSS] M. Furst, J. Saxe, M. Sipser, "Parity, circuits and the polynomial time hierarchy," *Mathematical Systems Theory*, Volume 17, pp. 13-27.
- [Ha] A. Haken, "The Intractability of Resolution," *Theoretical Computer Science*, 39, 1985, pp. 297-308.
- [H] J. Håstad, *Computational limitations of small-depth circuits*. The MIT Press, Cambridge, Massachusetts, 1987.
- [PW] J. Paris, A. Wilkie, "Counting problems in bounded arithmetic", *Methods in mathematical logic (Proceedings of the sixth Latin American symposium on mathematical logic, Caracas, 1983)*,

Lecture Notes in mathematics, v. 1130, Springer-Verlag, Berlin, 1985, pp. 317-340.

- [PWW] J. Paris, A. Wilkie, A. Woods, "Provability of the pigeonhole principle and the existence of infinitely many primes," *Journal of Symbolic Logic*, v. 53, Number 4, 1988.
- [Raz] A. A. Razborov, "Lower bounds on the monotone complexity of some Boolean functions," *Sov. Math. Doklady*, v. 31, 1985, pp. 354-357.
- [Sh] J. Shoenfield. *Mathematical Logic*. Addison-Wesley, Reading, Massachusetts, 1967.
- [Sp] J. Spencer. *Ten Lectures on the Probabilistic Method*. SIAM, Philadelphia, Pa., 1987.
- [Urq] A. Urquhart, "Hard Examples for Resolution," *JACM*, Volume 34, No. 1, January 1987, pp. 209-219.