Exponential Lower Bounds for the Pigeonhole Principle *

Paul Beame [†] Russell Impagliazzo Jan Krajíček
Toniann Pitassi Pavel Pudlák Alan Woods [‡]

Abstract

In this paper we prove an exponential lower bound on the size of bounded-depth Frege proofs for the pigeonhole principle (PHP). We also obtain an $\Omega(\log\log n)$-depth lower bound for any polynomial-sized Frege proof of the pigeonhole principle. Our theorem nearly completes the search for the exact complexity of the PHP, as Sam Buss has constructed polynomial-size, $\log n$-depth Frege proofs for the PHP. The main lemma in our proof can be viewed as a general Håstad-style Switching Lemma for restrictions that are partial matchings. Our lower bounds for the pigeonhole principle improve on previous superpolynomial lower bounds.

1 Introduction

In the last ten years, there has been significant progress in proving lower bounds for bounded-depth boolean circuits. One main technique for proving these results is the bottom-up method of restrictions, first described in [FSS], and later improved by Yao [Y], Håstad [H] and others. The strongest of these techniques is Håstad's Switching Lemma, which states that with high probability, a random restriction allows us to re-write an OR of small ANDs as an AND of small ORs.

A major drawback of this lemma and related ones is that they only apply when there is very little dependency between variables. There are many graph-based problems where the dependency between variables is too great to apply Håstad's Lemma, and there is no known reduction from a known hard problem in AC^0 to one of these problems. One graph-based problem for which a Håstad-style switching lemma has been shown is that of deciding whether or not a graph contains a clique on a small number of nodes (Lynch [Ly], Beame [Be]). However, the restrictions needed in that case still have very limited dependency.

In this paper, we prove a new switching lemma which applies to restrictions for which there is a great deal of dependency, namely those that represent partial matchings. A key feature that makes this more difficult is that after our restrictions are applied, the converted formula is only equivalent to the original one for certain classes of assignments.

We use this switching lemma to obtain the main result of this paper—an exponential bound on the size of bounded-depth Frege proofs for the pigeonhole principle. Frege systems are the typical propositional proof systems found in introductory textbooks. Besides their interest to logicians, they also arise in computer science due to their relationship to Resolution and other backtracking algorithms. Backtracking is a general technique to solve search problems in exponential-size domains. The fastest known algorithms for many NP-complete problems use backtracking techniques. Backtracking is also commonly used as a heuristic for many problems in artificial intelligence, particularly in automatic theorem-proving.

In a backtracking algorithm, one is searching a large space to find an element with a certain property P. The algorithm divides the space into those elements satisfying some property Q and those not satisfying Q. (This is called "branching on Q.") Of course, to be useful,

*Authors' addresses: Beame – University of Washington, beame@cs.washington.edu; Impagliazzo – UC San Diego, russell@cs.ucsd.edu; Krajíček – Mathematical Institute, Prague, krajicek@symcom.math.uiuc.edu; Pudlak – Mathematical Institute, Prague, pudlak@iam.unibe.ch; Pitassi – University of Toronto, toni@cs.toronto.edu; Woods – University of Western Australia, woods@maths.uwa.oz.au

[†] Research supported by NSF grants CCR-8858799 and CCR-8907960

[‡] Research partially supported by NSF grant DMS-8708657

24th ANNUAL ACM STOC - 5/92/VICTORIA, B.C., CANADA
© 1992 ACM 0-89791-512-7/92/0004/0200...$1.50

Q should be chosen so that the assumption $P \wedge Q$ narrows the search space significantly, as does $P \wedge \overline{Q}$. The algorithm continues to recurse until the search space is empty or a solution is found.

When run on a search space with no solutions, the transcript of a backtracking algorithm is a simple proof by contradiction that $\overline{P}$ is a tautology. This kind of proof is equivalent to that allowed in a Frege system. Thus, lower bounds on the complexity of Frege proofs show inherent limits on the backtracking technique. Since backtracking is a natural approach to solving problems like 3-Satisfiability, exponential lower bounds on Frege systems can be viewed as saying that a large class of natural approaches to solving NP-complete problems cannot run in sub-exponential worst-case time. Our bounds only apply to constant depth Frege systems, so the corresponding bound on backtracking algorithms applies to algorithms branching on properties that can be expressed in AC^0. This covers many suggested backtracking algorithms, but the moral for automatic theorem proving may be to develop heuristics to see when the proposition to be proved is of the type requiring a counting argument, in which case the heuristic should branch on formulas involving counting.

The complexity of Frege proofs of the pigeonhole principle has been studied extensively by many people in the last 20 years, beginning with an early paper by Tseitin [T]. In 1985, Haken [Ha] proved that any Resolution proof of the pigeonhole principle must have exponential size. The next major breakthrough was made by Ajtai [Ajt] who used nonstandard model theory to prove that any constant-depth Frege proof of the pigeonhole principle must have superpolynomial-size. Because Resolution is a particular depth-2 Frege system, Ajtai's proof yields a superpolynomial lower bound for Resolution as a special case. More recently, [BPU] obtained a new proof of Ajtai's theorem which eliminates the use of nonstandard models. While their techniques were more direct and more accessible, their improved bound was still barely superpolynomial. Then, using a different family of tautologies, an exponential lower bound on the size of constant-depth Frege proofs was established by Krajíček [K]. However, the hard examples used in Krajíček's proof do not have a fixed depth, independent of the depth of the Frege system in question, as the pigeonhole tautologies do. In addition, the examples are not as natural as the pigeonhole principle.

Our new exponential lower bound has several interesting consequences. As a corollary, we show that any polynomial-sized Frege proof of the pigeonhole principle must have depth $\Omega(\log\log n)$. Our theorem nearly completes the search for the exact complexity of the pigeonhole principle, as Sam Buss [Bu] has constructed polynomial-sized, logarithmic depth Frege proofs for the pigeonhole principle.

Constant-depth lower bounds are related to the power of weak systems of arithmetic (see [PW], [Bu]). This relationship together with our exponential lower bound for the propositional pigeonhole principle shows that relativized Bounded Arithmetic, $S_2(f)$, cannot prove the pigeonhole principle for f.

To see why this question is of interest in logic, consider the following two proof sketches that every non-zero residue modulo a prime has an inverse. Let p be a prime, and let $0 < a \leq p-1$. Then if we consider the map $F_a : \{0,...p-1\} \rightarrow \{0,..p-1\}$ defined by $F_a(b) = ab \ mod \ p$, it is easy to see that F_a is $1-1$. Therefore, (using the pigeonhole principle), it must also be onto, and so 1 must be in the image. Therefore, there exists a number b, $0 < b \leq p-1$, such that $ab = 1 \ mod \ p$. In the second proof, we would prove by induction on the length of numbers a, b that Euclid's Algorithm for extended gcd finds integers c, d so that $ca + db = gcd(a, b)$. Then applying this algorithm to a and p, we get $ca + dp = 1$, so $ca = 1 \ mod \ p$.

Both of the above proofs are simple, and only use basic facts of arithmetic. Both are constructive in the sense of intuitionistic logic. However, the first is combinatorially "non-constructive" in that it is based on a counting argument which yields no better way of finding the proven object than via exhaustive search. The second has "algorithmic content", and yields a good method for finding the object proven to exist. In this case, a counting argument was not necessary, and could be replaced by a more constructive computational argument. Our result can be phrased as saying that there is no *generic procedure* for converting a counting argument involving exponentially large but finite sets into an argument which only involves concepts in the polynomial-time hierarchy (relative to the object being counted). Thus, in general, one cannot automatically convert such an argument into a more algorithmic one, although in any particular case, this might be possible using special properties of the sets being counted.

In contrast with this negative result, Paris, Wilkie and Woods [PWW] showed that the weak pigeonhole principle, $WPHP_n$, is provable in $S_2(f)_*$ (PHP_n states that there is no 1-1 map from $[n+1]$ to $[n]$, while $WPHP_n$ states that there is no 1-1 map from $[2n]$ to $[n]$.) As a corollary, they show that $WPHP_n$ has quasi-polynomial size, constant-depth Frege proofs.

It is not hard to extend our results to weakenings of the pigeonhole principle that state the nonexistence of 1-1 mappings from sets of size $n+c$ to n (the lower bound is only minimally affected by c.) However, it is still an open problem whether $WPHP_n$ has constant-depth proofs of polynomial size. We can also extend our result quite easily to another weaker version of the pigeonhole principle, which states that there is no 1-1 *and onto* map from $[n+1]$ to $[n]$.

The main results of this paper were obtained independently, and first appeared in [PBI], and [KPW]. In this paper, we will first state and prove the common switching lemma, and then present two different proofs of the exponential lower bound. In section 2, we give some preliminary definitions. In Section 3, we state and prove the main combinatorial lemma. We present the proof appearing in [PBI]; an alternate proof can be found in [KPW]. In sections 4 and 5 we present the lower bound proof appearing in [PBI], and in section 6, we present the lower bound proof appearing in [KPW].

2 Definitions

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*. The *mapsize* of a map disjunction is the size of the largest map in the disjunction; if all the maps are of size at most t, then it will be called a *t-disjunction*. A truth assignment φ over D is any total assignment of $\{0,1\}$ to the variables over D. Let $D' = D'_0 \cup D'_1 \subseteq D$. A truth assignment φ over D is 1-1 over D' if for all $i \in D'_0$ there is a unique $j \in D_1$ such that $P_{ij} = 1$ and for all $j \in D'_1$ there is a unique $i \in D_0$ such that $P_{ij} = 1$.

If Y is a map or a set of variables, then $v(Y)$ denotes the set of vertices in $D_0 \cup D_1$ that are indexed by the variables in Y.

We will now define a probability space of partial 1-1 functions on D, where $D = D_0 \cup D_1$, and $|D_0| = |D_1|+1$. The probability space $\mathcal{R}_p^D$ is the set of all quadruplets $\rho =< i, S_0, S_1, \pi >$, where $i \in D_0$, $S_0 \subseteq D_0 \setminus \{i\}$, $S_1 \subseteq D_1$ and $|S_0| = |S_1|$. First, $i \in D_0$ is chosen uniformly and at random. The set S_0 is chosen as follows. For each $x \in D_0 \setminus \{i\}$, choose $x \in S_0$ with probability p and $x \notin S_0$ with probability $1-p$. After all elements, S_0, in D_0 have been selected, the set S_1 is obtained by selecting exactly $|S_0|$ elements of D_1 uniformly and at random. The third component in the triple, π, is a uniformly chosen bijection from D_1/S_1 to D_0/S_0. The quadruplet $< i, S_0, S_1, \pi >$ will sometimes be referred to as $< S, \pi >$, where $S = S_0 \cup S_1$. If $\rho =< i, S_0, S_1, \pi >$ then we will sometimes refer to i as *spare*(ρ).

Every $\rho =< S, \pi >$ in $\mathcal{R}_p^D$ determines a unique *restriction*, r, of the variables over D as follows.

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

In this way, the distribution $\mathcal{R}_p^D$ defines a probability distribution of restrictions. If r is a random restriction obtained by choosing a random ρ according to $\mathcal{R}_p^D$, we will refer to both the restriction and the random partial 1-1 function by ρ.

In order to prove the Switching Lemma (Lemma 3.2), we will first state a couple of useful properties of our distributions.

Lemma 2.1. Let $S_0 \subseteq D_0$, $S_1 \subseteq D_1$, $|S_0| = |S_1|$. Let $D' = D \setminus (S_0 \cup S_1)$. Then the subdistribution of $\mathcal{R}_p^D$ restricted to those ρ such that $\rho(S_0 \cup S_1) = *$ is equivalent to the distribution $\mathcal{R}_p^{D'}$. Similarly if A is any map defined on exactly those variables in $S_0 \cup S_1$, then the subdistribution of $\mathcal{R}_p^D$ restricted to those ρ such that $\rho(A) = 1$ is equivalent to the distribution $\mathcal{R}_p^{D'}$.

For a Boolean formula F and an element $\rho \in \mathcal{R}_p^D$, F restricted by ρ will be denoted by $F\lceil_\rho$.

There is an alternative experiment which can be used to obtain the same distribution on the variables. The probability space $\mathcal{P}_p^D$ is the set of all pairs $< \pi, S_1 >$, where π is a randomly chosen permutation from D_1 into D_0, and S_1 is a subset of D_1. The set S_1 is chosen as follows. For each $x \in D_1$, choose $x \in S_1$ with probability p and $x \notin S_1$ with probability $1-p$. Each $\rho =< \pi, S_1 >$ in $\mathcal{P}_p^D$ determines a unique restriction of the variables over D as follows.

$$r(P_{ij}) = \begin{cases} * & \text{if } \pi(j) = i \wedge j \in S_1 \\ 1 & \text{if } \pi(j) = i \wedge j \notin S_1 \\ 0 & \text{otherwise} \end{cases}$$

The following lemma states that the experiments $\mathcal{R}$ and $\mathcal{P}$ each define the same distribution of restrictions.

Lemma 2.2. The distributions $\mathcal{R}_p^D$ and $\mathcal{P}_p^D$ define the same probability distributions over restrictions.

Proof. For each element $\rho =< i, S_0, S_1, \pi >\in \mathcal{R}_p^D$, there is an associated unique set of elements $\rho' =< \pi', S_1' >$ from $\mathcal{P}_p^D$, which yields the same assignment to the variables P_{ij}. Namely, an element $\rho' =< \pi', S_1' >\in \mathcal{P}_p^D$ is associated with $\rho =< i, S_0, S_1, \pi >\in \mathcal{R}_p^D$ if the permutation, π' on $D/(S_0 \cup S_1)$ is identical to π and $S_1' = S_1$. Each element of $\mathcal{R}_p^D$ is associated with the same number of elements from $\mathcal{P}_p^D$; further, the probability over $\mathcal{R}$ of choosing a particular element, ρ, is equal to the probability over $\mathcal{P}$ of choosing an element in the set associated with ρ. Thus, the induced probability distributions on the setting of the variables, P_{ij} are identical. ▮

2.1 1-1 Decision Trees

A *1-1 decision tree* over domain $D = D_0 \cup D_1$ is defined as follows. It is a rooted tree where each interior node v is labelled by a query $i \in D_0$ or $j \in D_1$ and each edge is labelled by some pair $[i, j]$ where $i \in D_0$ and $j \in D_1$. Leaves are labelled with either "0" or "1". For each interior node v labelled by $i \in D_0$ ($j \in D_1$), there is exactly one out-edge labelled $[i, j]$ for each $j \in D_1$ ($i \in D_0$) that does not appear in any edge label on the path from the root to v. The label of an interior node v may not appear in any edge label on the path from the root to v. Thus the set of edge labels on any path defines a map.

A 1-1 decision tree T over D *represents* a function f over domain D if for all leaf nodes $v \in T$, if we let σ be the map defined by the path in T from the root to v then for all truth assignments α over D that are 1-1 on $v(\sigma)$ and consistent with σ, $f(\alpha)$ is equal to the label of v. For a boolean function f over domain D, we define $d_D(f)$ to be the minimum height of all 1-1 decision trees computing f.

If ρ is a partial 1-1 restriction over D and T is a 1-1 decision tree over D, then define $T\lceil_\rho$ to be the decision tree obtained from T by removing all paths which have a label that has been set to "0" by ρ, and contracting all edges whose labels are set to "1" by ρ.

Lemma 2.3 Let f be a boolean function over D and let T be a 1-1 decision tree representing f over D. If ρ is a partial 1-1 restriction over D, then $T\lceil_\rho$ is a 1-1 decision tree for $f\lceil_\rho$ over $D\lceil_\rho$.

Note that if T represents f over D then the tree T' obtained by by switching the 1's and 0's labelling the leaves of T represents $\neg f$. Also, given a 1-1 decision tree T over D of height d, we can obtain a d-disjunction $maps(T)$ over D whose maps consist of the labels of all the paths in T that end in leaves labelled 1. Notice that T represents $maps(T)$. Furthermore note that for any partial 1-1 restriction ρ over D, $maps(T\lceil_\rho) = maps(T)\lceil_\rho$. The lemmas in the next section actually is a switching lemma in the sense of Håstad because it will allow us to obtain a map disjunction that approximates the negation of f by representing f by a 1-1 decision tree T and then taking $maps(T')$.

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 The Switching Lemma

In this section we will assume that $D^n = D_0 \cup D_1$, where $|D_0^n| = |D_1^n| + 1 = n+1$, and the underlying probability distribution will be $\mathcal{R}$ (as defined in section 2). All other D, D', D'' will be bipartitions which are contained in D^n: $D = D_0 \cup D_1$, $|D_0| = |D_1| + 1$, and $D_0 \subseteq D_0^n$, $D_1 \subseteq D_1^n$.

Let $K \subseteq D = D_0 \cup D_1$. Then $Proj_D[K]$ is the set of all minimal partial 1-1 maps over D which involve all of the elements of K. A map $\sigma \in Proj_D[K]$ induces a restriction; we will refer to σ interchangeably as a restriction and as a map.

We define the *complete 1-1 tree* for $K \subseteq D$ over D inductively as follows. If K consists of a single node $k \in D_0$ ($k \in D_1$), then label the root "$k \in D_0$" ("$k \in D_1$"), and create n edges adjacent to the root, labelled by $[k, j]$, for all $j \in D_1$ ($[j, k]$ for all $j \in D_0$). Otherwise, $K = K' \cup \{k\} \subseteq D$. Assume that we have created the complete tree for K'; we will now extend it to a complete tree for K. This is done by extending each leaf node v_l as follows. Let p_l be the path from the root to v_l. The edge labellings along p_l define a partial 1-1 map involving all elements of K'. If this partial map does not include k, then label v_l by k, and add new edges leading out of v_l, one for every possible mapping for k that results in a 1-1 map extending the partial 1-1 map along p_l. Otherwise, if k is involved in the partial 1-1 map, leave v_l unlabelled. Note that each path of the complete tree over K will be labelled by some $\sigma \in Proj_D[K]$.

Lemma 3.1 Let f be a boolean function over the variables P_{ij}, $i \in D_0$, $j \in D_1$, where $|D_0| = |D_1| + 1$. For every $K \subseteq D_0 \cup D_1$, there exists a restriction, $\sigma \in Proj_D[K]$ such that $d_D(f) \leq |\sigma| + d_{D\restriction_\sigma}(f\restriction_\sigma)$.

Proof. The proof is very similar to that of Beame and Håstad [BH]. Fix $K \subseteq D$. We start with the complete 1-1 tree for K. As noted above the paths of this tree correspond exactly to elements of $Proj_D[K]$. Let v_σ be the leaf node corresponding to the path labelled by $\sigma \in Proj_D[K]$. For each σ, we replace the leaf node, v_σ, by a subtree that is a 1-1 decision tree for $f\restriction_\sigma$ over $D\restriction_\sigma$. The resulting tree clearly represents f over D. The depth of the resulting tree for K is at most $max_\sigma\{|\sigma| + d_{D\restriction_\sigma}(f\restriction_\sigma)\}$. ∎

If f is a map disjunction defined over a set D and ρ is a restriction on D then we will use the notation $\delta(f\restriction_\rho)$ for $d_{D\restriction_\rho}(f\restriction_\rho)$. We now state the main combinatorial lemma.

Lemma 3.2. (Switching Lemma) Let f be an r-disjunction over $D = D_0 \cup D_1$, $|D_0| = |D_1| = m+1$, $D_0 \subseteq D_0^n$, $D_1 \subseteq D_1^n$. Choose ρ at random from $\mathcal{R}_p^D$. For $s \geq 0$ and $p(m-s) \geq r$ we have

$$Pr[\delta(f\restriction_\rho) \geq s] < \alpha^s,$$

where $\alpha > 0$ satisfies $(1 + 9p^4n^3/\alpha^2)^r = 5/4$.

Fact: $\alpha < 8p^2n^{3/2}r^{1/2}$.

The proof of the switching lemma, like that of Håstad, proceeds by induction on the number of clauses in f. We work along the clauses one by one: if ρ falsifies a particular clause, then we are left with essentially the same problem as before; if ρ does not falsify the clause then, it is much more likely that ρ satisfies the clause (and thus ensures that the whole formula is set to true) than ρ leaves any variable in the clause unset. There are significant complications however in dealing with our partial 1-1 restrictions as opposed to fully independent ones. Once we know that a variable (edge) is unset we have information that biases incident variables towards being unset. Furthermore there is the subtler problem that having some variables set to 0 may bias other variables towards being unset. Both of these complicate the application of the inductive argument in the case that a given clause is not falsified. We handle the first problem by considering not only all possible assignments to the unset variables in the clause (as in Håstad's proof) but also to all variables that are incident to those unset variables. We get around the second problem by showing that, although setting variables to 0 may make a given variable more likely to be unset, it cannot bias the total number of unset variables to be larger and this turns out to be sufficient for our purposes.

We obtain Lemma 3.2 from the somewhat stronger Lemma 3.5 by setting $F = 0$ but first we prove a couple of technical lemmas.

Lemma 3.3. Let $D = D_0 \cup D_1$ such that $|D_0| = |D_1| + 1 = m+1$ and $U \subseteq D$ such that $|U \cap D_0| = |U \cap D_1| = k$.

If $pm \geq k$, for ρ chosen at random from $\mathcal{R}_p^D$,

$$Pr[\rho(U) = *] \leq 2 \cdot \frac{(2p^2m)^k(m+1-k)!}{(m+1)!}.$$

Proof. Let $U_0 = U \cap D_0$ and $U_1 = U \cap D_1$. We consider the choice of $\rho \in \mathcal{R}_p^D$ using the equivalent distribution $\mathcal{P}_p^D$. Thus ρ is chosen by selecting a random permutation $\pi : D_1 \rightarrow D_0$ and a set $S_1 \in D_1$ of starred endpoints chosen by selecting elements of D_1 independently with probability p. We split up the probability that $\rho(U) = *$ into separate cases depending on the image $\pi(U_1)$ of U_1 in D_0.

If $|\pi(U_1) \setminus U_0| = i$ then we divide the probability base on whether or not $spare(\rho) \in U_0$. Now $Pr[spare(\rho) \in U_0] = i/(m+1-k)$ in this case since $\pi(U_1)$ has already been ruled out. Given that $spare(\rho) \in U_0$, the probability that $\rho(U) = *$ is p^{k+i-1}, otherwise it is p^{k+i}. Thus if $|\pi(U_1) \setminus U_0| = i$ we have a total probability that $\rho(U) = *$ of

$$\begin{aligned} & p^{k+i-1}\frac{i}{m+1-k} + p^{k+i}\left(1 - \frac{i}{m+1-k}\right) \\ = \;& p^{k+i} + p^{k+i-1}\frac{(1-p)i}{m+1-k} \\ \leq \;& p^{k+i} + p^{k+i-1}\frac{(1-p)k}{m+1-k} \\ \leq \;& p^{k+i} + p^{k+i-1}\frac{pm-pk}{m+1-k} \\ < \;& 2p^{k+i} \end{aligned}$$

since $pm \geq k$ and $i \leq k$.

There are $\binom{m+1}{k}$ possible sets $\pi(U_1)$, all of which are equally likely, and $\binom{m+1-k}{i}\binom{k}{i}$ of these have $|\pi(U_1) \setminus U_0| = i$. Thus

$$\begin{aligned} Pr[\rho(U) = *] &= \sum_{i=0}^{k} \frac{\binom{m+1-k}{i}\binom{k}{i}}{\binom{m+1}{k}} 2p^{k+i} \\ &\leq \frac{2p^k}{k!\binom{m+1}{k}} \sum_{i=0}^{k} \binom{k}{i} k! \frac{m^i}{i!} p^i \\ &\leq \frac{2p^k}{k!\binom{m+1}{k}} \sum_{i=0}^{k} \binom{k}{i} k^{k-i}(pm)^i \\ &= \frac{2(pk)^k}{k!\binom{m+1}{k}} \sum_{i=0}^{k} \binom{k}{i} \left(\frac{pm}{k}\right)^i \\ &= \frac{2(pk)^k}{k!\binom{m+1}{k}} [(pm/k)+1]^k \\ &\leq \frac{2(pk)^k}{k!\binom{m+1}{k}} (2pm/k)^k \\ &= 2\frac{(2p^2m)^k}{k!\binom{m+1}{k}} \end{aligned}$$

since $pm \geq k$. ∎

Lemma 3.4 Suppose that $0 \leq \alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_n$, and for all $k \leq n$, $\sum_{j=k}^n a_j \leq \sum_{j=k}^n b_j$. Then for all $k \leq n$, $\sum_{j=k}^n \alpha_j a_j \leq \sum_{j=k}^n \alpha_j b_j$.

Proof. The proof is by downward induction on k. For $k = n$, the lemma holds. Now assume that the lemma holds for k. Consider $\sum_{j=k-1}^n \alpha_j b_j$. Either $b_{k-1} \geq a_{k-1}$ or $b_{k-1} < a_{k-1}$. In the first case, by the induction hypothesis, we know that $\sum_{j=k}^n \alpha_j b_j \geq \sum_{j=k}^n \alpha_j a_j$, thus because $b_{k-1} \geq a_{k-1}$, we also have $\sum_{j=k-1}^n \alpha_j b_j \geq \sum_{j=k-1}^n \alpha_j a_j$. In the second case, let $\delta = a_{k-1} - b_{k-1}$. Because $\sum_{j=k-1}^n b_j \geq \sum_{j=k-1}^n a_j$, we have that $\sum_{j=k}^n b_j \geq \sum_{j=k}^n a_j + \delta$. Applying the inductive hypothesis, with $a_k = a_k + \delta$, we have:

$$\begin{aligned} & \sum_{j=k}^{n} \alpha_j b_j \geq \sum_{j=k+1}^{n} \alpha_j a_j \; + \; \alpha_k(a_k + \delta) \\ \Rightarrow & \sum_{j=k-1}^{n} \alpha_j b_j \geq \sum_{j=k}^{n} \alpha_j a_j + \alpha_k(a_k + \delta) + \alpha_{k-1} b_{k-1} \\ \Rightarrow & \sum_{j=k-1}^{n} \alpha_j b_j \geq \sum_{j=k}^{n} \alpha_j a_j + \alpha_k(a_{k-1} - b_{k-1}) + \alpha_{k-1} b_{k-1} \\ \Rightarrow & \sum_{j=k-1}^{n} \alpha_j b_j \geq \sum_{j=k}^{n} \alpha_j a_j + \alpha_{k-1}(a_{k-1} - b_{k-1} + b_{k-1}) \\ \Rightarrow & \sum_{j=k-1}^{n} \alpha_j b_j \geq \sum_{j=k-1}^{n} \alpha_j a_j . \; ∎ \end{aligned}$$

Lemma 3.5. Let f be an r-disjunction over $D = D_0 \cup D_1$, $|D_0| = |D_1| + 1 = m + 1$, $D_0 \subseteq D_0^n$, $D_1 \subseteq D_1^n$, and let F be an arbitrary function over D^n. Let ρ be a random restriction chosen according to $\mathcal{R}_p^D$. Then for $s \geq 0$ and $p(m-s) \geq r$ we have

$$Pr[\delta(f\lceil_\rho) \geq s \mid F\lceil_\rho = 0] \leq \alpha^s,$$

where $\alpha > 0$ satisfies $(1 + 9p^4n^3/\alpha^2)^r = 5/4$.

Proof. The proof proceeds by induction on the total number of maps in f.

Base Case. There are no maps in f. In this case f is identically 0 and therefore f is represented by the tree consisting of the single node labelled 0. Hence $\delta(f\upharpoonright_\rho) = 0$ and the lemma holds.

Induction Step. Assume that the lemma holds for all map disjunctions with fewer maps than the map disjunction of f. We will write f as $f_1 \vee f_2 \vee \ldots$, where each f_i is a map of f. We will analyze the probability by considering separately the cases in which ρ does or does not force the map f_1 to be 0. The failure probability, the probability that $\delta(f\upharpoonright_\rho) \geq s$, is an average of the failure probabilities of these two cases. Thus

$$\begin{aligned} &Pr[\delta(f\upharpoonright_\rho) \geq s \mid F\upharpoonright_\rho = 0] \leq \\ &\quad max(Pr[\delta(f\upharpoonright_\rho) \geq s \mid F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho = 0], \\ &\qquad Pr[\delta(f\upharpoonright_\rho) \geq s \mid F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0]). \end{aligned}$$

The first term in the maximum is $Pr[\delta(f\upharpoonright_\rho) \geq s \mid (F \vee f_1)\upharpoonright_\rho = 0]$. Let f' be f with map f_1 removed; then $Pr[\delta(f\upharpoonright_\rho) \geq s \mid (F \vee f_1)\upharpoonright_\rho = 0] = Pr[\delta(f'\upharpoonright_\rho) \geq s \mid (F \vee f_1)\upharpoonright_\rho = 0]$. Because f' has one less map than f, this probability is no greater than α^s, by the inductive hypothesis.

Now we will estimate the second term in the maximum. Let T be the set of variables appearing in the first map, f_1. By hypothesis, $size(T) \leq r$. We will analyze the cases based on the subset Y of the variables in T to which ρ assigns $*$; we use the notation $*(\rho_T) = Y$ to denote the event that the variables in T which are assigned $*$ by ρ_T are exactly those in Y. Then

$$\begin{aligned} &Pr[\delta(f\upharpoonright_\rho) \geq s \mid F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0] \\ &= \sum_{Y \subseteq T} Pr[\delta(f\upharpoonright_\rho) \ \wedge \ *(\rho_T) = Y \mid F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0]. \end{aligned}$$

Consider the case in which $Y = \emptyset$. In this case the value of f_1 is forced to 1 by ρ. It follows that f is forced to 1 and hence $\delta(f) = 0$ so the term corresponding to $Y = \emptyset$ has probability 0. The sum then becomes

$$\sum_{\substack{Y \subseteq T, \\ Y \neq \emptyset}} Pr[\delta(f\upharpoonright_\rho) \geq s \wedge *(\rho_T) = Y \mid F\upharpoonright_\rho = 0 \wedge f_1\upharpoonright_\rho \neq 0],$$

which is equal to

$$\sum_{\substack{Y \subseteq T, \\ Y \neq \emptyset}} \begin{aligned} &Pr[\delta(f\upharpoonright_\rho) \geq s \mid \\ &\quad F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0 \ \wedge \ *(\rho_T) = Y] \qquad (1) \end{aligned}$$

$$\times \ Pr[*(\rho_T) = Y \mid F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0]. \quad (2)$$

We will first bound the latter term, (2), in each of these products. Given that $f_1\upharpoonright_\rho \neq 0$, the probability that $*(\rho_T) = Y$ is equal to the probability that $\rho(Y) = * \ \wedge \ \rho(T \setminus Y) = 1$. Thus term (2) is no greater than

$$\begin{aligned} &Pr[\rho(Y) = * \wedge \rho(T \setminus Y) = 1 \mid F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0] \\ &\leq Pr[\rho(Y) = * \mid F\upharpoonright_\rho = 0 \ \wedge \ \rho(T \setminus Y) = 1 \ \wedge \rho(Y) \neq 0] \end{aligned}$$

Let F' be $F \vee G$ where $G\upharpoonright_\rho = 0$ if and only if ρ sets all variables in $T \setminus Y$ to 1; then the above probability is equal to $Pr[\rho(Y) = * \mid F'\upharpoonright_\rho = 0 \ \wedge \ \rho(Y) \neq 0]$.

Claim A. $Pr[\rho(Y) = * \mid F'\upharpoonright_\rho = 0 \ \wedge \ \rho(Y) \neq 0] \leq Pr[\rho(Y) = * \mid \rho(Y) \neq 0]$.

Proof of Claim A. As in previous proofs, we will prove claim A by showing that

$$\begin{aligned} &Pr[F'\upharpoonright_\rho = 0 \mid \rho(Y) = * \wedge \rho(Y) \neq 0] \\ &\leq \ Pr[F'\upharpoonright_\rho = 0 \mid \rho(Y) \neq 0]. \end{aligned}$$

This proves the claim because for arbitrary events A and B, $Pr[A \mid B \wedge C] \leq Pr[A \mid C] \leftrightarrow Pr[B \mid A \wedge C] \leq Pr[B \mid C]$. Fix a particular ρ^* such that $\rho^*(Y) = *$. Then ρ^* represents an equivalence class of $\rho's$ such that $\rho(Y) \neq 0$. An element $\rho \in \mathcal{R}_n^D$ is in the equivalence class of ρ^* if and only if ρ is identical to ρ^* except for the variables of Y, which may be assigned the value 1 instead of $*$. Note that each such equivalence class is disjoint, has the same size, and the union of all equivalence classes is equal to the set of all ρ which satisfy $\rho(Y) \neq 0$. Now, consider a particular ρ^*. If $F'\upharpoonright_{\rho^*}$ is forced to 0, then so is $F'\upharpoonright_\rho$, for every ρ in the equivalence class of ρ^*. Thus the claim holds.

From Claim A it follows that the term (2) is at most $Pr[\rho(Y) = * \mid \rho(Y) \neq 0]$. Since $pm \geq r \geq |Y|$, by Lemma 3.3,

$$Pr[\rho(Y) = *] \leq \frac{2(2p^2m)^{|Y|}(m+1-|Y|)!}{(m+1)!}.$$

Also,

$$\begin{aligned} Pr[\rho(Y) \neq 0] \ &\geq \ Pr[\rho(Y) = 1] \\ &= \ \frac{(1-p)^{|Y|}}{(m+1)m..(m-|Y|+2)} \end{aligned}$$

$$= \frac{(1-p)^{|Y|}(m+1-|Y|)!}{(m+1)!}.$$

Therefore,

$$\begin{aligned} Pr[\rho(Y) = * \mid \rho(Y) \neq 0] &\leq 2 \cdot \left(\frac{2p^2m}{1-p}\right)^{|Y|} \\ &\leq 2 \cdot (3p^2m)^{|Y|} \\ &\leq 2 \cdot (3p^2n)^{|Y|}. \end{aligned}$$

Now we look at the first term, (1), in each product. Suppose that $2|Y| \leq s$. For each fixed Y, we will analyze the probability above by applying Lemma 3.1 with $K = v(Y)$ and $D = D\upharpoonright_\rho$. By this lemma, if $\delta(f\upharpoonright_\rho) \geq s$ then there is some $\sigma \in Proj_{D\upharpoonright_\rho}[v(Y)]$, such that $d_{(D\upharpoonright_\rho)\upharpoonright_\sigma}((f\upharpoonright_\rho)\upharpoonright_\sigma) \geq s - |\sigma|$. To use this requires that we consider all maps in $Proj_{D\upharpoonright_\rho}[v(Y)]$. One difficulty is that $D\upharpoonright_\rho$ is itself a random variable dependent on ρ. We handle this by considering all maps σ in $Proj_D[v(Y)]$ and including them only if $\rho(\sigma) = *$. For notational convenience let $P(D, Y) = Proj_D[v(Y)]$. When $\rho(\sigma) = *$, $(f\upharpoonright_\rho)\upharpoonright_\sigma = (f\upharpoonright_\sigma)\upharpoonright_\rho$ and applying the definition of $\delta(f\upharpoonright_\rho)$, the above probability is no greater than:

$$\begin{aligned} &\sum_{\sigma \in P(D,Y)} Pr[\delta((f\upharpoonright_\sigma)\upharpoonright_\rho) \geq s - |\sigma| \ \wedge \rho(\sigma) = * \mid \\ &\qquad\qquad F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0 \ \wedge \ *(\rho_T) = Y] \\ &\leq \sum_{\sigma \in P(D,Y)} Pr[\delta((f\upharpoonright_\sigma)\upharpoonright_\rho) \geq s - |\sigma| \mid F\upharpoonright_\rho = 0 \\ &\qquad\qquad \wedge \ f_1\upharpoonright_\rho \neq 0 \ \wedge \ *(\rho_T) = Y \ \wedge \ \rho(\sigma) = *] \\ &\qquad \times \ Pr[\rho(\sigma) = * \mid \\ &\qquad\qquad F\upharpoonright_\rho = 0 \ \wedge \ f_1\upharpoonright_\rho \neq 0 \ \wedge \ *(\rho_T) = Y] \\ &= \sum_{\sigma \in P(D,Y)} Pr[\delta((f\upharpoonright_\sigma)\upharpoonright_\rho) \geq s - 2|Y| \mid \\ &\qquad\qquad F\upharpoonright_\rho = 0 \ \wedge \rho(T \setminus Y) = 1 \ \wedge \ \rho(\sigma) = *] \\ &\qquad \times \ Pr[\rho(\sigma) = * \mid \\ &\qquad\qquad F\upharpoonright_\rho = 0 \ \wedge \ \rho(T \setminus Y) = 1 \ \wedge \ \rho(Y) = *] \end{aligned}$$

The last inequality above holds because $|\sigma| \leq 2|Y|$, the events $f_1\upharpoonright_\rho \neq 0 \ \wedge \ *(\rho_T) = Y$ are equivalent to the events $\rho(Y) = * \ \wedge \ \rho(T \setminus Y) = 1$, and the condition $\rho(Y) = *$ is implied by $\rho(\sigma) = *$. Recall that if Y is a map, $v(Y) \subseteq D$ denotes the set of underlying vertices which are contained in the map. We will split up the map σ into two maps, σ_1 and σ_2, where a variable, $P_{ij} \in \sigma$ is in σ_1 if both $i \in v(Y)$ and $j \in v(Y)$. Otherwise, $P_{ij} \in \sigma_2$. Note that for every $\sigma \in Proj_D[v(Y)]$, $0 \leq |\sigma_1| \leq |Y|$. We further divide the above probability into sums according to the size of σ_1 to get:

$$\begin{aligned} \sum_{i=0}^{|Y|} \sum_{\substack{\sigma \in P(D,Y), \\ |\sigma_1| = |Y| - i}} & Pr[\delta((f\upharpoonright_\sigma)\upharpoonright_\rho) \geq s - 2|Y| \mid F\upharpoonright_\rho = 0 \\ &\qquad \wedge \ \rho(T \setminus Y) = 1 \ \wedge \ \rho(\sigma) = *] \quad (3) \\ &\times \ Pr[\rho(\sigma) = * \mid F\upharpoonright_\rho = 0 \\ &\qquad \wedge \ \rho(T \setminus Y) = 1 \ \wedge \ \rho(Y) = *] \quad (4) \end{aligned}$$

For a fixed value of Y and $\sigma \in P(D, Y)$, we estimate the first term. Let f' be f with f_1 removed and consider the different possibilities for σ. Let f' be f with the variables in $T \setminus Y$ set to 1. Let F' be $F \vee G$ where $G\upharpoonright_\rho = 0$ if and only if ρ sets all variables in $T \setminus Y$ to 1. Then the first term is equal to

$$Pr[\delta((f'\upharpoonright_\sigma)\upharpoonright_\rho) \geq s - 2|Y| \mid F'\upharpoonright_\rho = 0 \ \wedge \ \rho(\sigma) = *].$$

Since $f'\upharpoonright_\sigma$ contains no variables which involve vertices of $v(\sigma)$ we can let $D' = D - v(\sigma)$ and conclude using Lemma 2.1 that the above probability is no greater than

$$Pr[\delta((f'\upharpoonright_\sigma)\upharpoonright_\rho) \geq s - 2|Y| \mid F'\upharpoonright_\rho = 0],$$

where the probability is for a ρ chosen from $\mathcal{R}_p^{D'}$. Now, if $\sigma = Y$ then f_1' is satisfied by σ and $f\upharpoonright_\rho$ is the constant 1 and this probability is $0 \leq \alpha^{s-2|Y|}$. Otherwise, $\sigma \neq Y$, the map f_1' is falsified by σ, so $f'\upharpoonright_\sigma$ has one fewer map than the original f that we started with. Furthermore, since $|\sigma| \leq 2|Y|$ and $p(m - s) \geq r$, $p(m - |\sigma| - (s - 2|Y|)) \geq r$ and we can apply the inductive hypothesis for D' and f'. It follows that the above quantity is no greater than $\alpha^{s-2|Y|}$.

Since the above calculation gives the same upper bound for term (3) for all values of σ, we can pull this quantity outside the sum to obtain:

$$\alpha^{s-2|Y|} \sum_{i=0}^{|Y|} \sum_{\substack{\sigma \in P(D,Y), \\ |\sigma_1| = |Y| - i}} Pr[\rho(\sigma) = * \mid F\upharpoonright_\rho = 0 \ \wedge \rho(T \setminus Y) = 1 \wedge \rho(Y) = *] \quad (5)$$

Now we will estimate the inner summation for a fixed value of i. As above, we replace the condition $F\upharpoonright_\rho = 0 \ \wedge \rho(T \setminus Y) = 1$ by the single condition $F'\upharpoonright_\rho = 0$. Also, for a particular σ, the event $\rho(\sigma) = *$ is equivalent to the events $\rho(\sigma_1) = * \ \wedge \rho(\sigma_2) = *$. Because $\rho(\sigma_1) = *$ is implied by $\rho(Y) = *$, the inner summation is equivalent to

$$\sum_{\substack{\sigma \in P(D,Y),\\ |\sigma_1| = |Y| - i}} Pr[\rho(\sigma_2) = * \mid F'\restriction_\rho = 0 \ \wedge \rho(Y) = *].$$

We would like to remove the conditioning on $F'\restriction_\rho = 0$ but it is not as simple as it was in Claim A. We have to consider the terms in this sum in the aggregate rather than individually. Let N_i be the number of σ's such that $|\sigma_1| = |Y| - i$. Then the above probability can be rewritten as:

$$N_i \cdot Pr_{(\sigma_2,\rho)}[\rho(\sigma_2) = * \mid F'\restriction_\rho = 0 \ \wedge \ \rho(Y) = *],$$

where the above probability is over all pairs (σ_2, ρ), such that $|\sigma_1| = |Y| - i$. For each σ_2, let u be the set of vertices in σ_2 which are not contained in $v(Y)$. Note that the number of domain vertices of u equals the number of range vertices of u and is equal to i. Also note that for σ_2 chosen at random, u is a uniformly distributed set over $D'' = D \setminus v(Y)$ having these properties. Applying Lemma 2.1 and letting V_i be the collection of all sets over D'' having both domain and range size i, this probability is equal to $N_i \cdot Pr_{(u,\rho)},[\rho(u) = * \mid F'\restriction_\rho = 0]$, where the probability is over all pairs (u, ρ), such that $u \in V_i$ and $\rho \in \mathcal{R}_p^{D''}$. This probability can be further divided according to $\#(\rho)$, the exact number of stars that are assigned to D_1 by ρ:

$$N_i \cdot \sum_{j=0}^{n} Pr_{(u,\rho)}[\rho(u) = * \mid F'\restriction_\rho = 0 \ \wedge \ \#(\rho) = j] \\ \times \ Pr_{(u,\rho)}[\#(\rho) = j \mid F'\restriction_\rho = 0].$$

Given that $\#(\rho) = j$, for a randomly chosen u the event $\rho(u) = *$ is independent of $F'\restriction_\rho = 0$. Thus the above probability is equal to

$$N_i \cdot \sum_{j=0}^{n} Pr_{(u,\rho)}[\rho(u) = * \mid \#(\rho) = j] \\ \times \ Pr[\#(\rho) = j \mid F'\restriction_\rho = 0],$$

where we have dropped the subscript on the probability in the second factor in each term since this probability only depends on ρ. For all $k \leq n$, $\sum_{j \geq k} Pr[\#(\rho) = j \mid F'\restriction_\rho = 0]$ equals $Pr[\#(\rho) \geq k \mid F'\restriction_\rho = 0]$, because the events are disjoint. Similarly, $\sum_{j \geq k} Pr[\#(\rho) = j]$ equals $Pr[\#(\rho) \geq k]$.

Claim B. For all k,
$Pr[\#(\rho) \geq k \mid F'\restriction_\rho = 0] \ \leq \ Pr[\#(\rho) \geq k]$.

Proof of Claim B. As in the proof of Claim A, we will prove this inequality by showing that for all k, $Pr[F'\restriction_\rho = 0 | \#(\rho) \geq k] \leq Pr[F'\restriction_\rho = 0]$. Let $F(C) = Pr[F'\restriction_\rho = 0]$. Then $F(C)$ is a weighted average of $F(A)$ and $F(B)$, where $F(A) = Pr[F'\restriction_\rho = 0 \mid \#(\rho) \geq k]$ and $F(B) = Pr[F'\restriction_\rho = 0 \mid \#(\rho) < k]$. We want to show that $F(A) \leq F(B)$, and then it follows that $F(A) \leq F(C)$, as desired. Let $F(i) = Pr[F'\restriction_\rho = 0 \mid \#(\rho) = i]$. Then $F(A)$ is a weighted average of terms $\{F(i),\ k \leq i \leq n\}$, and $F(B)$ is a weighted average of terms $\{F(i),\ 1 \leq i < k\}$. Thus, it suffices to show that for all k, $F(k) \leq F(k-1)$. Here we will consider ρ as being chosen from the alternative experiment, $\mathcal{P}_p^D$; recall that ρ is a pair $< \pi, S_1 >$, where π is a permutation from all of D_1 into D_0, and S_1 is the subset of D_1 which is set to $*$. We will divide the probability according to the particular permutation, π, chosen by ρ. Because each permutation is equally likely, it suffices to prove the above inequality conditional on the fact that the permutation is π. For all k, let the subdistribution A_π^k consist of those $\rho = (\pi, S_1)$ such that $|S_1| = k$, i.e. those ρ that were chosen by first choosing π and then choosing exactly k elements of D_1 to be $*$. We want to show that the probability that F' is forced to 0 over distribution A_π^{k-1} is greater than or equal to the probability that F' is forced to 0 over distribution A_π^k. Consider the collection C_k of sets S_1 with $|S_1| = k$, such that $\rho = (\pi, S_1)$, and $F'\restriction_\rho = 0$; similarly let C_{k-1} be those sets, S_1', $|S_1'| = k - 1$ such that $\rho' = (\pi, S_1')$ forces F' to 0. For any set $S_1' \subset S_1$, if $\rho = (\pi, S_1)$ forces F' to zero, then $\rho' = (\pi, S_1')$ also forces F' to zero; in particular, this holds for those subsets S_1' of size $k-1$. Thus, for each set in C_k there are k corresponding sets in C_{k-1} which are also forced to zero. Conversely, for each set in C_{k-1}, there are $(n-k+1)$ corresponding sets in C_k. The probability that a random ρ over A_π^k forces F' to 0 equals $\frac{|C_k|}{|A_\pi^k|}$; thus the probability that a random ρ' over A_π^{k-1} forces F' to 0 is at least $\frac{|C_k| \cdot k}{(n-k+1)|A_\pi^{k-1}|}$. Since $|A_\pi^{k-1}|$ is equal to $\frac{k|A_\pi^k|}{n-k+1}$, the probability that F' is forced to 0 over A_π^{k-1} is greater than or equal to the probability that F' is forced to 0 over A_π^k. ∎

Using Claim B and noting that $Pr_{(u,\rho)}[\rho(u) = * \mid \#(\rho) = j] \leq Pr_{(u,\rho)}[\rho(u) = * \mid \#(\rho) = j + 1]$ for all $j \geq 0$, we can apply Lemma 3.3 with $\alpha_j =$

$Pr_{(u,\rho)}[\rho(u) = * \mid \#(\rho) = j]$, $a_j = Pr[\#(\rho) = j \mid F'\lceil_\rho = 0]$, and $b_j = Pr[\#(\rho) = j]$ to show that the above probability is no greater than

$$N_i \cdot \sum_{j=0}^{n} Pr_{(u,\rho)}[\rho(u) = * \mid \#(\rho) = j] \cdot Pr[\#(\rho) = j]$$

which is equal to $N_i \cdot Pr_{(u,\rho)}[\rho(u) = *]$.

Since for each fixed value of $u \in V_i$, the probability that $\rho(u) = *$ is the same, the above probability is equal to $N_i \cdot Pr[\rho(u) = *]$, where the probability is now over the distribution $\mathcal{R}_p^{D''}$. Letting $m' = m - |Y|$ and using the fact that $pm' \geq p(m - s/2) \geq r \geq |Y|$, we can apply Lemma 3.4 to conclude that for $u \in V_i$, $Pr[\rho(u) = *] \leq 2 \cdot \frac{(2p^2m')^i(m'+1-i)!}{(m'+1)!} \leq 2 \cdot \frac{(2p^2n)^i(m'+1-i)!}{(m'+1)!}$.

Recall that N_i is equal to the number of σ's such that $|\sigma_1| = Y - i$. There are at most $\binom{|Y|}{i}^2(|Y| - i)!$ choices of σ_1 with $|\sigma_1| = |Y| - i$ and for each such σ_1 there are at most $\left(\frac{(m'+1)!}{(m'+1-i)!}\right)^2$ choices of σ_2. Thus there are a total of at most $\binom{|Y|}{i}^2(|Y| - i)!\left(\frac{(m'+1)!}{(m'+1-i)!}\right)^2$ choices of $\sigma \in P(D, Y)$ such that $|\sigma_1| = |Y| - i$.

Thus for all Y such that $2|Y| \leq s$, using the expression in (5), we have

$$\begin{aligned}
&Pr[\delta(f\lceil_\rho) \geq s \mid F\lceil_\rho = 0 \ \wedge \ f_1\lceil_\rho \neq 0 \ \wedge \ *(\rho_T) = Y] \\
&\leq 2\sum_{i=0}^{|Y|} \binom{|Y|}{i}^2 (|Y| - i)! \left(\frac{(m - |Y| + 1)!}{(m - |Y| + 1 - i)!}\right)^2 \\
&\qquad \times \alpha^{s-2|Y|} \frac{(2p^2n)^i(m - |Y| + 1 - i)!}{(m - |Y| + 1)!} \\
&\leq 2\sum_{i=0}^{|Y|} \binom{|Y|}{i}^2 (|Y| - i)!(m - |Y| + 1)^i \\
&\qquad \times \alpha^{s-2|Y|}(2p^2n)^i \\
&\leq \ 2\alpha^{s-2|Y|} \sum_{i=0}^{|Y|} \binom{|Y|}{i}^2 (|Y| - i)!(2p^2n^2)^i \\
&\leq \ 2\alpha^{s-2|Y|} \sum_{i=0}^{|Y|} \binom{|Y|}{i} (2p^2n^2)^i(|Y|)^{|Y|-i} \\
&= \ 2\alpha^{s-2|Y|}|Y|^{|Y|} \sum_{i=0}^{|Y|} \binom{|Y|}{i} (\frac{2p^2n^2}{|Y|})^i \\
&= \ 2\alpha^{s-2|Y|}|Y|^{|Y|}(\frac{2p^2n^2}{|Y|} + 1)^{|Y|} \\
&\leq \ 2\alpha^{s-2|Y|}|Y|^{|Y|}(\frac{3p^2n^2}{|Y|})^{|Y|} \\
&\leq \ 2\alpha^{s-2|Y|}(3p^2n^2)^{|Y|}.
\end{aligned}$$

For Y such that $2|Y| > s$ we cannot use the expansion in terms of (3) and (4) to estimate this probability. However in this case, since $\alpha \leq 1$ and $3p^2n^2 \geq 1$, $2\alpha^{s-2|Y|}(3p^2n^2)^{|Y|} > 1$ so it still is an upper bound on this probability.

Plugging in the bounds we have for the terms (1) and (2) we get

$$\begin{aligned}
&Pr[\delta(f\lceil_\rho) \geq s \mid F\lceil_\rho = 0 \ \wedge \ f_1\lceil_\rho \neq 0] \\
&\leq \ 4 \sum_{\substack{Y \subseteq T, \\ Y \neq \emptyset}} \alpha^{s-2|Y|}(3p^2n^2)^{|Y|}(3p^2n)^{|Y|} \\
&= \ 4\alpha^s \sum_{\substack{Y \subseteq T, \\ Y \neq \emptyset}} \left(\frac{9p^4n^3}{\alpha^2}\right)^{|Y|} \\
&\leq \ 4\alpha^s \sum_{i=1}^{r} \binom{r}{i} \left(\frac{9p^4n^3}{\alpha^2}\right)^i \\
&= \ 4\alpha^s \left[\left(1 + \frac{9p^4n^3}{\alpha^2}\right)^r - 1\right] \\
&\leq \ \alpha^s
\end{aligned}$$

The last inequality holds since α satisfies $(1 + 9p^4n^3/\alpha^2)^r \leq 5/4$. ∎

4 Critical Truth Assignments and Approximate Negation

For the pigeonhole variables, P_{ij}, $i \in D_0$, $j \in D_1$, where $size(D_0) = size(D_1) + 1$, we will consider the class of truth assignments which are *maximally* one-to-one. The set of *critical truth assignments* over D, CTA_D, is defined to be the class of all truth assignments over D which are one-to-one on all but one element of D_0: $CTA_D = \{\alpha \mid \exists x \in D_0 \text{ such that } \alpha \text{ is 1-1 on } D_0 \setminus \{x\} \cup D_1, \text{ and } \forall j \in D_1 \ P_{xj} = 0\}$. Given a map disjunction, f, over the pigeonhole variables, we want to apply the above switching lemma in order to obtain a new map disjunction which approximates $\neg f$.

Lemma 4.1. Let $D = D_0 \cup D_1$ where $|D_0| = n + 1$, $|D_1| = n$, and let T be a 1-1 decision tree of height k defined over the set D. At least a $1 - \frac{k}{|D_0|}$ fraction of all critical truth assignments α over D are consistent with some path in T.

Proof. We prove this claim by induction on the height of T, k. Consider a randomly chosen critical truth assignment α over D.

If $k = 0$ then T is just a single node and the Lemma is vacuously true.

Now suppose that the lemma is true for all trees of height at most k and suppose that T has height $k + 1$.

If the root of T is labelled by some $j \in D_1$ then α matches j with a unique $i \in D_0$. Let σ be the map consisting of P_{ij}. Then $T\lceil_\sigma$ is a 1-1 decision tree of height at most k defined over $D\lceil_\sigma$. Furthermore, the probability that α is consistent with some path in T is equal to the probability that it is consistent with some path in $T\lceil_\sigma$. By the induction hypothesis this is at least $1 - k/n \geq 1 - (k + 1)/(n + 1)$ as required.

If the root of T is labelled by some $i \in D_0$ then either $i = spare(\alpha)$ or $spare(\alpha) \neq i$ and α matches i with a unique $j \in D_0$. Let E be the event that α is not consistent with any path in T. Thus we have

$$\begin{aligned} Pr[E] \leq \; & Pr[spare(\alpha) = i] + \\ & Pr[spare(\alpha) \neq i] \times Pr[E \mid spare(\alpha) \neq i]. \end{aligned}$$

Since the induced distribution on $spare(\alpha)$ is uniform over D_0, $Pr[spare(\alpha) = i] = 1/(n + 1)$. Given that $spare(\alpha) \neq i$ we can argue, as in the case that the label was $j \in D_1$, that the probability of E is at most k/n. Thus we get a total probability of E of

$$\frac{1}{n+1} + \left(1 - \frac{1}{n+1}\right)\frac{k}{n} = \frac{1}{n+1} + \frac{k}{n+1} = \frac{k+1}{n+1}$$

as required. ▮

Corollary 4.2. Let $D = D_0 \cup D_1$ where $|D_0| = n + 1$, $|D_1| = n$, and let T be a 1-1 decision tree of height k representing f over the set D. Then $maps(T')$ and $\neg f$ agree on at least a $(1 - \frac{k}{n+1})$ fraction of all critical truth assignments over D.

5 Exponential Lower Bounds – Proof 1.

5.1 Overview

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, $\{\mathrm{PHP}_n : n \in N\}$, 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. We encode PHP_n using $(n+1)n$ propositional variables, $\{P_{ij} : i \in D_0 \wedge j \in D_1\}$, 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 $\neg\mathrm{PHP}_n$, which can be written as the conjunction of the following *pigeonhole clauses*:

$$\begin{gathered} \bigvee\{P_{ij} : j \in D_1\},\ i \in D_0; \\ \bigvee\{\neg P_{ik}, \neg P_{jk}\},\ i \neq j,\ i, j \in D_0,\ k \in D_1. \end{gathered}$$

In a refutation, one starts with the negated clauses $\neg PHP_n$ as axioms and then derives $\bigvee\{\}$, i.e. False.

As in the paper by Bellantoni, Pitassi and Urquhart ([BPU]), we proceed by induction on the depth of the Frege proof. Assume that we have a small, depth d Frege proof of the pigeonhole principle. Without loss of generality, we also assume that each formula in the proof consists of ORs and NOTs, except for the bottom two levels which are ORs of small ANDs. Applying a random restriction to each formula in 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 allows us to reduce the depth of each formula to $d - 1$, but now each depth $d - 1$ formula only approximates the original depth d formula on the reduced domain. Due to this approximation, instead of obtaining a depth $d - 1$ refutation of the pigeonhole principle (on the reduced domain) which is completely sound, we obtain a depth $d - 1$ *approximate* refutation which is only approximately sound.

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. The approximation is obtained by a new method which will be described in the next section. The key property of the approximation is that the pseudo-complement has the property that it is identical to the actual complement on a large fraction of the assignments that are maximally 1-1, namely the critical truth assignments.

We repeat the restriction argument $d-2$ times to obtain an approximate depth-2 Frege refutation of the pigeonhole principle, *i.e.* a refutation in which each formula is an OR of small ANDs. We then apply a separate base case argument which shows that there can be no good approximation to a Frege proof of small size and with this special form.

5.2 Definitions

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 simulation.

The Frege refutation system that we will use is the system H described in [BPU]. H is slightly nonstandard in that it is formulated as a propositional proof system for unbounded fan-in formulas. More precisely, the formulas of H are unordered rooted trees defined inductively by the rules: (1) if γ is a set of variables then $\bigvee\{\bigwedge\gamma\}$ is a formula; (2) if A is a formula then $\neg A$ is a formula; and (3) 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. The system H has one axiom: Excluded Middle Axiom $A\bigvee\neg A$, and two rules: (1) Weakening Rule $A \Rightarrow A\bigvee B$; (2) Cut Rule $(\neg A\vee B),(A\vee C) \Rightarrow (B\vee C)$, where A, B and C represent formulas. In addition, associativity and merging and unmerging of $\vee$ are implicit. The crucial property of H that we will exploit is that each rule and axiom involves at most one negation.

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 the 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. A Frege refutation of $A_1\wedge A_2\wedge\ldots\wedge A_k$ can be viewed as a directed acyclic graph, where each node in the graph is a formula of the proof. The leaves of the graph are the formulas A_i, the root of the graph is the empty (false) formula, and two formulas, A and B are parents of another formula C if C follows by some inference rule from A and B. A Frege refutation has ***height*** h if the directed acyclic graph which describes the proof has height no greater than h.

We will relax our proof rules to yield a new approximate proof system, H', as follows. The Weakening Rule does not change; the approximate Cut Rule is: $(A\vee B)(A'\vee C)\rightarrow(B\vee C)$, and the approximate Excluded Middle Axiom is: $A\vee A'$. An application of the approximate cut rule is γ-sound if A' is equal to $\neg A$ on a fraction γ of all critical truth assignments. Similarly, an application of the approximate excluded middle axiom is γ-sound if A' is equal to $\neg A$ on a fraction γ of all critical truth assignments. All applications of the weakening rule are 1-sound. A proof in H' is γ-sound if all inferences in the proof are γ'-sound, for some $\gamma'\geq\gamma$. Note that a γ-sound proof has the property that for each inference there exists a a subset S of all critical truth assignments, CTA_D of size at least $\gamma|CTA_D|$, such that for each assignment $s\in S$, if s makes all precedents of the inference true, then s also makes the antecedent of the inference true. (Note that an axiom can be viewed as a rule with one precedent, the "true" formula.) All rules and axioms in H are 1-sound rules; a completely unsound rule such as $[(1,1)\rightarrow 0]$ is 0-sound.

5.3 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. Let P be a sequence of formulas over D, $|D_0| = n+1$, $|D_1| = n$, each of depth at most d ($d>2$) and let $\rho\in\mathcal{R}_p^D$. Suppose that ρ leaves exactly those variables in $D'\subseteq D$ unset, where $|D_0'| = n'+1$, $|D_1'| = n'$. P is converted into a sequence of depth $d-1$ formulas over D' in the following three steps. When P' is obtained by applying the conversion process to P with ρ, we say that P' is P converted by ρ.

(1) Apply ρ to each formula of P, obtaining $P\upharpoonright_\rho$.

(2) Let $G_0 ... G_m$ be the distinct map disjunctions appearing in formulas of $P\upharpoonright_\rho$. Represent each G_i by some 1-1 decision tree T_i over D'. Define the *pseudo-complement* of G_i, $c_{D'}(G_i) = maps(T_i')$.

Replace each occurrence of $\neg G_i$ by $c_{D^i}(G_i)$, uniformly throughout $P\lceil_\rho$.

(3) For each formula of $P\lceil_\rho$, merge together OR gates appearing at heights 2 and 3.

Definition. A refutation of PHP_n over D in H' is (n, d, t, γ, S)-approximate if: each formula has depth at most d, each map disjunction has mapsize t, the total size of all formulas in the proof is at most S, each inference is γ-sound, and the proof was obtained from a (1-sound) proof in H of the pigeonhole principle over a larger universe, by applying the above conversion process (to the sequence of formulas in the proof) a finite number of times.

The following lemma shows that if we choose the right restrictions, then successive applications of the above conversion process results in an approximately sound refutation. The main idea behind the proof of this lemma is that while each formula may not be approximated well at all (since every negation is approximated, and there may be many negations in each formula), each inference will still remain approximately sound because each rule and axiom of H involves at most one negation.

Lemma 5.1. (Conversion Lemma) Let P^0 be a refutation in H of PHP_n over D, of depth d and size S. Let $k+1 \leq d-2$. Let $\rho = \rho^0, \rho^1, \rho^2, ..., \rho^k$ be a sequence of restrictions such that ρ^i leaves all variables over D^{i+1} unset, and $D^{k+1} \subseteq D^k \subseteq ... \subseteq D^1 \subseteq D$. Also, let $|D_1^i| = n_i$, and $|D_0^i| = n_i + 1$. Let $P^1, P^2, ..., P^{k+1}$ be a sequence of proofs in H' where P^{i+1} is equal to P^i converted by ρ^i. Suppose also that for every i, every map disjunction in P^i has mapsize at most t_i, and $t_i \leq t_{i+1}$ for all $i \leq k$. Let $\gamma_i = 1 - \frac{t_i}{n_i+1}$. If for all i, $1 \leq i \leq k$, P^i is a proof in H' which is $(n_i, d-i, t_i, \gamma_i, S)$-approximate, then P^{k+1} is a refutation of $PHP_{n_{k+1}}$ in H' which is $(n_{k+1}, d-(k+1), t_{k+1}, \gamma_{k+1}, S)$-approximate.

Proof. The conversion process, applied to any proof in H' of depth d yields a new proof in H' of depth $d-1$ and size at most S. Applying the conversion process $k+1$ times thus yields a new proof in H' of depth $d-(k+1)$ and size at most S. Because ρ leaves exactly those variables in D^{k+1} unset, where $|D_0^{k+1}| = n_{k+1}+1$ and $|D_1^{k+1}| = n_{k+1}$, it follows that P^{k+1} is a proof of $PHP_{n_{k+1}}$ in H' over D^{k+1}. Also, since $size[c_{D^{k+1}}(G\lceil_{\rho_k})] \leq t_{k+1}$ for every map disjunction G in P^k, step (3) of the conversion process insures that P^k converted by ρ_k will have mapsize t_{k+1}. It is left to show that every inference in P^{k+1} is γ_{k+1}-sound. Fix a particular formula f^0 in P^0. Let f^i be the formula which results from f^0 after i conversion steps – f^i is the corresponding formula in P^i. We want to show that f^{k+1} follows from a γ_{k+1}-sound inference. There are three cases to consider: either f^0 is an application of the approximate excluded middle axiom, or f^0 follows from the cut rule, or f^0 follows from the weakening rule. Here we assume that f^0 follows from the cut rule; the other two possibilities are handled similarly. Assume that $f^0 = B \vee C$, where f^0 follows from $g^0 = A \vee B$ and $h^0 = \neg A \vee C$. Then for all proofs P^i, $1 \leq i \leq k$, g^i and h^i are the two formulas in P^i which imply f^i. The inference $(g^k, h^k) \rightarrow f^k$ has one of two forms, depending on the depth of g^k h^k.

(1) If the inference has the form $(A' \vee B'), (\neg A' \vee C') \rightarrow (B' \vee C')$ then there are two cases to consider. If A' has depth greater than 2, then the new inference will have the same form since the negation in front of A' will not yet be converted; hence the new inference will be 1-sound. On the other hand, if A' is a map disjunction, then $\neg A'\lceil_{\rho^k}$ will be replaced by $c_{D^{k+1}}(A'\lceil_{\rho^k})$. Because $size[c_{D^{k+1}}(G\lceil_{\rho^k})] \leq t_{k+1}$ for all map disjunctions G in P^k, by Corollary 4.2, we know that $c_{D^{k+1}}(A'\lceil_{\rho^k})$ will equal $\neg A'\lceil_{\rho^k}$ for at least $1 - \frac{t_{k+1}}{n_{k+1}+1}$ of the critical truth assignments over D^{k+1}. Hence this inference will be γ_{k+1}-sound.

(2) Otherwise, some previous f_i, $i < k$, which follows from g^i and h^i, has the form $(A' \vee B'), (c_{D^i}(A') \vee C') \rightarrow (B' \vee C')$. Let $\rho' = \rho^i \rho^{i+1} .. \rho^k$. The inference $(g^k, h^k) \rightarrow f^k$ thus has the form $(A'\lceil_{\rho'} \vee B'\lceil_{\rho'}), (c_{D^i}(A')\lceil_{\rho'} \vee C'\lceil_{\rho'}) \rightarrow (B'\lceil_{\rho'} \vee C'\lceil_{\rho'})$. Because every map disjunction in P^i has mapsize t_i, the map disjunction $c_{D^i}(A')$ has mapsize t_i, which by assumption is less than or equal to t_{k+1}. By Corollary 4.2, this implies that $c_{D^i}(A')\lceil_{\rho'}$ equals $\neg A'\lceil_{\rho'}$ for at least a fraction $1 - \frac{t_{k+1}}{n_{k+1}+1}$ of the critical truth assignments over D^{k+1}, and hence the new inference will be γ_{k+1}-sound. ∎

5.4 The Lower Bound

Theorem 5.2 (Lower Bound on Size) For sufficiently large n, any Frege refutation of PHP_n of depth d must have size at least $\exp\left(n^{\frac{3}{4}4^{-(d+3)}}/24\right)$.

Corollary 5.3 (Lower Bound on Depth) For sufficiently large n, any Frege refutation of PHP_n of polynomial-size must have depth at least $\Omega(\log\log n)$.

Theorem 5.2 will be proven by induction on d, the depth of the Frege refutation. To facilitate the proof of the base case, we would like to restrict attention to Frege proofs which are *balanced.* Recall that the height of a Frege refutation is the height of the proof, viewed as a directed acyclic graph. A Frege refutation is *balanced* if the height of the refutation is logarithmic in the size of the refutation. The following lemma states that any Frege refutation can be efficiently converted into an equivalent, balanced one.

Lemma 5.4.(The Simulation Lemma) Any Frege refutation of size S and depth d can be transformed into another Frege refutation of size S^2, depth $d+2$ and height $O(\log S)$.

To prove the simulation lemma, we first show that any Frege proof can be converted into a proof in *tree form,* while preserving the size and depth of the proof to within small factors. A proof in *tree form* is a proof where each intermediate formula is only used once in the derivation. The main idea behind this proof is keep around all intermediate formulas that have been generated at each step in the derivation. Ie., If $P = f_0, f_1, ..., f_q$ is the original proof, then we construct the new proof, $P' = f'_0, f'_1, ..., f'_1$, where f'_i is the conjunction of $f_0, f_1, ..., f_i$. Secondly, we show that any size S, depth d Frege proof in tree form can be efficiently converted into a *balanced* Frege proof – one which has size S^2 and height $O(\log S)$. The full proof of the Simulation Lemma can be found in [PBI].

By the above Simulation lemma, theorem 5.2 is a corollary to the following theorem.

Theorem 5.5.(Lower Bound on Size for balanced Frege refutations) For sufficiently large n, any balanced Frege refutation of PHP_n of depth d must have size at least $S = \exp\left(n^{\frac{3}{4}4^{-(d+1)}}/12\right)$.

Proof. The proof is by induction on d. Suppose that there were such a refutation, P, of PHP_n in our system H, of size S, depth d, and height at most $\log S$. Let $t = 2^{1/4}\log S$. Define $\lambda(n) = (n/256t)^{1/4}$. If λ^i is the i-fold composition of λ with itself then it is easy to show that $\lambda^i(n) \geq n^{4^{-i}}/(256t)^{1/3}$. If $S < \exp(n^{\frac{3}{4}4^{-(d+1)}}/12)$ then straightforward calculation shows that $t < \frac{1}{2}\lambda^{d+1}(n)$. Because the system H is sound, and each map disjunction has mapsize 1, P is a refutation in H' which is (n_0, d, t, γ_0, S)-approximate, where $n_0 = n$ and $\gamma_0 = \left(1 - \frac{t}{n_0+1}\right)$. Applying the Induction and Base lemmas below, we show that that for sufficiently large n_0, that there is no proof in H' of PHP_{n_0} which is (n_0, d, t, γ_0, S)-approximate.

Suppose that $n_i \geq \lambda(n_{i-1}) \geq ... \geq \lambda^i(n_0)$ for all i, $0 \leq i \leq d-2$. Let $p_i = \lambda(n_i)/n_i$ and $\gamma_i = \left(1 - \frac{t}{n_i+1}\right)$ for all such i.

Lemma 5.6. (Induction Lemma) Let P^i be a refutation of PHP_{n_i} in H' which is $(n_i, d-i, t, \gamma_i, S)$-approximate, where t, γ_i, and n_i are as above. Then there is a restriction ρ such that P^i converted by ρ is a refutation of $PHP_{n_{i+1}}$ in H' which is $(n_{i+1}, d-(i+1), t, \gamma_{i+1}, S)$-approximate, where t, γ_{i+1} and n_{i+1} are as above.

Proof. Let D be the domain of the formulas in P^i. Since $t \leq \frac{1}{2}\lambda(n_i)$ for any $i \leq d$, $p_i(n_i - t) \geq t$ so we can apply the Switching Lemma, for ρ drawn at random from $\mathcal{R}^D_{p_i}$ to get our desired result. The probability that P^i converted by ρ does not result in new proof, P^{i+1}, where each map disjunction has size at most t is at most $S\alpha^t$ where $0 < \alpha < 8p_i^2 n_i^{3/2} t^{1/2}$. Since $p_i = \lambda(n_i)/n_i$ we see that α is no greater than 1/2. It follows that since $t = 2^{1/4}\log S$, $S\alpha^t$ is no greater than 1/6.

The expected number of stars after applying the restriction ρ is $n_i p_i = \lambda(n_i)$. Since the number of stars is binomially distributed, for sufficiently large n_0, a random ρ leaves at least the expected number of stars with probability greater than 1/3. (See, for example, Lemma 4.1 of [BH]). Thus, there exists a restriction, ρ, leaving n_{i+1} stars, $n_{i+1} \geq \lambda(n_i)$, such that P^i converted by ρ results in a new proof P^{i+1} of $PHP_{n_{i+1}}$, where each map disjunction has size at most t and the depth is $d-(i+1)$. Now by the Conversion Lemma, P^{i+1} is also γ_{i+1}-sound. ∎

Lemma 5.7. (Base Case Lemma) For $t \leq n^\epsilon$, $\epsilon < 1/2$, there is no balanced, approximate proof of PHP_n, which is $1 - \frac{t}{n+1}$-sound, where each formula is a t-disjunction and the total size of the proof is 2^t. In particular, there is no proof of $PHP_{n_{d-2}}$ in H' which is $(n_{d-2}, 2, t_{d-2}, \gamma_{d-2}, S)$-approximate.

Proof of Lemma 5.7. Recall that a γ-sound proof of $\neg PHP_n$ has the property that each inference is sound with respect to at least the fraction γ of the total number of critical truth assignments. The idea is to hit the proof with another restriction of size no greater than $2t \log S = 2t^2$, to obtain an approximate proof of the pigeonhole principle on a subset, $[m]$, of $[n]$, with an inference of the form $[(1,1) \rightarrow 0]$. Such an inference is 0-sound. But this will be shown to contradict the lemma which states that a $(1 - \frac{t}{n+1})$-sound proof of PHP_n, when hit by a small restriction leaving m holes unset, should yield a $(1 - \frac{t}{m+1})$-sound proof of PHP_m.

We will obtain the restriction constructively, by walking up the proof, from the root (the "false" formula) to the leaves (the pigeonhole clauses), setting variables as we go along until we eventually force an inference $[(1,1) \rightarrow 0]$. The bottom formula is 0. Consider the precedents (there are at most two of them.) If both are 1, then we are done. Otherwise, either both are unset or (at least) one is a zero. If either is zero, then continue up that side. If both are unset, then force one of the two antecedents to 1 by setting t variables; this is possible because all formulas in the proof are t-disjunctions. If this forces the other antecedent to 0, then continue up this side; otherwise, we can force this antecedent to 1 by setting t additional variables. Continue in this fashion until we force an inference $[(1,1) \rightarrow 0]$. It is left to argue that this will eventually happen since any two clauses at the leaves can always be simultaneously forced to 1. Note that each leaf formula is either an approximate excluded middle axiom, or a pigeonhole clause, or a formula which has already been set to "1". There are 3 nontrivial cases to consider: (1) both leaf formulas are instances of the approximate excluded middle axiom; (2) one leaf formula is a pigeonhole clause and the other is an instance of the approximate excluded middle axiom; and (3) both leaf formulas are pigeonhole clauses. Assume case (1): both formulas, f_1 and f_2 are instances of the approximate excluded middle axiom. We first force f_1 to 1 by setting at most t variables. Because the proof has height $\log S = t$, at this point we have set at most $2t^2$ variables. Now, consider the resulting proof, over the new universe of size at least $n - 2t^2 > n/2$. By Corollary 4.2, each remaining approximate excluded middle axiom is $1 + \frac{2t}{n}$-sound. In particular, because f_2 is $1 + \frac{2t}{n}$-sound, it has not been forced to 0. By setting t additional variables, we can also force f_2 to 1. We can apply the same argument for case (2). Now assume case (3): both f_1 and f_2 are pigeonhole clauses. By examining the pseudo-complement applied to a pigeonhole clause, it is clear that any two such clauses can simultaneously be set to 1.

Since the proof has height $\log S = t$, and at each step in our ascent up to the leaves we have set at most $2t$ variables, we eventually force an inference $[(1,1) \rightarrow 0]$ by setting at most $2t^2$ variables. By lemma 5.1, we should now have an approximate refutation of $\neg PHP_m$, where $m = n - 2t^2 \geq \frac{n}{2}$, and which is $(1 - \frac{t}{m+1})$-sound. Because $\frac{t}{m+1} \leq 1/2$, we know that each inference in the approximate refutation of $\neg PHP_m$ is greater than 1/2-sound. However, we have forced a 0-sound inference, and hence we have reached a contradiction. ▮

6 Exponential Lower Bounds – Proof 2.

We shall use PHP_n in the form

$$\bigvee_{i \neq j \in D_0, k \in D_1} \neg(\neg p_{ik} \vee \neg p_{jk}) \vee \bigvee_{i \in D_0} (\neg \bigvee_{k \in D_1} p_{ik}).$$

Thus, the size of PHP_n is $O(n^3)$ and the depth is 4.

6.1 Complete systems of partial maps.

Let n be a natural number and D_0 and D_1 two sets of cardinalities $n+1$ and n respectively. 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_0 to D_1,

$$M := \{h :\subseteq D_0 \longrightarrow D_1 | h \text{ injective}\}.$$

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

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

(As h is injective, $|h| = |\text{dom}h| = |\text{rng}h|$.)

(b) A subset $S \subseteq M$ is k-complete iff it satisfies four conditions:

(i) $S \neq \emptyset$,

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

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

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

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

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:

$$S \times T = \{\delta \cup \tau \in M | \text{some}\delta \in S, \tau \in T\text{s.t} \\ \forall \delta' \in S, \tau' \in T, \neg(\delta' \cup \tau' \subset \delta \cup \tau)\}.$$

In other words, it is the set of $\subseteq$-minimal 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:

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

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

Proof.

(a) Let $\delta \in S$. As $|\delta| + l \leq ||S|| + l \leq n$ and T is l-complete, $\delta \cup \tau \in M$ for some $\tau \in T$. Assume $\delta' \cup \tau' \subseteq \delta \cup \tau$ for some $\delta' \in S, \tau' \in T$. Then $\delta \cup \delta' \in M$ and, by the k-completeness of $S, \delta = \delta'$. Analogously $\tau = \tau'$. Hence $\delta \cup \tau$ is $\subseteq$-minimal and thus in $S \times T$, so $S \times T \neq \emptyset$.

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)(ii) of Definition A.

To verify (b)(iii) 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. $h \cup \delta \cup \tau \in M$. Then clearly $h \cup (\delta' \cup \tau') \in M$ for some $\delta' \cup \tau' \subseteq \delta \cup \tau$, an element of $S \times T$.

Finally, as obviously $||S \times T|| \leq ||S|| + ||T|| \leq k + l$, condition (b)(iv) 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:

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

Lemma C1. Let $H, S, T \subseteq M$, let S be k-complete, $||T|| + k \leq n$ and $H \lhd S \lhd T$. 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, as in the proof of Lemma A, for some $\delta \in S, \delta \subseteq \tau$. Hence $h \cup \delta \in M$ and, as $H \lhd S, h' \subseteq \delta$ 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 $H, S, T \subseteq M, H \lhd S \lhd T, ||S|| + l \leq n, ||T|| + k \leq n$, and let S be k-complete and let 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 $H \lhd S, h' \subseteq \delta$ some other $h' \in H$. Therefore $\delta \in S(H)$.

Lemma C3. For any $S, H_i \in M, i \in I$,

$$S\left(\bigcup_I H_i\right) = \bigcup_I S(H_i).$$

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, $||T|| + k \leq n, S \lhd T$ 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, comming from boolean complexity.

Definition D.

(a) Let $0 < p < 1$. Then $\mathcal{R}_p^D$ is the probability space of maps $\rho \in M$, as defined in section 2.

(b) For $\rho, h \in M, h^\rho$ is undefined if $h \cup \rho \notin M$ and, if $h \cup \rho \in M$, $\mathrm{dom} h^\rho = \mathrm{dom} h \backslash \mathrm{dom} \rho$ and $h^\rho = h | \mathrm{dom} h^\rho$. Also, $D_0^\rho = D_0 \backslash \mathrm{dom} \rho, D_1^\rho = D_1 \backslash \mathrm{rng} \rho$ and $(n)^\rho = |D_1^\rho|$. For $H \subseteq M, H^\rho = \{h^\rho \mid h \in H \text{ and } h^\rho \text{ is defined}\}$.

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_0, D_1, M and some $H, S, T, \ldots \subseteq M$ to a situation with $(n)^\rho, D_0^\rho, D_1^\rho, M^\rho$ and $H^\rho, S^\rho, T^\rho, \ldots$, by choosing random $\rho \in \mathcal{R}_p^D$, 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) $H \lhd S$ implies $H^\rho \lhd S^\rho$,

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

(c) $K = S(H)$ and $H \lhd S$ 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) $S^\rho \neq \emptyset$ by k-completeness of S and by $|\rho| + k \leq n$. 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 third 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 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 $H \lhd S, h' \subseteq \delta$ 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. Let $H \subseteq M, ||H|| \leq t \leq s$ and $0 < p < 1$. Assume that $p(n - 5s^2) \geq 2$ and $t \leq p(n+1)$. Then for random $\rho \in \mathcal{R}_p^D$ the statement: "there is $2s$-complete $S \subseteq M^\rho$ such that $H^\rho \lhd S$" holds with probability at least $1 - (64p^4n^3t)^s$.

For the choice of $p = n^{\varepsilon - 1}$ and $t = s = n^\delta$ such that $0 < \delta < \varepsilon < \frac{1}{5}$ and n sufficiently large this probability is at least $1 - 2^{-n^\delta}$ even if we add the condition $|\rho| \leq n - \frac{1}{2}pn$ (using Chernoff inequality).

Proof. H can be viewed as a t-disjunction. Also, note that for any 1-1 decision tree T representing $H \restriction_\rho$

over $D\restriction_\rho$, the maps defined by the paths of T is a k-complete set S such that $H^\rho \lhd S$ where $k = \delta(H\restriction_\rho)$. Thus, by Lemma 3.2, with probability at least $1-\alpha^{2s}$, there exists a $2s$-complete $S \subseteq M^\rho$ where α satisfies $0 < \alpha \leq 8p^2n^{3/2}t^{1/2}$. Plugging in this term, we obtain the desired probability $1-(64p^4n^3t)^s$. ▮

6.2 The Lower Bound.

In this section, we will prove the following theorem.

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 for all sufficiently large n every depth d F-proof of PHP_n must have the size at least $\exp\left(n^{6^{-d}}\right)$.

Recall that we consider only formulas in the language $\vee, \neg, 0, 1, p_{ij}, i \in D_0, j \in D_1, |D_0 = n+1, |D_1| = n$. 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 (H, S).

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

such that:

(1) for every $\varphi \in \Gamma, H_\varphi \subseteq S_\varphi \subseteq M$ and S_φ is k-complete;

(2) $H_0 = \emptyset, H_1 = \{\emptyset\}, S_0 = S_1 = \{\emptyset\}$; $H_{p_{ij}} = \{\{(i,j)\}\}$;
$S_{p_{ij}} = \{\{(i,j')\}|i' \neq i, j' \neq j\} \cup \{\{(i,j)\}\}$;

(3) if $\neg\varphi \in \Gamma$, then $H_{\neg\varphi} = S_\varphi \backslash H_\varphi$, $S_{\neg\varphi} = S_\varphi$;

(4) if $\varphi \in \Gamma$ and $\bigvee_{i\in I} \varphi_i$ is the reduced form of φ, then

$$\bigcup_{i\in I} H_{\varphi_i} \lhd S_\varphi \text{ and } H_\varphi = S_\varphi\left(\bigcup_{i\in I} H_{\varphi_i}\right).$$

Let $\rho \in M$. We define

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

If φ is a formula, then φ^ρ is obtained by applying ρ to all variables of φ; if Γ is a set of formulas, then $\Gamma^\rho = \{\varphi^\rho | \varphi \in \Gamma\}$.

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

Proof. - use Lemma D1(a) and (b).

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 $H_{p_{ij}}$ and $S_{p_{ij}}$ by (2) of the definition. Clearly $S_{p_{ij}}$ 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 $d+1$, Γ 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 (H', 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 $D_0^{\rho'}$ and $D_1^{\rho'}$. By Lemma F1 the restrictions of M' and 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_0, D_1 replaced by $m, D_0^{\rho'}, D_1^{\rho'}$, 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 $n^\delta = n^{\varepsilon^{d-1}\cdot\frac{\delta}{\varepsilon^{d-1}}} \leq m^{\frac{\delta}{\varepsilon^{d-1}}}$ 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 $\bigcup_i (H'_{\varphi i})^{\rho''} \lhd S$ and $|\rho''| \leq m - 2pm = m - m^\varepsilon$. If this is the case, we extend (H', S') to φ by defining

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

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, (H, S) is a k-evaluation of the set of all subformulas of the proof and $k \leq n/f$, then $H_{\gamma_i} = S_{\gamma_i}$ 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. 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 (H, 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 $H_\xi = S_\xi$ for $\xi = \varphi_i(\psi_1, \ldots, \psi_m)$, $i = 1, \ldots, r$. Then $H_\xi = S_\xi$ for $\xi = \varphi_0(\psi_1, \ldots, \psi_m)$.

Proof of Claim. Let (H, 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 $S_\xi \lhd T$ for every $\xi \in \Gamma$. Such a system exists by Lemma B. Note also that $||S_\xi|| + ||T|| \leq n$ for every $\xi \in \Gamma$. Suppose that $\neg\xi \in \Gamma$. Then $H_{\neg\xi} = S_\xi \backslash H_\xi$, hence, by Lemma C5, $T(H_{\neg\xi}) = T \backslash T(H_\xi)$.

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,

$$H_{\alpha\vee\beta} = S_{\alpha\vee\beta}\left(\bigcup_{i\in A} H_{\gamma_i}\right) \cup S_{\alpha\vee\beta}\left(\bigcup_{i\in \beta} H_{\gamma_i}\right).$$

Now using Lemma C3 and using Lemma C1 twice we get

$$\begin{aligned} T(H_{\alpha\vee\beta}) &= T(S_{\alpha\vee\beta}\left(\bigcup_{i\in A} H_{\gamma_i}\right)) \cup T(S_{\alpha\vee\beta}\left(\bigcup_{i\in B} H_{\gamma_i}\right)) \\ &= T\left(\bigcup_{i\in A} H_{\gamma_i}\right) \cup T\left(\bigcup_{i\in B} H_{\gamma_i}\right) \\ &= T(S_\alpha\left(\bigcup_{i\in A} H_{\gamma_i}\right)) \cup T(S_\beta\left(\bigcup_{i\in b} H_{\gamma_i}\right)) \\ &= T(H_\alpha) \cup T(H_\beta). \end{aligned}$$

Furthermore, by Lemma C1 we have $T(H_\xi) = T(S_\xi) = T$, for $\xi = \varphi_i(\psi_1, \ldots, \psi_m), i = 1, \ldots, r$, since $H_\xi = S_\xi$ and S_ξ is complete.

Thus we have shown that the mapping $\xi \to T(H_\xi)$ 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, we must have $T(H_\xi) = T$ for $\xi = \varphi_0(\psi_1, \ldots, \xi_m)$, hence by definition and Lemma C2, $H_\xi = S_\xi(H_\xi) = S_\xi$, which concludes the proof of the claim and, consequently, of the lemma.

Lemma F4.

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

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

Proof.

(i) PHP_n is a disjunction of formulas $\neg\varphi$ where φ ranges over $\neg p_{ik} \vee \neg p_{jk}, i \neq j \in D_0, k \in D_1$, $\bigvee_{k\in D_1} p_{ik}, i \in D_0$. We shall show that $H_\varphi = S_\varphi$

for each such formula, hence $H_{\neg\varphi} = \emptyset$, thus

$$H_{PHP_n} = S_{PHP_n}(\bigcup H_{\neg\varphi}) = S_{PHP_n}(\emptyset) = \emptyset.$$

There are two cases. First, suppose that φ is $\neg p_{ik} \vee \neg p_{jk}$. By definition

$$\begin{aligned} H_{\neg p_{ik}} &= \{\{(i,k'),(i',k)\} | i' \neq i, k' \neq k\}; \\ H_{\neg p_{jk}} &= \{\{(j,k'),(j',k)\} | j' \neq j, k' \neq k\}. \end{aligned}$$

Let T be the 3-complete set

$$\begin{aligned} &\{\{(i,k'),(j,k''),(l,k)\} | i \neq l \neq j, k' \neq k'' \neq k\} \\ &\cup\{\{(i,k),(j,k'\} | k \neq k'\} \\ &\cup\{\{(i,k'),(j,k)\} | k \neq k'\}. \end{aligned}$$

It can be easily verified that $T(H_{\neg p_{ik}} \cup H_{\neg p_{jk}}) = T$. By Lemma B, we have some $n/2$-complete W which is a common refinement of S_φ and T. Hence, by Lemma C2, $W(H_{\neg p_{ik}} \cup H_{\neg p_{jk}}) = W$, and again by Lemma C2,

$$H_{\neg p_{ik} \vee \neg p_{jk}} = S(H_{\neg p_{ik}} \cup H_{\neg p_{jk}}) = S_{\neg p_{ik} \vee \neg p_{jk}}.$$

The second case is when φ is $\bigvee_{k \in D_1} p_{ik}$. By definition $H_\varphi = S_\varphi(\{(i,k) | k \in D_1\})$. But, clearly, $\{(i,k) | k \in D_1\}$ is 1-complete, hence

$$S_\varphi(\{(i,k) | k \in R\}) = S_\varphi.$$

(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 $H_{\neg\varphi} = \emptyset$.

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 $(\gamma_1, \ldots, \gamma_t)$ 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 (H,S) of Γ^ρ, where Γ is the set of subformulas of the proof $(\gamma_1, \ldots, \gamma_t)$. Clearly $(\gamma_1^\rho, \ldots, \gamma_t^\rho)$ is an F-proof with variables $p_{ij}, i \in D_0^\rho, j \in D_1^\rho$ 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 $H_{\gamma_i^\rho} = S_{\gamma_i^\rho}$, for $i = 1, \ldots, t$; by the second one, $H_{PHP_n^\rho} = \emptyset$, which is different from $S_{PHP_n^\rho}$. Thus, $(\gamma_1, \ldots, \gamma_t)$ cannot be a proof of PHP_n, i.e. any proof of PHP_n of depth d must have size at least 2^{n^δ}.

7 Acknowledgements

We wish to thank Alasdair Urquhart, Sam Buss, Stephen Cook and Stephen Bellantoni for many valuable conversations.

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.

[Be] P. Beame, "Lower Bounds for Recognizing Small Cliques on CRCW PRAM's," *Discrete Applied Mathematics*, v. 29 (1990), pp. 3-20.

[BH] P. Beame, J. Håstad, "Optimal Bounds for Decision Problems on the CRCW PRAM," *Journal of the ACM*, v. 36 (1989), pp.643-670.

[BPU] S. Bellantoni, T. Pitassi, A. Urquhart, "Approximation and small-depth Frege proofs," to appear in SIAM Journal of Computing. Also appeared in Proceedings from Sixth Structures in Complexity Theory, 1991.

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

[Cai] J. Cai, "With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy," *18th Annual Symposium on the Theory of Computing* (1986), pp.21-29.

[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.

[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.

[K] J. Krajíček, "Lower bounds to the size of constant-depth propositional proofs," (1991), preprint.

[KPW] J. Krajíček, P. Pudlák, A. Woods, "Exponential lower bounds to the size of bounded-depth Frege proofs of the pigeonhole principle," preprint (1991).

[Ly] J. Lynch, "A Depth-Size Tradeoff for Boolean Circuits with Unbounded Fan-in," *Structure in Complexity Theory Proceedings*, Lecture Notes in Computer Science 223, Springer 1986, pp. 234-248.

[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.

[PBI] T. Pitassi, P. Beame, R. Impagliazzo, "Exponential lower bounds for the pigeonhole principle," University of Toronto TR 257/91, (1991).

[T] Tseitin, G.S. "On the complexity of derivation in the propositional calculus," *Studies in Constructive Mathematics and Mathematical Logic*, Part II, A.O. Slisenko, 1968.

[Urq] A. Urquhart, "Hard Examples for Resolution," *JACM*, Volume 34, No. 1, January 1987, pp. 209-219.

[Wo] A. Woods, "Approximating truth tables of constant depth circuits," To appear.

[Y] A. C. Yao, "Separating the polynomial-time hierarchy by oracles," *Proceedings of 26th Annual IEEE Symposium on Foundations of Computer Science*, pp.1-10.