Bounded-depth Frege lower bounds for weaker pigeonhole principles

Josh Buresh-Oppenheim
Computer Science Department
University of Toronto
Toronto, ON M5S 3G4
bureshop@cs.toronto.edu

Paul Beame*
Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350
beame@cs.washington.edu

Toniann Pitassi†
Computer Science Department
University of Toronto
Toronto, ON M5S 3G4
toni@cs.toronto.edu

Ran Raz†
Faculty of Math and Computer Science
Weizmann Institute of Science
Rehovot 76100, Israel
ranraz@wisdom.weizmann.ac.il

Ashish Sabharwal*
Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350
ashish@cs.washington.edu

Abstract

We prove a quasi-polynomial lower bound on the size of bounded-depth Frege proofs of the pigeonhole principle PHP_n^m *where* $m = (1 + 1/\text{polylog}\, n)n$. *This lower bound qualitatively matches the known quasi-polynomial-size bounded-depth Frege proofs for these principles. Our technique, which uses a switching lemma argument like other lower bounds for bounded-depth Frege proofs, is novel in that the tautology to which this switching lemma is applied remains random throughout the argument.*

1 Introduction

The propositional pigeonhole principle asserts that m pigeons cannot be placed in n holes with at most one pigeon per hole whenever m is larger than n. It is an exceptionally simple fact that underlies many theorems in mathematics, and is the most extensively studied combinatorial principle in proof complexity. (See [19] for an excellent survey on the proof complexity of pigeonhole principles.) It can be formalized as a propositional formula, denoted PHP_n^m, in a standard way; by convention, this formalization rules out relational as well as functional mappings of m pigeons to n holes.

Proving super-polynomial lower bounds on the length of propositional proofs of the pigeonhole principle when $m = n+1$ has been a major achievement in proof complexity. The principle can be made weaker (and hence easier to prove) by increasing the number of pigeons relative to the number of holes, or by considering fewer of the possible mappings of pigeons to holes. Two well-studied examples of the latter weakenings, the onto-PHP and the functional-PHP, only rule out, respectively, surjective and functional mappings from pigeons to holes. In this paper, we will prove lower bounds that apply to all of these variations of the basic PHP.

For all $m > n$, Buss [9] has given polynomial-size Frege proofs of PHP_n^m. He uses families of polynomial-size formulas that count the number of 1's in an N-bit string and Frege proofs of their properties to show that the number of pigeons successfully mapped injectively can be at most the number of holes.

In weaker proof systems, where such formulas cannot be represented, the proof complexity of the pigeonhole principle depends crucially on the number of pigeons, m, as a function of the number of holes, n. As m increases, the principle becomes weaker (easier to prove) and in turn the proof complexity question becomes more difficult. We review the basics of what is known for Resolution and bounded-depth Frege systems below. Generally, the weak pigeonhole principle (WPHP) has been used to refer to PHP_n^m whenever m is at least a constant factor larger than n. We will be primarily concerned with forms of the pigeonhole principle that are significantly weaker than the usual pigeonhole principle but somewhat stronger than these typical weak forms.

For the Resolution proof system, the complexity of the pigeonhole principle is essentially resolved. In 1985, Haken proved the first super-polynomial lower bounds for unrestricted Resolution proofs of PHP_n^m, for $m = n+1$ [10]. This lower bound was generalized by Buss

*Research supported by NSF grant CCR-0098066
†Research supported by US-Israel BSF grant 98-00349

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

and Turan [8] for $m < n^2$. For the next 10 years, the resolution complexity of PHP_n^m for $m \geq n^2$ was completely open. A recent result due to Raz [17] gives exponential Resolution lower bounds for the weak pigeonhole principle, and subsequently Razborov has resolved the problem for most interesting variants of the PHP [20].

Substantially less is known about the complexity of the pigeonhole principle in bounded-depth Frege systems, although strong lower bounds are known when the number of pigeons m is close to the number of holes n. Ajtai proved super-polynomial lower bounds for PHP_n^{n+1} with an ingenious blend of combinatorics and nonstandard model theory [1]. This result was improved to exponential lower bounds in [4]. It was observed in [5] that the above lower bounds can in fact be applied to PHP_n^m for $m \leq n + n^{\varepsilon}$, for some ε that falls off exponentially in the depth of the formulas involved in the proof.

For the case of larger m (the topic of this paper), the complexity of bounded-depth Frege proofs of PHP_n^m is slowly emerging, with surprising and interconnected results. There are several deep connections between the complexity of the weak pigeonhole principle and other important problems. First, lower bounds for bounded-depth Frege proofs of the weak pigeonhole principles suffice to show unprovability results for the P versus NP statement (see [19]). Secondly, the long-standing question of whether or not the existence of infinitely many primes has an $I\Delta_0$ proof is closely related to the complexity of WPHP in bounded-depth Frege systems [16]. Thirdly, the question is closely related to the complexity of approximate counting [15].

In bounded-depth Frege systems more powerful than resolution, there are two significant prior results concerning the proof complexity of weak pigeonhole principles: There are bounded-depth Frege proofs of PHP_n^m for m as small as $n + n/\text{polylog}\, n$ of quasi-polynomial size [16, 13, 14]; thus exponential lower bounds for the weak pigeonhole principle are out of the question. In fact, this upper bound is provable in a very restricted form of bounded-depth Frege where all lines in the proof are disjunctions of polylog n-sized conjunctions, a proof system known as $Res(\text{polylog}\, n)$. On the other hand, [2] shows exponential lower bounds for weak pigeonhole principles in $Res(2)$, a proof system which allows lines to be disjunctions of size-2 conjunctions.

In this paper we prove quasi-polynomial lower bounds for the weak pigeonhole principle whenever $m \leq n + n/\text{polylog}\, n$. More precisely, we show that given integers c and h such that c is sufficiently large compared to h, there exists an integer $a > 1$ such that any depth-h proof of PHP_n^m, where $m \leq n + n/\log^c n$, requires size $2^{\log^a n}$. This is a substantial improvement over previous lower bounds. Our proof technique applies a switching lemma to a weaker tautology based on certain bipartite graphs. This type of tautology was introduced in [7]. Although we rely heavily on the simplified switching lemma arguments presented in [3, 21], in a major difference from previous switching-lemma-based proofs, both the tautologies themselves and the restrictions we consider remain random throughout most of the argument.

2 Overview

The high-level schema of our proof is not new. Ignoring parameters for a minute, we start with an alleged proof of PHP_n^m of small size. We then show that assigning values to some of the variables in the proof leaves us with a sequence of formulas, each of which can be represented as a particular type of decision tree of small height. This part of the argument is generally referred to as the switching lemma. We then prove that the leaves of any such short tree corresponding to a formula in the proof must all be labelled 1 if the proof is to be sound. Finally, we show that the tree corresponding to PHP_n^m has leaves labelled 0, which is a contradiction since it must appear as a formula in the alleged proof. We now overview the lower bound components in more detail.

The lower bounds for bounded-depth Frege proofs of PHP_n^{n+1} [1, 4] used *restrictions*, partial assignments of values to input variables, and iteratively applied "switching lemmas" with respect to random choices of these restrictions. The first switching lemmas showed that after one applies a randomly chosen restriction that assigns values to many, but far from all, of the input variables with high probability one can convert an arbitrary DNF formula with small terms into a CNF formula with small clauses (hence the name). More generally, such switching lemmas allow one to convert arbitrary DNF formulas with small terms into small height decision trees (which implies the conversion to CNF formulas with small clauses). The basic idea is that for each level of the formulas/circuits, one proves that a randomly chosen restriction will succeed with positive probability for all sub-formulas/gates at that level. One then fixes such a restriction for that level and continues to the next level. To obtain a lower bound one chooses a family of restrictions suited to the target of the analysis. In the case of PHP_n^m, the natural restrictions to consider correspond to partial matchings between pigeons and holes.

The form of the argument by which switching lemmas are proven generally depends on the property that the ratio of the probability that an input variable remains unassigned to the probability that it is set to 0 (respectively, to 1) is sufficiently less than 1. In the case of a random partial matching that contains $(1-p)n$ edges applied to the variables of PHP_n^m, there are pn unmatched holes and at least pm unmatched pigeons. Hence, the probability that any edge-variable remains unassigned (i.e. neither used nor ruled out by the partial matching) is at least p^2. However, the partial matching restrictions set less than a $1/m$ fraction of variables to 1. Thus the proofs required that $p^2n < p^2m < 1$ and thus $p < n^{-1/2}$. This compares with choices of $p = n^{-O(1/h)}$ for depth h circuit lower bounds in the best arguments for parity proven in [11]. Hence, the best known lower bounds on the size of depth-h circuits computing parity is of the form $2^{n^{\Omega(1/h)}}$, while the best known lower bound on the size of depth-h proofs of PHP_n^{n+1} is of the form $2^{n^{2^{-O(h)}}}$.

A problem with extending the lower bounds to PHP_n^m for larger m is that, after a partial matching restriction is applied, the absolute difference between the number pigeons and holes does not change but the number of holes is dramatically reduced. This can qualitatively change the ratio between pigeons and holes. If this is too large then the probability that variables remain unassigned grows dramatically and, in the next level, the above argument does not work at all. For example, with the above argument, if the difference between the number of pigeons and holes is as large as $n^{3/4}$ then after only one round the above argument will fail. The extension in [5] to lower bound proofs for $PHP_n^{n+n^{\varepsilon_h}}$ for formulas of depth h relies on the fact that even after h rounds of restrictions the gap is small enough that there is no such qualitative change; but this is the limit using the probabilities as above.

We are able to resolve the above difficulties for m as large as $n + n/\text{polylog}\, n$. In particular, we increase the probability that variables are set to 1 to $1/\text{polylog}\, n$ from $1/m$ by restricting the matchings to be contained in bipartite graphs G of polylog n degree. Thus we can keep as many as $n/\text{polylog}\, n$ of the holes unmatched in each round. Therefore, by choosing the exponents in the polylog n carefully as a function of the depth of the formulas, we can tolerate gaps between the number of pigeons and the number of holes that are also $n/\text{polylog}\, n$.

A difficulty with this outline is that one must be careful throughout the argument that the restrictions one chooses do not remove all the neighbors of a node without matching it, which would simplify the pigeonhole principle to a triviality. It is not at all clear how one could explicitly construct low degree graphs such that some simple additional condition on the restrictions that we choose at each stage could enforce the desired property. It is unclear even how one might do this nonconstructively because it is not clear what property of the random graph would suffice.

Instead, unlike previous arguments, we do not fix the graph in advance; we keep the input graph random throughout the argument, and consider for each such graph G its associated proof of the pigeonhole principle restricted to G. Since we do not know what G is at each stage we cannot simply fix the restriction as we deal with each level; we must keep that random as well. Having done this, we can use simple Chernoff bounds to show that, for almost all combinations of graphs and restrictions, the degree at each level will not be much smaller than the expected degree, so the pigeonhole principle will remain far from trivial. We adjust parameters to reduce the probability that a restriction fails to simplify a given level so that it is much smaller than the number of levels. Then we apply the probabilistic method to the whole experiment involving the graph G as well as the sequence of restrictions.

There is one other technical point that is important in the argument. In order for the probabilities in the switching lemma argument to work out it is critical that the degrees of vertices in the graph after each level of restriction is applied are decreased significantly at each step as well as being small in the original graph G. Using another simple Chernoff bound we show that the degrees of vertices given almost all combinations of graphs and restrictions will not be much larger than their expected value and this suffices to yield the decrease in degree.

Overall, our argument is expressed in much the same terms as those in [3, 21], although we find it simpler to omit formally defining k-evaluations as separate entities. One way of looking at our technique is that we apply two very different kinds of random restrictions to a proof of PHP_n^m: first, one that sets many variables to 0, corresponding to the restriction of the problem to the graph G, and then, one that sets partial matchings for use with the switching lemma.

3 Frege proofs and $WPHP(G)$

A *formula* is a tree whose internal nodes are labelled by either $\vee$ (fanin 2) or $\neg$ (fanin 1) and whose leaves are labelled by variables. Given a node in this tree, the full tree rooted at that node is called a (not necessarily proper) *subformula* of the original formula. If a formula contains no connectives, then it has *depth* 0. Otherwise, the *depth* of a (sub)formula A is the maximum number of alternations of connectives along any path from the root to leaf, plus one. The *merged form* of a formula A is the tree such that all $\vee$'s labelling adjacent vertices of A are identified into a single node of unbounded fanin, also labelled $\vee$.

A *Frege proof system* is specified by a finite set of sound and complete *inference rules*, rules for deriving new propositional formulas from existing ones by consistent substitution of formulas for variables in the rule.

A typical example is the following, due to Schoenfield, in which p,q,r are variables that stand for formulas and $p,q \vdash r$ denotes that p and q yield r in one step:
Excluded Middle: $\vdash \neg p \vee p$, Expansion Rule: $p \vdash q \vee p$, Contraction Rule: $p \vee p \vdash p$, Associative Rule: $p \vee (q \vee r) \vdash (p \vee q) \vee r$, Cut Rule: $p \vee q,\ \neg p \vee r \vdash q \vee r$.

We will say that the *size* of a Frege rule is the number of distinct subformulas mentioned in the rule. For example, the size of the cut rule above is 7; the subformulas mentioned are: $p,q,r,\neg p, p \vee q, \neg p \vee r, q \vee r$.

DEFINITION 3.1. A *proof* of a formula A in Frege system $\mathcal{F}$ is a sequence of formulas $A_1,\ldots,A_r = A$ such that $\vdash A_1$ and for all $i > 1$ there is some (possibly empty) subset $\mathcal{A} \subset \{A_1,\ldots,A_{i-1}\}$ such that $\mathcal{A} \vdash A_i$ is a substitution instance of a rule of $\mathcal{F}$.

DEFINITION 3.2. For a Frege proof Π, let $cl(\Pi)$ denote the closure of the set of formulas in Π under subformulas. The *size* of a Frege proof Π is $|cl(\Pi)|$, the total number of distinct subformulas that appear in the proof. The *depth* of a proof is the maximum depth of the formulas in the proof.

Let $G = (V_1 \cup V_2, E)$ be a bipartite graph where $|V_2| = n$ and $|V_1| = m > n$. We use $L(G)$ to denote the language built from the set of propositional variables $\{X_e : e \in E\}$, the connectives $\{\vee, \neg\}$ and the constants 0 and 1.

The following is a formulation of the onto and functional weak pigeonhole principle on the graph G. Note that if G is not the complete graph $K_{m,n}$, then this principle is weaker than the standard onto and functional weak pigeonhole principle.

DEFINITION 3.3. $WPHP(G)$ is the OR of the following four (merged forms of) formulas in $L(G)$. In general, i,j,k represent vertices in G and $\Gamma(i)$ represents the set of neighbors of i in G.

1. $\bigvee_{(e,e')\in I} \neg(\neg X_e \vee \neg X_{e'})$ for $I = \{(e,e') : e,e' \in E; e = \{i,k\}, e' = \{j,k\}; i,j \in V_1; i \neq j; k \in V_2\}$: two different pigeons go to the same hole.
2. $\bigvee_{(e,e')\in I} \neg(\neg X_e \vee \neg X_{e'})$ for $I = \{(e,e') : e,e' \in E; e = \{k,i\}, e' = \{k,j\}; i,j \in V_2; i \neq j; k \in V_1\}$: one pigeon goes to two different holes.
3. $\bigvee_{i\in V_1} \neg \bigvee_{j\in\Gamma(i)} X_{\{i,j\}}$: some pigeon has no hole.
4. $\bigvee_{j\in V_2} \neg \bigvee_{i\in\Gamma(j)} X_{\{i,j\}}$: some hole remains empty.

In fact, we consider an arbitrary orientation of the above formula whereby each $\vee$ is binary.

4 Representing matchings by trees

In this section we make minor modifications to standard definitions from [3, 21] to apply to the edge variables given by bipartite graphs and not just complete bipartite graphs.

Let G be a bipartite graph as in the last section and let D denote the set of Boolean variables X_e in $L(G)$. Assume there is an ordering on the nodes of G.

DEFINITION 4.1. Two edges of G are said to be *inconsistent* if they share exactly one endpoint. Two partial matchings ρ_1, ρ_2 on the graph G are said to be *consistent* if no edge in ρ_1 is inconsistent with an edge in ρ_2. For a partial matching ρ, let $\mathrm{Im}(\rho)$ denote the set of nodes of V_2 that are matched by ρ.

DEFINITION 4.2. For ρ a partial matching on the graph G that matches nodes $V_1' \subset V_1$ to nodes $V_2' \subset V_2$, we define $G|_\rho$ as the bipartite graph $((V_1 \setminus V_1') \cup (V_2 \setminus V_2'), E - (V_1' \times V_2 \cup V_1 \times V_2'))$.

DEFINITION 4.3. A *matching decision tree* T for G is a tree where each internal node u is labelled by a node of G, v, and each edge from a node u is labelled by an edge of G that touches v. Furthermore, given any path in the tree from the root to a node u, the labels of the edges along the path constitute a partial matching on G, called $path(u)$. Let $path(T) = \{path(u) : u \text{ is a leaf of } T\}$. If v is a node of G that appears as a label of some node in T, then T is said to *mention* v.

Furthermore, each leaf of T is labelled by 0 or 1 (if a tree satisfies the above conditions but its leaves remain unlabelled, we will call it a *leaf-unlabelled* matching decision tree). Let T^c be the same as T except with the value of each leaf-label flipped. If U is the set of leaves of T labelled 1, let $disj(T)$ be the DNF formula $\bigvee_{u\in U} \bigwedge_{e\in path(u)} X_e$.

DEFINITION 4.4. A *complete* (leaf-unlabelled) matching decision tree for G is one in which, for each internal node u labelled v, the set $\{path(u') : u' \text{ a child of } u\}$ constitutes all matchings in G of the form $path(u) \cup \{\{v,v'\}\}$ for all v' such that $\{v,v'\} \in E$.

DEFINITION 4.5. Let K be a subset of the nodes in G. The *full matching tree* for K over G is a leaf-unlabelled matching decision tree for G defined inductively: if $K = \{k\}$, then the root of the tree is labelled by k and, for each edge e in G that touches k, there is an edge from the root of the tree labelled e. If K contains more than one node, let k be its largest node under the ordering of nodes and assume we have a full matching tree for $K \setminus \{k\}$. For each (unlabelled) leaf u of this tree, let p be the path from the root to u. The labels of the edges along p constitute a partial matching on G. If this partial matching touches k, leave u unlabelled. Otherwise, label u by k and and attach an edge to u for each edge in G that touches k and that extends the partial matching.

Note that the full matching tree for any subset K is complete. If the degree of each node in K is at least $|K|$, then the full matching tree for K is guaranteed to mention all nodes in K. Otherwise, it might not.

Lemma 1. *Let T be a complete matching tree for G and let ρ be any partial matching on G. Let d be the minimal degree of any node in G mentioned by T. If $d > \max\{|\rho|, height(T)\}$, then there is a matching in $path(T)$ that is consistent with ρ.*

Proof. Assume we have found an internal node u in T labelled by v in G such that $path(u)$ is consistent with ρ. We will find a child u' of u such that $path(u')$ is still consistent with ρ. Since the degree of v is greater than the size of ρ, there is an edge $\{v, v'\}$ in G such that $\{v, v'\}$ is either included in ρ (if ρ touches v) or extends ρ (if ρ does not touch v). Since T is complete and the degree of v is greater than $height(T)$, the edge $\{v, v'\}$ appears as a label of an edge from u in T. □

DEFINITION 4.6. We call F a *matching disjunction* if it is one of the constants 0 or 1, or it is a DNF formula with no negations over the variables D such that the edges of G corresponding to the variables in any one term constitute a partial matching. In the latter case, order the terms lexicographically based on the nodes they touch and the order of the nodes in G.

DEFINITION 4.7. For F a matching disjunction, the *restriction* $F|_\rho$ for ρ a partial matching is another matching disjunction generated from F as follows: set any variable in F corresponding to an edge of ρ to 1 and set any variable corresponding to an edge not in ρ but incident to one of ρ's nodes to 0. If a variable in term t is set to 0, remove t from F. Otherwise, if a variable in term t is set to 1, remove that variable from t.

The DNF $disj(T)$ for a matching decision tree T is always a matching disjunction.

DEFINITION 4.8. A matching decision tree T is said to *represent* a matching disjunction F if, for every leaf l of T, $F|_{path(l)} \equiv 1$ when l is labelled 1 and $F|_{path(l)} \equiv 0$ when l is labelled 0.

A matching decision tree T always represents $disj(T)$. Furthermore, if ρ extends some matching $path(l)$ for l a leaf of T, then $disj(T)|_\rho \equiv 0$ (1, respectively) if l is labelled 0 (1).

DEFINITION 4.9. Let F be a matching disjunction. We define a tree $Tree_G(F)$ called the *canonical decision tree* for F over G: if F is constant, then $Tree_G(F)$ is one node labelled by that constant. Otherwise, let C be the first term of F. Let K be the nodes of G touched by variables in C. The top of $Tree_G(F)$ is the full matching tree on K over G. We replace each leaf u of that tree, with the tree $Tree_{G|_{path(u)}}(F|_{path(u)})$.

The tree $Tree_G(F)$ will have all of its leaves labelled. It is designed to represent F and to be complete.

DEFINITION 4.10. For T a matching decision tree and ρ a matching, T *restricted by* ρ, written $T|_\rho$, is a matching decision tree obtained from T by first removing all edges of T that are inconsistent with ρ, and retaining only those nodes of T that remain connected to the root of T. Each remaining edge that corresponds to an element of ρ is then contracted (its endpoints are identified and labelled by the label of the lower endpoint).

Lemma 2. *([21], Lemma 4.8) For T a matching decision tree and ρ a matching: (a) $disj(T)|_\rho \equiv disj(T|_\rho)$, (b) $(T|_\rho)^c = T^c|_\rho$, and (c) If T represents a matching disjunction F, then $T|_\rho$ represents $F|_\rho$.*

5 The lower bound

Let $m = n + n/\log^c n$ for some integer $c > 0$ and let $h > 0$ be an integer. We assume for simplicity that n is large compared to c and that all subsequent expressions are integers. We will show that for any a such that $8^h(a+3) < c$, any proof of $PHP^m_n = WPHP(K_{m,n})$ of depth h is of size greater than $2^{\log^a n}$. To do this we do not work directly with proofs of $WPHP(K_{m,n})$ but rather we work with proofs of $WPHP(G)$ for randomly chosen subgraphs G of $K_{m,n}$.

More precisely, let $b = 8^h(a+3)$, define $d = \log^b n$ and observe that $a < b < c$.

Let $\mathcal{G}(m,n,d/n)$ be the uniform distribution on all bipartite graphs from m nodes to n nodes where each edge is present independently with probability d/n.

Let $H = (V_1 \cup V_2, E)$ be a fixed bipartite graph. Define $M^\ell(H)$ to be the set of all partial matchings of size ℓ in H and for $I \subseteq V_2$ with $|I| = \ell$ let $M^\ell_I(H)$ be the set of all $\rho \in M^\ell(H)$ with $\mathrm{Im}(\rho) = I$. Define a partial distribution $\mathcal{M}^\ell(H)$ on $M^\ell(H)$ by first choosing a set $I \in V_2$ uniformly at random among all subsets of V_2 of size ℓ, then choosing a $\rho \in M^\ell_I(H)$ uniformly at random; if $M^\ell_I(H)$ is empty then no matching is chosen and the experiment fails.

We now define several sequences of parameters for a probabilistic experiment. The meanings of these parameters will be explained after the definition of the experiment. For initial values, let $m_0 = m$, $n_0 = n$, $b_0 = b$, and $k_0 = 7b_0/8$, $\ell_0 = n_0 - n_0/\log^{k_0} n$. Then, for $1 \leq i \leq h$, we define recursively:
$m_i = m_{i-1} - \ell_{i-1}$, $n_i = n_{i-1} - \ell_{i-1}$, $b_i = b_{i-1} - k_{i-1}$, $k_i = 7b_i/8$, and $\ell_i = n_i - n_i/\log^{k_i} n$. In closed form,
$n_i = n/(\log n)^{\sum_{j=0}^{i-1} k_j} = n/(\log n)^{b-b/8^i}$,

Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)
0272-5428/02 $17.00 © 2002 IEEE

$m_i = n_i + (m-n)$, $b_i = b - \sum_{j=0}^{i-1} k_j = b/8^i$, $k_i = 7b/8^{i+1}$, and $\ell_i = (1 - 1/\log^{k_i} n)(n/(\log n)^{b-b/8^i})$.

Now we are ready to define the experiment: let $G_0 = G$ be a graph chosen randomly from the distribution $\mathcal{G}(m,n,d/n)$. For $0 \le i \le h-1$, let $\rho_i \sim \mathcal{M}^{\ell_i}(G_i)$ and define $G_{i+1} = G_i|_{\rho_i}$. (We say that the experiment fails during stage $i+1$ if the partial distribution $\mathcal{M}^{\ell_i}(G_i)$ fails to return an element ρ_i.) Observing that the choice of ρ_i depends only on the edges of G_i that are incident to $\mathrm{Im}(\rho_i)$ and these are among the edges of G_i that are removed to produce G_{i+1} we have:

Proposition 3. *If this experiment succeeds up to stage i then the distribution induced on G_i is $\mathcal{G}(m_i,n_i,d/n)$.*

Thus, the expected degree of any pigeon in G_i is $n_i d/n = \log^{b_i} n$. The expected degree of any hole in G_i is $m_i d/n$, which is between $\log^{b_i} n$ and $2\log^{b_i} n$ since $n_i < m_i < 2n_i$ (because $c > b$).

We make several observations about "bad" events in this experiment; the first two follow from simple Chernoff bounds.

Lemma 4. *For $0 \le i \le h$, the probability, given that the experiment succeeds up to stage i, that any node in G_i has degree greater than $\Delta_i \stackrel{def}{=} 6\log^{b_i} n$ is at most $(m_i + n_i)2^{-\log^{b_i} n} < 2^{-\log^{b_i - 1} n}$.*

Lemma 5. *For $0 \le i \le h$ and sufficiently large n, the probability, given that the experiment succeeds up to stage i, that any node in G_i has degree less than $\frac{1}{2}\log^{b_i} n$ is at most $(m_i + n_i)2^{-\frac{1}{16}\log^{b_i} n} < 2^{-\log^{b_i - 1} n}$.*

Lemma 6. *For $0 \le i \le h-1$, the probability that the experiment fails at stage $i+1$, given that it has succeeded up to stage i is at most $2^{-\log^{b_i - 2} n}$.*

Proof Sketch. This is less than the probability that a random graph $\mathcal{G}(m_i,n_i,d/n)$ does not contain a matching of all holes. The bound follows by standard calculations using $n_i d/n = \log^{b_i} n$. □

We now develop the switching lemma argument. The overall structure uses the simplified counting techniques of [18] and [3], however the statement and proof are both complicated by the need to use probabilistic properties of the formulas themselves as well as the relationship of those properties to the restrictions under consideration. We first need some definitions:

DEFINITION 5.1. For a bipartite graph $H = (V_1 \cup V_2, E)$ and integers ℓ and Δ, let $N^{\ell,\Delta}(H)$ be the set of all ρ in $M^\ell(H)$ such that all nodes of $H|_\rho$ have degree at most Δ. For a set $I \subseteq V_2$ with $|I| = \ell$ let $N_I^{\ell,\Delta}(H)$ be the set of elements $\rho \in N^{\ell,\Delta}(H)$ with $\mathrm{Im}(\rho) = I$.

Lemma 7. *Let $0 \le i < h$ and suppose that the experiment succeeds up to stage $i+1$. Then the probability that $\frac{|N^{\ell_i,\Delta_{i+1}}_{\mathrm{Im}(\rho_i)}(G_i)|}{|M^{\ell_i}_{\mathrm{Im}(\rho_i)}(G_i)|} < 1 - 2^{-\log^{b_{i+1}-2} n}$ is at most $1/n$.*

Proof. Observe that the expectation of $\frac{|N^{\ell_i,\Delta_{i+1}}_{\mathrm{Im}(\rho_i)}(G_i)|}{|M^{\ell_i}_{\mathrm{Im}(\rho_i)}(G_i)|}$ conditional on success up to stage $i+1$ is precisely the probability that $\rho_i \in N^{\ell_i,\Delta_{i+1}}(G_i)$ conditional on success up to stage $i+1$ which is $> 1 - 2^{-\log^{b_{i+1}-1} n}$ by Lemma 4. The result follows Markov's inequality and by observing that $n \cdot 2^{-\log^{b_{i+1}-1} n} < 2^{-\log^{b_{i+1}-2} n}$. □

We are now ready to state the switching lemma.

Lemma 8 (Switching Lemma). *Let i,s,r be any integers such that $0 \le i < h$, $0 < s \le \Delta_{i+1}/\log^3 n$ and $r > 0$. Suppose that the experiment above succeeds up to stage $i+1$, consider G and $\rho_0,\ldots,\rho_i$ resulting from this experiment, and suppose that G_i has maximum degree at most Δ_i. Finally, let F be any matching disjunction with conjunctions of size $\le r$ over the edge-variables of G_i. The probability that $Tree_{G_{i+1}}(F|_{\rho_i})$ has height $\ge s$ conditioned on the events $\rho_i \in N^{\ell_i,\Delta_{i+1}}(G_i)$ and $\frac{|N^{\ell_i,\Delta_{i+1}}_{\mathrm{Im}(\rho_i)}(G_i)|}{|M^{\ell_i}_{\mathrm{Im}(\rho_i)}(G_i)|} \ge 1 - 2^{-\log^{b_{i+1}-2} n}$ is at most $2\left(720r/\log^{b_i/2} n\right)^{s/2}$.*

DEFINITION 5.2. Let $stars(r,j)$ be the set of all sequences $\beta = (\beta_1,\ldots,\beta_k)$ such that for each i, $\beta_i \in \{*,-\}^r \setminus \{-\}^r$ and the total number of $*$'s in β is j.

Lemma 9 ([3]). $|stars(r,j)| < (r/\ln 2)^j$.

Lemma 10. *For H a fixed bipartite graph with an ordering on its nodes, let F be a matching disjunction with conjunctions of size $\le r$ over the edge-variables of H and let S be the set of matchings $\rho \in N^{\ell,\Delta}(H)$ such that $Tree_{H|_\rho}(F|_\rho)$ has height $\ge s$. There is an injection from the set S to the set $\bigcup_{s/2 \le j \le s} M^{\ell+j}(H) \times stars(r,j) \times [\Delta]^s$. Furthermore, the first component of the image of $\rho \in S$ is an extension of ρ.*

Proof. Let $F = C_1 \vee C_2 \vee \ldots$. If $\rho \in S$, then let π be the partial matching labelling the first path in $Tree_{H|_\rho}(F|_\rho)$ of length $\ge s$ (actually, we consider only the first s edges in π, starting from the root, and hence we assume $|\pi| = s$). Let C_{ν_1} be the first term in F not set to 0 by ρ and let K_1 be the variables of C_{ν_1} not set by ρ. Let σ_1 be the unique partial matching over K_1 that satisfies $C_{\nu_1}|_\rho$ and let π_1 be the portion of π that touches K_1.

Now define $\beta_1 \in \{*,-\}^{|K_1|} \setminus \{-\}^{|K_1|}$, so that the p-th component of β_1 is a $*$ if and only if the p-th variable in C_{ν_1} is set by σ_1.

Continue this process to define π_i, σ_i, K_i, etc. (replacing ρ with $\rho\pi_1\dots\pi_{i-1}$ and π with $\pi\setminus\pi_1\dots\pi_{i-1}$ until some stage k when we've exhausted all of π. Let σ be the matching $\sigma_1\dots\sigma_k$, and β be the vector $(\beta_1,\dots,\beta_k)$. Let $j=|\sigma|$ be the number of edges in σ. Note that $s/2\le j\le s$. Observe that $\beta\in stars(r,s)$ and $\rho\sigma\in M^{\ell+j,\Delta}(H)$ and is an extension of ρ.

We now encode the differences between all the corresponding π_i and σ_i pairs in a single vector δ consisting of $|\pi|=s$ components, each in $\{1,\dots,\Delta\}$. Let u_1 be the smallest numbered node in K_1 and suppose that π (in particular π_1) matches u_1 with some node v_1. Then the first component of δ is the natural number x such that v_1 is the x-th neighbor (under the ordering of nodes) of u_1 in the graph $H|_{\rho\sigma_2\sigma_3\dots\sigma_k}$. More generally, until the mates of all nodes in K_1 under π_1 have been determined, we determine the p-th component of δ by finding the smallest numbered node u_p of $K_1\setminus\{u_1,\dots,u_{p-1},v_1,\dots,v_{p-1}\}$ and then we find its mate v_p under π_1 and encode the position x of v_p in the order of the neighbors of u_p in $H|_{\rho\sigma_2\sigma_3\dots\sigma_k}$. Once K_1 (and thus π_1) has been exhausted the next component is based on the mates of the smallest numbered nodes in K_2 under π_2, until that is exhausted, etc. where the ordering about each vertex when dealing with K_i is with respect to the graph $H|_{\rho\sigma_{i+1}\sigma_{i+2}\dots\sigma_k}$.

Finally, we define the image of $\rho\in S$ under the injection to be $(\rho\sigma,\beta,\delta)$. To prove that this is indeed an injection, we show how to invert it: Given $\rho\sigma_1\dots\sigma_k$, we can identify ν_1 as the index of the first term of F that is not set to 0 by it. Then, using β_1 we can reconstruct σ_1 and K_1. Next, reading the components of δ and the graph $H|_{\rho\sigma_2\dots\sigma_k}$, until all of K_1 is matched, we can reconstruct π_1. Then we can derive $\rho\pi_1\sigma_2\dots\sigma_k$.

At a general stage i of the inversion, we will know $\pi_1,\dots,\pi_{i-1}$ and $\sigma_1,\dots,\sigma_{i-1}$ and $K_1,\dots,K_{i-1}$. We use $\rho\pi_1\dots\pi_{i-1}\sigma_i\dots\sigma_k$ to identify ν_i and, hence, σ_i and K_i (using β). Then we get π_i from δ, K_i, and $\rho\sigma_{i+1}\dots\sigma_k$. After k stages, we know all of σ and can recover ρ. $\square$

Proof of Lemma 8. Let R_i be the set of $\rho_i\in N^{\ell_i,\Delta_{i+1}}(G_i)$ such that $\frac{|N^{\ell_i,\Delta_{i+1}}_{\mathrm{Im}(\rho_i)}(G_i)|}{|M^{\ell_i}_{\mathrm{Im}(\rho_i)}(G_i)|}\ge 1-2^{-\log^{b_{i+1}-2}n}$. By Lemma 7, the total probability of R_i under distribution $\mathcal{M}^{\ell_i}(G_i)$ is at least $(1-1/n)(1-2^{-\log^{b_{i+1}-2}n})\ge 1-2/n$.

By Lemma 10 with $H\leftarrow G_i$, $\ell\leftarrow\ell_i$, and $\Delta\leftarrow\Delta_{i+1}$, a *bad* $\rho_i\in R_i$, for which $Tree_{G_{i+1}}(F|_{\rho_i})$ has height at least s, can be mapped uniquely to a triple $(\rho',\beta,\delta)\in M^{\ell_i+j}(G_i)\times stars(r,j)\times[\Delta_{i+1}]^s$ where ρ' extends ρ_i, for some integer $j\in[s/2,s]$. We compute the probability of such $\rho_i\in R_i$ associated with a given j and then sum up the probabilities and divide by the probability of R_i to compute the desired probability.

We analyze the total probability of bad $\rho_i\in R_i$ associated with a given j by comparing the probability of ρ_i under $\mathcal{M}^{\ell_i}(G_i)$ and the probability of ρ' under $\mathcal{M}^{\ell_i+j}(G_i)$. Since the total probability of all $\rho'\in M^{\ell_i+j}(G_i)$ under $\mathcal{M}^{\ell_i+j}(G_i)$ is at most 1 this will allow us to compute the desired bound.

Let $I=\mathrm{Im}(\rho_i)$ and $I'=\mathrm{Im}(\rho')$. By definition, $I\subset I'$. Also, by definition, the ratio of the probability of ρ_i under $\mathcal{M}^{\ell_i}(G_i)$ to that of ρ' under $\mathcal{M}^{\ell_i+j}(G_i)$ is precisely

$$\frac{\binom{n_i}{\ell_i+j}|M^{\ell_i+j}_{I'}(G_i)|}{\binom{n_i}{\ell_i}|M^{\ell_i}_{I}(G_i)|}.$$

Now any matching $\tau'\in M^{\ell_i+j}_{I'}(G_i)$ is an extension of some unique matching $\tau\in M^{\ell_i}_I(G_i)$. If $\tau\in N^{\ell_i,\Delta_{i+1}}_I(G_i)$ then the degrees of all nodes in $G_i|_\tau$ are at most Δ_{i+1} and so at there are at most Δ^j_{i+1} matchings $\tau'\in M^{\ell_i+j}_{I'}(G_i)$ extending τ. If $\tau\notin N^{\ell_i,\Delta_{i+1}}_I(G_i)$ then the degrees of all nodes in $G_i|_\tau$ are at most Δ_i because that is true of G_i itself by assumption. Therefore there are at most Δ_i^j extensions $\tau'\in M^{\ell_i+j}_{I'}(G_i)$ of τ. Since $\rho_i\in R_i$, $|N^{\ell_i,\Delta_{i+1}}_I(G_i)|/|M^{\ell_i}_I(G_i)|$ is at least $1-2^{-\log^{b_{i+1}-2}n}$ so the probability ratio is at most

$$\begin{aligned}
&\frac{\binom{n_i}{\ell_i+j}}{\binom{n_i}{\ell_i}}[(1-2^{-\log^{b_{i+1}-2}n})\Delta^j_{i+1}+2^{-\log^{b_{i+1}-2}n}\Delta^j_i]\\
\le\quad&\left[1+2^{1-\log^{b_{i+1}-2}n}\left(\frac{\Delta_i}{\Delta_{i+1}}\right)^j\right]\left(\frac{\Delta_{i+1}(n_i-\ell_i)}{\ell_i}\right)^j\\
<\quad&\left[1+2^{1-\Delta_{i+1}/(6\log^2 n)}(\log n)^{k_i s}\right]\left(\frac{\Delta_{i+1}n_i}{\ell_i\log^{k_i}n}\right)^j\\
<\quad&\left[1+2^{1-\Delta_{i+1}/(6\log^2 n)}(\log n)^{k_i\Delta_{i+1}/\log^3 n}\right]\left(\frac{\Delta_{i+1}n_i}{\ell_i\log^{k_i}n}\right)^j\\
<\quad&\left(\frac{2\Delta_{i+1}}{\log^{k_i}n}\right)^j=\left(\frac{12\log^{b_{i+1}}n}{\log^{k_i}n}\right)^j.
\end{aligned}$$

The first two inequalities follow from $j\le s\le\Delta_{i+1}/\log^3 n$ and the definitions of Δ_i and Δ_{i+1}. The third inequality follows since $12k_i\log\log n<\log n$ for n sufficiently large and the fact that $n_i/\ell_i=1/(1-1/\log^{k_i}n)$ which is close to 1. Therefore the total probability of bad $\rho_i\in R_i$ associated with a given j is at most

$$\begin{aligned}
&(12\log^{b_{i+1}-k_i}n)^j\times(r/\ln 2)^j\times\Delta^s_{i+1}\\
&\quad\le\quad(20r\log^{b_{i+1}-k_i}n)^j\times(6\log^{b_{i+1}}n)^s.
\end{aligned}$$

Thus the total probability in question is at most

$$(1-2/n)^{-1}(6\log^{b_{i+1}}n)^s\times\sum_{s/2\le j\le s}(20r\log^{b_{i+1}-k_i}n)^j.$$

Since $b_{i+1}=b_i-k_i$ and without loss of generality $20r\log^{b_i-2k_i}n<1/3$ (otherwise the probability bound in the lemma statement is meaningless), this quantity is at most $2(720r\log^{3b_i-4k_i}n)^{s/2}\le 2(720r/\log^{b_i/2}n)^{s/2}$ since $3b_i-4k_i=-b_i/2$. $\square$

DEFINITION 5.3. Let S_G be a set of formulas of depth at most h that is closed under subformulas and defined over the graph G. For $\rho = \rho_0 \dots \rho_{h-1}$, we define, for every $0 \leq i < h$, $\mathcal{T}_{\rho_0 \dots \rho_i}$, a mapping from formulas with depth $\leq i+1$ in S_G to matching decision trees. It is defined inductively as follows:

For a variable X_e, $\mathcal{T}_{\rho_0}(X_e)$ is $Tree_G(X_e)|_{\rho_0}$. Furthermore, $\mathcal{T}_{\rho_0}(\neg X_e)$ is $(\mathcal{T}_{\rho_0}(X_e))^c$. For A a depth-1 formula with merged form $\bigvee_{e \in I} X_e$, $\mathcal{T}_{\rho_0}(A)$ is $Tree_{G_1}((\bigvee_{e \in I} X_e)|_{\rho_0})$.

For $0 < i < h$, for all formulas A of depth $< i+1$, $\mathcal{T}_{\rho_0 \dots \rho_i}(A)$ is $\mathcal{T}_{\rho_0 \dots \rho_{i-1}}(A)|_{\rho_i}$. For a formula A of depth $i+1$, if $A = \neg B$, then $\mathcal{T}_{\rho_0 \dots \rho_i}(A)$ is $(\mathcal{T}_{\rho_0 \dots \rho_i}(B))^c$, and otherwise, if the merged form of A is $\bigvee_{j \in J} B_j$, let F be the matching disjunction $\bigvee_{j \in J} disj(\mathcal{T}_{\rho_0 \dots \rho_{i-1}}(B_j))$ and let $\mathcal{T}_{\rho_0 \dots \rho_i}(A)$ be the canonical matching tree $Tree_{G_{i+1}}(F|_{\rho_i})$.

From the definition of $\mathcal{T}_\rho$, we have that if $\neg A$ is a formula in S_G, then $\mathcal{T}_\rho(\neg A) = (\mathcal{T}_\rho(A))^c$. Also, by lemma 2, if $\bigvee_{i \in I} A_i$ is the merged form of some formula A in S_G, then $\mathcal{T}_\rho(A)$ represents $\bigvee_{i \in I} disj(\mathcal{T}_\rho(A_i))$.

Lemma 11. *Let a and h be positive integers. For each graph G, let S_G be a set of formulas closed under subformulas defined on the variables of G such that $|S_G| \leq 2^{\log^a n}$ and each formula $A \in S_G$ has depth at most h. For n sufficiently large in a and h, there exists a choice of G and $\rho = \rho_0, \dots, \rho_{h-1}$ as defined above such that the following conditions hold:*
1. $\mathcal{T}_\rho(A)$ has height at most $\log^a n$ for all $A \in S_G$, and
2. every node in G_h has degree at least $\log^{a+1} n$.

Proof. We proceed using the probabilistic method and the experiment above. For $0 \leq i \leq h$, define the following events:
E_i: The experiment succeeds up to stage i.
A_i: Every node in G_i has degree at most $\Delta_i = 6 \log^{b_i} n$.
B_i: Every node in G_i has degree at least $(1/2) \log^{b_i} n$.
C_i: $\frac{|N^{\ell_i, \Delta_{i+1}}_{\mathrm{Im}(\rho_i)}(G_i)|}{|M^{\ell_i}_{\mathrm{Im}(\rho_i)(G_i)}|} \geq 1 - 2^{-\log^{b_{i+1}-2} n}$. Here $i < h$.
$D_i(A)$: $\mathcal{T}_{\rho_0 \dots \rho_{i-1}}(A)$ has height at most $\log^a n$ for some formula $A \in S_G$ of depth at most i. Here $i \geq 1$.
D_i: for all formulas $A \in S_G$ of depth at most i, $D_i(A)$ holds. Here $i \geq 1$.

We compute an upper bound on the probability that any of these events fails to be true and prove that this probability is strictly less than 1. Since $b_h = a+3$, if both E_h and B_h occur and $D_i(A)$ occurs for each $i = 1, \dots, h$ and each $A \in S_G$ of depth i then the claims of the lemma are satisfied for that (G, ρ), so this probability bound suffices.

Now by Lemma 6, $\mathbf{Pr}[\neg E_{i+1} \mid E_i] < 2^{-\log^{b_i - 2} n}$, Lemma 4, $\mathbf{Pr}[\neg A_i \mid E_i] < 2^{-\log^{b_i - 1} n}$ and by Lemma 5, $\mathbf{Pr}[\neg B_i \mid E_i] < 2^{-\log^{b_i - 1} n}$. Furthermore, by Lemma 7, $\mathbf{Pr}[\neg C_i \mid E_{i+1}] \leq 1/n$. Let $A \in S_G$ be of depth $i < h$ with the merged form of A equal to $\bigvee_{j \in J} Q_j$ and let F be the matching disjunction $\bigvee_{j \in J} disj(\mathcal{T}_{\rho_0 \dots \rho_{i-1}}(Q_j))$. Observing that $b_h = b/8^h = (a+3)$, by Lemma 8 applied to F with $r = s = \log^a n \leq \Delta_h / \log^3 n$, we have

$$\begin{aligned} &\mathbf{Pr}[\neg D_{i+1}(A) \mid E_{i+1} \wedge A_i \wedge D_i \wedge A_{i+1} \wedge C_i] \\ &\quad \leq \quad 2(720/\log^{b_i/2 - a} n)^{(\log^a n)/2} \\ &\quad \leq \quad 2(720/\log^{b_{h-1}/2 - a} n)^{(\log^a n)/2} \\ &\quad \leq \quad 2(720/\log^{3a+3} n)^{(\log^a n)/2} < 2^{-\log^a n}/n \end{aligned}$$

for n sufficiently large. Therefore, $\mathbf{Pr}[\neg D_{i+1} \mid E_{i+1} \wedge A_i \wedge D_i \wedge A_{i+1} \wedge C_i] \leq 1/n$ since each S_G contains at most $2^{\log^a n}$ disjunctions of depth $i+1$.

Therefore the total probability that some E_i, A_i, B_i, C_i, or D_i fails is at most:

$$\begin{aligned} &\sum_{i=0}^{h-1} \mathbf{Pr}[\neg E_{i+1} \mid E_i] + \sum_{i=0}^{h} \mathbf{Pr}[\neg A_i \mid E_i] \\ &\quad + \sum_{i=0}^{h} \mathbf{Pr}[\neg B_i \mid E_i] + \sum_{i=0}^{h-1} \mathbf{Pr}[\neg C_i \mid E_{i+1}] \\ &\quad + \mathbf{Pr}[\neg D_1 \mid E_1 \wedge A_0 \wedge A_1 \wedge C_0] \\ &\quad + \mathbf{Pr}[\neg D_2 \mid E_2 \wedge A_1 \wedge D_1 \wedge A_2 \wedge C_1] + \cdots \\ &\quad + \mathbf{Pr}[\neg D_h \mid E_h \wedge A_{h-1} \wedge D_{h-1} \wedge A_h \wedge C_{h-1}]. \end{aligned}$$

In total there are $5h+2$ terms in this sum, each of which is at most $1/n$, and thus the whole probability is < 1. $\square$

From now on, we fix a graph G and a restriction $\rho = \rho_0, \dots, \rho_{h-1}$ obeying the conditions of Lemma 11 when applied to the sets of formulas in $S_G = cl(\Pi_G)$ where each Π_G is a proof of $WPHP(G)$ in a proof system $\mathcal{F}$ whose largest rule has size f.

The following three lemmas are adapted from [21].

Lemma 12. *Let C be a line in a Frege proof Π. Let $\mathcal{A}$ be the immediate ancestors of C in the proof (if there are any), so that $\mathcal{A} \vdash C$. Let $\mathcal{B}$ be the subformulas of $\mathcal{A}$ and C mentioned in the application of the rule which derives C from $\mathcal{A}$. Let $\Gamma = \mathcal{A} \cup \mathcal{B} \cup \{C\}$. Note that by our bound on the size of rules in $\mathcal{F}$, $|\Gamma| + 1 \leq f$. Finally, let σ be a matching which extends soundly some $\sigma_A \in path(\mathcal{T}_\rho(A))$ for each $A \in \mathcal{A}$, some $\sigma_B \in path(\mathcal{T}_\rho(B))$ for each $B \in \mathcal{B}$, and some $\sigma_C \in path(\mathcal{T}_\rho(C))$. If $disj(\mathcal{T}_\rho(A))|_\sigma \equiv 1$ for all $A \in \mathcal{A}$, then $disj(\mathcal{T}_\rho(C))|_\sigma \equiv 1$.*

Proof. First note the following facts, where $\alpha, \beta \in \Gamma$ and $D(\alpha)$ is an abbreviation for $disj(\mathcal{T}_\rho(\alpha))$:

- $D(\alpha)|_\sigma \equiv 0$ or $D(\alpha)|_\sigma \equiv 1$
- If $\neg\alpha \in \Gamma$, then $D(\neg\alpha)|_\sigma \equiv 1$ iff $D(\alpha)|_\sigma \equiv 0$.
- If $(\alpha \vee \beta) \in \Gamma$, then $D(\alpha \vee \beta)|_\sigma \equiv 1$ iff $D(\alpha)|_\sigma \equiv 1$ or $D(\beta)|_\sigma \equiv 1$

Now consider the rule R used to derive C formulated as in the examples from section 3. The application of R substitutes subformulas $A_p, A_q, A_r, \ldots$ in Γ for each of the atoms $p, q, r, \ldots$ in R and there is a derived correspondence mapping subformulas F appearing in R to formulas $A_F \in \Gamma$. Define a function τ on the atoms of R by $\tau(p) = D(A_p)|_\sigma$ for each such atom p. By the first property, τ is a truth assignment to these atoms. Furthermore, by the other two properties, the truth assignment τ extends to all subformulas F in R so that $\tau(F) = D(A_F)|_\sigma$. Since R is sound, if τ satisfies all formulas in $\mathcal{A}$ it will satisfy C and thus $D(C)|_\sigma \equiv 1$. □

Lemma 13. *Let $\mathcal{F}$ be a Frege system with maximum rule size f. Let n be sufficiently large w.r.t. f. Let $a, h > 0$. For each G, assume that Π_G is a proof in $\mathcal{F}$ of $WPHP(G)$ of size at most $2^{\log^a n}$ and depth at most h. Let ρ and G be as defined in Lemma 11 applied with $S_G = cl(\Pi_G)$. If C is an arbitrary line in proof Π_G then all leaves of $\mathcal{T}_\rho(C)$ are labelled by 1.*

Proof. We proceed by (complete) induction on the lines in the proof. Assume every leaf of $\mathcal{T}_\rho$ for any line preceding C is labelled 1. Let $\mathcal{A}$, $\mathcal{B}$, Γ be as in Lemma 12. For any leaf l of $\mathcal{T}_\rho(C)$, we use Lemma 1 to find σ that extends $path(l)$ and extends a matching in each of the sets $path(\mathcal{T}_\rho(A))$ for all $A \in \mathcal{A}$ and $path(\mathcal{T}_\rho(B))$ for all $B \in \mathcal{B}$. This is possible since there are at most f trees to consider and by Lemma 11 the sum of their heights is at most $f \log^a n \leq \log^{a+1} n$ which is the degree of G_h.

By assumption, $disj(\mathcal{T}_\rho(A))|_\sigma \equiv 1$ for all A in $\mathcal{A}$. Hence, by Lemma 12, $disj(\mathcal{T}_\rho(C))|_\sigma \equiv 1$, so l must be labelled 1. □

Lemma 14. *All leaves of $\mathcal{T}_\rho(WPHP(G))$ have label 0.*

Proof. It suffices to show that $\mathcal{T}_\rho$ applied to each of the following types of formulas has all leaves labelled 0:

1. $\neg(\neg X_e \vee \neg X_{e'})$ for $e, e' \in E; e = \{i,k\}, e' = \{j,k\}; i, j \in V_1; i \neq j; k \in V_2$.
2. $\neg(\neg X_e \vee \neg X_{e'})$ for $e, e' \in E; e = \{k,i\}, e' = \{k,j\}; i, j \in V_2; i \neq j; k \in V_1$.
3. $\neg \bigvee_{j \in \Gamma(i)} X_{\{i,j\}}$ for $i \in V_1$.
4. $\neg \bigvee_{i \in \Gamma(j)} X_{\{i,j\}}$ for $j \in V_2$.

In fact, we will show that $\mathcal{T}_\rho$ applied to the complement of each of these formulas has all leaves labelled 1.

For a formula of the first type, $T = \mathcal{T}_\rho(\neg X_e \vee \neg X_{e'})$ must represent $disj(\mathcal{T}_\rho(\neg X_e)) \vee disj(\mathcal{T}_\rho(\neg X_{e'}))$. If ρ sets the value of either X_e or $X_{e'}$ then it must set one of $\neg X_e$ or $\neg X_{e'}$ to 1 and thus all leaves of $\mathcal{T}_\rho(\neg X_e \vee \neg X_{e'})$ are certainly labelled 1. Otherwise, for l a leaf of T, $path(l)$ cannot contain both e and e'. Without loss of generality it does not contain e. By Lemma 1 applied to graph G_h we can find σ that extends $path(l)$ and is an extension of some matching in $\mathcal{T}_\rho(\neg X_e)$. But then $disj(\mathcal{T}_\rho(\neg X_e))|_\sigma \equiv 1$, so l must be labelled 1. The argument is the same for formulas of the second type.

For a formula of the third type, $T = \mathcal{T}_\rho(\bigvee_{j \in \Gamma(i)} X_{\{i,j\}})$ must represent $\bigvee_{j \in \Gamma(i)} disj(\mathcal{T}_\rho(X_{\{i,j\}}))$. Hence, if ρ sets $X_{\{i,j\}}$ to 1 for some $j \in \Gamma(i)$, then all leaves of T are certainly labelled 1. Otherwise, for a leaf l of T, if $path(l)$ touches node i, then $\bigvee_{j \in \Gamma(i)} disj(\mathcal{T}_\rho(X_{\{i,j\}}))|_{path(l)} \equiv 1$. Finally, if $path(l)$ does not touch node i, extend it to $\sigma = path(l) \cup \{i,j\}$ for some j such that $X_{\{i,j\}}$ is not set by ρ. Then $disj(\mathcal{T}_\rho(X_{\{i,j\}}))|_\sigma \equiv 1$, so l is labelled 1. Formulas of the fourth type follow in the same way. □

Theorem 15. *Given any c sufficiently large, there exists a bipartite graph G from $m = n + n/\log^c n$ pigeons to n holes such that there is no depth-h, $2^{\log^a n}$-size $\mathcal{F}$-proof of $WPHP(G)$ provided that $8^h(a+3) < c$.*

Proof. Assume that for all such G, there is a proof Π_G of the required depth and size. For the G in Lemma 11 and its corresponding proof Π_G of $WPHP(G)$, there exists a ρ such that $\mathcal{T}_\rho(A)$ has all leaves labelled 1 for any $A \in cl(\Pi_G)$, but $\mathcal{T}_\rho(WPHP(G))$ has all leaves labelled 0. If Π_G is to be a proof of $WPHP(G)$, then $WPHP(G)$ must appear in Π_G, so we have a contradiction. □

Corollary 16. *Given any c sufficiently large, there is no depth-h, $2^{\log^a n}$-size $\mathcal{F}$-proof of $WPHP = WPHP(K_{m,n})$ from $m = n + n/\log^c n$ pigeons to n holes, provided that $8^h(a+3) < c$.*

6 Open questions

Among the many unresolved proof complexity questions regarding the pigeonhole principle (see [19]) the most important open problem is to resolve the complexity of the weak pigeonhole principle with $2n$ or more pigeons, and n holes. This would have many implications for: the metamathematics of the P versus NP statement, the complexity of approximate counting, and the proof-theoretic strength underlying elementary number theory.

In the proof presented here, we derived a switching-lemma using simple restrictions that limit the space of truth assignments to a subcube where certain variables are set to 0 or to 1. While this fails with $2n$ pigeons, a more general class of restrictions may suffice. Possible generalizations include the projections suggested in [22], which also allow identification of variables, or restrictions given by linear equations. Two important results ([12] and [6]) for bounded-depth Frege systems already employ such generalized switching lemmas in cases where direct restrictions fail (although the latter use is implicit). Bounded-depth Frege reductions, such as those in [6] may also be useful for resolving the $2n$

to n case; conversely, via reductions, generalizing our weak pigeonhole principle bounds to a class of graphs with more pigeons and smaller degree would yield lower bounds for random CNFs.

A potentially simpler problem that still gets to the heart of the matter is to prove quasipolynomial lower bounds for $Res(\text{polylog}\, n)$ proofs of the weak pigeonhole principle which would match the upper bounds in [14]. It is conceivable that this could be achieved by proving lower bounds for $Res(k)$ proofs of the weak pigeonhole principle for larger and larger k, extending the exponential lower bound for $Res(2)$ in [2]; but new techniques seem to be needed.

Acknowledgments

We are very grateful to Alan Woods for suggesting this problem a while back and for the many valuable comments and insights that he shared with us. We also want to thank Mike Molloy for some helpful discussions. Finally, we are grateful to Miki Ajtai for hinting that there may just not be a polynomial-size bounded-depth proof.

References

[1] M. Ajtai. The complexity of the pigeonhole principle. *Combinatorica*, 14(4):417–433, 1994.

[2] A. Atserias, M. L. Bonet, and J. Esteban. Lower bounds for the weak pigeonhole principle beyond resolution. In J. Orejas, P. G. Spirakis, and J. van Leeuwen, editors, *Automata, Languages, and Programming: 28th International Colloquium*, volume 2076 of *Lecture Notes in Computer Science*, pages 1005–1016, Heraklion, Crete, July 2001. Springer-Verlag.

[3] P. Beame. A switching lemma primer. Technical Report UW-CSE-95–07–01, Department of Computer Science and Engineering, University of Washington, November 1994.

[4] P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, P. Pudlák, and A. Woods. Exponential lower bounds for the pigeonhole principle. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, pages 200–220, Victoria, B.C., Canada, May 1992.

[5] P. Beame and S. Riis. More on the relative strength of counting principles. In P. W. Beame and S. R. Buss, editors, *Proof Complexity and Feasible Arithmetics*, volume 39 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 13–35. American Mathematical Society, 1998.

[6] E. Ben-Sasson. Hard examples for bounded-depth Frege. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 563–572, Montreal, Quebec, Canada, May 2002.

[7] E. Ben-Sasson and A. Wigderson. Short proofs are narrow – resolution made simple. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 517–526, Atlanta, GA, May 1999.

[8] S. Buss and G. Turán. Resolution proofs of generalized pigeonhole principles. *Theoretical Computer Science*, 62:311–317, 1988.

[9] S. R. Buss. Polynomial size proofs of the pigeonhole principle. *Journal of Symbolic Logic*, 57:916–927, 1987.

[10] A. Haken. The intractability of resolution. *Theoretical Computer Science*, 39:297–305, 1985.

[11] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 6–20, Berkeley, CA, May 1986.

[12] R. Impagliazzo and N. Segerlind. Counting axioms do not polynomially simulate counting gates. In *Proceedings 42nd Annual Symposium on Foundations of Computer Science*, pages 200–209, Las Vegas, Nevada, October 2001. IEEE.

[13] J. Krajíček. *Bounded Arithmetic, Propositional Logic and Complexity Theory*. Cambridge University Press, 1996.

[14] A. Maciel, T. Pitassi, and A. Woods. A new proof of the weak pigeonhole principle. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 368–377, Portland,OR, May 2000.

[15] J. Paris and A. Wilkie. Counting problems in bounded arithmetic. In *Methods in Mathematical Logic: Proceedings of the 6th Latin American Symposium on Mathematical Logic 1983*, volume 1130 of *Lecture notes in Mathematics*, pages 317–340, Berlin, 1985. Springer-Verlag.

[16] J.B. Paris, A. J. Wilkie, and A. R. Woods. Provability of the pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic*, 53:1235–1244, 1988.

[17] R. Raz. Resolution lower bounds for the weak pigeonhole principle. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 553–562, Montreal, Quebec, Canada, May 2002.

[18] A. A. Razborov. Bounded arithmetic and lower bounds in Boolean complexity. In P. Clote and J. Remmel, editors, *Feasible Mathematics II*, pages 344–386. Birkhauser, 1995.

[19] A. A. Razborov. Proof complexity of pigeonhole principles. In *Proceedings of the Fifth International Conference on Developments in Language Theory*, pages 100–116, Vienna, Austria, July 2001.

[20] A. A. Razborov. Resolution lower bounds for perfect matching principles. In *Proceedings Seventeenth Annual IEEE Conference on Computational Complexity*, pages 17–26, Montreal, PQ, Canada, May 2002.

[21] A. Urquhart and X. Fu. Simplified lower bounds for propositional proofs. *Notre Dame Journal of Formal Logic*, 37(4):523–544, 1996.

[22] Leslie G. Valiant. Reducibility by algebraic projections. *L'Enseignement Mathématique*, XXVIII:253–268, 1982.