Electronic Colloquium on Computational Complexity, Report No. 56 (2001)

An Exponential Separation between Regular and General Resolution

Michael Alekhnovich
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA USA
mike@mccme.ru

Jan Johannsen*
Institut für Informatik
Ludwig-Maximilians-Universität
München, Germany
jjohanns@informatik.uni-muenchen.de

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

Alasdair Urquhart‡
Department of Philosophy
University of Toronto
Toronto, ON CANADA
urquhart@cs.toronto.edu

August 6, 2001

Abstract

This paper gives two distinct proofs of an exponential separation between regular resolution and unrestricted resolution. The previous best known separation between these systems was quasi-polynomial.

1 Introduction

Propositional proof complexity is a currently very active area of research. In the realm of the theory of feasible proofs it plays a role analogous to the role played by Boolean circuit complexity in the theory of computational complexity.

The motivation to study the complexity of propositional proof systems comes from two sides. First, it was shown in the seminal paper of Cook and Reckhow [7] that the existence of "strong" systems in which all tautologies have proofs of polynomial size is tightly connected to the NP vs. $co-NP$ problem. This direction explains the considerable efforts spent in proving super-polynomial lower bounds for proof systems that are as strong as possible.

The second motivation concerns automated theorem proving. The main goal is to give some theoretical justification for and investigate the efficiency of heuristics for testing satisfiability. It turns out that the more sophisticated our propositional proof system is the harder it is to find

* Research supported by the DFG Emmy-Noether-Programme under grant No. Jo 291/2-1.
† Research partially supported by an NSF grant, a US-Israeli BSF grant, and an NSERC University Faculty Award.
‡ Research supported by NSERC grant -9700239.

ISSN 1433-8092

proofs of size close to optimal. The most thoroughly studied and oldest class of algorithms for satisfiability is based on simple proof systems such as resolution, or even on restricted subsystems of resolution. This partially explains why it is of interest to study such simple and weak proof systems and investigate how they relate to each other.

A lot of recent research was concentrated on the separation between different variants of resolution. In a series of papers [9], [10], [4],[2] the following relationships were studied: tree-like resolution vs. resolution, Davis-Putnam resolution vs. general resolution, regular resolution vs. general resolution. For all pairs except the latter exponential gaps have been produced.

The regularity restriction was first introduced by Grigory Tseitin in a ground-breaking article [15], the published version of a talk given in 1966 at a Leningrad seminar. This restriction is very natural, in the sense that algorithms such as that of Davis, Logemann and Loveland [8] (often described as the "Davis-Putnam procedure" and the prototype of almost all satisfiability algorithms used in practice today) can be understood as a search for a regular refutation of a set of clauses. If refutations are represented as trees, rather than dags, then minimal-size refutations are regular, as can be proved by a simple pruning argument [16, p. 436].

The main result of Tseitin's paper [15] is an exponential lower bound for regular resolution refutations of contradictory CNF formulas based on graphs. Subsequently, some authors tried to extend Tseitin's results to general resolution by showing that regular resolution can simulate general resolution efficiently. The results of Goerdt [10] and the present paper show that these attempts were doomed to failure. However, it remains an open question whether this simulation might not hold for some special cases. In the conclusion of the paper, we make a conjecture to the effect that for the formulas of Tseitin, as well as other well-known families of examples, there is always a minimal-size regular refutation.

The first example of a contradictory CNF formula whose shortest resolution refutation is irregular was given by Wenqi Huang and Xiangdong Yu [11]. Andreas Goerdt [10] gave the first super-polynomial separation between regular resolution and unrestricted resolution by constructing a family of formulas that have polynomial-size resolution proofs, but require quasipolynomial-size regular resolution refutations.

In this paper, we present two new families of formulas, and prove that they have simple polynomial-size resolution refutations, but require exponential-size regular resolution refutations. Our first example is technically simpler, and results in a stronger lower bound, but looks like a cheating trick designed especially for the separation. The second example is far more natural, and the corresponding lower bound is of interest in itself.

The paper is organized in the following way. Section 2 contains definitions necessary for both examples. We give these examples independently in Sections 3 and 4-5.

2 Preliminaries

The *resolution principle* says that if C and D are clauses and x is a variable, then any assignment that satisfies both of the clauses $C \vee x$ and $D \vee \neg x$ also satisfies $C \vee D$. The clause $C \vee D$ is said to be a *resolvent* of the clauses $C \vee x$ and $D \vee \neg x$ derived by *resolving on* the variable x. A *resolution derivation* of a clause C from a CNF formula F consists of a sequence of clauses in which each clause is either a clause of F, or is a resolvent of two previous clauses, and C is the last clause in the sequence; it is a *refutation* of F if C is the empty clause Λ. The *size* of a refutation is the number of resolvents in it. We can represent it as a directed acyclic graph (dag) where the nodes are the clauses in the refutation, each clause of F has out-degree 0, and any other clause has two arcs pointing to the two clauses that produced it. The arcs pointing to $C \vee x$ and $D \vee \neg x$ are labeled with the literals x and $\neg x$ respectively. It is well known that resolution is a *sound* and *complete* propositional proof system, i.e., a formula F is unsatisfiable if and only if there is a resolution refutation for F.

A resolution refutation is *regular* if on any path from Λ to a clause in F (in the directed acyclic graph associated with the refutation), each variable is resolved on at most once along the path.

An *assignment* for a formula F (sometimes we call it also a *restriction*) is a Boolean assignment to some of the variables in the formula; the assignment is *total* if all the variables in the formula are assigned values. If C is a clause, and σ an assignment, then we write $C\lceil\sigma$ for the result of applying the assignment to C, that is, $C\lceil\sigma = 1$ if $\sigma(l) = 1$ for some literal l in C, otherwise, $C\lceil\sigma$ is the result of removing all literals set to 0 by σ from C (with the convention that the empty clause is identified with the Boolean value 0). If F is a CNF formula, then $F\lceil\sigma$ is the conjunction of all the clauses $C\lceil\sigma$, C a clause in F.

If $R = C_1, \ldots, C_k$ is a resolution derivation from a formula F, and σ an assignment to the variables in F, then we write $R\lceil\sigma$ for the sequence $C_1\lceil\sigma, \ldots, C_k\lceil\sigma$.

Lemma 2.1. *If R is a regular resolution derivation of C from a formula F, and σ an assignment, then there is a subsequence of $R\lceil\sigma$ that is a regular resolution derivation of $C\lceil\sigma$ from $F\lceil\sigma$.*

Proof. This is a straightforward induction on the length of the derivation from F. □

Every regular resolution refutation can be represented by a read-once branching program [12]. Although we prefer to speak in terms of refutations rather than branching programs, we nevertheless need some ideas from the latter framework. A path in a resolution refutation can be considered as determined by the answers to a series of queries. That is to say, starting at the root of the refutation, let us follow a path in the proof according to the following recipe. If a node ν in the refutation is labeled with a clause $C \vee D$ derived from clauses $C \vee x$ and $D \vee \neg x$, then we say that the variable x is *queried* at ν. If we extend the path to the node labeled with $C \vee x$, then we say that we have answered the query with "x," while in the other case, we have answered with "$\neg x$." Thus every path π in the refutation from the root to a node in the proof corresponds to a set of literals constituting the set of answers to the queries in the path; conversely the set of literals constituting the answers uniquely determines the path. The assignment defined by setting all of the literals in this set to 0 falsifies all the clauses labeling nodes in the path.

For any node in a regular refutation we can define an important set of *forgotten* variables:

Definition 2.2. *For every node ν in a regular refutation labeled with a clause C, we say that a variable is* forgotten *at ν if it does not occur in C, but is queried on some path from the root to ν.*

The intuition behind this definition is given by the following lemma:

Lemma 2.3. *If a variable is forgotten at a node ν then it is also forgotten at all successors (that go after ν on a path from the root).*

Proof. Assume that a variable x is forgotten at the node ν. By definition there exists a path from the root to ν on which this variable is queried and the clause labeling ν does not contain x. All we need to prove is that the clauses labeling successors of ν do not contain x either. This follows from the regularity condition; if some successor contains x then this variable has been resolved twice on the path from the root. $\square$

Corollary 2.4. *If the variable x is forgotten at a node ν in a regular refutation, then no axiom reachable from ν can contain x. In particular, the clause labeling ν must be inferred from initial clauses that do not contain x.*

Clearly this result only applies to the case of a regular refutation. This corollary will be a central point in our lower bounds for regular resolution. Our strategy will be to find a node ν with a forgotten variable x and show that all initial clauses free of x do not imply the clause C_ν even semantically; to show this we produce an assignment σ that is falsified only on clauses containing x and $C_\nu \lceil \sigma = 0$.

3 First example: GT'_n formulas

Our first example is based on the ordering principle first considered by Krishnamurthy[13].

Definition 3.1. *Let X be the set of variables x_{ij} for $i, j \in [n]$, $i \neq j$. The contradiction GT_n consists of the following axioms:*

$$
\begin{array}{ll}
x_{i,j} = \neg x_{j,i} & 1 \leq i < j \leq n \\
x_{i_1,i_2} \wedge x_{i_2 i_3} \rightarrow x_{i_1 i_3} & \textit{for any distinct } i_1, i_2, i_3 \in [n] \\
\bigvee_{k \in [n],\, k \neq j} x_{k,j} & j \in [n]
\end{array}
$$

This contradictory principle can have several semantical interpretations, our proof will essentially depend upon the following one. Consider the variable x_{ij} as a predicate $i \succ j$. The first two groups of axioms assure that $\succ$ is a total linear ordering on the set $[n]$. The principle GT_n claims that any such ordering never has a maximum. This contradiction was used by Bonet and Galesi [3] to show the optimality of size-width relationship. It was shown there that:

Theorem 3.2 (Bonet and Galesi [3]). *There exist polynomial-size regular resolution refutations of GT_n.*

We modify GT_n to make it harder for regular resolution, but preserve its unrestricted resolution complexity. For that we replace some axioms C with a pair $C \vee x_C$, $C \vee \neg x_C$. The variable x_C should be chosen in a certain precise way to simplify our lower bound.

Definition 3.3. *Let X be the set of variables x_{ij} $i, j \in [n]$, $i \neq j$; clearly the cardinality of X is less than n^2. Let us fix an arbitrary enumeration (with repetitions) of X as a pair of functions $i(r), j(r)$ so that $X = \{x_{i(r),j(r)} | \; r \in [0 \ldots n^2 - 1]\}$.*

The set of clauses GT'_n consists of the following axioms:

$$
\begin{array}{ll}
x_{i,j} = \neg x_{j,i} & 1 \leq i < j \leq n \\
\bigvee_{k \in [n],\, k \neq j} x_{k,j} & j \in [n]
\end{array}
$$

$\neg x_{i_1,i_2} \vee \neg x_{i_2 i_3} \vee \neg x_{i_3 i_1} \vee x_{i(r),j(r)}$	*for any distinct* $i_1, i_2, i_3 \in [n]$ *and* $r = n(i_1 + i_2) + i_1 + i_3 \pmod{n^2}$
$\neg x_{i_1,i_2} \vee \neg x_{i_2 i_3} \vee \neg x_{i_3 i_1} \vee \neg x_{i(r),j(r)}$	*for any distinct* $i_1, i_2, i_3 \in [n]$ *and* $r = n(i_1 + i_2) + i_1 + i_3 \pmod{n^2}$

Thus GT'_n is defined up to an arbitrary enumeration of the set X. It is more convenient to us to write the axioms $x_{i_1,i_2} \wedge x_{i_2 i_3} \to x_{i_1 i_3}$ in the symmetric form $\neg x_{i_1,i_2} \vee \neg x_{i_2 i_3} \vee \neg x_{i_3 i_1}$, these representations are equivalent modulo axioms $x_{ij} = \neg x_{ji}$.

Corollary 3.4. *There is a polynomial size refutation of GT'_n in general resolution.*

Proof. It is easy to see that the principle GT_n can be inferred from GT'_n in general resolution in a polynomial number of steps. Hence, the Corollary follows by Theorem 3.2. $\square$

The refutation of the preceding corollary is irregular. By contrast, the main theorem of this section shows that any regular refutation of GT'_n has exponential size. Before proving this lower bound, we require some definitions and lemmas.

Definition 3.5. *For an assignment α on X let $Supp(\alpha)$ be the set of all $i \in [n]$ such that α assigns a value to either x_{ij} or x_{ji} for some j.*

Recall the notion of *critical assignment* for GT_n [3]. We generalize it to the case of *partial* critical assignments:

Definition 3.6. *For a subset of vertices $S \subset [n]$ a* partial critical assignment *for S is an arbitrary assignment α that gives values to all variables x_{ij}, $i, j \in S$ and for any clause C of the original principle GT_n, $C\lceil\alpha \not\equiv 0$.*

Thus an assignment is critical iff it does not violate the properties of the linear ordering (recall that we associate variable x_{ij} with a predicate $i \succ j$). Hence partial critical assignments α are in one-to-one correspondence with all linear orderings on the set $Supp(\alpha)$. This semantical interpretation is essential and we will use it throughout the proof.

Lemma 3.7. *Assume that α is a critical partial assignment with $|Supp(\alpha)| < n-2$ and the variable x_{ij} is unassigned by α. Then for any $\epsilon \in \{0, 1\}$, α can be extended to a critical assignment α' with $|Supp(\alpha')| \leq |Supp(\alpha)| + 2$ such that $\alpha'(x_{ij}) = \epsilon$.*

Proof. The proof becomes trivial after the decoding of the definitions: we have a linear ordering on a set S with $|S| < n - 2$. We need to extend this ordering on one or two new elements (depending on whether i or j is already contained in S). Clearly we can set $i \prec j$ as well as $i \succ j$ and extend our ordering in the correct way. $\square$

Lemma 3.8. *Assume that α is a critical partial assignment with $|Supp(\alpha)| \leq n/5 + 1$ and i_1, i_2, i_3 are distinct elements from $[n] \setminus Supp(\alpha)$. Then α can be extended to a total assignment for X so that all axioms of the original principle GT_n except the axiom $x_{i_1 i_2} \wedge x_{i_2 i_3} \to x_{i_1 i_3}$ are satisfied.*

Proof. We need to extend the linear ordering on $Supp(\alpha)$ to some total (contradictory) ordering on $[n]$. For that we extend α to some arbitrary partial critical assignment on $[n] \setminus \{i_1, i_2, i_3\}$, then set $i_1 \succ i_2 \succ i_3 \succ i_1$ and make i_1, i_2, i_3 greater than the rest of the vertices. $\square$

Lemma 3.9. *Assume that S is a ("bad") subset of $[n]$ of the size at most $n/5+1$. Then for any $r \in [0 \dots n^2-1]$ there exist three distinct elements $i_1, i_2, i_3 \in [n] \setminus S$ such that $n(i_1+i_2)+i_1+i_3 = r \pmod{n^2}$.*

Proof. Consider some $r \in [0 \dots n^2-1]$. Let us first assume that $S = \emptyset$. Clearly the equation $n \cdot i + j = r (\mathrm{mod}\ n^2)$; $i, j \in [0 \dots 2n-1]$ can have at most 4 solutions. Thus there exist at least $n-12$ triples of distinct $i_1, i_2, i_3 \in S$ such that $n(i_1+i_2)+i_1+i_3 = r \pmod{n^2}$. Whenever we add an element i to S we throw away at most 4 solutions. □

Theorem 3.10. *Any regular resolution refutation of GT'_n has size greater than $2^{n/10}$.*

Proof. Let R be a regular resolution refutation of GT'_n. We will single out a particular set of distinct paths in the refutation, and then show that this set has exponential size. These paths are defined by successive extensions; at each node ν along these paths we define an auxiliary critical assignment α_ν that falsifies the clause in ν.

Initially we start with a single path at the root and α is empty. Now assume that we have defined l distinct paths up to the nodes $\nu_1, \dots, \nu_l$. For each node ν_k with $|Supp(\alpha_{\nu_k})| < n/5$ we extend the corresponding path in the following way:

- If a variable x_{ij} is queried at ν_k and its value is already defined by α then we extend our path according to this value; α does not change.
- If a variable x_{ij} is queried at ν_k and either i or j does not belong to $Supp(\alpha_{\nu_k})$ then ν_k is a *branching node* and we proceed in the following way. Assume that the answer 0 to x_{ij} leads to the vertex ν_{k_0} and 1 to ν_{k_1}. We extend the existing path up to both ν_{k_0} and ν_{k_1}. For each $\epsilon \in \{0,1\}$ we extend α_{ν_k} by Lemma 3.7 to an arbitrary partial critical assignment $\alpha_{\nu_{k_\epsilon}}$ such that $\alpha_{\nu_{k_\epsilon}}(x_{ij}) = \epsilon$ and $|Supp(\alpha_{\nu_{k_\epsilon}})| \le |Supp(\alpha_{\nu_k})| + 2$.

We use this strategy to extend paths till every path leads to a node with $|Supp(\alpha)| \ge n/5$. Since the value of $|Supp(\alpha)|$ can increase at most by 2 in branching nodes, every path has at least $n/10$ branching nodes, hence there are at least $2^{n/10}$ distinct paths. It is left to show that they do not intersect each other.

For the sake of contradiction, assume that two distinct paths diverge in the node ν_1 and then merge again in the node ν_2. Assume that the variable x_{ij} is queried in ν_1. The key observation is that x_{ij} is forgotten in ν_2 (because the clause in ν_2 cannot contain both literals x_{ij} and $\neg x_{ij}$). Choose any r such that $i = i(r)$ and $j = j(r)$ (see Definition 3.3). By Lemma 3.9 we can choose three vertices $i_1, i_2, i_3 \notin Supp(\alpha_{\nu_2})$ such that $n(i_1+i_2)+i_1+i_3 = r \pmod{n^2-1}$. Now let us set by Lemma 3.8 the rest of variables so that all axioms of GT'_n except $A_0 = \neg x_{i_1 i_2} \vee \neg x_{i_2 i_3} \vee \neg x_{i_3 i_1} \vee x_{ij}$ and $A_1 = \neg x_{i_1 i_2} \vee \neg x_{i_2 i_3} \vee \neg x_{i_3 i_1} \vee \neg x_{ij}$ are satisfied. We produce a total assignment that falsifies the clause C_{ν_2} labeling ν_2 (because it is an extension of α_{ν_2}) and all violated clauses contain the forgotten x_{ij}. The contradiction with Corollary 2.4 proves the Theorem. □

In the end of this section we remark that the same procedure can be applied to the set of clauses MGT_n that results from GT_n if we replace all wide clauses with equivalent 3-CNF's, thus yielding a bounded width separation between regular and general resolution.

Namely, the set MGT_n as defined by Bonet and Galesi [3] contains, instead of the clauses $\bigvee_{k=1,k\neq j}^{n} x_{k,j}$, the following axioms over the auxiliary variables $y_{0,j}, \ldots, y_{n,j}$:

$$\neg y_{0,j} \wedge \bigwedge_{i=1,i\neq j}^{n} (y_{i-1,j} \vee x_{i,j} \vee \neg y_{i,j}) \wedge y_{n,j} \ .$$

Let us substitute 0 for all $y_{0,j}$ so that in all we have a set Y of n^2 auxiliary variables. Now we replace each axiom $C = \neg x_{i_1,i_2} \vee \neg x_{i_2 i_3} \vee \neg x_{i_3 i_1}$ with a tuple $C \vee x_r^{\epsilon_1} \vee y_r^{\epsilon_2}$, where $r = n(i_1 + i_2) + i_1 + i_3 \pmod{n^2}$. Obviously the resulting principle MGT_n' has a polynomial-size refutation in resolution, and its hardness for regular resolution can be shown essentially in the same way as in the proof of Theorem 3.10.

4 Second example: the stone formulas

4.1 Definitions

Our second example will be a generalization of the implication graph formulas, first introduced by Raz and McKenzie [14], and also used in subsequent papers [2, 4, 5]. Let G be a directed, acyclic graph, with fan-in 2, n vertices, and a single sink vertex; we shall call a graph satisfying these conditions a *pointed graph.* We shall use the phrase "decorated graph" to refer to a pair (G, U), where $G = (V, E)$ is a pointed graph, and U is a subset of V not containing the sink.

The implication graph formulas encode the following contradictory statement: "All of the source vertices and the vertices in U are colored red, the sink is colored blue, and if both the predecessors of a vertex are red, so is the vertex itself." In order to make the formulas difficult for regular resolution, we express this statement somewhat indirectly, so that instead of speaking directly of colored vertices, we introduce extra variables speaking of the placement of colored stones on the vertices.

The decorated graph (G, U) can be viewed as a board for a board game such as Go or Hex. Additionally we assume that we have a set S of $m \geq n$ stones that are to be placed on the board, each of which can be colored red or blue. (The reader might picture these stones as similar to the discs in the game of Othello, being red on one side and blue on the other; thus they can be red or blue depending on which side is up.)

Our "stone formulas," *Stone*(G, U, S), are defined as follows. Let (G, U) be a decorated graph, where $G = (V, E)$, $|V| = n$, and S is a set of $m \geq n$ stones. Let Sources(G) be the source vertices of G. The variables of the formula are $P_{i,j}$, $i \in (V \setminus U)$ and $j \in S$, and R_j, $j \in S$. The variable $P_{i,j}$ says "Vertex i contains stone j," or "Stone j is placed on vertex i," while R_j says "Stone j is colored red," and $\neg R_j$ says "Stone j is colored blue." We call a variable $P_{i,j}$ an *edge variable*, and a variable R_j a *stone variable.* For a vertex $i \in V$, and a stone $t \in S$, let $D_{i,t}$ be 1 if $i \in U$, otherwise, $D_{i,t}$ is the formula $(P_{i,t} \wedge R_t)$. The clauses are as follows.

(i) $\bigvee\{P_{i,j} | \ j \in S\}$, $i \in V \setminus U$. These clauses express the fact that every vertex contains some stone (the vertices in U should be thought of as containing a particular red stone).

(ii) For all vertices $k \in \text{Sources}(G) \setminus U$ and stones j, $(P_{k,j} \rightarrow R_j)$.

(iii) For the sink vertex s and stone j, $(P_{s,j} \rightarrow \neg R_j)$.

(iv) For all vertices i, j, k such that (i,k) and (j,k) are edges in G, where $k \in V \setminus U$, and for all stones t,u,v, $(D_{i,t} \wedge D_{j,u} \wedge P_{k,v} \to R_v)$. That is, if the stones t,u,v are placed on the vertices i,j,k, and t and u are both red, then the stone v must also be red. Since this clause expresses an induction rule, we shall refer to it as an "induction clause" associated with the vertex k.

We write $Stone(G,S)$ for the special case of the stone formulas where $U = \emptyset$. It is not hard to check that the number of variables in $Stone(G,S)$ is $(n+1)m$, and the number of clauses in the formula is $O(m^3 n)$.

Lemma 4.1. *For all m and all directed acyclic graphs G, $Stone(G,S)$ has a resolution refutation of size $O(m^3 n)$.*

Proof. For a vertex k with predecessors i and j, we will say that k is colored red if we have derived all clauses of type (ii) for k. The refutation proceeds inductively from the source vertices to the sink, deriving all clauses of type (ii) for every k in the graph.

Let us assume that both predecessors i and j of a vertex k are colored red. We first derive all of the clauses $(\neg P_{i,t} \vee \neg P_{j,u} \vee \neg P_{k,v} \vee R_v)$ for $t,u,v \in S$, by resolving clauses of type (ii) against appropriate induction clauses. There are m^3 such clauses, and each can be derived in two steps, so this part of the proof takes $O(m^3)$ steps. Next, we derive all clauses of the form $(\neg P_{j,u} \vee \neg P_{k,v} \vee R_v)$, by resolving the previously obtained clauses against clauses of type (i). There are m^2 such clauses, each of which takes m steps to obtain, so this part of the proof also takes $O(m^3)$ steps. We then repeat this pattern to derive all clauses of the form $(\neg P_{kv} \vee R_v)$, in $O(m^2)$ steps, completing the induction step.

Finally, when the sink s is colored red, we can derive the empty clause from R_s and $\neg R_s$ in $O(m)$ steps. The entire refutation takes $O(m^3 n)$ steps. ☐

The refutation constructed in the preceding lemma is highly irregular, since at each induction step, we resolve on all the stone variables, so that there are paths in the derivation in which a stone variable is resolved on over and over again. The regularity restriction rules out a refutation of this type, and (as we shall see below), any regular refutation has to be exponentially large.

4.2 Graph Pebbling

In order to prove an exponential lower bound for the stone formulas, we will need to begin with a graph G with high pebbling number. The next definition describes a slight generalization of the usual concept of pebbling number.

Definition 4.2. *Let (G,U) be a decorated graph. A configuration is a subset of V. A* legal pebbling *of a vertex v from U in G is a sequence $C_0, C_1, \ldots, C_\ell$ of configurations, with $C_0 = \emptyset$ and $v \in C_\ell$, in which each configuration C_{i+1} follows from C_i by one of the following rules:*

1. *Any vertex $u \in U \cup \mathrm{Sources}(G)$ can be added to C_i, i.e. $C_{i+1} = C_i \cup \{u\}$.*

2. *A vertex u can be added to C_i to get $C_{i+1} = C_i \cup \{u\}$, if all immediate predecessors of u are in C_i.*

3. *Vertices can be removed, so that $C_{i+1} \subset C_i$.*

The complexity *of a legal pebbling of v from U is the maximal size of any configuration in the sequence. The* pebbling number $\mathsf{Peb}(G,U)$ of a decorated graph (G,U) *is the minimal number n for which there exists a legal pebbling of the sink of G from U with complexity n. The* pebbling number $\mathsf{Peb}(G)$ *of a pointed graph is the pebbling number of the decorated graph* $(G,\emptyset)$.

Lemma 4.3. *There is a constant $\beta > 0$ such that for all sufficiently large n, there is a pointed graph G with n vertices that has pebbling number* $\mathsf{Peb}(G) \geq \beta n/\log n$.

Proof. Graphs satisfying these conditions are constructed in a paper by Celoni, Paul and Tarjan [6]; their proof is based on an earlier construction of Valiant. ☐

The following definition is useful in describing the effect of restrictions on stone formulas.

Definition 4.4. *For a pointed graph $G = (V,E)$ and $v \in V$, let $G[v]$ denote the induced subgraph of G on those vertices u from which v is reachable, i.e. there is a directed path from u to v.*

The next lemma shows that if we add a new "free" vertex to a decorated graph with high pebbling number, then we can always find a subgraph with high pebbling number.

Lemma 4.5. *Let (G,U) be a decorated graph with pebbling number $p+1$, and i an vertex of G not in U. Then either $(G, U \cup \{i\})$ or $(G[i], U \cap G[i])$ has pebbling number at least p.*

Proof. Assume that both $(G, U \cup \{i\})$ and $(G[i], U \cap G[i])$ have pebbling number at most $p-1$. Then there is a legal pebbling $C_0, \ldots, C_l = \{i\}$ of $G[i]$ from $U \cap G[i]$ and a legal pebbling $D_0, \ldots, D_m$ of G from $U \cup \{i\}$ each of complexity at most $p-1$. Then $C_0, \ldots, C_l, D_1 \cup \{i\}, \ldots, D_m \cup \{i\}$ is a legal pebbling of (G,U) from U of complexity at most p, contradicting the assumption that (G,U) has pebbling number $p+1$. ☐

The next definition picks out a type of vertex that plays a central role in the restrictions described in the following subsection.

Definition 4.6. *Let (G,U) be a decorated graph. A vertex $v \in G$ is* important *if there is a path from v to s not containing any vertex in U, otherwise v is* unimportant.

Lemma 4.7. *If $\mathsf{Peb}(G,U) = p$, then there exist at least $p-2$ important vertices in G.*

Proof. We prove the lemma by showing that if (G,U) is a decorated graph, then there is a legal pebbling of the sink of G from U in which at most two unimportant vertices occur in every configuration. Let $C_0, \ldots, C_\ell$ be a legal pebbling of the sink s from U, and $D_0, \ldots, D_\ell$ the sequence obtained by setting $D_i = (C_i \setminus J)$, where J is the set of unimportant vertices in the pebbling. We argue by induction, proceeding backwards from D_ℓ to D_0, that the resulting sequence can be converted into a legal pebbling of s from U, by adding at most two vertices from U to each configuration in the sequence.

By definition, s is important, so we only need to examine the case where C_{i+1} is obtained from C_i by an appropriate rule. Let us suppose that C_{i+1} was obtained from C_i by the second rule in Definition 4.2, and that the vertex u is important, but one or both of the predecessors of u is unimportant. This can only happen if the predecessor in question is in U. Hence, we can obtain D_{i+1} from D_i by one or two applications of Rule 1 in Definition 4.2, followed by an application of Rule 2. ☐

4.3 Critical assignments for stone formulas

In the remainder of the paper, we concentrate on certain special assignments for the formulas $Stone(G, U, S)$. These assignments are determined by two items, a mapping from (some of the) vertices to stones, and a coloring of some of the stones.

Definition 4.8. *Let $G = (V, E)$ be a pointed graph, and S a set of stones, where $|S| \geq |V|$. A* restriction *$\rho = \langle \mu, \kappa \rangle$ for (G, S) is determined by*

1. *A bijective map μ from a subset of V to S;*
2. *A coloring κ assigning colors from the set $\{R, B\}$ to a subset of S (i.e., we assign the colors red and blue to some of the stones).*

The assignment σ determined by ρ *is defined by setting:*

- *$\sigma(P_{i,j}) = 1$ if $\langle i, j \rangle \in \mu$, $\sigma(P_{i,j}) = 0$ if i or j are in $Dom(\mu) \cup Ran(\mu)$, but $\langle i, j \rangle \notin \mu$;*
- *$\sigma(R_j) = 1$ if $\kappa(j) = R$, $\sigma(R_j) = 0$ if $\kappa(j) = B$;*
- *If j is in the domain of κ, but not in the range of μ, then $P_{i,j} = 0$, for all $i \in V$.*

If $|\mu| = r$ and $|\kappa| = s$, then we say that σ is a restriction with parameters *r and s. To simplify notation, we shall identify a restriction with the assignment that it determines.*

There are two special types of restrictions that are important in what follows. The first type is one in which none of the stones placed on vertices are assigned colors, that is to say, $\mathrm{Ran}(\mu) \cap \mathrm{Dom}(\kappa) = \emptyset$, and furthermore, the parameters r and s are equal. In this case, we shall describe $\rho = \langle \mu, \kappa \rangle$ as an *r-restriction.* The second special type of restriction is one in which *all* of the stones placed on vertices are assigned colors, that is, $\mathrm{Ran}(\mu) \subseteq \mathrm{Dom}(\kappa)$. This second type of restriction, we call a *clamping.*

Definition 4.9. *Let (G, U) be a decorated graph, k a vertex in G, and χ a partial coloring of G, that is, χ is a map from a subset of V to $\{R, B\}$. We say that χ is a k-*based coloring *of (G, U) if the following conditions hold:*

1. *If $i \in U$, then $\chi(i) = R$;*
2. *There is a path π in G from k to the sink of G so that the vertices in the path are exactly those to which χ assigns the color blue;*
3. *If a vertex i is not in $G[k] \cup \pi$, then $\chi(i) = R$.*

*We say that χ is k-*critical *if it is a total coloring of G.*

Lemma 4.10. *If (G, U) is a decorated graph and k an important vertex of G, then there is a k-critical coloring of (G, U).*

Proof. By assumption, there is a path from k to the sink not passing through any vertices in U. Color all the vertices in this path blue, and all other vertices in G red. Since k is the only blue vertex in G, all of whose predecessors are colored red, this is a k-critical coloring of (G, U). ☐

If χ is a coloring of a decorated graph (G, U), then a clamping $\rho = \langle \mu, \kappa \rangle$ is said to be *compatible* with χ if $\mathrm{Dom}(\chi) = \mathrm{Dom}(\mu)$, and for any vertex in $\mathrm{Dom}(\chi) = \mathrm{Dom}(\mu)$, $\kappa(\mu(i)) = \chi(i)$. If χ is a k-based coloring, then we say that ρ is a *k-based* clamping, and that it is *k-critical* if it is compatible with a k-critical coloring. The reader can easily check that if ρ is a k-critical clamping of (G, U), then it forces all the clauses in $\mathit{Stone}(G, U, S)$ to true, except for a single induction clause, which is forced to false. In addition, we say that a restriction is compatible with a k-based coloring χ if it can be extended to a clamping compatible with χ.

The next two lemmas, which are straightforward to prove, are used repeatedly in the rest of the proof.

Lemma 4.11. *Let (G, U) be a decorated graph, and χ a coloring of $G = (V, E)$. In addition, let S be a set of stones with $|S \setminus \mathit{Dom}(\kappa)| \geq |V|$, and $\rho = \langle \mu, \kappa \rangle$ a restriction for (G, S) so that $\mathit{Dom}(\mu) \subseteq \mathit{Dom}(\chi)$, and $\kappa(\mu(i)) = \chi(i)$, whenever $i \in \mathit{Dom}(\mu) \cap \mathit{Dom}(\chi)$ and $\kappa(\mu(i))$ is defined. Then ρ is compatible with χ.*

Proof. Extend μ to a bijection μ' from $\mathrm{Dom}(\chi)$ to $S \setminus \mathrm{Dom}(\kappa)$, and then set the colors of stones in the range of μ' by setting $\kappa'(\mu(i)) = \chi(i)$ for all $i \in \mathrm{Dom}(\chi) = \mathrm{Dom}(\mu')$. Then $\rho' = \langle \mu', \kappa' \rangle$ is a clamping compatible with χ. □

Lemma 4.12. *If $\rho = \langle \mu, \kappa \rangle$ is a k-based clamping of $\mathit{Stone}(G, U, S)$ with parameters r, s, then*

$$\mathit{Stone}(G, U, S)\lceil\rho = \mathit{Stone}(G[k], U', S'),$$

where $U' = G[k] \cap (U \cup \mathit{Dom}(\mu))$ and $S' = S \setminus \mathit{Dom}(\kappa)$.

Proof. This follows readily from the definitions. □

Lemma 4.13. *Let $\rho = \langle \mu, \kappa \rangle$ be an r-restriction for (G, S), where G has pebbling number N, and $|S| \geq |V|$. Then there is a vertex $k \in G$ and $\rho' = \langle \mu, \kappa' \rangle$ so that:*

1. *$\rho' = \langle \mu, \kappa' \rangle$ is a k-based clamping of (G, S), with $\kappa \subseteq \kappa'$;*
2. *$(G[k], G[k] \cap \mathit{Dom}(\mu))$ has pebbling number at least $N - r$.*

Proof. We construct ρ' by first constructing an appropriate k-based coloring χ of G, with $\mathrm{Dom}(\chi) = \mathrm{Dom}(\mu)$, and then using Lemma 4.11 to find the required k-based clamping. At each stage in the construction, we are given a designated vertex $k \in G$, and a k-based coloring χ of G. Initially, no vertices in G are assigned colors, and we choose the root as the designated vertex. In successive stages, we choose a new designated vertex $k' \in G$, and a new coloring χ', and we make these choices in such a way as to maximize the pebbling number of the decorated graph $(G[k'], G[k'] \cap U(k', \chi'))$, where $U(k', \chi')$ is the set of vertices in $G[k']$ mapped to R by χ'.

To be more precise, let us suppose that we have $\mu(i) = t$, and that the stone t has not yet been assigned a color. Let $U = U(k, \chi)$. Compare the two values

$$p_1 = \mathsf{Peb}(G[k], U \cup \{i\}) \qquad \text{and} \qquad p_2 = \mathsf{Peb}(G[i], U \cap G[i]) .$$

If $p_1 \geq p_2$, then set $k' = k$, extend χ by setting $\chi'(i) = R$ and $\chi'(i') = R$ for all i' that are unimportant in $(G[k], U \cup \{i\})$. If $p_1 < p_2$ then set $k' = i$, choose a path in G from i to k, and extend χ to an i-based coloring of $(G, U \cap G[i])$.

Let k and χ be the designated vertex and k-based coloring produced at the end of this process, that is, when $\mathrm{Dom}(\chi) = \mathrm{Dom}(\mu)$. By Lemma 4.11, ρ is compatible with χ, so that there is a a k-based clamping $\rho' = \langle \mu, \kappa' \rangle$ compatible with χ. By Lemma 4.5, $(G[k], G[k] \cap \mathrm{Dom}(\mu))$ has pebbling number at least $N - r$. $\square$

5 The lower bound for stone formulas

The general structure of the proof of the lower bound is similar to that of Beame and Pitassi's [1] simplified lower bound proof for the pigeonhole principle. That is, we will assume for the sake of contradiction that we have a short (sub-exponential) refutation of a stone formula. We will first show that we can apply a restriction to some (but not too many) of the variables such that the resulting refutation, after the restriction, is still a refutation of a stone formula on a reduced set of variables, that contains no complex clauses. Secondly, we argue separately that any regular resolution refutation of the formula must contain a complex clause, and thus we reach a contradiction.

Definition 5.1. *A clause C is called d-*complex *if one of the following holds:*

1. *C contains at least d distinct stone variables, or*
2. *There is a matching of size at least d from vertices to stones such that C contains the negative edge literals $\neg P_{i,j}$ for each pair (i, j) in the matching.*
3. *There is a subset W of at least d vertices, such that for all $i \in W$ there is a subset P_i of at least d stones, such that all variables $P_{i,j}$, $i \in W$, $j \in P_i$ are present in C.*

There are three parameters in our lower bound, γ, δ, and ϵ. The lower bound is of the form $2^{\delta n/(\log n)^3}$; the parameter associated with a complex clause is $\epsilon n/\log n$ and the size of the restriction is $\gamma n/\log n$.

Lemma 5.2 (Restriction Lemma). *Let γ, δ, ϵ be constants less than 1 such that $3\delta/\epsilon^2 \leq \gamma \leq \epsilon/2$, and let R be a resolution refutation of Stone(G, S), $|S| = 3n$, of size at most $S = 2^{\delta n/(\log n)^3}$. Then there exists a restriction ρ of size $r = \gamma n/\log n$, such that $R\lceil\rho$ has no $(\epsilon n/\log n)$-complex clauses.*

Proof. We will divide up the complex clauses into those of the first type and those of type two or three. For those of type two or three, we will greedily choose a matching between vertices and stones that forces all complex clauses of these types to true.

The total number of pairs in $G \times S$ to choose from is $3n^2$; each complex clause of the third type can be forced to true by choosing a $P_{i,j}$ from $(\epsilon n/\log n)^2$ pairs; each complex clause of the second type can be forced to true by choosing a $P_{i,j}$ from $(\epsilon(3n-1)n)/\log n \geq (\epsilon n/\log n)^2$ pairs. Thus, by averaging, there exists a $P_{i,j}$ such that at least $S(\epsilon n/\log n)^2/3n^2 \geq S(\epsilon^2/3(\log n)^2)$ complex clauses of the second and third type are forced to true.

Hence to force all the complex clauses of these types to true we need to choose a matching of size

$$\frac{\log S}{\log\left(1/\left(1 - \frac{\epsilon^2}{3(\log n)^2}\right)\right)} \leq \frac{\delta n \,/\, (\log n)^3}{\epsilon^2 \,/\, 3(\log n)^2} \leq \frac{3\delta n}{\epsilon^2 \log n} \leq \frac{\gamma n}{\log n}\ .$$

Now for the clauses of the first type, notice that there are $3n - \gamma n/\log n \geq 2n$ stones left that are untouched by the matching chosen above, and any complex clause of the first type still contains $(\epsilon - \gamma)n/\log n \geq \epsilon n/2\log n$ literals corresponding to these stones, since $\gamma \leq \epsilon/2$.

Since the refutation has size at most S, there are at most S clauses of the first type, each containing $\epsilon n/2\log n$ distinct stone variables that have not been set by the restriction already chosen. Each of these is thus set to one by $\epsilon n/2\log n$ out of at least $4n$ choices of stone literals.

By averaging, there must be one stone literal that forces $S(\epsilon/8\log n)$ of the complex clauses of type one to true. Thus to force all complex clauses of this type to true we need to set

$$\frac{\log S}{\log\left(1\,/\,\left(1 - \frac{\epsilon}{8\log n}\right)\right)} \leq \frac{\delta n\,/\,(\log n)^3}{\epsilon\,/\,8\log n} \leq \frac{8\delta n}{\epsilon(\log n)^2} \leq \frac{\gamma n}{\log n}$$

stone literals. $\square$

Our last lemma shows the existence of a complex clause for any initial graph with sufficiently large pebbling number. This is the only part of the proof in which the regularity restriction is used.

Lemma 5.3 (Complex Clause Lemma). *Let (G, U) be a decorated graph with $\mathsf{Peb}(G, U) = N = \Omega(n/\log n)$, where G has n vertices. Then for $|S| = m = 2n$, and n sufficiently large, any regular resolution refutation of Stone(G, U, S) contains an $(N/16)$-complex clause.*

Proof. Let R be a regular resolution refutation of *Stone*(G, U, S). We will follow a particular path π partway through the refutation; this path will be determined by giving a strategy for answering queries. The path is defined by successive extensions; at each stage in the definition, we define three auxiliary objects. First, at each stage, we single out a *designated vertex* k in G, second, we define a k-based coloring χ of G, third, we define a restriction ρ compatible with χ. The answers given along the path at each stage are compatible with ρ, that is to say, if the literal l occurs as a label on the path, then $\rho(l) = 0$. At each stage, the designated node k and the coloring χ determine the subgraph $G[k]$ of G, together with a subset U of $G[k]$, consisting of the vertices i of $G[k]$ with $\chi(i) = R$.

Initially, we start the path at the root, the designated vertex is the sink s of G, $\chi(u) = R$ for $u \in U$, and $\chi(s) = B$. Now assume that we have defined the path up to a node ν, together with a designated vertex k, a k-based coloring χ and a restriction ρ compatible with χ. In addition, let (G_ν, U_ν) be the decorated graph $(G[k], U)$, where U is the set of vertices in $G[k]$ colored red by χ. We want to extend the path by providing an appropriate answer to the variable queried at the node ν; we need also to define a new designated vertex k', a new k'-based coloring χ' and a new restriction ρ' compatible with χ'. In each case, the answers to the queries will be given as an extension of the current restriction ρ – unless the current restriction answers the query already, as can happen in the last case below.

- If a stone variable R_j is queried at ν and it is not currently placed on a vertex, i.e. for no vertex i do we have $\rho(P_{i,j}) = 1$, then extend ρ by coloring j red or blue arbitrarily. Set $k' = k$, $\chi' = \chi$.
- If a stone variable R_j is queried at ν, and it is already placed on a vertex i, i.e. $\rho(P_{i,j}) = 1$, then we answer as follows:

– If i is colored by χ then answer accordingly. That is, if $\chi(i) = R$, then set $\rho(R_j) = 1$, and if $\chi(i) = B$, then set $\rho(R_j) = 0$. Then set $k' = k$ and $\chi' = \chi$.

– If i is not assigned a color by χ, but is not important with respect to (G_ν, U_ν), then set $\rho(R_j) = 1$. Extend χ by setting $\chi'(i) = R$, and set $k' = k$.

– Otherwise, answer to maximize the pebbling number of the associated decorated graph. That is, compare the two values

$$p_1 = \mathsf{Peb}(G_\nu, U_\nu \cup \{i\}) \qquad \text{and} \qquad p_2 = \mathsf{Peb}(G_\nu[i], U_\nu \cap G[i]) \ .$$

If $p_1 \geq p_2$, then set $k' = k$, extend χ by setting $\chi'(i) = R$ and $\chi'(i') = R$ for all i' that are unimportant in $(G_\nu, U_\nu \cup \{i\})$, and set $\rho(R_j) = 1$.

If $p_1 < p_2$ then set $G_\xi = G_\nu[i]$, choose a path in G_ν from i to k, extend χ to an i-based coloring of G, and set $\rho(R_j) = 0$.

- If an edge $P_{i,j}$ is queried at ν, but $\rho(P_{i,j})$ is not defined, then set $\rho(P_{i,j}) = 1$. Set $k' = k$ and $\chi' = \chi$.

The so defined path can only end in an initial clause of type (i) or an induction clause of type (iv). The former case can only occur if all the stones have received colors, and thus all stones must have been queried. In the latter case, note that at the root, we have $\mathsf{Peb}(G_\nu, U_\nu) = N$, by Lemma 4.5 the pebbling number decreases by at most 1 with each answer to a query, and only at nodes where a stone is queried. But when we reach an induction clause, the pebbling number has decreased to 0. Therefore the path followed by the above strategy will not finish until at least N stones are asked about.

Thus, there is some node ξ on the path where exactly $N/2$ stones are queried. The pair (G_ξ, U_ξ) associated with ξ has pebbling number at least $N/2$. This follows from the fact that the pebbling number decreases by at most 1 at each stone query, and we have queried exactly $N/2$ stones. By Lemma 4.7, there are at least $N/2 - 2$ important vertices in (G_ξ, U_ξ).

We will now show that the clause C_ξ attached to the node ξ must be $N/16$ complex. Let I be the set of important vertices in (G_ξ, U_ξ). If at least $N/8$ of the vertices in I are mapped by μ and the mapping is remembered, i.e., the corresponding edge variables set to 1 by ρ occur negated in C_ξ, then C_ξ is an $N/8$-complex clause of type two. Hence, in the remainder of the proof, we shall assume that there is a subset $I' \subset I$ of at least $N/8$ vertices that are either unmapped by μ or are mapped by μ but all of the edge variables corresponding to this part of the mapping are forgotten.

There are several cases. We will introduce some terminology to help with these cases. Let $\rho =< \mu, \kappa >$ be the restriction associated with ξ. We partition the set of $N/2$ stones queried on the path to ξ into the *free* stones, F, and the *attached* stones, A. F consists of those stones queried along the path to ξ that are not in the range of μ; A are those stones queried on the path to ξ that are in the range of μ.

There are two general cases to consider. The first case (which is slightly easier) is when $|F| \geq N/4$ and the second case is when $|A| \geq N/4$.

Assume first that $|F| \geq N/4$. We will show that the clause C_ξ attached to the node ξ must be $N/16$-complex. The first subcase is when at least $N/8$ of the stones in F occur in C_ξ. In this case, C_ξ is an $N/8$-complex clause of type one. If this subcase does not occur, then at least $N/8$ of the stones in F are forgotten at the node ξ. Let $F' \subset F$ be this set of at least $N/8$ forgotten free stones.

Now we claim that for every $i \in I'$ and $t \in F'$, the variable $P_{i,t}$ must occur in C_ξ, and thus C_ξ is an $N/8$-complex clause of type three.

Suppose that $P_{i,t}$ does not occur in C_ξ for some $i \in I'$, $t \in F'$. Then we can modify the restriction ρ to ρ' by mapping i to t, and if $\rho(i) = u$, where $u \neq t$, then we assign a color to u, if it was not colored already. Since we have not tampered with any of the variables in C_ξ, it follows that $\rho'(C_\xi) = 0$. Since i is important, there is an i-based coloring extending the coloring χ associated with ξ, and this i-based coloring can be extended to an i-critical coloring. Let σ be an i-critical clamping compatible with this coloring, extending the restriction ρ', in which $\sigma(P_{i,t}) = 1$ and $\sigma(C_\xi) = 0$. The only initial clause falsified by σ is an induction clause associated with i, containing the variable R_t. However, by Corollary 2.4, C_ξ was inferred from clauses none of which contain the forgotten variable R_t. This is a contradiction, so it follows that C_ξ must contain the variables $P_{i,t}$.

We will now argue the second case, when $|A| \geq N/4$. Let B be the subset of vertices that are mapped by ρ to stones in A. Clearly, $|B| = |A|$. The first subcase, 2a, is when at least $N/8$ of the stones in A are remembered. In this case, C_ξ is an $N/8$-complex clause of type one.

The second subcase, 2b, is when for at least $N/16$ of the vertices i in B, the mapping of i into A is remembered, i.e., the negative edge literal $\neg P_{i,j}$ where $j = \mu(i)$ occurs in C_ξ. In this case, C_ξ is an $N/16$-complex clause of type two.

The third subcase, 2c, is when for at least $N/16$ of the vertices i in B, at least $N/2$ zero-edges containing i are remembered. That is, the edge variable $P_{i,j}$ occurs in C_ξ for at least $N/2$ distinct j's. In this case, C_ξ is an $N/16$-complex clause of type three.

If none of the subcases 2a, 2b or 2c occur, then there are sets $B' \subset B$ and $A' \subset A$ each of size exactly $N/8$ such that ρ defines a mapping from B' to A' but the mapping is forgotten; the color of each stone in A' is forgotten, and for every $i \in B'$, at most $N/2$ zero-edges out of i are remembered at ξ. We claim that for every $i \in I'$ and $t \in A'$, the variable $P_{i,t}$ must occur in C_ξ.

Suppose that this is not the case and let $P_{i,t}$ be some variable that does not occur in C_ξ, where $i \in I'$, $t \in A'$ and let $i' \in B'$ be the vertex that is mapped to t by ρ. Then we will modify the restriction ρ to obtain ρ' as follows. First, ρ' will map i to t. Secondly, ρ' will map i' to some unqueried stone t' such that the edge from i' to t' has not already been remembered to be 0. Further, the color associated with t' will be chosen to be the same as the color given to t by ρ; that is, the underlying coloring of the graph will be the same for ρ and for ρ'. It will still be the case that $\rho'(C_\xi) = 0$ since we have not tampered with any of the variables that are remembered (i.e., that occur in C_ξ.) Now as in Case 1, we will extend ρ' to an i-critical coloring, σ, extending the restriction ρ'. Since the only initial clause falsified by σ contains the variable R_t, and R_t is not remembered, we have reached a contradiction, and thus C_ξ must contain all such variables $P_{i,t}$, showing that it is an $N/8$-complex clause of the third type.

This completes the proof of the complex clause lemma. □

We are ready to state the main theorem of this section.

Theorem 5.4. *Let G be a pointed graph with n vertices and pebbling number $N = \Omega(n/\log n)$. Then any regular resolution refutation of Stone(G, S), where $|S| = 3n$ requires size $2^{\Omega(n/(\log n)^3)}$.*

Proof. We start with a pointed graph G with n vertices and pebbling number $\mathsf{Peb}(G) \geq \beta n/\log n$, in accordance with Lemma 4.3. Set $\epsilon = \beta/32$, $\gamma = \epsilon/2 = \beta/32$ and $\delta = \epsilon^3/6 = \beta^3/199608$, so that we have $3\delta/\epsilon^2 \leq \gamma \leq \epsilon/2$. Now assume that there is a regular resolution refutation R of $Stone(G, S)$, with size at most $S = 2^{\delta n/(\log n)^3}$. By Lemma 5.2 (the Restriction Lemma), there is a restriction ρ of size $r = \gamma n/\log n$, such that $R\lceil\rho$ has no $(\epsilon n/\log n)$-complex clauses.

By Lemma 4.13, there is a vertex $k \in G$ and $\rho' = \langle \mu, \kappa' \rangle$ so that ρ' is a k-based clamping of (G, S), with $\kappa \subseteq \kappa'$, and $(G[k], G[k] \cap \mathrm{Dom}(\mu))$ has pebbling number at least $N - r \geq N/2$. By Lemma 5.3 (the Complex Clause Lemma), the refutation $R\lceil\rho$ must contain a $(\beta n/32 \log n)$-complex clause, contradicting the conclusion of the previous paragraph. □

6 Open problem

The present paper gives a separation between regular and general resolution that is essentially optimal. Nevertheless, there are still mysteries surrounding the exact effect of the regularity restriction. As mentioned in the introduction, it was widely believed in the early years of research on the complexity of resolution that optimal proofs are always regular. This belief would be justified, in a sense, if the conjecture formulated below were proved to be true.

The most deeply investigated family of tautologies are those based on matchings in graphs, of which the best known are the pigeonhole formulas. These examples are based on graphs G for which no perfect matching exists; the corresponding contradictory CNF formula $F(G)$ asserts that G has a perfect matching. For example, the pigeonhole formula PHP_n can be understood as the formula $F(H)$, where H is the complete bipartite graph $K(n, n-1)$. Other well-studied examples are the graph-based formulas of Tseitin [15]. For both of these families of examples, the shortest known resolution refutations are regular.

Conjecture 6.1. *For contradictory formulas expressing matching principles in graphs, and also for the graph-based examples of Tseitin, there is always a regular refutation of minimal size.*

A proof (or disproof) of this conjecture would shed light on the question of exactly when the regularity restriction helps in searching for short refutations.

7 Acknowledgments

The first author is grateful to A. A. Razborov for his helpful comments.

References

[1] P. Beame and T. Pitassi. Simplified and improved resolution lower bounds. In *Proc. 28th ACM Symposium on Theory of Computing*, 1996.

[2] E. Ben-Sasson, R. Impagliazzo, and A. Wigderson. Near-optimal separation of tree-like and general resolution. ECCC TR00-005, 2000.

[3] M. Bonet and N. Galesi. A study of proof search algorithms for resolution and polynomial calculus. In *Proc. 40th Symposium on Foundations of Computer Science*, pages 422–432, 1999.

[4] M. L. Bonet, J. L. Esteban, N. Galesi, and J. Johannsen. On the relative complexity of resolution restrictions and cutting planes proof systems. *SIAM Journal of Computing*, 30:1462–1484, 2000.

[5] J. Buresh-Oppenheim, M. Clegg, R. Impagliazzo, and T. Pitassi. Homogenization and the polynomial calculus. In *Proceedings of ICALP 2000*, pages 926–937, 2000.

[6] J. Celoni, W. Paul, and R. Tarjan. Space bounds for a game on graphs. *Mathematical Systems Theory*, 10:239–251, 1977.

[7] S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 6:169–184, 1979.

[8] M. Davis, G. Logemann, and D. Loveland. A machine program for theorem proving. *Communications of the Association for Computing Machinery*, 5:394–397, 1962.

[9] A. Goerdt. Davis-Putnam resolution versus unrestricted resolution. *Annals of Mathematics and Artificial Intelligence*, 6:1–3, 1992.

[10] A. Goerdt. Regular resolution versus unrestricted resolution. *SIAM Journal of Computing*, 22:661–683, 1993.

[11] W. Huang and X. Yu. A DNF without regular shortest consensus path. *SIAM Journal on Computing*, 16:836–840, 1987.

[12] J. Krajíček. *Bounded Arithmetic, Propositional Logic and Complexity Theory.* Cambridge University Press, 1995.

[13] B. Krishnamurthy. Short proofs for tricky formulas. *Acta Informatica*, 22:253–274, 1985.

[14] R. Raz and P. McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19:403–435, 1999. Preliminary Version in: *Proc. 38th Symposium on Foundations of Computer Science*, 1997.

[15] G. Tseitin. On the complexity of derivation in propositional calculus. In A. O. Slisenko, editor, *Studies in Constructive Mathematics and Mathematical Logic, Part 2*, pages 115–125. Consultants Bureau, New York, 1970.

[16] A. Urquhart. The complexity of propositional proofs. *The Bulletin of Symbolic Logic*, 1:425–467, 1995.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'