Time-Space Tradeoffs in Resolution: Superpolynomial Lower Bounds for Superlinear Space

Paul Beame[*]
University of Washington
Seattle, WA 98195
beame@cs.washington.edu

Chris Beck[†]
Princeton University
Princeton, NJ 08544
cbeck@princeton.edu

Russell Impagliazzo[‡]
Institute for Advanced Study
Princeton, NJ 08540
and the University of California
San Diego, CA 92093
russell@cs.ucsd.edu

ABSTRACT

We give the first time-space tradeoff lower bounds for Resolution proofs that apply to superlinear space. In particular, we show that there are formulas of size N that have Resolution refutations of space and size each roughly $N^{\log_2 N}$ (and like all formulas have Resolution refutations of space N) for which any Resolution refutation using space S and length T requires $T \geq (N^{0.58 \log_2 N}/S)^{\Omega(\log\log N/\log\log\log N)}$. By downward translation, a similar tradeoff applies to all smaller space bounds.

We also show somewhat stronger time-space tradeoff lower bounds for Regular Resolution, which are also the first to apply to superlinear space. Namely, for any space bound S at most $2^{o(N^{1/4})}$ there are formulas of size N, having clauses of width 4, that have Regular Resolution proofs of space S and slightly larger size $T = O(NS)$, but for which any Regular Resolution proof of space $S^{1-\epsilon}$ requires length $T^{\Omega(\log\log N/\log\log\log N)}$.

Categories and Subject Descriptors

F.0 [**Theory of Computation**]: General; F.2.3 [**Analysis of Algorithms and Problem Complexity**]: Tradeoffs between Complexity Measures

General Terms

Theory

Keywords

Proof Complexity, Time-Space Tradeoffs

1. INTRODUCTION

For many modern SAT solvers, memory use is as much a bottleneck as time. Earlier DPLL-based SAT-solvers used very little memory, just needing to keep track of the stack in recursion. However, a major reason for the improved performance of recent SAT solvers is the inclusion of clause learning [18], which adds a large number of derived clauses to the input clauses. Clause learning uses considerable extra space corresponding to the number (and size) of derived clauses that are active at any one time. This creates opportunities for finding smaller refutations of unsatisfiable formulas, but at the cost of potentially requiring huge amounts of memory. Balancing these two factors is an art that determines the success of the SAT-solver. (See e.g, [29] for a discussion of how one successful SAT-solver handles this.)

This raises the question of whether such tradeoffs between total time and memory is an inherrent limitation, or just an artifact of the known algorithms. To make this question precise while also being general enough to handle the wide variety of SAT-solving algorithms, we can use the well-known correspondence between these algorithms and the Resolution proof system. All of the DPLL-based SAT solvers, when run on unsatisfiable formulas, implicitly or explicitly find resolution refutations of the input formulas. Thus, the minimum size of such a refutation is a lower bound on the time taken by any such algorithm. This observation has been part of the motivation for proof complexity studies of resolution for many years. More recently, an analogous measure of the *space* of a resolution proof has been introduced, which lower bound the number of derived clauses that must be remembered throughout the algorithm, and hence bounds the memory requirements of resolution-based SAT solvers.

If we use a pure backtracking approach to SAT-solving, like many of the earlier generation of SAT-solvers did, the proofs generated have at most linear space. Thus, we cannot hope to prove a super-linear space lower bound in isolation. Our goal is instead to show that the small space of these proofs come at the expense of being substantially larger than proofs using more space, i.e., to give time-space tradeoffs for proof complexity. There has been substantial work devoted to understanding space in proof complexity, proving almost linear lower bounds on the space required, and giving sharp time-space tradeoffs for refutations. However, previous work always hit a barrier at linear space. Until this current paper, no known time-space trade-off lower bound for proof complexity was meaningful when the space allowed was greater than that of the input formula. Since SAT solvers have

[*] Research supported by NSF grants CCF-0830626, DMS-0835373, and by the Ellentuck Fund. Much of this research was done at the Institute for Advanced Study, Princeton.

[†] Research supported by NSF grant CCF-0832797.

[‡] Research supported by NSF grants DMS-0835373, CCF-0832797, and The Oswald Veblen Fund.

STOC'12, May 19–22, 2012, New York, New York, USA.

ample memory to hold the input formulas, this meant that such work was not of direct benefit to understanding time-memory trade-offs for SAT-solvers. Ben-Sasson (see [22, 21]) posed the question of whether such trade-offs existed as follows: do all CNF formulas have resolution proofs of linear space that are at most polynomially larger than the resolution proof length possible when space is unrestricted?

We give a strong negative answer to this question. We give explicit CNF tautologies of size N that have proofs of size $T_{UB} \leq N^{O(\log_2 N)}$ (and hence space at most T_{UB}), but when space is restricted to $S_{LB} \leq T_{UB}^{1/2}$, any resolution proof requires size $T_{LB} \geq (T_{UB})^{\Omega(\log\log N/\log\log\log N)}$. In other words, restricting memory to any polynomial in the input formula size has a super-polynomial cost in terms of time the SAT-solver will take. Our formulas are Tseitin graph tautologies for complete bipartite graphs connected along a path and so are totally explicit, as are the proofs in the upper bound on T_{UB}. In fact, [2] has shown that proofs of such tautologies are derivable by standard SAT-solving techniques in time proportional to T_{UB}, so the trade-off applies to actual SAT-solvers, not just the theoretical optimal SAT-solver.

We obtain an even stronger tradeoff for Regular Resolution, a subclass of Resolution proofs that includes most natural Resolution proofs. Using Tseitin formulas on a family of grid graphs, we show that for any integer m, there is a size $N = O(m^4)$ formula with unbounded space regular resolution proof size at most $T_{UB} = \mathit{poly}(m)2^m$ so that any space $S_{LB} \leq 2^{m(1-\Omega(1))}$ proof requires size $T_{LB} \geq (T_{UB})^{\Omega(\log\log m/\log\log\log m)}$. This is an improvement over the previous result in that space restrictions provably cost in terms of size even for (weak) exponential amounts of space. Furthermore, the lower bound holds for space almost equal to the upper bound on total size. Finally, the tautologies where we prove this lower bound are constant width.

These results show that restricting space can increase size by more than a polynomial amount. It seems possible that space restrictions could have an exponential cost; however, that would require a different kind of tautology than the ones we consider. The quasipolynomial time-space tradeoff lower bound we achieve for Resolution proofs of these Tseitin tautologies is qualitatively not far from optimal. In particular, the Tseitin tautologies to which our bounds apply have Regular Resolution refutations of space $O(\log T_{UB} \log n)$ and size $T_{UB}^{O(\log n)}$. A recent result [12] shows a general purpose resolution based SAT algorithm which finds this proof as well, answering a question of [2]; see also [4]. [12] also gives an algorithm for high space which finds a refutation matching our high space refutation – this is also achieved by the algorithm of [2].

1.1 History and Related Work

Much initial work on proof space concentrated on the space required by Resolution proofs as a parameter on its own. Early work [13] showed that every unsatisfiable formula has a Resolution refutation in which the total space required is at most the number of variables (plus one), and the focus moved to finding matching space lower bounds. Atserias and Dalmau [3] showed that Resolution space is at least Resolution *width*, the length of the longest clause in any Resolution proof, for which $\Omega(n)$ lower bounds were already known for many formulas because width lower bounds are the most widely used method for proving Resolution proof size lower bounds [10]. (Despite its utility, Resolution width alone can be far from characterizing Resolution space [19].)

Lately, there has been great theoretical interest in understanding the interplay between the resources of Resolution proof length (time) and space [22, 7], and in showing trade-offs between them [8]. For example, Nordström [20] shows that there are formulas with linear size proofs but for which reducing the total space used by a constant increases the length required to exponential[1]. At higher space levels, Ben-Sasson and Nordström [8] recently showed the existence of families of formulae that have linear size (and hence linear space) proofs, but which require exponentially large proofs when the space is constrained to be $O(n/\log n)$.

1.2 Overview of our techniques

Many of the earlier papers on bounds for sub-linear space proofs modified arguments taken from time-space tradeoffs for pebbling games. We also use arguments based on time-space tradeoffs for straight-line programs such as [27, 23]. In such a bound, one typically defines a notion of "progress", and divides up the program into intervals called "epochs". Then one shows that, starting from any given point in the program, on a random input, one is unlikely to make significant progress in a short amount of time. The conclusion is that either there are many values that are in memory at those points, or the length of epochs is large, giving a time-space trade-off.

We follow a similar strategy, combined with a "bottleneck counting" argument as introduced in [15]. We consider the time steps to be the derived clauses in order, and divide these into equal sized intervals, or *epochs*. We consider the sets of clauses in memory at the start and end of epochs. We use a simple measure of the complexity of clauses, akin to those in [15, 10], the number of vertices of the graph G associated with the input clauses of the Tseitin formula that are required to derive them. The input clauses have complexity 1, the final contradiction has complexity $|V(G)|$, and each resolution step can at most double the complexity. Thus, clauses of each approximate (to within factors of 2) complexity between 1 and $|V(G)|$ must occur in the proof. The current complexities of clauses in memory is our measure of progress.

We show that clauses of intermediate complexities involve many variables and are hence intuitively are "unlikely" to be false, and we show that the likelihoods of very different clauses are in some sense independent. Because of the small space, the clauses in memory at the frontiers of epochs are "bottlenecks" ; it is unlikely that clauses of many distinct complexities are in memory at these times. If this does not happen, we must make a great deal of progress during some epoch, in that the largest complexity of a clause in memory has increased substantially. We then apply the argument recursively to that epoch.

While this intuitive picture is the same for the two lower bounds, the techniques used to formalize them are quite different. For general resolution, we formalize "likelihood" by looking at the proof after a random restriction. We show that if we apply a restriction to a Tseitin formula on a suit-

[1] This was claimed in [16] which used a more complicated approach, but this has been retracted as has the claim that determining the minimum space required for Resolution refutations is PSPACE-hard.

able graph G, then the probability that a clause has medium complexity after the restriction is quite small, and the probability that k clauses simultaneously have different intermediate complexities is a $\Theta(k)$-th power of this small probability. We use this to show that, with high probability, there are not many distinct complexities in memory at the frontiers of epochs after the restriction. Thus, after the restriction, there is one epoch that contains clauses of many approximate complexities within a large interval. We then use a recursive version of the same argument to show that this is also unlikely.

In the case of regular resolution, we define a probabilistic process that works its way backwards from the contradiction to one of the input clauses. It always visits clauses of each approximate complexity. "Likelihood" is then the probability of being visited in this process. We show that this probability is small for intermediate complexity clauses, and, with a few exceptions, uncorrelated for clauses of very different complexities. A bottleneck counting argument then shows that it is unlikely that this process visits many different clauses of distinct complexities that are each in memory at the frontiers of epochs, and thus must almost always visit a large interval of complexities within some epoch. The argument is repeated recursively within the chosen epoch.

2. BASIC DEFINITIONS AND NOTATION

We consider Boolean formulas over a set of variables $X = \{x_1, ..x_n\}$. As usual, a literal is a Boolean variable or its negation, a clause is a disjunction of literals, and a CNF is a conjunction of clauses. We think of clauses as being specified by their sets of literals, and CNFs as specified by their sets of clauses. For a clause C, we write $vars(C)$ to be the set of variables appearing in C. The *width* $w(C)$ of a clause C is $|vars(C)|$ and the width of a set or sequence of clauses F, is the maximum width of clauses in F. The *size* of a CNF formula F is the total number of literal occurrences in the formula, i.e., $\sum_{C\in F} w(C)$.

One of the simplest and most widely studied propositional proof systems is resolution which operates with clauses and has one rule of inference, the resolution rule: $\frac{A\vee x \qquad B\vee \overline{x}}{A\vee B}$. We say that the variable x is *resolved* in this instance of the resolution rule. A *resolution refutation* of a CNF formula (a set of clauses) is a sequence of clauses ending in the unsatisfiable empty clause $\bot$, each of which is either a clause of from the formula (an "axiom") or follows from two previous clauses via the resolution rule. (The term *resolution proof* is used more generally to refer to any inference of this sort that may not necessarily result in $\bot$.) Every resolution proof naturally corresponds to a directed acyclic graph (DAG), known as the proof DAG, in which every clause derived via the resolution inference rule has a directed edge "backwards" from a derived clause to each of its antecedents. (Note that, formally, a resolution proof corresponds to one of possibly many topological sorts of its proof DAG.)

A resolution proof is *regular* if along each path in the proof DAG, each variable is resolved at most once. (Regular) resolution is *sound and complete* in that every CNF formula is unsatisfiable if and only if it has a regular resolution refutation.

The *size* or *length* of a resolution proof is the total number of clauses in the proof. The usual definition of the *space* (*clause space*) used by a resolution proof is the maximum number of clauses which need to be held in memory at any one time when carrying out the proof. Say that a clause is active at time step t if it has been derived before t but is still needed for an inference step to be made at time t or later. Then the clause space is the maximum number of clauses active at any time step. In the most straightforward model, the initial clauses are made available at $t = 0$. To consider sublinear space, this model is modified, and the input clauses become available only as needed, accessed by "axiom download" steps, and may be cleared from local memory and derived again later, in analogy with the usual off-line definition of Turing machine space. In our results, we will be considering superlinear space and hence it is unnecessary to treat the input clauses this way. Doing so does not increase the space bound by even a constant factor. We allow clauses to be derived multiple times in a proof, since that may allow fewer clauses to be active at any one time.

A *restriction* is a mapping $\rho : X \to \{0, 1, \star\}$. Restrictions on X can be identified with partial assignments on V by viewing unassigned inputs as being mapped to $\star$ and vice versa. We will use the two terms interchangeably, depending on the context. Given two partial assignment π and ρ that are *compatible*, i.e., they agree on $\mathrm{dom}(\pi) \cap \mathrm{dom}(\rho)$, we use $\rho \cup \pi$ to denote the partial assignment that applies both assignments. The restriction of a clause C by ρ, denoted by $C|_\rho$ is the clause obtained from C by setting the value of each $x \in \rho^{-1}(\{0, 1\})$ to $\rho(x)$, and leaving each $x \in \rho^{-1}(\star)$ as a variable. The restriction of a set of clauses is defined by restricting each one.

2.1 Tseitin Tautologies

Following [28] we use the following formula to represent the principle that every undirected graph has even total degree.

Definition 2.1. *Let $G = (V, E)$ be an undirected graph, and $\chi : V \to \{0, 1\}$ an* odd charge function*, i.e. one such that $\bigoplus_{v\in V} \chi(v)$ is odd. For each edge $e \in E$, we have a variable x_e. For $v \in V, \epsilon \in \{0, 1\}$, let*

$$PARITY_{v,\epsilon} := \bigwedge\{\bigvee_{e\sim v} x_e^{a(e)} | \bigoplus_{e}(a(e) \oplus 1) \not\equiv \epsilon\}$$

be the CNF representation of the parity constraint $\bigoplus_{e\sim v} x_e \equiv \epsilon$. The Tseitin tautology *on (G, χ) is defined by the conjunction*

$$\tau(G, \chi) := \bigwedge_{v} PARITY_{v,\chi(v)} \ .$$

Also, independent of whether or not G is connected and χ has odd charge, the resulting formula is known as a Tseitin formula. Frequently we will suppress reference to χ, since when G is connected, any two odd charge functions χ give essentially equivalent formulae.

By a simple counting argument, if charge function χ is odd, then the parity constraints cannot all be satisfied – this would correspond to an undirected graph such that the sum of its degrees is odd.

Observation 2.2. *When χ is odd, $\tau(G, \chi)$ is unsatisfiable.*

When the degree of the graph is d, each constraint can be written as a CNF formula of $O(2^d)$ clauses of width d. It is easy to see that the formula $\tau(G)$ has size $O(2^d|V|)$.

Definition 2.3. *Let $G = (V, E)$ be a graph. For $E' \subset E$ let $G|E'$ denote the graph $G = (V, E - E')$. When π is a partial assignment to E, we overload the notation $G|\pi$ to mean $G|dom(\pi)$. (Alternatively, if ρ is a restriction then $G|\rho$ denotes $G|\rho^{-1}(\{0,1\})$.)*

Definition 2.4. *If χ is a charge function on a graph $G = (V, E)$ and π is a partial assignment to E then let $\chi \oplus \pi$ denote the new charge function χ' on V given by $\chi'(v) = \chi(v) \oplus \bigoplus_{v \in e \in dom(\pi)} \pi(e)$.*

The following formalizes how restrictions affect Tseitin formulas.

Proposition 2.5. *If $\tau(G, \chi)$ is a Tseitin formula and ρ is a restriction on the edges of G, then $\tau(G, \chi)|_\rho$ is $\tau(G|\rho, \chi \oplus \rho)$, and the parity of χ and $\chi \oplus \rho$ are the same.*

2.2 Resolution refutations of Tseitin formulas

For the graphs we will consider, some generic resolution refutations of odd charged Tseitin formulas follow from bounds on their *cut width*. We give two such refutations. One uses large space and implies that our lower bound results yield tradeoffs of the form that restricting the space does increase the minimum proof length required. The other uses a very small amount of space (indeed, exponentially less) and shows that there is a quasi-polynomial upper limit on the kind of tradeoff one can prove for the formulas we consider.

Definition 2.6. *The* cut width *of a graph G is the smallest W such that there is a linear ordering of the vertices $v_1, \dots, v_n$ such that, for every $1 \le t \le n$, there are no more than W edges crossing the cut $(\{v_1, \dots, v_t\}, \{v_{t+1}, \dots, v_n\})$.*

Lemma 2.7. *Let G be a graph with n vertices, maximum degree d, and cut width W. Then there is a regular resolution refutation of a Tseitin tautology on G using $\le n \cdot 2^d \cdot 2^{(W-1)}$ resolution steps, and space $\le 2 \cdot 2^{(W-1)} + d$, plus space for the initial axioms.*

Note that the number of resolution steps is within a constant factor of the proof size.

This refutation described by this lemma works by maintaining clauses corresponding to parities of the edges in the cuts and slowly moving those cuts from one end to the other end of the graph. It mirrors a refutation for similar formulas given in [11]. A detailed proof is given in the appendix.

The following lemma shows that for the graphs we consider here, even radically restricted space can only increase the size required for Tseitin tautologies with small cut width by an $O(\log n)$ power.

Lemma 2.8. *Under the same conditions as Lemma 2.7, the Tseitin tautology on $\tau(G)$ has a tree-like Resolution refutation using space $W\lceil \log n \rceil$ and $2^{W\lceil \log n \rceil}$ resolution steps.*

The refutation in this case involves simulating a binary search for a charge violation over the vertex ordering that achieves the cut width. At each cut in the search, the proof maintains all the clauses in the parities of the edges crossing that cut. A detailed proof is given in the appendix.

Though the graphs we consider all have small cutwidth, there are versions of both these upper bounds that can be expressed in terms of a smaller and more precise parameter, *carving width*. Moreover, some of these refutations can also be found algorithmically with similar complexity. Details are in the appendix.

2.3 A measure of complexity of clauses

First we need some preliminary definitions.

Definition 2.9. *Consider assignments to the variables of a Tseitin formula τ. A* critical assignment *is a total assignment that violates exactly one clause of τ. The vertex associated with that constraint is said to be its* bad vertex*. For any partial assignment π, we denote by $Crit(\pi)$ the set of critical assignments extending π.*

Definition 2.10. *Given a partial assignment π to a Tseitin formula τ over graph G, we say that a vertex v is a* critical vertex for π *iff v is bad for some critical assignment to the edges of G consistent with π. We define $crit_\tau(\pi)$ to be the set of critical vertices for π. We drop the subscript τ when it is understood from the context. We also define critical vertices for clauses by letting $crit(C)$ be the set of critical vertices of the partial assignment associated with $\neg C$.*

Fortunately, for Tseitin tautologies, the set $crit(\pi)$ has a nice, well-known characterization:

Proposition 2.11. *$crit(\pi)$ is non-empty if and only if $G|\pi$ has exactly one component of odd charge, in which case $crit(\pi)$ is equal to that component.*

Let S be a subset of nodes. We write $\delta(S)$ for the edges on the boundary of set S.

Corollary 2.12. *If π is a partial assignment with $crit(\pi) = S$, then every edge on the boundary of S is assigned by π.*

Finally, we note how partial assignments impact critical sets.

Proposition 2.13. *Let $\tau(G, \chi)$ be a Tseitin formula and ρ be a restriction (partial assignment) on the edges of G. Then for any partial assignment π compatible with ρ, $crit_{\tau(G,\chi)^\rho}(\pi) = crit_{\tau(G,\chi)}(\pi \cup \rho)$.*

Proof. This follows immediately from the fact that the set of critical assignments for $\tau(G, \chi)$ that are consistent with ρ is precisely the set of critical assignments for $\tau(G, \chi)^\rho$ with assignment ρ appended. □

We use $|crit(C)|$ as a measure of the complexity of clause C.

Proposition 2.14. *Fix any odd-charged Tseitin formula τ over a connected graph G.*

(a) $|crit_\tau(\bot)| = |V(G)|$ where $\bot$ is the empty clause.

(b) For any clause C of τ, $|crit_\tau(C)| = 1$.

(c) If clause C follows from clauses C_1, C_2 by resolution then $|crit_\tau(C)| \le |crit_\tau(C_1)| + |crit_\tau(C_2)|$.

Proof. Part (a) and (b) follow immediately from Proposition 2.11. Part (c) follows from the soundness of resolution since any critical assignment that falsifies C must falsify either C_1 or C_2, hence $Crit(C) \subseteq Crit(C_1) \cup Crit(C_2)$. □

Subadditivity of $|crit(C)|$ from Proposition 2.14(c), immediately implies the following.

Corollary 2.15. *For any t, any resolution derivation of a clause D with $|crit(D)| > 2t$ from a collection of clauses $\mathcal{C}$ having $|crit(\mathcal{C})| \leq t$, must contain a clause C' such that $t < |crit(C')| \leq 2t$.*

Proof. If not, then at the first point where such a D is derived, it must be derived from two clauses with complexity $\leq t$, contradicting subadditivity. □

3. TIME-SPACE TRADEOFF FOR GENERAL RESOLUTION

To show lower bounds in General Resolution, we consider Tseitin tautologies on a graph $G = (V, E)$ that is a path of length ℓ of complete bipartite graphs $K_{n,n}$, where the left side of one is the right side of the next. Equivalently, this is the tensor product of the complete graph K_n with the path graph P_ℓ. For $c > 0$ an absolute constant, our results hold for any $2^{cn} > \ell > n^4$, but are most impressive for larger ℓ. We picture the vertices of this graph as lying in an $n \times \ell$ grid, with each column corresponding to a vertex of P_ℓ. We let V_i denote the set of vertices in the i-th column, and E_i denote the edges between the i-th column and the $(i+1)$-st column.

Proposition 3.1. *Any Tseitin tautology over $G = K_n \otimes P_\ell$ has (regular) resolution size at most $O(2^{n^2+2n} n\ell)$.*

Proof. It is easy to see that G has cut width n^2 so this is immediate from Lemma 2.7. □

The main result of this section is the following time-space tradeoff lower bound for general resolution.

Theorem 3.2. *If n is sufficiently large and $2n^4 \leq \ell < (4n)^{-1}\left(\frac{9}{8}\right)^n$ for any resolution refutation of a Tseitin tautology on the graph $G = K_n \otimes P_\ell$ of size T and space S*

$$T \geq \left(\frac{2^{0.58n^2}}{S}\right)^{\frac{\log_2 L}{\log_2 \log_2 L}}$$

where $L = \log_2(\ell/(2n^3))$.

Corollary 3.3. *There is a $c > 0$ such that for sufficiently large N and any S satisfying $N \leq S \leq N^{\frac{1}{9}\log_2 N}$, there is a CNF formula of size N that has a resolution refutation of size at most $T_{UB} \leq N \cdot S$ and space at most S such that any resolution refutation using space $S^{1/2}$ requires size $T_{LB} \geq T_{UB}^{c \log_2 \log_2 N / \log_2 \log_2 \log_2 N}$.*

Proof. Choose the largest n such that $2^{n^2} \leq S$. Since $N \leq S \leq N^{\frac{1}{9}\log_2 N}$, we must have $2^{3n} \leq N \leq 2^{n^2}$ and hence $log_2 \log_2 N$ is $\Theta(\log n)$. Set $\ell = 2^{n/8} < (9/8)^n/(4n)$. With this value of ℓ, L is at least $n/9$ for N sufficiently large and hence $\log_2 L / \log_2 \log_2 L$ is $\Theta(\log\log N / \log\log\log N)$. Since the Tseitin formula for $K_n \otimes P_\ell$ has size at most $\ell n 2^{2n} < N$. We can pad this formula by dummy clauses until it has N clauses. Applying the above theorem to this formula with space $S^{1/2}$ instead of S yields the claimed lower bound. □

We use the "progress argument" sketched in the introduction, together with a bottleneck counting based on random restrictions. One subtle point is that, although our argument is applied recursively, we only apply the random restriction once, not after each step of the recursion.

We use the complexity measure on clauses defined in the previous section, which is 1 for the input clauses, is $|V(G)|$ for the final contradiction $\perp$, and grows slowly throughout a resolution proof (possibly not monotonically). We define many different medium complexity levels of clauses.

Definition 3.4. *Let $t_0 = n^4$. Define $\mathcal{L}_i = \{C \in \Pi : 2^i t_0 < |crit(C)| \leq 2^{i+1} t_0\}$ for $0 \leq i < \log_2(\ell n/t_0) - 1$. Each $\mathcal{L}_i$ represents a different* complexity level. *We say that a clause of* Π *has* medium complexity *if it is in one of the $\mathcal{L}_i$.*

We will use the fact that the restriction of a proof of a Tseitin tautology is itself a proof of a Tseitin tautology on a subgraph of the original graph; hence the proof for a graph G contains proofs of Tseitin tautologies for all of the subgraphs of G. We choose such a restriction randomly from a suitable distribution that is overwhelmingly likely to keep the complexity of the final contradiction high.

Following standard notation, let $\mathcal{R}_{1/3}$ be the probability distribution on restrictions $\rho : E \to \{0, 1, *\}$ such that for each e independently, $\Pr_{\rho\sim R_{1/3}}[\rho(e) = 0] = \Pr_{\rho\sim R_{1/3}}[\rho(e) = 1] = \Pr_{\rho\sim R_{1/3}}[\rho(e) = *] = 1/3$.

After we apply a random $\rho \sim R_{1/3}$ to a refutation of $\tau(G, \chi)$, we get a refutation of the Tseitin formula on the smaller graph $G|\rho$ with properties as in Proposition 2.5. We will measure the progress of the proof using the medium complexity levels of its clauses. We therefore will compute bounds on the probability that a clause becomes a medium complexity clause for the Tseitin tautology on the graph $G|\rho$ after such a restriction; we will do the same for the probability that a set of clauses have different medium complexity levels after restriction.

The graph $G|\rho$ and the charge $\chi \oplus \rho$ that results from applying the restriction ρ to $\tau(G, \chi)$ is not fixed. Therefore, it is easier to analyze the critical sets and complexity of the restricted clauses using Proposition 2.13 which says that for any partial assignment π compatible with ρ, $crit_{\tau(G,\chi)^\rho}(\pi) = crit_{\tau(G,\chi)}(\pi \cup \rho)$. For any clause C over G, this means that

$$crit_{\tau(G,\chi)^\rho}(C|_\rho) = crit_{\tau(G,\chi)}(C \vee \overline{\rho})$$

where $\overline{\rho}$ is the clause that is falsified precisely on those assignments consistent with ρ. This allows us simply to work with clauses $C \vee \overline{\rho}$ defined over the original graph G, so for the remainder of this section we write *crit* for $crit_{\tau(G,\chi)}$.

Now, we view a refutation of total size T and space S as ordered into epochs and sub-epochs as follows: Let L be the number of different medium complexity levels, $L = \log_2((n\ell)/(2t_0)) \geq \log_2 n$. Let $k = \lfloor \log_2 L \rfloor - 1$ and $h = \lfloor \log_2 L / \log_2 \log_2 L \rfloor < (k+2)/\log_2 k$. With these values we have $(k+1)k^{h-1} < L$. Now choose

$$m \leq \left\lfloor \frac{(3/2)^{n^2} \ell^{-1} L^{-1} 2^{-4n}}{S} \right\rfloor^{1-3/\log_2 k}.$$

One can easily verify that $m \geq \max(8, 2^{0.58n^2}/S)$ for sufficiently large n.

The theorem follows immediately if $T \geq m^h$ so assume, by contradiction, that $T < m^h$. An *epoch at the leaf level* is an interval of m consecutive clauses in the proof, beginning at a time step that is a multiple of m. We consider all clauses in the leaf level epoch to be *frontier* clauses. An *epoch at level* $2 \leq i \leq k$ is an interval of m^i consecutive clauses in the proof, beginning at a time step that is a multiple of m^i.

Its sub-epochs are the epochs of level $i-1$ within the epoch, and the *frontier* clauses for the epoch are the ones that are active at the end of any of its sub-epochs. Thus, leaf level epochs have m frontier clauses and higher level epochs have at most mS frontier clauses, since at most S clauses are active at each of the m times when a sub-epoch ends.

LEMMA 3.5. *For any refutation of a Tseitin tautology on a connected graph with $n\ell$ nodes, at least one epoch has k frontier clauses of distinct medium complexity levels.*

PROOF. Assume that no such epoch exists, at any level of epochs. We will inductively find an epoch E_j at level $k-j$ and medium complexity levels i_{min} and $i_{max} \geq i_{min} + L/k^j$ so that all active clauses at the start of E have complexity at most that of medium complexity level i_{min} and some clause at the end of E has complexity at least that of complexity level i_{max}.

At the start, we can choose E_0 to be the entire proof, since at the beginning, we have only clauses of complexity 1 (input clauses) and at the end, we have only the empty clause, $\perp$, of complexity $n\ell$.

Assume that we have E_j, i_{min} and i_{max} as above. Consider the partition of E_j into m sub-epochs at level $k-j-1$. Because there are at most $k-1$ distinct complexities among E_j's frontier clauses, there must be a sub-interval i'_{min} to i'_{max} of length at least $(i_{max} - i_{min})/k \geq L/k^{j+1}$ where no frontier clause is of medium complexity level between i'_{min} and i'_{max}. Consider the first sub-epoch of E_j that ends with an active clause of medium complexity level at least i'_{max}; since E_j ends with a clause of medium complexity level $i_{max} \geq i'_{max}$, such a sub-epoch must exist. Since no clause of medium complexity level at least i'_{max} was active at the start of this sub-epoch, and no clause of medium complexity level between i'_{min} and i'_{max} is active at the start of any sub-epoch, we can choose this first sub-epoch as E_{j+1}.

Inductively, we get a leaf-level epoch E_{h-1} and an interval i_{min} to i_{max} of length at least $L/k^{h-1} \geq k+1$ so that all active clauses at its start have complexity at most that of medium complexity level i_{min} and at least one clause has medium complexity level i_{max} or greater at the end. Then clauses of all k medium complexity levels between i_{min} and i_{max} appear in the leaf-level epoch. Since all lines in a leaf-level epoch are frontier clauses, this is a contradiction. □

To use the above to get a contradiction, we show that, for $G = K_n \otimes P_\ell$s, the graph is very likely to stay connected after a random restriction, and any small set of clauses is very unlikely to contain many different medium complexity levels of clauses after the restriction. We then apply the above lemma to the restricted proof.

The following lemmas, proved in the next section, summarize our bounds on these probabilities.

DEFINITION 3.6. *Let $CONN(\rho)$ denote the event that each bipartite graph $(V_i, V_{i+1}, E_i \setminus dom(\rho))$ is connected.*

PROPOSITION 3.7. $\Pr_\rho[\neg CONN(\rho)] \leq \ell \cdot 2(n-1)(8/9)^n$.

LEMMA 3.8. *For $k \leq 1/3L$, any k clauses $C_1 \dots C_k$, and any k medium complexity levels $\ell_1, \dots, \ell_k$ with $\ell_{i+1} \geq \ell_i + 3$ for $1 \leq i \leq k-1$,*

$$\Pr_{\rho \sim \mathcal{R}_{1/3}}[CONN(\rho) \wedge \forall i, (C_i \vee \overline{\rho}) \in \mathcal{L}_{\ell_i}] \leq \left(\ell \cdot 2^{4n} \cdot (2/3)^{n^2}\right)^k$$

COROLLARY 3.9. *Let $\mathcal{C}$ be a collection of $\leq M$ clauses. The probability that $\mathcal{C}|_\rho$ contains k clauses of distinct medium complexity levels is at most $\left(ML(2/3)^{n^2}\ell 2^{4n}\right)^{\lceil k/3 \rceil}$.*

PROOF. If there are k distinct medium complexity levels, a subset of $\lceil k/3 \rceil$ of them are mutually separated by 3. There are at most $L^{\lceil k/3 \rceil}$ choices for this subset of medimum complexity levels $\ell_1, .., \ell_{\lceil k/3 \rceil}$, and for each element of the subset, at most M choices for a clause $C_i \in \mathcal{C}$ to become this complexity. Since the complexity of $C_i|_\rho$ is the same as that of $C_i \vee \overline{\rho}$ we can apply the previous lemma with a union bound to get the probability bound in the corollary. □

PROOF OF THEOREM 3.2. Assume that m, k, and h are defined as above and that there is a refutation of the Tseitin tautology on $K_n \otimes P_\ell$ of space S and size $T < m^h$. Under any restriction ρ, either $G|\rho$ is disconnected, or after the restriction ρ some epoch in the proof contains frontier clauses of k distinct medium complexity levels. By proposition 3.7, the first probability is $< 1/2$. There are fewer than $2m^{h-1}$ epochs in all, and each has a frontier set of size at most $M = mS$. Thus, from the above corollary, the probability that there is an epoch that has k distinct medium complexity levels among the restriction of its frontier clauses is at most

$$\begin{aligned}
&2m^{h-1}\left(mSL(2/3)^{n^2}\ell 2^{4n}\right)^{\lceil k/3 \rceil} \\
&< \frac{1}{2}\left(m^{1+3/\log_2 k} SL(2/3)^{n^2}\ell 2^{4n}\right)^{\lceil k/3 \rceil} \\
&\qquad\qquad \text{since } m \geq 8 \text{ and } h < (k+2)/\log_2 k \\
&\leq \frac{1}{2}\left[\left(\frac{(3/2)^{n^2}\ell^{-1}L^{-1}2^{-4n}}{S}\right)^{1-9/\log_2^2 k} SL(2/3)^{n^2}\ell 2^{4n}\right]^{\lceil k/3 \rceil} \\
&\leq \ \frac{1}{2}.
\end{aligned}$$

This is a contradiction to Lemma 3.5 and the theorem follows. □

3.1 Medium complexity clauses and random restrictions

In this section, we prove the lemmas about the effect of random restrictions on Tseitin tautology clauses. We begin with the proof of Proposition 3.7 that the graph remains connected after a random restriction is applied.

PROOF OF PROPOSITION 3.7. We apply a union bound over the columns and compute an explicit bound for the well-known connectivity properties of random bipartite graphs $G(n, n, \frac{1}{3})$.

Consider the left side of $G(n, n, 1/3)$. For any two vertices here, the probability that they are not connected with a two-step path in $G(n, n, 1/3)$ is at most $(8/9)^n$, since there are n possible disjoint two-step paths. By a union bound, the left side and right side are both connected to themselves except with probability $2(n-1)(8/9)^n$. Since these imply that an edge exists, the $G(n, n, 1/3)$ is connected except with this probability. □

Note that the condition $\ell < (4n)^{-1}(9/8)^n$ implies that $\Pr_\rho[\neg CONN(\rho)] < 1/2$.

LEMMA 3.10. *For every clause C and every fixed set of edges $E' \subseteq E$, the probability over $\rho \sim \mathcal{R}_{1/3}$ that $C \vee \overline{\rho}$ is non-trivial and $E' \subseteq vars(C \vee \overline{\rho})$ is at most $(2/3)^{|E'|}$.*

Proof. For $C \vee \overline{\rho}$ to be non-trivial, ρ must not set any edge to falsify C. In particular, this applies to each edge in E' that appears in $vars(C)$. In order for $E' \subseteq vars(C \vee \overline{\rho})$, each edge in E' not in $vars(C)$ must be set by ρ. For each edge in E', these successes happen with probability 2/3 and the probabilities are independent, so the total probability that $E' \subseteq \delta(crit(C \vee \overline{\rho}))$ is at most $(2/3)^{|E'|}$. □

To prove a non-trivial tradeoff lower bound using our outline above, we need a stronger lower bound than the upper bound for unrestricted space of roughly 2^{n^2} given in Proposition 3.1. Therefore, we need to be able to argue that after such a restriction the probability that a clause will become a medium complexity clause is exponentially small in n^2.

Intuitively, a set of vertices S of size at least n^2 and less than $\ell n/2$ will have a boundary of size at least n^2 for the following reasons. If it has as many as n "partial columns" in which it contains some, but not all, of the vertices, then each such column will contribute at least n boundary edges. If it has fewer than n partial columns, it must contain a full column and an empty column, and its not hard to see that deleting the elements in the partial columns cannot decrease the size of the boundary, so this configuration has a boundary of at least n^2 as well.

By Corollary 2.12, for every clause C, $crit(C \vee \overline{\rho}) = S \subset V$ only if $\delta(S) \subseteq vars(C \vee \overline{\rho})$. Therefore, by Lemma 3.10, $\Pr_{\rho \sim \mathcal{R}_{1/3}}[crit(C \vee \overline{\rho}) = S] \leq (2/3)^{n^2}$ for any C and any S of size between n^2 and $\ell n/2$. Though the bound for a single S is of the form we want, the number of such S is huge, so a simple union bound over all choices of S is insufficient to derive the bound we need. Nonetheless, we can show a bound very close to this one for the probability that $crit(C \vee \overline{\rho})$ is *any* medium sized set. We also generalize it to show bounds on the probability that, after applying ρ, k clauses appear from k different $\mathcal{L}_i$

Definition 3.11. *Let S denote a connected subset of the vertices V. Define $S_i := S \cap V_i$ and $s_i := |S_i|/|V_i|$. We say that*

- *i is a* full column *of S if $s_i = 1$, an* empty column *if $s_i = 0$, and a* partial column *otherwise;*
- *i is a* transition point *of S if $i-1, i, i+1, i+2 \in \{1, \ldots, \ell\}$ are columns, and either both $s_i \geq 1/2$ and $s_{i+1} \leq 1/2$, or both $s_i \leq 1/2$ and $s_{i+1} \geq 1/2$;*
- *S has a* transition point with signature *(i, A_0, A_1, A_2, A_3) if i is a transition point of S, $S_{i-1} = A_0$, $S_i = A_1$, $S_{i+1} = A_2$, and $S_{i+2} = A_3$.*

The transition point concept is useful because of the following lemma which says that the existence of a transition point for a set S implies that S has a large boundary.

Lemma 3.12 (Transition Point Lemma). *If i is a transition point of S, then $|\delta(S)| \geq n^2$. Moreover, for any signature (i, A_0, A_1, A_2, A_3), there exists a set of at least n^2 edges, $E^* \subseteq E_{i-1} \cup E_i \cup E_{i+1}$, depending only on (i, A_0, A_1, A_2, A_3), such that if S has a transition point with signature (i, A_0, A_1, A_2, A_3) then $E^* \subseteq \delta(S)$.*

There are only $\ell \cdot 2^{O(n)}$ possible signatures, but, by this lemma, the probability that $crit(C \vee \overline{\rho})$ has a transition point with a specific signature is at most $2^{-\Omega(n^2)}$ for any particular C, so this will allow us to obtain a very strong upper bound on the probability that C has any transition point at all.

Proof of Transition Point Lemma. Let S be a connected set of vertices and have a transition point with signature (i, A, B, C, D). We would like to find, from this information only, a set of E^* edges on the boundary of S. We write $a = |A|/n$, $b = |B|/n$, $c = |C|/n$ and $d = |D|/n$. Without loss of generality we can assume that $b \geq 1/2, c \leq 1/2$.

We will find a set of n^2 edges in the columns E^{i-1}, E^i, E^{i+1} on the boundary of any S with this signature. Since the signature determines exactly which edges in these columns are boundary edges, and the graph between each pair of layers is the complete bipartite graph, only the numbers (a, b, c, d) matter. The number of boundary edges may be computed exactly as

$$n^2(a(1-b)+b(1-a)+b(1-c)+c(1-b)+c(1-d)+d(1-c)) .$$

However,

$$\begin{aligned} a(1-b)+b(1-a) &= a(1-2b)+b \\ &= 1-b+(2b-1)(1-a) \\ &\geq 1-b \qquad \text{since } b \geq 1/2 \end{aligned}$$

and

$$\begin{aligned} c(1-d)+d(1-c) &= c+d(1-2c) \\ &\geq c \qquad \text{since } c \leq 1/2. \end{aligned}$$

So, the expression above is at least

$$\begin{aligned} n^2(1-b+b(1-c)+c(1-b)+c) &= n^2(1+2c-2bc) \\ &= n^2+n^2 2c(1-b) \\ &\geq n^2 \end{aligned}$$

as claimed. □

In the informal argument that medium sized sets of vertices have at least n^2 edges on their boundaries, we identified two cases to handle – those with many partial columns and those with few. If the set is in the right size range and has few partial columns, it can be shown to have both a full column and an empty column, and it is easy to see that such sets must have a transition point somewhere between the full and empty column. We handle the other case, with many partial columns, by showing that it holds conditioned on the graph being connected.

Lemma 3.13. *For any clause C and restriction ρ, if $CONN(\rho)$ holds then $|C|$ is at least the number of partial columns in $crit(C \vee \overline{\rho})$.*

Proof. Let $S = crit(C \vee \overline{\rho})$. If i is a partial column of S, then both $S_i \cup S_{i+1}$ and its complement in $V_i \cup V_{i+1}$ are nonempty. If $CONN(\rho)$ holds, then these two sets of vertices are adjacent, and in fact there is an edge between them that is not assigned by ρ. By Corollary 2.12, all edges on the boundary of the critical set must be assigned, so C must assign some edge in E_i. Thus C contains at least one edge for each partial column of S. □

Now we will show our bound for the probability that multiple clauses come to occupy many separated medium complexity levels. We will use the separation in complexity levels to argue that there must be many transition points for these clauses rather than just one. This will be based on bounding the number of distinct endpoints of sets of intervals of increasing size on a line or circle for which we will use the following lemma proved in Section 4.

LEMMA 3.14. *Suppose that we have k points of the n point circle $a_i \in Z_n$, and k natural numbers $d_i \leq n/2$, determining k intervals on the circle $(a_i, b_i = a_i + d_i)$ each of length d_i. Let U denote the set of endpoints of these intervals; i.e., $U := \bigcup_i \{a_i, b_i\}$. It follows that*

1. *if $\forall i,\ d_{i+1} \geq 2d_i$, and $1 \leq d_1, d_k \leq n/2$ then $|U| \geq k+1$.*
2. *Further, if for some a, $k \cdot a < d_1$, then there exists a subset U' of U such that $|U'| \geq k+1$ and any two points of U' are at distance greater than a from each other.*

The same argument we give can also be used to prove a more general statement, but we limit ourselves to the following for simplicity.

PROOF OF LEMMA 3.8. If any C_i is as large as kn^2, it has probability at most $(2/3)^{kn^2}$ to survive the restriction and we are done, so suppose that this is not the case. Using Lemma 3.13, we can restrict attention to candidate critical sets S^i that have fewer than $kn^2 \leq n^3/3$ partial columns, containing fewer than $n^4/3$ vertices. Because $t_0 = n^4$, for each S^i there must be many full columns. There are also many empty columns since the largest critical sets in medium complexity levels have size at most $\ell n/2$.

So, we can already see that each S^i has a transition point. What we would like to see is that there must be at least k transition points among all of them, and that these transition points are all far apart from each other. We use Lemma 3.14 for this.

For each S^i, let a_i denote the first column from the left so that $|S^i_{a_i}| \geq n/2$, and let b_i similarly denote the first such column from the right. So long as b_i is not one of $\{1, 2, \ell-1, \ell\}$, it is a transition point, and the same is true for $a_i - 1$. So, it suffices to prove that $\left|\bigcup_i \{a_i - 1, b_i\} \setminus \{1, 2, \ell-1, \ell\}\right| \geq k$. In fact it suffices to meet the requirements for the strong form of Lemma 3.14 with $a \geq 3$, since then at most one point of U' is in the set $\{1, 2, \ell-1, \ell\}$.

Let d_i denote the difference $b_i - (a_i - 1)$. Since C_i has medium complexity level ℓ_i, d_i is at least the number of full columns, i.e., $d_i \geq 2^{\ell_i} t_0/n - n^3/3$. We can also upper bound the total number of full and partial columns as $2^{\ell_i+1} t_0/n + n^3/3$, which is an upper bound on d_i, since S^i is connected. Since $\ell_{i+1} \geq \ell + 3$ and $t_0 \geq n^3$, we have $2^{\ell_{i+1}} \geq 4 \cdot 2^{\ell_i+1}$ and

$$\frac{d_{i+1}}{d_i} \geq \frac{4 - \frac{1}{3}}{1 + \frac{1}{3}} > 2 \ .$$

Further, $d_k \leq \ell n/4 + n^3/3 < \ell n/2$, as required, and $d_1 \geq \frac{2}{3}n^3$, while $k < n/3$ implies $3k+1 < n+1 << d_1$. Therefore, Lemma 3.14 may be applied to say that there are k distinct transition points that are far apart.

If we fix a particular sequence of signatures for these transition points then the Transition Point Lemma implies that there is a fixed set E' of kn^2 edges, which depends only on the signatures, contained in $\bigcup_{i=1}^{k} \delta(C_i \vee \overline{\rho})$, consisting of the union of the k disjoint sets of edges for each transition point. As in the proof of Lemma 3.10, any such edge must either be set by ρ, or appear in some C_i and hence cannot be set by ρ to violate that C_i. Each such event occurs with probability at most $2/3$ and therefore, by the independence over edges, the total probability is at most $(2/3)^{kn^2}$. The number of sequences of k signatures is at most $(\ell \cdot 2^{4n})^k$ and the bound follows. □

4. ENDPOINTS OF INTERVALS

This section is devoted to the proof of Lemma 3.14.

Suppose that we have a collection of k intervals on the line. How many distinct endpoints are there? If nothing else is known about the intervals, there can be as few as $\mathbf{O}(\sqrt{k})$ endpoints. However, if each interval is at least twice as long as the interval before it, then intuitively the intervals cannot be packed together so nicely; we will see that under this assumption there are at least $k+1$ distinct endpoints.

Although we only care about the line, in Lemma 3.14 it is more convenient to analyze the case of the circle.

PROOF OF LEMMA 3.14. Suppose that we have k points of the n point circle $a_i \in Z_n$, and k natural numbers $d_i \leq n/2$, determining k intervals on the circle $(a_i, b_i = a_i + d_i)$ each of length d_i. Let U denote the set of endpoints

$$U := \bigcup_i \{a_i, b_i\} \ .$$

Suppose that $d_{i+1} \geq 2d_i$ for all i and $1 \leq d_1, d_k \leq n/2$. Let us define a graph G on vertex set is U, and for each i, an edge from a_i to b_i. Thus there are k total edges. For the first claim, the strategy is to show that this graph is a forest, and hence has at least $k+1$ vertices. To show the second claim, we will show a process which constructs an acceptable set of $k+1$ vertices.

We will still use the terms "distance" and "displacement" to refer to distances in the circle.

Consider any walk in G which does not use any edges more than once. (Note that this is not the same as a path, as it may include cycles.) The total displacement as we move along from vertex to vertex can be written

$$\sum_{i \in S} \pm d_i$$

where S is the set of indices corresponding to edges appearing on the walk.

By induction on $|S|$, we see that for any nonempty S

$$\sum_{i \in S, i < \max S} d_i \leq d_{\max S} - d_1.$$

The base case is trivial, and if the claim holds for S, we may add $d_{\max S}$ to both sides, and we can apply the assumption $2d_{\max S} \leq d_{\max S+1} \leq d_{s'}$ for any $s' > \max S$, and so deduce the claim for $S \cup \{s'\}$, thus the inductive hypothesis also holds.

Suppose S is nonempty. By the triangle inequality,

$$d_1 \leq \left|\sum_{i \in S} \pm d_i\right| \leq 2d_{\max S} - d_1 \leq n - d_1 \ ,$$

using the $d_k \leq n/2$ assumption.

Now we can deduce the first claim. Suppose the graph contains a cycle. Then there is a walk in G as before which starts and ends at the same vertex, thus the total displacement is $0 \mod n$. But since the $1 \leq d_1$, the above inequality contradicts this. Thus G contains no cycles and hence is a forest with k edges – this implies there are at least $k+1$ vertices, or $|U| \geq k+1$ as desired.

To see the second claim, we will show a greedy algorithm to find U'.

1. Begin with $U' := \emptyset$. We will process the components of the graph in some order. The first one we can just add entirely to U'.

2. While a component we have not yet processed contains some vertex v within a of anything in U', choose such a v and add every other vertex in that component C_v to U'.

3. If all of the remaining components have every vertex at distance greater than a from U', call the algorithm recursively on the remaining unprocessed components.

Our first observation is that the algorithm omits at most one vertex from any component from U', and for at least one component omits none. Since a forest on k edges and c components has at least $k+c$ vertices, U' will always contain at least $k+c-(c-1) = k+1$ vertices at the end.

It suffices to show the algorithm never adds any two vertices at distance less than a to U'. Call such a pair of vertices an "unsafe pair". The idea is that when $d_1 > k \cdot a$, the argument before that G is cycle-free will hold even if we add as many as k unsafe pairs as edges to the graph G, since these new edges can only change the sum above by at most $k \cdot a$, and d_1 is large enough that we can tolerate that. Our strategy is to view the $(k-1)$ unsafe pairs identified in step 2 as edges which we add to G. For any vertices of U' which become connected by this, we will argue that the path connecting them contains at least one original edge of G and so the previous calculation shows their distance is at least a. For any vertices which were not connected by this, we will see that they could not have formed an unsafe pair – this can only happen when step 3 occurs, and we will handle this scenario as follows.

It suffices to assume that step (3) never happens. When step 3 happens, there are no unsafe pairs between the remaining components and the U' obtained from the components processed already. If the algorithm didn't add any unsafe pairs when run just on the first group of components, and also doesn't add any unsafe pairs when just run on the second group of components, it will succeed when run on all of the components, so we may break up the current instance of the problem into two subinstances and restrict attention to these subinstances. Appealing to this argument sufficiently many times, it suffices to show correctness in cases where step (3) never occurs. (To make this completely precise, one could remove the arbitrary choice in the definition of the algorithm by assigning a priority to each component, and specify that the algorithm deterministically chooses to process the highest priority component with an unsafe pair to U'. Then, when we split an instance on which step (3) occurs into two subinstances, the subinstances inherit a priority ordering, and the U' returned by the algorithm on the initial instance will be exactly the union of the U''s returned by the algorithm on the two subinstances, so correctness indeed follows from correctness of the runs on the two subinstances.)

Therefore without loss, step 1 occurs once and then step 2 occurs $k-1$ times. When step 2 occurs, by definition there is some vertex $u \in U'$ and some vertex v which is chosen so that uv is an unsafe pair, and component C_v is the next processed component. Let G' be the graph G, plus an edge for each such uv. The edge uv connects the component C_v to the previously processed components. When we start, G is a forest, and at the end, every component has been processed, so G' is a tree.

Now consider a walk in G' which does not use any edge more than once. As long as it contains at least one edge of G, by the triangle inequality the total displacement is between $d_1 - (k-1) \cdot a$ and $n - (d_1 - (k-1) \cdot a)$, so since $d_1 > k \cdot a$, the start and end point of such a walk do not form an unsafe pair.

The tree G' contains a path connecting every $u, u' \in U'$, so it suffices to see that this path contains at least one edge of G. What vertices are connected using only the edges of G' that don't appear in G? By construction this set of edges forms a collection of stars with vertices of U' at the centers, since for each unsafe pair found in step 2, the vertex v corresponds to a distinct component, and v may not be chosen as u at some later step since it is not added to U'. Thus for no two distinct points of U' does the unique path in G' consist only of unsafe pair edges – each such pair has a path in G' with at least one edge of G, so they do not form an unsafe pair. □

5. TIME-SPACE TRADEOFF FOR REGULAR RESOLUTION

For regular resolution, it is possible to use a fairly different argument to get super-polynomial size lower bounds for space that is polynomially large for Tseitin formulas on a family of constant-degree graphs (and hence constant clause size) in contrast to the large degree of the graphs we used in the general case. Though not qualitatively different, the tradeoffs themselves are slightly sharper than in the general case and also apply in the case of substantially larger space bounds (up to sub-exponential space rather than up to quasipolynomial space).

To achieve this, the random restriction argument from the case of general resolution is replaced with a random adversary argument and the analysis now takes a top-down flavor. This gives much more freedom in the choice of hard graphs since we no require connectivity under random restrictions that we needed in the general resolution case. We still must use graphs that are "long and skinny" as before. For simplicity and for applicability to a wider range of space bounds than for the graphs we used previously, we will choose the $n \times \ell$ grid graph, $G_{n,\ell}$ of n rows and ℓ columns.

We first note that the Tseitin tautologies over grid graphs have short regular resolution refutations.

PROPOSITION 5.1. *Any Tseitin tautology over $G_{n,\ell}$ has a regular resolution refutation size at most $32\ell n \cdot 2^n$ and space at most $2^{n+1} + 5 + |\{axioms\}|$, and $|\{axioms\}| \leq 8\ell n$.*

PROOF. $G_{n,\ell}$ has cut width $n+1$ and maximum degree 4 so this follows from Lemma 2.7. The number of axioms is easily counted as $\leq 2^{d-1}|V|$. □

By contrast, the main result that we will prove in this section is that if the space allowed is reduced significantly below 2^n, and ℓ is not too large then such a formula requires regular resolution refutations of superpolynomial size.

THEOREM 5.2. *For any ℓ such that $n^3 \leq \ell < 2^n$, any size T and space S regular resolution refutation of the Tseitin tautology on a n by ℓ grid must satisfy*

$$T \geq \left(\frac{2^{(1-o(1))n}}{S}\right)^{\frac{\log_2 L}{2\log_2\log_2 L}}$$

where $L = \log_2(\ell/(4n^2))$.

To prove Theorem 5.2, we will begin with a regular resolution refutation Π, and repeatedly subdivide Π into polynomially many epochs, at each point choosing an epoch in which enough "progress" happens that we may continue subdividing the epoch. The process peters out after $\mathbf{O}\left(\frac{\log L}{\log\log L}\right)$ subdivisions.

In the next three subsections we develop the main technical tools that allow us to derive this lower bound. In section 5.1 we modify the complexity measure for clauses from general resolution to a complexity measure specific to regular resolution. This measure has the advantage of growing monotonically with the inferences in the proof, which makes it a much more precise tool. We use this measure to define the complexity levels that are the indicators for progress through the proof. In section 5.2 we describe a probabilistic adversary that follows the inferences in the proof. One can think of this adversary as observing the progress of the proof. We show that the probability that adversary reaches a clause (or set of clauses) can be analyzed in terms of a potential function Φ applied to a set of edges associated with that clause (or set of clauses). Φ turns out to be the rank function of a certain matroid and we characterize several properties of Φ that we need in the overall argument. In section 5.3, we give fairly straightforward proofs that the potential Φ will be large for sets that correspond to clauses of many different medium complexity levels.

In section 5.4, we put the pieces together and give the inductive argument that yields Theorem 5.2. The base case shows simply that any sub-epoch in the proof in which the adversary has a reasonable probability of visiting clauses of many different complexity levels must have length exponential in the width n of the grid graph. The inductive step shows that if one divides an epoch of the proof through which the adversary passes into sub-epochs then either the adversary visits clauses of many different complexity levels at the boundaries between sub-epochs or the adversary sees a large fraction of the overall progress in complexity levels within a single sub-epoch. If the space is small then the first case is very unlikely; if the second case holds then we can subdivide that sub-epoch and apply the argument recursively.

5.1 Read-once branching programs and common information

It is well known (cf. [17]) that a resolution refutation yields a branching program for the "clause search problem" and this branching program is read-once if and only if the resolution refutation is regular. In this construction, we start with the proof DAG and labeled the edges by the following rule: If $C \vee D$ is derived by resolving $C \vee x$ with $D \vee \neg x$, then the node for clause $C \vee D$ is also labeled by a query to x, the edge directed from $C \vee D$ to $C \vee x$ is labeled $x = 0$, and the edge from $C \vee D$ to $D \vee \neg x$ is labeled $x = 1$. Like the proof DAG, the single-source, or root, of the branching program is the contradiction $\perp$ and the sinks, or leaves, are the input clauses or axioms. Note that paths p in this branching program correspond to partial assignments. We identify a regular resolution proof with both its DAG and its associated branching program. The following is immediate from the definition.

PROPOSITION 5.3. *Every node reached by an assignment starting at the root of the branching program for any regular resolution proof* Π *is labeled by a clause that is falsified by that assignment.*

For purposes of a top-down analysis, more relevant than the literals appearing in a clause (node) in a proof is a closely related concept we call its *common information.* A similar definition has appeared previously, e.g. in [24].

DEFINITION 5.4. *For C a clause in a regular resolution refutation Π, define the* (common) information at C in Π *to be a partial assignment $I_{C,\Pi}$ in terms of the proof Π as follows:*

If every directed path in Π from the contradiction $\perp$ to C contains the label $x = 1$, then $I_{C,\Pi}$ assigns $x = 1$; if every such path contains the label $x = 0$, then $I_{C,\Pi}$ assigns $x = 0$; Otherwise $I_{C,\Pi}$ does not assign x.

Note that, by definition, $I_{C,\Pi}$ assigns C to false since every literal in C must be made false on any path between C and $\perp$. The regularity assumption permits us to prove the following crucial consistency lemma.

LEMMA 5.5. *Let τ be a Tseitin tautology on graph G with regular resolution refutation Π. If a clause C is reachable from the root $\perp$ by a path p in Π consistent with a critical assignment, then $crit_\tau(I_{C,\Pi}) = crit_\tau(p)$, and p and $I_{C,\Pi}$ assign all edges incident to this critical set identically.*

PROOF. Let p be such a path from $\perp$ to C, consistent with a critical assignment σ with bad vertex b. Let e be any edge incident to b, and p' be any other path in Π from $\perp$ to C. Then we claim that p, p' assign e identically.

The assignment σ corresponds to a root-to-leaf path in Π extending p. Let q denote the portion of this path which occurs after it reaches C, so that pq is the entire root-to-leaf path. By the regularity assumption, $p'q$ is also consistent with some assignment, and by correctness of the proof, $p'q$ violates the axiom labeling the leaf that $p'q$ and pq both reach, which is also violated by pq. e is a variable of this axiom, so $p'q(e) = pq(e)$, and both do assign e. If p assigns e, then q does not assign e, so p' must assign e, and in the same way as p does. If p does not assign e, then q must, so p' cannot, by the regularity assumption.

By definition of $I_{C,\Pi}$, we conclude that $I_{C,\Pi}$ and p assign all edges incident to $crit_\tau(p)$ identically, and so $crit_\tau(I_{C,\Pi}) \subseteq crit_\tau(p)$ since $I_{C,\Pi}$ and p assign the boundary of $crit_\tau(p)$ in the same way and with the same parity. But since p also extends $I_{C,\Pi}$ then $crit_\tau(p) \subseteq crit_\tau(I_{C,\Pi})$, and so they are equal. □

For the rest of this section we will drop the subscript τ because there will be one fixed τ for our arguments. For regular resolution, our measure of complexity for a clause C will be $|crit(I_{\Pi,C})|$.

Definition 5.6. *Let $\mathcal{L}_i^* = \{C \in \Pi : 2^i n^3 < |crit(I_{\Pi,C})| \le 2^{i+1}n^3\}$ for $0 \le i \le \log_2(\ell/(4n^2)) - 1$. Each $\mathcal{L}_i^*$ represents a different complexity level. We say that a clause of* Π *has* medium complexity *if it is in one of the $\mathcal{L}_i^*$.*

The following is a key property that makes regular resolution refutations much each easier for us to analyze than general resolution ones, namely that the complexity of clauses decreases monotonically throughout the proof, insofar as the critical assignments are concerned.

Lemma 5.7. *Let Π be a regular resolution refutation of a Tseitin tautology. If there is a path consistent with a critical assignment from the root to clause C to D in Π then $crit(I_{D,\Pi}) \subseteq crit(I_{C,\Pi})$*

Proof. By Lemma 5.5, $crit(I_{C,\Pi}) = crit(p)$ for any path p from the root to C consistent with a critical assignment. Let q be any path from C to D in Π. Again, $crit(I_{D,\Pi}) = crit(pq)$, and the result follows immediately from the fact that by definition *crit* is monotonically decreasing over partial assignments and that pq extends p. □

5.2 A probabilistic adversary

For this section we fix a connected graph $G = (V, E)$, a Tseitin tautology τ over G, and a regular resolution refutation Π of τ. Rather than fix a distribution over partial assignments or total assignments depending only on G as we did for general resolution, for regular resolution we can use a probabilistic adversary which navigates Π to define a distribution on assignments that depends on Π itself.

Definition 5.8. *The* probabilistic adversary A_Π *responds to the queries in the proof Π in the following way. Let σ be the assignment that the adversary has given to the queries already asked, and suppose that Π queries edge e at the proof node reached by following the path labeled σ,*

- *If e is not a cut edge of $G|\sigma$, then the adversary responds randomly.*
- *If e is a cut edge of $G|\sigma$, then the adversary responds so as to maximize the size of the critical vertex set, breaking ties arbitrarily. More precisely, letting σ_0 and σ_1 denote the extensions of σ that assign 0 or 1 to e, respectively, the adversary answers according to the larger of $|crit(\sigma_0)|$ and $|crit(\sigma_1)|$.*

It is notable that our adversary strategy is very similar to one used in studying Prover-Delayer games and lower bounds for tree-like resolution as in [6, 26, 25].

The sizes of the critical sets associated with the clauses visited by the adversary satisfy some basic properties analogous to ones shown in Proposition 2.14 and Lemma 2.15 for general resolution, as well as an extra monotonicity property.

Proposition 5.9. *Let p be any root-leaf path in Π in the support of the distribution induced by A_Π and let $C_0 = \bot, C_1, \dots, C_r$ be the clauses labeling the nodes of p in order.*

1. *$crit(I_{\bot,\Pi}) = V$.*
2. *$|crit(I_{C_r,\Pi})| = 1$.*
3. *$|crit(I_{C_{i+1},\Pi})| \le |crit(I_{C_i,\Pi})|$ for every i with $0 \le i < r$.*
4. *$|crit(I_{C_{i+1},\Pi})| \ge |crit(I_{C_i,\Pi})|/2 > 0$ for every i with $0 \le i < r$.*

Proof. Let p_i be the partial assignment given by the length i prefix of path p. Observe that $crit(I_{C_i,\Pi}) = crit(p_i)$. The first two are immediate, the third follows immediately because by definition *crit* is monotone decreasing over partial assignments and the fourth follows from sub-additivity of *crit* and the maximizing choice of the adversary. □

Corollary 5.10. *For any t with $1 \le t \le |V|/2$, the path taken by adversary A_π must pass through a clause C with $t < |crit(I_{C,\Pi})| \le 2t$. In particular, the path taken by A_π must pass through a clause from every $\mathcal{L}_i^*$.*

To make use of the probability distribution given by the adversary, we will need to understand the probability that the adversary reaches a given in Π, as well as the conditional probability of reaching a clause given that the adversary has already followed some partial path Π. The following measure will allow us to bound these probabilities.

Definition 5.11. *Let $\#(H)$ denote the number of connected components of a subgraph H of G. Define a function Φ on (pairs of) subsets of the edges E of G by*

- *$\Phi(A|B) = |A\backslash B| - \#(G|(A \cup B)) + \#(G|B)$, for A, $B \subseteq E$, and*
- *$\Phi(A) = \Phi(A|\emptyset)$ for $A \subseteq E$.*

For ρ, σ, both partial assignments to the edges of G, we write $\Phi(\rho|\sigma)$ for $\Phi(dom(\rho)|dom(\sigma))$.

The following gives some simple intuition about Φ.

Proposition 5.12. *Let $A \subseteq E$. For any edge set A and edge $e \notin A$,*

$$\Phi(A \cup \{e\}) - \Phi(A) = \begin{cases} 0 & \textit{if } e \textit{ cuts } G|A \\ 1 & \textit{if not.} \end{cases}$$

Proof. By definition,

$$\begin{aligned} &\Phi(A \cup \{e\}) - \Phi(A) \\ &= |A \cup \{e\}| - |A| - \#(G|A \cup \{e\}) + \#(G|A) \\ &= 1 - \#(G|A \cup \{e\}) + \#(G|A). \end{aligned}$$

and the result follows immediately. □

Definition 5.13. *We call a set of edges in G* independent *if and only if their removal from G does not increase the number of connected components of G.*

Corollary 5.14. *$\Phi(A)$ is the maximum size of an independent set $A' \subseteq A$.*

Proof. Apply the Proposition 5.12 inductively starting with $\Phi(\emptyset)$ and adding the edges in A one at a time beginning with the edges in A'. □

Remark 5.15. *$\Phi(A)$ is the rank of edge set A in the* cut matroid *(also known as on the* bond matroid*) of the graph G. This is the dual of the more well-known cycle matroid of G. The independent sets of the cut matroid are those sets of edges that do not separate any component of the graph, as in our definition. The rank of a set of edges A in the cut matroid is the the cardinality of the largest independent $A' \subseteq A$. In the literature, $\Phi(\cdot|\cdot)$ is called the* conditional rank *of the cut matroid.*

Our reason for using this definition is the following lemma, which is the key to understanding the distribution on assignments given by our probabilistic adversary arguments. Though the formulation looks fairly different from the arguments based on the sizes of boundaries of critical sets that we used in the case of general resolution, it deals with a similar property. In particular, in the case of a connected set of vertices S (such as a critical set) whose complement also happens to be connected, it is easy to check that $\Phi(\delta(S)) = |\delta(S)| - 1$.

LEMMA 5.16. *[Main Adversary Lemma] Let Π be a regular resolution refutation of a Tseitin tautology. For any clause C in Π of with information $I_{C,\Pi} = \rho$, the probability that adversary A_Π reaches C conditioned on following a path labeled σ is at most $2^{-\Phi(\rho|\sigma)}$. In particular, the probability that the adversary A_Π reaches a clause C with information $I_{C,\Pi} = \rho$ is at most $2^{-\Phi(\rho)}$.*

PROOF. For a fixed ρ, we prove the bound by induction on the length of σ, with large σ as the base case.

If the assignment given by σ contains ρ, the result holds trivially, since the bounding expression is 1. Similarly, for σ not consistent with ρ the probability of the event is 0, so the bound holds.

Suppose inductively that the result is true for every assignment strictly larger than σ. If the adversary has followed the path labeled σ in Π, at which point edge e is resolved on, then, by the induction hypothesis, the bound holds for $\sigma_0 = \sigma \cup \{e = 0\}$ and $\sigma_1 = \sigma \cup \{e = 1\}$. We always have $\Phi(\rho|\sigma_0) = \Phi(\rho|\sigma_1)$ since σ_0 and σ_1 make assignments to the same set of edges.

Suppose that $\Phi(\rho|\sigma) \le \Phi(\rho|\sigma_0)$. Since σ is not as large as ρ, the probability that the adversary reaches C after following σ is a weighted average of the probabilities that the adversary reaches C after following σ_0 and σ_1, respectively. Therefore, by induction, the conditional probability of reaching C after following σ is at most $2^{-\Phi(\rho|\sigma_0)} \le 2^{-\Phi(\rho|\sigma)}$ as required.

Suppose now that $\Phi(\rho|\sigma_0) < \Phi(\rho|\sigma)$. There are two subcases, depending on whether e is assigned by ρ. Consider first the subcase that e is assigned by ρ. Then $\mathrm{dom}(\rho) \cup \mathrm{dom}(\sigma) = \mathrm{dom}(\rho) \cup \mathrm{dom}(\sigma_0)$. Therefore, since $\Phi(\rho|\sigma_0) < \Phi(\rho|\sigma)$, we have

$$\begin{aligned} 0 &< \Phi(\rho|\sigma) - \Phi(\rho|\sigma_0) \\ &= |\mathrm{dom}(\rho)\backslash\mathrm{dom}(\sigma)| - |\mathrm{dom}(\rho)\backslash\mathrm{dom}(\sigma_0)| \\ &\quad -[\#(G|(\rho\cup\sigma)) - \#(G|(\rho\cup\sigma_0))] \\ &\quad +\#(G|\sigma) - \#(G|\sigma_0) \\ &= 1 + \#(G|\sigma) - \#(G|\sigma_0). \qquad (1) \end{aligned}$$

Hence $\#(G|\sigma) - \#(G|\sigma_0) > -1$. Therefore, since $\#(G|\sigma) \le \#(G|\sigma_0)$ we must have $\#(G|\sigma) = \#(G|\sigma_0)$, which means that e is not a cut edge of $G|\sigma$. Moreover, (1) implies that $\Phi(\rho|\sigma_0) = \Phi(\rho|\sigma) - 1$. By the definition of the adversary, the value the adversary assigns to e is chosen randomly in this case. However, one of σ_0 and σ_1 is inconsistent with ρ because ρ assigns some particular value to e. So, without loss of generality,

$$\Pr_{adv}[\rho|\sigma] = \frac{1}{2}\Pr_{adv}[\rho|\sigma_0] \le 2^{-\Phi(\rho|\sigma)} ,$$

as desired. Now consider the subcase that e is not assigned by ρ. In this case we follow the definitions of $\Phi(\rho|\sigma_0)$ and $\Phi(\rho|\sigma)$ and observe that $|\mathrm{dom}(\rho)\backslash\mathrm{dom}(\sigma_0)| = |\mathrm{dom}(\rho)\backslash\mathrm{dom}(\sigma)|$. Since $\Phi(\rho|\sigma_0) < \Phi(\rho|\sigma)$ we must have

$$\begin{aligned} &0 < \Phi(\rho|\sigma) - \Phi(\rho|\sigma_0) \\ &\quad = -[\#(G|(\rho\cup\sigma)) - \#(G|(\rho\cup\sigma_0))] + \#(G|\sigma) - \#(G|\sigma_0) \qquad (2) \end{aligned}$$

and hence

$$\#(G|(\rho\cup\sigma_0)) - \#(G|(\rho\cup\sigma)) > \#(G|\sigma_0) - \#(G|\sigma).$$

The left hand side is equal to 1 or 0 indicating whether e cuts $G|(\rho\cup\sigma)$, and the right hand side indicates whether e cuts $G|\sigma$. The inequality implies the former is one and the latter is zero, hence e is not a cut edge of $G|\sigma$ but it is a cut edge of $G|(\rho\cup\sigma)$. Plugging these facts back in (2), we see that the potential drop is exactly one; that is, $\Phi(\rho|\sigma_0) = \Phi(\rho|\sigma) - 1$. Again, by the definition of the adversary, e is assigned randomly here.

CLAIM 5.17. *Only one of σ_0, σ_1 can permit the adversary to reach ρ.*

Suppose to the contrary that π_0 and π_1 extend σ_0 and σ_1 respectively, each corresponding to paths that can be travelled by the adversary and that meet at the target clause C with information ρ. Then, $crit(\pi_0) = crit(\pi_1) \ne \emptyset$, by Lemma 5.5 and Proposition 5.9 (4). But, by the definition of common information, π_0 and π_1 extend $\sigma_0 \cup \rho$ and $\sigma_1 \cup \rho$ respectively, and by Proposition 2.11 these two assignments have disjoint critical sets since e cuts $G|(\rho\cup\sigma)$. Since by definition $crit$ is monotone decreasing over partial assignments, $crit(\pi_0)$ and $crit(\pi_1)$ are also disjoint, a contradiction.

The claim now implies the bound we need exactly as before, and thus completes the proof. □

If C is a clause and p is a path from the root of Π, it is most convenient to abbreviate the conclusion of the lemma as

$$\Pr_{adv}[C|p] \le 2^{-\Phi(I_{C,\Pi}|p)} .$$

For our superpolynomial bounds we need to use the following properties of Φ, all of which are consequences of Φ being a matroid rank function. While these can be proved easily using matroid theory, we sketch simple proofs for our specific case without relying on matroids.

PROPOSITION 5.18. *For all non-empty sets of edges A, B, and C in G,*

(a) $\Phi(\cdot)$ is non-negative and monotone increasing.

(b) $\Phi(A|B) = \Phi(A\cup B) - \Phi(B)$.

(c) (Non-negativity and Monotonicity) $\Phi(\cdot|\cdot)$ is non-negative, increasing in its first argument, and decreasing in its second argument.

(d) (Chain Rule) $\Phi(A) \le \Phi(A|B) + \Phi(B)$. More generally, $\Phi(A|C) \le \Phi(A|B) + \Phi(B|C)$ when $B \supseteq C$.

(e) (Subadditivity) $\Phi(A\cup B|C) \le \Phi(A|C) + \Phi(B|C)$.

PROOF. Part (a) is immediate from Corollary 5.14. By definition,

$$\begin{aligned} &\Phi(A\cup B) - \Phi(B) \\ &\quad = |A\cup B| - |B| - \#(G|(A\cup B)) + \#(G|B) \\ &\quad = |A\backslash B| - \#(G|(A\cup B)) + \#(G|B) \end{aligned}$$

so part (b) follows. The nonnegativity and monotonicity in the first argument given in (c) follow from (b) and the monotonicity from (a). The monotonicity in the second argument given in (c) follows from (b), Proposition 5.12, and the fact that if e is not a cut edge with respect to $B \supseteq B'$ then e is not a cut edge with respect to B' either. For part (d), since $C \subseteq B$, we have $B \cup C = B$ and $\Phi(A \cup C) \leq \Phi(A \cup B)$ by the monotonicity of Φ. Therefore by (b),

$$\begin{aligned}\Phi(A|C) &= \Phi(A \cup C) - \Phi(C) \\ &\leq \Phi(A \cup B) - \Phi(B) + \Phi(B \cup C) - \Phi(C) \\ &= \Phi(A|B) + \Phi(B|C)\end{aligned}$$

which yields the chain rule. By the chain rule,

$$\Phi(A \cup B|C) \leq \Phi(A \cup B|B \cup C) + \Phi(B \cup C|C) \ ,$$

but by definition $\Phi(B \cup C|C) = \Phi(B|C)$, and also $\Phi(A \cup B|B \cup C) = \Phi(A|B \cup C)$ which, by monotonicity, is bounded above by $\Phi(A|C)$. Hence subadditivity (e) follows. □

We will also need this fact, specific to cut matroids.

LEMMA 5.19. *[Matroid Cut Argument] Suppose that S is a set of vertices of G, and that $\delta(S) \subseteq B$. Then for any A,*

$$\Phi(A|B) = \Phi(A \sqcap S|B \sqcap S) + \Phi(A \sqcap \overline{S}|B \sqcap \overline{S}) \ ,$$

where $E' \sqcap V' := \{e \in E' : e$ is incident to $V'\}$, for E' a set of edges and V' a set of vertices.

PROOF. We can easily prove this by induction on $|A|$. Suppose that $e \notin A$. We would like to show that when e is added to A, the equality still holds. If $e \in B$, then none of $\Phi(A|B)$, $\Phi(A \sqcap S|B \sqcap S)$, or $\Phi(A \sqcap \overline{S}|B \sqcap \overline{S})$ can change; so, without loss of generality $e \notin B$. This implies that $e \notin \delta(S)$ as well. Without loss of generality, e is adjacent to S but not $\overline{S}$, so the only terms that can change are $\Phi(A|B)$ and $\Phi(A \sqcap S|B \sqcap S)$. It is enough to show that these terms must change in the same way when e is added; so, in fact, it is enough to see that $\Phi(A \cup B)$ changes the same way as $\Phi((A \cup B) \sqcap S)$ does when e is added. In light of Proposition 5.12, this holds because an edge e is a cut edge in a graph G if and only if it is a cut edge in the component containing it. □

5.3 Isoperimetric inequality for grid graphs

In light of Lemma 5.16, for regular resolution we replace the "isoperimetric inequality" used for general resolution which showed that medium sized sets S have have many edges on their boundaries (large $\delta(S)$) by one showing that such sets have large $\Phi(\delta(S))$, i.e. large independent subsets in $\delta(S)$.

For every medium size set of vertices S of $G_{n,\ell}$, we will show that $\delta(S)$ has large rank ($\Phi(\delta(S))$ is large) rather than large cardinality. Moreover, we show that for several medium-sized sets of vertices $S_1 \ldots S_k$ of sufficiently separated cardinalities, the rank of $\bigcup_i \delta(S_i)$ grows linearly with k. We will be able to prove a tight lower bound on these ranks without much more difficulty, and without giving up much compared to cardinality.

The following is an analog of the properties we proved in the case of general resolution using transition points and Lemma 4 about endpoints of intervals of increasing sizes. The proof here uses the same lemma on intervals but directly uses common information rather than transition points.

LEMMA 5.20. *Let $S_1 \ldots S_k$ be sets of vertices in $G_{n,\ell}$. If $2kn^2 \leq |S_1|$, for each $i \in [k-1]$, $4|S_i| \leq |S_{i+1}|$, and $|S_k| \leq n\ell/4$, then $\Phi(\bigcup_i \delta(S_i)) \geq k(n-1)$.*

PROOF. If any S_i contains $k(n-1)$ partial columns, then each such column contributes a vertical edge to $\bigcup_i \delta(S_i)$, and these vertical edges together form an independent set of the necessary size, so we are done.

If no S_i contains $k(n-1)$ partial columns, then by the size bounds on S_i, each contains a full column, and an empty column. In particular, each S_i contains a horizontal edge in every row. We will show that each row contains at least k edges of $\bigcup_i \delta(S_i)$. Choosing any $n-1$ rows and the k edges from each such row give an independent set of size $k(n-1)$, so this will finish the proof.

Fix a row r. Let a_i be the column number of the leftmost vertex in S_i and in the r-th row, let b_i be the column number of the rightmost such vertex. Then S_i has boundary edges with left endpoints in columns $a_i - 1$ and b_i, provided that these are not 0 or ℓ. Thus all we need to show is that

$$\left| \bigcup_i \{a_i - 1, b_i\} \setminus \{0, \ell\} \right| \geq k \ .$$

Since 0 and ℓ are the same modulo ℓ, it suffices to show that these intervals meet the conditions for the first part of Lemma 3.14.

Since S_i has fewer then $k(n-1)$ partial columns, $b_i - (a_i - 1) > |S_i|/n - k(n-1)$. On the other hand, the number of full columns in S_i is at most $|S_i|/n$, so $b_i - (a_i - 1) < |S_i|/n + k(n-1)$. Thus the ratio of successive differences is at least

$$\begin{aligned}\frac{b_{i+1} - (a_{i+1} - 1)}{b_i - (a_i - 1)} &> \frac{|S_{i+1}| - kn(n-1)}{|S_i| + kn(n-1)} \\ &\geq \frac{4|S_i| - kn(n-1)}{|S_i| + kn(n-1)} \ .\end{aligned}$$

Since $|S_i| \geq |S_1| \geq 2kn^2$, the ratio is at least $\frac{7}{3} > 2$ and we can apply Lemma 3.14 as desired. □

It is convenient to have a lower bound on $|S_1|$ that does not depend on k. To do this we need to upper bound ℓ.

COROLLARY 5.21. *Suppose that $\ell \leq 2^n$. Let $S_1 \ldots S_k$ be sets of vertices in $G_{n,\ell}$. If $n^3 \leq |S_1|$, $4|S_i| \leq |S_{i+1}|$ for each i, and $|S_k| \leq n\ell/4$, then $\Phi(\bigcup_i \delta(S_i)) \geq k(n-1)$.*

PROOF. Since $\ell \leq 2^n$, and $|S_1| \geq n^3$, we have $k \leq 1 + \log_4(|S_k|/|S_1|) \leq \log_2 n/2 \leq n/2$ and Lemma 5.20 applies. □

5.4 Regular resolution time-space tradeoff for grid graphs

To prove Theorem 5.2, we will begin with a regular resolution refutation Π, and repeatedly subdivide Π into polynomially many epochs, at each point choosing an epoch in which enough progress happens that we may continue subdividing it. The process peters out after $\mathbf{O}\left(\frac{\log L}{\log\log L}\right)$ steps.

By Corollary 5.10, the adversary path will pass through clauses of every medium complexity level. In this analysis, an epoch in the refutation is viewed as representing a lot of progress if there is a path p to a clause appearing in that epoch such that, conditioned on having followed p, the adversary *probably* reaches a *much smaller* complexity clause by the end of that epoch than it began with. The major technical step is to show how to divide an epoch that represents a significant amount of progress into much smaller epochs, one of which does a comparable amount of work.

LEMMA 5.22. *Let $0 < \epsilon < 1$ and Π be a regular resolution refutation of a Tseitin tautology on the grid graph $G_{n,\ell}$ for $\ell \le 2^n$. Suppose that p is a path in Π from the root to some clause C. Let $0 \le \ell_1 < \cdots < \ell_k$ satisfy $\ell_{i+1} \ge \ell_i + 3$ for each $i \in [k]$. If for each $i \in [k]$ there is some $C_i \in \mathcal{L}^*_{\ell_j}$ appearing in Π with $\Pr_{adv}[C_i|p] \ge 2^{-(1-\epsilon)(n-1)}$ then $\Phi(I_{C,\Pi}) \ge k\epsilon(n-1)$.*

PROOF. Fix the choices of ℓ_i and the associated clauses C_i. In order to make it relatively likely to reach each of the C_i from C as in the hypothesis, we show that the extra information at C_i over that at C cannot be too large. On the other hand we can show, using the isoperimetric properties of the grid graph, that the clauses C_i in total have a large amount of information and hence C must also. We start with the latter.

Let $S_i = crit(I_{C_i,\Pi})$ for $i = 1, \ldots, k$ and $S = crit(p)$. Since $\ell_{i+1} \ge \ell_i + 3$ we have $|crit(I_{C_{i+1},\Pi})| \ge 4|crit(I_{C_i,\Pi})|$ and Corollary 5.21 implies that

$$\Phi(\bigcup_i \delta(S_i)) \ge k(n-1) \tag{3}$$

On the other hand, by Lemma 5.16, the hypothesis that $\Pr_{adv}[C_i|p] \ge 2^{-(1-\epsilon)(n-1)}$ implies that $\Phi(I_{C_i,\Pi}|p) \le (1-\epsilon)(n-1)$ where, as usual, we have identified p with the partial assignment labeling it. Moreover, since $\Pr_{adv}[C_i|p] > 0$, there is some path q_i from C to C_i in Π. Therefore,

$$S_i = crit(I_{C_i,\Pi}) = crit(pq_i) \subseteq crit(p) = S,$$

where the third equality follows from Lemma 5.5 and the containment follows since by definition *crit* is monotone decreasing over partial assignments. Note that Lemma 5.5 also implies that $p \sqcap S = I_{C,\Pi} \sqcap S$ and Corollary 2.12 implies that $\delta(S) \subseteq \mathrm{dom}(p)$. Therefore, we may apply Proposition 5.19 with $B = \mathrm{dom}(p)$ to obtain

$$\begin{aligned}
\Phi(I_{C_i,\Pi}|p) &\ge \Phi(\delta(S_i)|p) \qquad \text{by monotonicity}\\
&= \Phi(\delta(S_i) \sqcap S|p \sqcap S) + \Phi(\delta(S_i) \sqcap \overline{S}|p \sqcap \overline{S})\\
&\qquad\qquad \text{by Lemma 5.19}\\
&\ge \Phi(\delta(S_i) \sqcap S|p \sqcap S) \qquad \text{by nonnegativity}\\
&= \Phi(\delta(S_i)|p \sqcap S) \qquad \text{since } S_i \subseteq S\\
&= \Phi(\delta(S_i)|I_{C,\Pi} \sqcap S) \qquad \text{by Lemma 5.5}\\
&\ge \Phi(\delta(S_i)|I_{C,\Pi}) \qquad \text{by monotonicity.}
\end{aligned}$$

It follows that $\Phi(\delta(S_i)|I_{C,\Pi}) \le (1-\epsilon)(n-1)$ for each $i \in [k]$. Hence, by the subadditivity of Φ given in Proposition 5.18 we have

$$\Phi(\bigcup_i \delta(S_i)|I_{C,\Pi}) \le (1-\epsilon)k(n-1).$$

On the other hand, by the chain rule of Proposition 5.18 and (3) above, we have

$$\begin{aligned}
\Phi(I_{C,\Pi}) &\ge \Phi(\bigcup_i \delta(S_i)) - \Phi(\bigcup_i \delta(S_i)|I_{C,\Pi})\\
&\ge k(n-1) - (1-\epsilon)k(n-1) = \epsilon k(n-1)
\end{aligned}$$

which is what we needed to prove. □

The real value of Lemma 5.22 is in its contrapositive. If the proof is small, then by Lemma 5.16, the vast majority of the time the adversary path will only reach clauses C with $\Phi(I_{C,\Pi})$ small. Moreover, there can only be a few medium complexity levels of clauses that the adversary is now relatively likely to visit. For most of the levels, the adversary on visiting C has essentially no greater chance to reach any clauses of that complexity level than the chance when the adversary began at the $\bot$ clause.

With this lemma in hand we can now try to implement the overall plan for the proof of Theorem 5.2. Based on some parameters that we will set later, we first specify a property of clauses, which, by Lemma 5.16 and a union bound, are rarely encountered by an adversary in any refutation that is not too large.

DEFINITION 5.23. *Let $0 < \epsilon < 1$ and m be parameters, which we will fix later. We say that a clause C has* high potential *if $\Phi(I_C) \ge \epsilon m(n-1)$.*

Recall that a refutation is a *sequence* of clauses ending in $\bot$, and that the adversary walks down the proof DAG starting at $\bot$, which involves moving backwards through this sequence, possibly jumping over intermediate clauses. The ordering on the sequence of clauses gives a sequence of *time steps.* Write $time(C)$ for the time step in the proof sequence in which clause C appears. Recall that every arc in the proof DAG that crosses from one time step to an earlier one corresponds to a clause that must be in memory during all intervening time steps. Moreover, by Corollary 5.10, the adversary path must pass through clauses in every $\mathcal{L}^*_i$ starting from large values of i and ending with $i = 0$. For a medium complexity clause C we write $level(C)$ to denote the i such that $C \in \mathcal{L}^*_i$. These properties motivate us to define a way to measure the progress of a portion of the proof.

DEFINITION 5.24. *Let C be a medium complexity clause in Π. Say that a path p in Π from the root to clause C is a* (T, gap, δ)-path in Π *if, conditioned on following p, with probability at least $1-\delta$, the adversary A_Π*

- *does not reach any high potential clauses, and*
- *reaches a clause C' with $level(C') \le level(C) - gap$ such that $time(C') \ge time(C) - T$.*

LEMMA 5.25 (INDUCTIVE STEP). *Let Π be a space S regular resolution refutation of a Tseitin tautology on $G_{n,\ell}$ for $n^3 \le \ell \le 2^n$. Suppose that p is a (T, gap, δ)-path in Π, $gap \ge 1$, and B is any natural number. Then there exists a (T', gap', δ')-path p' in Π extending p such that*

- *$T' = \lceil T/B \rceil$,*
- *$gap' = \lfloor \frac{gap}{3m} \rfloor$, and*
- *$\delta' = \delta + B \cdot S \cdot 2^{-(1-\epsilon)(n-1)}$.*

PROOF. If $\delta \ge 1$ then the claim is vacuous, so assume that $\delta < 1$. Let C be the clause reached by p. By hypothesis, $level(C) \ge gap$ and C must not have high potential.

Therefore, by the contrapositive of Lemma 5.22 with $k = m$, there do not exist m complexity levels $\ell_1 < \ldots < \ell_m$ with $\ell_{i+1} \ge \ell_i + 3$ such that there is a clause $D_i \in \mathcal{L}^*_{\ell_i}$ with $\Pr_{adv}[C_i|p] \ge 2^{-(1-\epsilon)(n-1)}$. In particular, this means that there do not exist $3m$ distinct complexity levels of clauses $\mathcal{L}_j^*$ such that there is a clause $D \in \mathcal{L}_j^*$ with $\Pr_{adv}[D|p] \geq 2^{-(1-\epsilon)(n-1)}$.

It follows that there exists a sequence of at least $gap' = \lfloor \frac{gap}{3m} \rfloor$ *consecutive* complexity levels between $level(C)$ and $level(C) - gap$, such that for any D in one of these levels, $\Pr[D|p] \leq 2^{-(1-\epsilon)(n-1)}$. Let i' be the largest level in this sequence, and $i' - gap' + 1$ be the smallest.

Divide the T time steps between $time(C)$ and $time(C) - T$, into B epochs of length at most $T' = \lceil T/B \rceil$, and let $\mathcal{M}$ denote the union over the B breakpoints between these epochs (including the start of the first epoch), of the sets of clauses in memory that are of complexity level between i' and $i' - gap' + 1$. By a union bound,

$$\Pr[\text{adversary reaches some } C'' \in \mathcal{M}] \leq B \cdot S \cdot 2^{-(1-\epsilon)(n-1)}.$$

Thus, except with probability at most δ', the adversary A_Π, conditioned on following p,

1. does not hit any high potential clauses,
2. reaches a clause C' with $level(C') \leq level(C) - gap$ and $time(C') \geq time(C) - T$, and
3. does not hit any clause in $\mathcal{M}$.

By averaging, there must exist a path p' extending p and reaching a clause $C' \in \mathcal{L}_{i'}^*$ such that this also holds, except with probability at most δ'.

Then, we claim that p' is a (T', gap', δ')-path. Let C' denote the clause reached by p'. C' falls in some epoch of length at most T'. If no clause in $\mathcal{M}$ is reached and the adversary has followed p', then by the beginning of the epoch containing C', the complexity level of the clause reached by the adversary must be less than or equal to $i' - gap'$. Thus, by construction, p' is indeed a (T', gap', Δ')-path in Π. □

Additionally, whenever we have a (T, gap, δ)-path p in Π with nontrivial gap and δ parameters, we can show that T is nontrivial.

Lemma 5.26 (Base Case). *Let Π be a regular resolution refutation of a Tseitin tautology on $G_{n,\ell}$ for $n^3 \leq \ell \leq 2^n$. If p is a (T, gap, δ)-path in Π with $gap > 3m$, then $T \geq (1-\delta)2^{(1-\epsilon)(n-1)}$.*

Proof. If $\delta \geq 1$, the claim is trivial. Suppose that $\delta < 1$. Let C be the clause reached by p and consider the clauses in the epoch between time steps $time(C) - T$ and $time(C)$ in Π.

By Lemma 5.22, there exists some complexity level $\mathcal{L}_i^*$ with $i > level(C) - gap$ such that, conditioned on following p, the adversary A_Π

- reaches a clause $C' \in \mathcal{L}_i^*$, with $time(C') \geq step(C) - T$ with probability at least $(1-\delta)$, and
- does not reach any fixed clause D in $\mathcal{L}_i^*$, except with probability at most $2^{-(1-\epsilon)(n-1)}$.

By a union bound over the clauses in this epoch, we conclude that $T \geq (1-\delta)2^{(1-\epsilon)(n-1)}$. □

We now have the ingredients needed to complete the proof of our time-space tradeoff for regular resolution.

Proof of Theorem 5.2. We begin with a regular resolution proof Π of length T that uses space S and refutes a Tseitin tautology on $G_{n,\ell}$. There are $L = \log_2(\ell/(4n^2))$ distinct medium complexity levels of clauses with respect to T. By Corollary 5.10, the path followed by adversary A_Π must pass through clauses with each of these complexity levels.

The outline of the remainder of the proof is as follows:

- First we show, via a union bound over the number of steps in the proof, that there is a (T, L, δ_0)-path in Π for some suitably small δ_0. This step only involves calculating the probability that the adversary passes through some high potential clause. This calculation will require that $\epsilon \cdot m$ not be small compared to $\frac{\log_2 T}{\log_2 n}$.
- Next we choose an appropriate value for B and continually apply the inductive step from Lemma 5.25, until either the accumulated error δ becomes too large, or the gap becomes too small. (We use the same value of B at each step.)
- Finally, some number r of rounds, just before gap has become too small after r rounds, we apply the base case Lemma 5.26 to deduce that the T' at the last step is reasonably large, and that $T \geq B^r \cdot T'$.

Let r denote the number of rounds that we will apply the inductive step and let gap_i denote the value of gap after each round and δ_i denote the value of δ after each round. We will set $gap_0 = L = \log_2(\ell/(4n^2))$. By Lemma 5.16, the probability that A_Π reaches a high potential clause is at most $2^{-\epsilon m(n-1)}$. Therefore if

$$T \cdot 2^{-\epsilon m(n-1)} \leq \delta_0 \tag{4}$$

the adversary must reach some clause of complexity L. Fix any path to such a clause. We obtain that this path is a (T, L, δ_0)-path,

We will choose B so that gap, rather than the error δ, is the limiting resource, so we take

$$r = \frac{\log_2 L}{\log_2(3m)} - 1 \tag{5}$$

Observe that by Lemma 5.25, $\delta_r = \delta_0 + r \cdot B \cdot S \cdot 2^{-(1-\epsilon)(n-1)}$. and applying Lemma 5.26 after the r-th round point yields $T' \geq (1-\delta_r)2^{(1-\epsilon)(n-1)}$. By choosing

$$\delta_0 = \frac{1}{3} \text{ and } B = \frac{1}{3} \cdot \frac{2^{(1-\epsilon)(n-1)}}{S \cdot r}\ ,$$

we obtain that $\delta_r = 2/3$ and hence

$$T \geq \frac{1}{3} 2^{(1-\epsilon)(n-1)} \cdot \left(\frac{2^{(1-\epsilon)(n-1)}}{3 \cdot S \cdot r} \right)^r\ . \tag{6}$$

Together with the $n^3 \leq \ell \leq 2^n$ and the values of δ_0, L, and r, inequalities (4) and (6) provide the only constraints on our parameters. It remains to choose m and ϵ to optimize them and derive a convenient tradeoff lower bound. It is convenient to choose $m = \log_2 L$ so that $r = \log_2 L / \log_2(3m) - 1$ is between $\frac{1}{2}\log_2 L / \log_2 \log_2 L$ and $\log_2 L / \log_2 \log_2 L$. For convenience we also choose ϵ to be $2/\log_2 m = 2/\log_2 \log_2 L$. With these choices, the constraint (4) is satisfied whenever T is at most $\frac{1}{3} 2^{\frac{2 \log_2 L}{\log_2 \log_2 L}(n-1)}$ and this upper bound is strictly larger than the lower bound from (6). On the other hand, since $r \leq \log_2 L / \log_2 \log_2 L$ and $L \leq n$, the $3^{r+1} r^r$ in the denominator of the expression in (6) is at most $2^{(1-\epsilon)(n-1)}$. Hence

$$T \geq \left(\frac{2^{(1-\frac{2}{\log_2 \log_2 L})(n-1)}}{S}\right)^{\frac{\log_2 L}{2\log_2 \log_2 L}}.$$

Since L is grows at least logarithmically with n, the statement of the theorem follows. □

6. CONCLUSION

We have shown superpolynomial size-space tradeoff lower bounds for resolution proofs which are the first to apply for superlinear space. The two different methods for deriving these bounds are based on a similar recursive decomposition of proofs into epochs. Our results suggest a number of open questions: Can this decomposition framework be applied to show size-space tradeoffs for stronger proof systems? With very small space, resolution size upper bounds for the Tseitin formulas we consider are $\log_2 n$ powers of their size for unlimited space. Is this tight? Can we increase the exponent in our lower bound from $\Theta(\log\log n/\log\log\log n)$ to $\Theta(\log n)$? More generally, is it true that for every k and every formula of size n with a proof of size n^k, there exists a proof in space $O(n)$ with size $n^{O(\log n)}$? with size 2^{n^ϵ} for any $\epsilon > 0$? Finally, our tradeoff lower bound can be viewed as a separation between two search paradigms: dynamic programming vs. divide and conquer. Can we find other settings in which these paradigms may be separated?

7. REFERENCES

[1] M. Alekhnovich, J. Johannsen, T. Pitassi, and A. Urquhart. An exponential separation between regular and general resolution. Technical Report TR01-56, Electronic Colloquium in Computation Complexity, `http://www.eccc.uni-trier.de/eccc/`, 2001.

[2] M. Alekhnovich and A. A. Razborov. Satisfiability, branch-width and tseitin tautologies. In *Proceedings 43nd Annual Symposium on Foundations of Computer Science*, pages 593–603, Vancouver, BC, November 2002. IEEE.

[3] A. Atserias and V. Dalmau. A combinatorial characterization of resolution width. In *Proceedings Eighteenth Annual IEEE Conference on Computational Complexity*, pages 239–247, Aarhus, Denmark, July 2003.

[4] Fahiem Bacchus, Shannon Dalmao, and Toniann Pitassi. Solving #sat and Bayesian inference with backtracking search. *jair*, 34:391–442, 2009.

[5] P. Beame and T. Pitassi. Simplified and improved resolution lower bounds. In *Proceedings 37th Annual Symposium on Foundations of Computer Science*, pages 274–282, Burlington, VT, October 1996. IEEE.

[6] E. Ben-Sasson, R. Impagliazzo, and A. Wigderson. Near-optimal separation of treelike and general resolution. *Combinatorica*, 24(4):585–603, 2004.

[7] E. Ben-Sasson and J. Nordström. Short proofs may be spacious: An optimal separation of space and length in resolution. In *Proceedings 49th Annual Symposium on Foundations of Computer Science*, pages 709–718, Philadelphia,PA, October 2008. IEEE.

[8] E. Ben-Sasson and J. Nordström. Understanding space in resolution: Optimal lower bounds and exponential tradeoffs. Technical Report TR09-034, Electronic Colloquium in Computation Complexity, `http://www.eccc.uni-trier.de/eccc/`, 2009.

[9] E. Ben-Sasson and J. Nordström. Understanding space in proof complexity: Separations and trade-offs via substitutions. Technical Report arXiv:1008.1789, arxiv, 2010.

[10] 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.

[11] S. R. Buss. Polynomial-size Frege and resolution proofs of *st*-connectivity and Hex tautologies. *Theoretical Computer Science*, 357(1–3):35–52, 2006.

[12] S. Chen, T. Lou, P. Papakonstantinou, and B. Tang. Width-parameterized SAT: Time-space tradeoffs. Technical Report arXiv:1108.2385, CoRR, 2011.

[13] J. L. Esteban and J. Toran. Space bounds for resolution. In *(STACS) 99: 16th Annual Symposium on Theoretical Aspects of Computer Science*, volume 1563 of *Lecture Notes in Computer Science*, pages 530–539, Trier, Germany, March 1999. Springer-Verlag.

[14] Z. Galil. On the complexity of regular resolution and the Davis-Putnam procedure. *Theoretical Computer Science*, 4:23–46, 1977.

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

[16] P. Hertel and T. Pitassi. Exponential time/space speedups for resolution and the pspace-completeness of black-white pebbling. In *Proceedings 48th Annual Symposium on Foundations of Computer Science*, pages 137–149, Berkeley, CA, October 2007. IEEE.

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

[18] João P. Marques-Silva and Karem A. Sakallah. Grasp – a new search algorithm for satisfiability. In *Proceedings of the International Conference on Computer-Aided Design*, pages 220–227, San Jose, CA, November 1996. ACM/IEEE.

[19] J. Nordström. Narrow proofs may be spacious: Separating space and width in resolution. *SIAM Journal on Computing*, 39(1):59–121, 2009.

[20] J. Nordström. A simplified way of proving trade-off results for resolution. *Information Processing Letters*, 109(18):1030–1035, 2009.

[21] J. Nordström. Pebble games, proof complexity, and time-space trade-offs. *Logical Methods in Computer Science*, 2011. To appear.

[22] J. Nordström and J. Håstad. Towards an optimal separation of space and length in resolution. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 701–710, Victoria, BC, May 2008.

[23] Wolfgang J. Paul and Robert E. Tarjan. Time-space trade-offs in a pebble game. *Acta Informatica*, 10:111–115, 1978.

[24] T. Pitassi and R. Raz. Lower bounds for regular

resolution proofs of the weak pigeonhole principle. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, pages 347–355, Hersonissos, Crete, Greece, July 2001.

[25] Pavel Pudlák. Proofs as games. *American Math. Monthly*, June-July:541–500, 2000.

[26] Pavel Pudlák and Russell Impagliazzo. A lower bound for DLL algorithms for k-SAT. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 128–136, 2000.

[27] Martin Tompa. Time-space tradeoffs for computing functions, using connectivity properties of their circuits. *Journal of Computer and System Sciences*, 20:118–132, April 1980.

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

[29] Lintao Zhang, Conor F. Madigan, Matthew H. Moskewicz, and Sharad Malik. Efficient conflict driven learning in a boolean satisfiability solver. In *Proceedings of the International Conference on Computer-Aided Design*, pages 279–285, San Jose, CA, November 2001. ACM/IEEE.

APPENDIX

A. SOME UPPER BOUNDS

For the graphs we will consider, some generic upper bounds follow from bounds on their *cut width.* The first will form the high space upper bound in our time space trade-off result, and the second is a low space upper bound we use to argue about the tightness of the lower bound we get for low space.

Definition A.1. *The* cut width *of a graph G is the smallest W such that there is a linear ordering of the vertices $v_1, \ldots, v_n$ such that, for every $1 \leq t \leq n$, there are no more than W edges crossing the cut $(\{v_1, \ldots, v_t\}, \{v_{t+1}, \ldots, v_n\})$.*

Throughout this section we will use the following well-known fact which we prove for completeness:

The *rank* of a resolution proof is the height of its proof DAG.

Observation A.2. *Any tree-like proof of rank r has size at most 2^{r+1} and clause space at most $r+1$.*

Proof. The first is an elementary fact about binary trees of height r. The second, we prove by induction. In the base case, suppose that a proof has rank 1. Then the proof consists of one step and so we need 2 units of space to hold both resolvents. Now suppose the inductive hypothesis holds for rank r, and that we have a proof of rank $r+1$. Suppose C is the last clause of the proof, and A, B are its children. Then the derivations leading up to A, B are rank r. We then derive C as follows – run the derivation of A using r clause space, then A is the only remaining active clause by the tree-like assumption, so clear all of this space and store only A. Then, derive B using r clause space. When we have B, resolve it with A to finish the derivation of C. We only needed $r+1$ space in total, as desired. □

Lemma A.3 (Lemma 2.7). *Let G be a graph with n vertices, maximum degree d, and cut width W. Then there is a resolution refutation of a Tseitin tautology on G using $\leq n \cdot 2^{d+W}$ resolution steps, and space $\leq \cdot 2^W + d + 1$, plus space for the initial axioms.*

Proof. We will specify a sequence of $n+1$ collections $\mathcal{C}_i$ of clauses, such that

1. $\mathcal{C}_1$ is a subset of the axioms .
2. $|\mathcal{C}_i| \leq 2^{W-1}$ for all i .
3. $\mathcal{C}_n = \{\bot\}$.
4. Given a configuration with $\mathcal{C}_i$ and the axioms in memory, we derive any clause in $\mathcal{C}_{i+1}$ using at most 2^{d+1} proof steps and $d+1$ extra work cells.

This construction mirrors a construction in [11]. For each $1 \leq i < n$, let E_i be the collection of at most W edges crossing the cut $(\{v_1 \ldots v_i\}, \{v_{i+1}, \ldots, v_n\})$ as described in the assumption. Let $\mathcal{C}_i$ be the collection of $2^{|E_i|-1}$ clauses whose conjunction is semantically equivalent to the parity constraint $\bigoplus_{e \in E_i} x_e \equiv \bigoplus_{1 \leq j \leq i} \chi(j)$.

We've defined everything so that (1), (2) and (3) hold immediately. Now we will show that item (4) holds for $i < n-1$.

The symmetric difference of E_i and E_{i+1} is, by definition, the edges incident tohas v_{i+1}. Let A_{i+1} denote the axioms associated with vertex v_{i+1}. By definition A_{i+1} is logically equivalent to the constraint $\bigoplus_{e \sim v_{i+1}} x_e \equiv \chi(i+1)$. If we think of parity constraints as $\mathcal{GF}(2)$-linear equations, and we add the two equations associated with A_{i+1} and $\mathcal{C}_i$, we obtain the equation associated with $\mathcal{C}_{i+1}$, so we conclude that

$$A_{i+1}, \mathcal{C}_i \models \mathcal{C}_{i+1} \ .$$

This move from $\mathcal{C}_i$ to $\mathcal{C}_{i+1}$ thus corresponds to adding two linear equations together and deleting one of them from memory.

Since resolution is implicationally complete, it is clear that it is possible to complete this move. We would like to now say that with at most 2^{d+1} size and $d+1$ space, we can derive any particular clause of $\mathcal{C}_{i+1}$, so that we can complete the move using $|\mathcal{C}_{i+1}| \cdot 2^{d+1}$ size and $d+1$ work space in addition to the space needed just to hold $\mathcal{C}_i, \mathcal{C}_{i+1}$ and the axioms.

Let C be an arbitrary clause in $\mathcal{C}_{i+1}$. It is well known that we may convert any derivation

$$A_{i+1}|_{\neg C} \ , \ \mathcal{C}_i|_{\neg C} \ \vdash \bot$$

into a derivation

$$A_{i+1}, \mathcal{C}_i \ \vdash C \ ,$$

using exactly the same proof DAG and possibly with some weakening steps added. It should be clear that weakening steps can be eliminated at the end of our construction without increasing the space or size used.

Since $\mathcal{C}_{i+1}$ is logically equivalent to a parity constraint, every $C \in \mathcal{C}_{i+1}$ assigns every variable E_{i+1}. Thus the only variables remaining in any of the $\mathcal{C}_i$ after the restriction $\neg C$ is applied correspond to edges incident to v_{i+1}. Thus there are at most d variables in any such derivation, and this derivation may be carried out trivially in a tree-like fashion using only 2^{d+1} size and $d+1$ space.

So, by completing a sequence of 2^{W-1} derivations each of length 2^{d+1}, reusing the $d+1$ space for each, we may fill 2^{W-1} new cells with the clauses in $\mathcal{C}_{i+1}$. We may then safely delete the cells used to hold $\mathcal{C}_i$, so that we never need more than $2 \cdot 2^{W-1} + d + 1$ space, plus the space for the axioms.

Finally we will see (4) holds for $i = n$, that is, from $\mathcal{C}_{n-1}$ and the axioms, we may derive a contradiction using 2^{d+1} size and $d + 1$ space.

Since E_{n-1} is just the set of neighbors of v_n, $\mathcal{C}_{n-1}$ is logically equivalent to $\bigoplus_{e \sim v_n} x_e \equiv \bigoplus_{1 \le j \le n-1} \chi(j)$. But A_n is logically equivalent to $\bigoplus_{e \sim v_n} x_e \equiv \chi(n)$, which by assumption that χ is odd-charged, is of different parity. So $\mathcal{C}_{n-1}$ and A_n are contradictory sets of clauses, and on at most d variables. Thus a contradiction can be derived in tree-like fashion as claimed.

In total we carried out n phases, each with $2^{W-1}2^{d+1} = 2^{d+W}$ steps, and each using $2 \cdot 2^{W-1} + d + 1$ workspace in addition to the axioms. If we include the axioms in the space budget we will need additional space for $n2^{d-1}$ clauses. □

A little thought shows that this proof may be carried out in regular resolution, since within each phase, the derivation is tree-like, and the phases can be seen to operate on disjoint sets of variables.

We will appeal to this lemma to establish upper bounds for large space, against which we will prove size-space tradeoffs. The following lemma shows that for Tseitin graphs satisfying the conditions of Lemma 2.7, which include the ones that we consider here, even radically restricted space can only increase the size required by a $O(\log n)$ power.

Lemma A.4 (Lemma 2.8). *Under the same conditions as Lemma 2.7, the Tseitin tautology on $\tau(G)$ has a tree-like Resolution refutation using space $W\lceil \log n \rceil + 1$ and $\le 2^{W\lceil \log n \rceil + 1}$ resolution steps.*

Proof. We prove that when n is a power of two, $\tau(G)$ has a tree-like refutation of rank $W \log n$, which implies the claim. Suppose that G is such a graph of size n. Suppose inductively that the claim holds on graphs of size $n/2$. As before let $E_{n/2}$ denote the edges crossing the cut

$$(\{v_1, \ldots, v_{n/2}\}, \{v_{n/2+1}, \ldots, v_n\}) \ .$$

Let C be any clause containing every variable in $E_{n/2}$. Then $\tau(G)|_{\neg C}$ can be written as a pair of disjoint Tseitin formulae, one from each side of the cut, one of which is odd and therefore unsatisfiable. The induced subgraphs on either side of have cut width at most W, therefore $\tau(G)|_{\neg C}$ has a tree-like resolution refutation of rank at most $W(\log n - 1)$, which can be lifted to a tree-like derivation of C from the axioms $\tau(G)$ in the same rank.

It is easy to see that there is a tree-like refutation of rank $|E_{n/2}| \le W$ using the set of all such C as axioms. By replacing the appearances of these axioms in that refutation with their $W(\log n - 1)$ rank tree-like refutations from the $\tau(G)$ axioms, we obtain a $W \log n$ rank tree-like refutation of $\tau(G)$. □

Generalizations

These ideas can be generalized to the stronger notion of carving width, yielding both a high space version and a low space version for any graph. For the high space version, we essentially just restate part of the main result of Aleknovich and Razborov [2]:

Definition A.5. *A* carving *of a graph G is a rooted binary tree $\mathcal{T}$ whose leaves are in bijection with the vertices of G. For t a node of $\mathcal{T}$, we let $v(t)$ denote the vertices corresponding to leaves which are under t, and $Cut(t)$ denote the edges on the boundary of $v(t)$. The size of the largest $Cut(t)$ is the* width *of a particular carving, and the* carving width *of a graph G is the width of the narrowest carving of G.*

Observation A.6. *The carving width is always less than the cut width – given any linear ordering of the vertices yielding cut width W, we can obtain a carving of width W by taking a maximally unbalanced binary tree, and assigning vertices to leaves as dictated by the linear ordering. The cuts corresponding to internal nodes then correspond exactly to the cuts occurring in the definition of cut width.*

Corollary A.7. *[2] For every n vertex graph G of carving width W, the corresponding Tseitin tautology has a regular resolution refutation of size $poly(n) \cdot 2^{O(W)}$, and uses comparable space.*

Proof Sketch:. There is a natural way to convert a carving into a refutation – for each node t, we will have an associated set of clauses on the variables $Cut(t)$:

For t a leaf corresponding to vertex v, we will simply take all the axioms associated to vertex v.

For t an internal node, we will take the clauses associated to its children t_1, t_2, which are on variables $Cut(t_1)$ or $Cut(t_2)$, and make all derivations possible which yield clauses on $Cut(t)$ and only which resolve on variables from $Cut(t_1) \cup Cut(t_2) \setminus Cut(t)$.

To analyze this, we observe there are at most $3^{|Cut(t)|}$ clauses derived for any node, and each one can be derived from clauses at the children by resolving on at most $2W$ variables, hence in 2^{2^W} time. There are only $2n$ nodes in total since there are n leaves, and so there is $poly(n) \cdot exp(O(W))$ total work, and at most $poly(n) \cdot exp(O(W))$ active clauses at any point. We omit the proof that this actually does result in deriving a contradiction at the end – Aleknovich and Razborov refer to (Krajicek 1992, Theorem 4.2.1). It is straightforward to see this when considering only Tseitin tautologies, since then the clauses associated to t will always correspond to a parity constraint on the variables $Cut(t)$, and the derivation of t from t_1, t_2 corresponds to addition of linear equations; Aleknovich and Razborov in fact establish this claim for any tautology. □

Lemma A.8. *For every graph G on n nodes with carving width W, the corresponding Tseitin tautology has a tree-like refutation of rank at most $W \log_{\frac{3}{2}} n$. In particular, it has a refutation with size $n^{W \log \frac{3}{2}}$, using at most $W \log_{\frac{3}{2}} n + 1$ clause space.*

Proof. Given an optimal carving T of G, we use the classic $\frac{1}{3}, \frac{2}{3}$ lemma for binary trees to find an edge which represents a balanced cut in T. Suppose that t is the lower vertex of this edge. (More explicitly: observe that if we define a function f on nodes of the tree s.t. $f(t) = |v(t)|$, then f is a subadditive function on the tree, that is $f(t) \le f(t_1) + f(t_2)$, where t_1, t_2 are the children of t. f of each leaf is 1, and f of the root is n, therefore there exists t such that $\frac{n}{3} \le f(t) \le \frac{2n}{3}$.)

First observe that the induced subgraph on $v(t)$, and the induced subgraph on $V \setminus v(t)$, both have carving width at most W, and at most $\frac{2}{3}n$ vertices by assumption. Therefore by induction hypothesis, Tseitin tautologies on these graphs have refutations of rank at most $W(\log_{\frac{3}{2}} n-1)$. This implies that every clause containing exactly the variables of $Cut(t)$ has a tree-like derivation of rank at most $W(\log_{\frac{3}{2}} n-1)$, since any such clause falsifies at least one component of $G\backslash Cut(t)$.

If we may start with all clauses on exactly $Cut(t)$, it is easy to see that there is a tree-like refutation of rank $|Cut(t)| \leq W$, by branching on each variable of $Cut(t)$ in succession. By replacing the leaves in this refutation with corresponding small rank tree-like derivations of these clauses, we get a tree-like refutation of rank at most $W \log_{\frac{3}{2}} n$ as desired. □

It is worth pointing out that while this lemma shows that small space proofs exist, it generally does not yield a small space method for finding them. The results of [2] show that it is possible to find a good carving or branch decomposition in small space; however, they left open the question of how to exploit this to solve small width SAT instances efficiently in small space. Important progress has been made by [4], [12].

Some algorithmic upper bounds from the literature

While we have described two different resolution *proofs* for Tseitin tautologies, it remains to be seen whether these upper bounds really correspond to the executions of actual algorithms. There has been a series of SAT algorithms [2, 4, 12] based on branch width or tree width that achieve this.

CLAIM A.9. *The branch-width based algorithm of Aleknovich and Razborov [2] gives an efficient implementation of Lemma 2.7, matching the space and time usage up to constants in the exponent. A recent algorithm of [12] achieves this as well.*

CLAIM A.10. *The small space variation of [12] achieves a runtime matching the size of the proof in Lemma 2.8, up to constants in the exponent, and achieving memory usage which is polynomial in $n = |G|$. The branch width-based algorithm [4] achieves memory usage which is near linear in $|G|$, after a near optimal branch decomposition has been computed. Both results essentially appeal to the branch decomposition routine of [2] to achieve their results.*

A further contribution of [12] is to provide a smooth interpolation between the high space parameters and the low space parameters. Their results are phrased in terms of treewidth parameterized SAT, and they show that time T and space S are feasible when

$$\alpha\left(\log T/TW(\phi)\log|\phi|\right) + \beta\log S \geq TW(\phi) + \gamma\log|\phi|,$$

for appropriate constants α, β, and γ. They conjecture that this essentially cannot be improved. Specifically, they conjecture that any SAT algorithm that runs in time T with $\log T = o(TW(\phi)\log(|\phi|))$ requires space exponential in $TW(\phi)$.

We note that one way to prove that this is true, at least for backtracking algorithms, would be to improve the lower bound result given in Lemma 3.2 for Tseitin tautologies, by improving the exponent to match the small space upper bounds.