Electronic Colloquium on Computational Complexity, Report No. 86 (2023)

Random $(\log n)$-CNF are Hard for Cutting Planes (Again)

Dmitry Sokolov

École Polytechnique Fédérale de Lausanne

June 8, 2023

Abstract

The random Δ-CNF model is one of the most important distribution over Δ-SAT instances. It is closely connected to various areas of computer science, statistical physics, and is a benchmark for satisfiability algorithms. Fleming, Pankratov, Pitassi, and Robere [Fle+22] and independently Hrubeš and Pudlák [HP17] showed that when $\Delta = \Theta(\log n)$, any Cutting Planes proof for random Δ-CNF on n variables requires size $2^{n/\mathsf{polylog} n}$ in the regime where the number of clauses guarantees that the formula is unsatisfiable with high probability. In this paper we show tight lower bound $2^{\Omega(n)}$ on size CP-proofs for random $(\log n)$-CNF formulas. Moreover, our proof is much simpler and self-contained in contrast with previous results based on Jukna's lower bound for monotone circuits.

1 Introduction

Proof complexity studies whether there are efficient certificates (or proofs) for the unsatisfiability of boolean formulas. The non-existence of such proofs in any proof system would separate classes **NP** and **coNP**. According to Cook's program, the idea is to prove lower bounds for stronger and stronger proof systems, so eventually, we would be able to do it in a general case.

At the current moment we do not have explicit candidates of hard families of unsatisfiable formulas for all proof systems. And the important problem here is that explicit unsatisfiable formulas are usually accompanied by *mathematical reasoning* of unsatisfiability, and these reasonings one may translate into formal proofs in some strong enough proof system. However, the situation is different in the case of distribution over formulas that are unsatisfiable with high probability. Candidates of this form are actively studied [CS88; Gri01; Bea+02; AR03; Ale+04; FKO06; Ale11; MT14; Raz15; HP17; Ats+18; Sok20; SS22; Fle+22]. And one of the most popular candidates distribution that generates hard formulas for all proof systems are random Δ-CNF formulas.

Definition 1.1

Let $\mathfrak{F}(m, n, \Delta)$ denote the distribution of random Δ-CNF on n variables obtained by sampling m clauses (out of the $\binom{n}{\Delta}2^{\Delta}$ possible clauses) uniformly at random with repetitions.

The famous result of Chvátal–Szemerédi says if we pick a formula from with distribution with proper parameters the resulting formula will be unsatisfiable with high probability.

Theorem 1.2 [Chvátal–Szemerédi, [CS88]]

For any $\Delta \geq 3$ whp $\varphi \sim \mathfrak{F}(m, n, \Delta)$ is unsatisfiable if $m \geq \ln 2 \cdot 2^{\Delta} n$.

Formal conjectures were formulated by Feige [Fei02]: no polynomial time algorithm may *prove* whp the unsatisfiability of a random $\mathcal{O}(1)$-CNF formula with arbitrary large constant clause density. Assuming Feige's conjecture it is known that some problems are hard to approximate: vertex covering, DNF PAC learning, etc.

ISSN 1433-8092

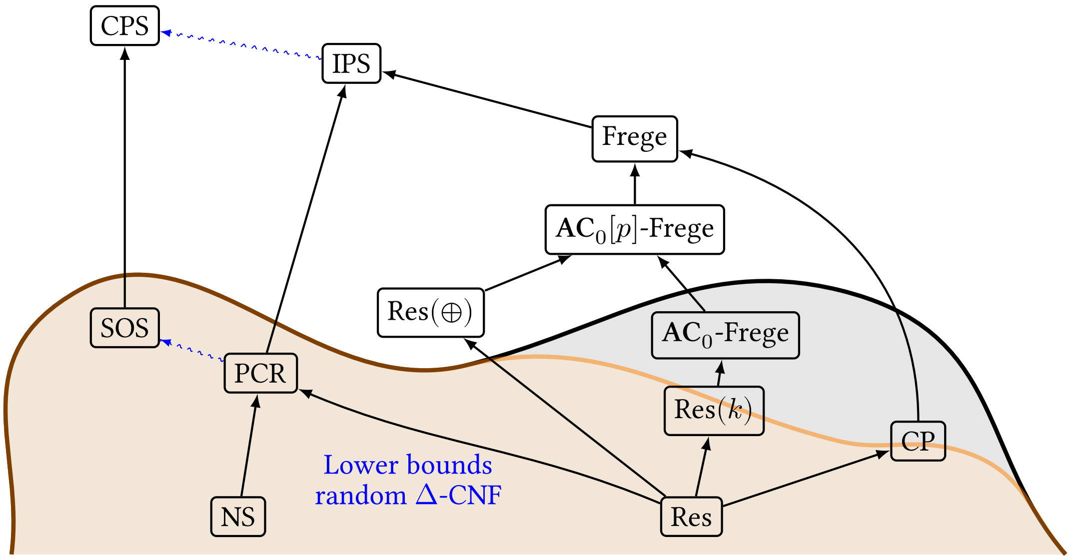

Figure 1: State of the art of lower bounds in proof complexity.

Related results. The state of art results in proof complexity are presented in figure 1. We know some lower bounds for proof systems in shaded region. The second line indicates proof systems for which we know also lower bounds for random Δ-CNF formulas. In particular, we know lower bounds for random Δ-CNF formulas in:

- Resolution (Chvátal–Szemerédi, [CS88]);
- Polynomial Calculus (Ben-Sasson, Impagliazzo [BI99]);
- Sum-of-Squares (Grigoriev, [Gri01]).

The notable exception is an $\mathbf{AC}_0$-Frege proof system for which current techniques require extremely structured formulas and do not allow to deal with random CNFs.

There are two proof systems on the frontier: Cutting Planes and Res$[k]$ (k-DNF Resolution). Exponential lower bounds on Res$[k]$-proofs for random CNF formulas was given by Atserias, Bonet and Esteban [ABE02], the result was significantly improved by Segerlind, Buss, Impagliazzo [SBI04] and by Alekhnovich [Ale11]. Recently these results were unified and generalized by Sofronova and Sokolov [SS22]. To show lower bounds for $k \gg \sqrt{\log n}$ is an open problem. Lower bound for Cutting Planes was shown by Hrubeš and Pudlák [HP17] and independently by Fleming, Pankratov, Pitassi, and Robere [Fle+22]. We give an overview of the technique that was used in these papers.

1.1 Technique and Results

All current techniques for proving lower bounds on Cutting Planes proofs can be divided into two classes: interpolation and lifting.

Lifting works for structured formulas [Gar+18; Göö+20] and at the current moment this approach is not useful for random Δ-CNFs and many other classes of formulas. We refer readers to papers [Gar+18] and [Lov+22]. At the same time we notice that proving lifting-like theorem for symmetric gadgets like *inner product* may also allow advance with random formulas.

Interpolation technique based on Craig's Interpolation Theorem. In other words: let $\varphi(x, y, z) := A(x, z) \vee B(z, y)$ be an unsatisfiable CNF formula, then one can define a function $f(z)$ that for given assignment to variables z says which formula A or B is unsatisfiable (we skip the question of what happens if both A and B are unsatisfiable for simplicity). Krajíček [Kra97] showed that if $\varphi(x, y, z)$ has *monotone encoding* and has an efficient proof in Cutting Planes with bounded coefficients then $f(z)$ can be computed by an efficient monotone circuit. And based on known lower bounds for monotone circuits Krajíček showed lower bounds for Cutting Planes with bounded coefficients. The restriction was removed by Pudlák [Pud97] who showed the first lower bound for the full version of Cutting Planes.

Random CNF formula φ does not have the structure required by the interpolation technique, however, in both papers [HP17] and [Fle+22] authors suggested an adaptation of this technique. The general plan for proving lower bounds in these papers consists of the following steps.

1. Choose a monotone function f_φ associated with the formula φ.
2. Show that if there is a small CP-proof for φ then there is a small *real monotone circuit* for f_φ.
3. Use Jukna's criteria to show that there are no small real monotone circuits for chosen f_φ.

There are several known ways how to associate a monotone function with a formula in a natural way, see for example [GP18]. In terms of communication complexity these methods are based on the fact that "monotone Karchmer–Wigderson relation is complete" (we refer readers to [RGR22] for more details). Usage of Jukna's criteria in [HP17; Fle+22] requires a precise description of the f_φ that requires some technical job and ideas. In these papers, authors showed that whp smallest CP-proof of random $\mathcal{O}(\log n)$-formula has size $2^{n/\mathsf{polylog} n}$. In this paper we show a much simpler proof of the stronger result.

Theorem 1.3 [See also Theorem 4.2]

There is a constant $c > 0$ such that if $\varphi \sim \mathfrak{F}(m, n, \Delta)$ where $m = \mathcal{O}\left(n 2^{\Delta}\right)$ and $\Delta \geq c \log n$, then whp every semantic CP-proof of φ has size $2^{\Omega(n)}$.

To show this Theorem we modify the general plan from papers [HP17] and [Fle+22]. First, we notice, that all extractions of the function f_φ utilize *dag-like communication protocols* as an intermediate constructions either in explicit [Fle+22] or implicit way [HP17; Pud97]. The notion of dag-like protocols formally was introduced by Sokolov in [Sok17] as a simplification of communication PLS games introduced by Razborov [Raz95] and simplified by Pudlák [Pud10] (the restricted version was independently introduced by Hrubeš and Pudlák [HP18]). Instead of extraction of function the f_φ we stop at the intermediate step, create the dag-like communication protocol, and give a combinatorial analysis of this protocol. Hence the full plan is the following.

1. We use the lemma from [Sok17] and say that if there is a small CP-proof for φ then there is a small *real dag-like communication protocol* for the *Unsatisfied clause search problem* (for the sake of completeness we give the proof in the Appendix A.4).
2. We use general idea of the *bottleneck counting argument* [HC99] (see also [Sok17]) and show that there is no small real dag-like communication protocol for the Unsatisfied clause search problem.

As an intermediate step inside bottleneck counting we introduce a 2-*dimensional width* measure for the Cutting Planes proofs. We believe that this measure is of independent interest.

Remarks. In fact, in Theorem 1.3 it is allowed to have slightly large clause density, namely $m = n^{1+\delta} 2^{\Delta}$ for small enough δ. But due to saving the simplicity of computations we do not try to reach the optimal parameters.

The notion of 2-width is related to the notion of *fences* from the papers [HC99]. But this connection passed through a reduction between unsatisfied clause search problem and monotone Karchmer–Wigderson relation.

2 Preliminaries

Denote by $\mathrm{H}(x) := x \log x - \frac{1}{x} \log \frac{1}{x}$ the **binary entropy** function. We use the symbol $\sqcup$ for the disjoint union.

With an unsatisfied CNF formula φ on variables of disjoint union of sets V_x and V_y we associate an **unsatisfied clause search problem** $\mathsf{Search}_\varphi \subseteq X \times Y \times \mathcal{O}$ where:

- X is a set of assignments with support V_x, Y is a set of assignments with support V_y;
- $\mathcal{O}$ is a set of clauses of φ;
- $(x, y, o) \in \mathsf{Search}_\varphi$ iff clause $o \in \varphi$ is not satisfied by assignments x and y.

Communication protocols and triangles. Consider a bipartite input domain $X \times Y$. A **triangle** $T \subseteq X \times Y$ is a set that can be written as $T := \{(x, y) \in X \times Y \mid a_T(x) < b_T(y)\}$ for some labeling $a_T : X \to \mathbb{R}$ of the set X and labelling $b_T : Y \to \mathbb{R}$ of the set Y by real numbers.

For a triangle $T \subseteq X \times Y$ and $x \in X$ let $T^x := \{(x, y) \in T \mid y \in Y\}$ be a **horizontal cut** and for $y \in Y$ let $T^y := \{(x, y) \in T \mid x \in Y\}$ be a **vertical cut**.

A **triangle-dag** (aka **real dag-like communication protocol**) for a search problem $S \subseteq X \times Y \times \mathcal{O}$ is a directed acyclic graph H of fan-out at most 2 where each node h is associated with a triangle $T_h \subseteq X \times Y$ satisfying the following:

root: there is a distinguished root node r (fan-in 0), and $T_r = X \times Y$;

non-leaves: for each non-leaf node h with children u, u', we have $T_h \subseteq T_u \cup T_{u'}$;

leaves: each leaf node h is labeled with an output $o_h \in \mathcal{O}$ such that $T_h \subseteq S^{-1}(o_v)$.

Expanders. We use the following notation: $\mathrm{N}_G(S)$ is the set of neighbours of the set of vertices S in the graph G, i.e. the set $\{v \in V \mid v \text{ share an edge with some } u \in S\}$ where V is the set of vertices of G. We omit the index G if the graph is evident from the context.

A bipartite graph $G := (L, R, E)$ is an (r, Δ, c)-**expander** if all vertices $u \in L$ have degree at most Δ and for all sets $S \subseteq L$, $|S| \le r$, it holds that $|\mathrm{N}(S)| \ge c \cdot |S|$.

Cutting Planes. We consider a semantic version of the Cutting Planes (CP) proof system [CCT87; Hru13].

A proof in **semantic** CP for CNF formula φ is a sequence of linear inequalities with real coefficients $C_1, C_2, \dots, C_\ell$, such that C_ℓ is the trivially unsatisfiable inequality $0 \ge 1$ and C_i can be obtained by one of the following rules:

- C_i is a linear inequality that encodes a clause of formula φ;
- C_i semantically follows on $\{0, 1\}$ values from $C_j \wedge C_k$ where $j, k < i$.

The **size of proof** is the number of inequalities ℓ.

3 Formulas and Partitions

Witl a boolean formula φ we associate a **dependency graph** $G := (U, V, E)$ in a natural way. There are two bijections: between clauses of φ and verties from U, and between variables of φ and vetices from V. Edge $(u, v) \in E$ iff variable v is appear in the clause u.

A well-known fact is the that dependency graph of a random CNF formula is an expander.

Lemma 3.1

Let $\Delta := c \log n$, $m \le \alpha n 2^\Delta$, for some constants $\alpha > 0, c > 0$. For any constant $\varepsilon > 0$ there is a constant $\kappa > 0$ such that whp for $r := \kappa \cdot \frac{n}{\Delta}$ a dependency graph of $\varphi \sim \mathfrak{F}(m, n, \Delta)$ is an $(r, \Delta, (1 - \varepsilon)\Delta)$-expander.

Proof. For proof see Appendix A.2. □

The following notion is of technical nature, but it will be useful for the main theorem. Let φ be a Δ-CNF formula over boolean variables from a set Z. We say that a partition $Z := V_x \sqcup V_y$ is δ-**good** iff φ can be represented as $\psi \land \psi_x \land \psi_y$ such that:

- $||V_x| - |V_y|| \leq 10\sqrt{|Z|}$;
- each clause $C \in \psi$ contains at least $\delta\Delta$ variables from V_x and at least $\delta\Delta$ variables from V_y;
- $\Pr[\psi_x|_\rho = 1] \geq 0.9$, where ρ is taken uniformly at random over all assignments with support V_x;
- $\Pr[\psi_y|_\rho = 1] \geq 0.9$, where ρ is taken uniformly at random over all assignments with support V_y.

Lemma 3.2

For every constants $\alpha > 0, c > 1$, if $\varphi \sim \mathfrak{F}(m, n, \Delta)$ where $m = \alpha n 2^\Delta$ and $\Delta \geq c \log n$ there exists δ-good partition of variables $V_x \sqcup V_y$ of φ for any δ such that $c > (1 - \delta - \mathrm{H}(\delta))^{-1}$.

Proof. For proof see Appendix A.3. ☐

4 Main Theorem

In this section we show the main result. We start with a technical theorem.

Theorem 4.1

Let φ be a CNF formula over variables from a set Z where $|Z| = n$. If dependency graph of φ is $(r, \Delta, (1 - \delta/2)\Delta)$-expander and there is δ-good partition $Z := V_x \sqcup V_y$ then any triangle-dag for Search_φ has size at least $2^{k\Delta\delta/4 - 10\sqrt{n}}$ where $k := \min(r, 2^{\Delta\delta/8})$.

Here we may think that dependency graph of a given formula φ defines the expansion parameter and hence it defines δ. For fixed δ we know how good the partition of variables should be. And parameter k depends on parameter of the partition.

We defer the proof of this Theorem to the Section 4.1. And start with the application to random formulas.

Theorem 4.2 [Reformulation of Theorem 1.3]

For constants $\alpha > \ln 2$ and $c > 1$, if $\varphi \sim \mathfrak{F}(m, n, \Delta)$ for $m = \alpha n 2^\Delta$ and $\Delta \geq c \log n$, then every semantic CP-proof of φ has size $2^{n^{\Omega(1)}}$ whp over the choice of φ. Moreover, if $c > 800$ then every semantic CP-proof of φ has size $2^{\Omega(n)}$ whp over the choice of φ.

Proof. By Lemma A.3 instead of considering CP-proofs we show lower bound for triangle-dags for Search_φ.

Fix $\delta > 0$ such that $c(1 - \delta - \mathrm{H}(\delta)) > 1$. Whp by Lemma 3.1 the dependency graph of φ is an $(r, \Delta, (1 - \delta/2)\Delta)$-expander where $r = \kappa\frac{n}{\Delta}$ for some constant $\kappa > 0$. By Lemma 3.2 there exists δ-good partition of variables, hence the statement follows from Theorem 4.1 since $\min(r, 2^{\Delta\delta/8}) = n^{\Omega(1)}$.

For the "moreover" part we fix $\delta := \frac{1}{100}$. Note that $c(1 - \delta - \mathrm{H}(\delta)) > 1$ and $\min(r, 2^{\Delta\delta/8}) = r = \kappa\frac{n}{\Delta}$ for some constant $\kappa > 0$ by Lemma 3.1. Hence the statement follows from Theorem 4.1. ☐

Let us informally describe the bottneck counting argument for proving Theorem 4.1.

1. Fix some triangle-dag H for Search_φ. Partition of variables gives a representation of φ as a conjunction $\psi \land \psi_x \land \psi_y$. On clauses of ψ the partition is well-behaved, what cannot be said about clauses of ψ_x and ψ_y. By using properties of good partitions we do a *pruning* step and get rid of assignments that do not satisfy ψ_x and ψ_y. Since we deal with the unsatisfied clause search problem then we also can switch from φ to ψ.

2. For each assignement z either with support V_y or with support V_y and each node $h \in H$. We define a measure $w(h,z)$ that we call 2-*width* such that if $w(h,z) > k$ for some chosen threshold k then triangle T_h contain some *useful information* about z. Informally speaking T_h contain useful information about z iff in the formula ψ restricted by z is still hard to find an unsatisfied clause even on set of assignments T_h^z.

3. We show that for most of assignments z there should be some node h that contains useful information about z.

4. At the same time we consider the bottomest node h such that T_h contain useful information about some z. We show that T_h may contain useful information only about few assignmets z' (since it is the bottomest for z).

4.1 Proof of Theorem 4.1

4.1.1 Pruning

Let C' be a set of clauses appearing in φ. Remind that we fix some δ-good partition of variables $V_x \sqcup V_y$. Let X' be a set of assignments with support V_x, and Y' be a set of assignments with support V_y.

Consider a triangle-dag H' for $\text{Search}_\varphi \subseteq X' \times Y' \times \mathcal{C}'$. We start by *pruning* the protocol and erasing all assignments with support V_x that do not satisfy ψ_x and all assignments with support V_y that not satisfy ψ_y. To be more formal, let $X \subseteq X'$ be a set of assignments with support V_x that satisfy ψ_x and $Y \subseteq Y'$ be a set of assignments with support V_y that satisfy ψ_y. We define a protocol H by taking H' and replacing each triangle T'_h by a new triangle $T_h := T'_h \cap (X \times Y)$. Note that H is a triangle-dag that solves Search_φ on $X \times Y$, and moreover all leaves are marked by clauses $\mathcal{C} \subseteq \mathcal{C}'$ that correspond to ψ (since all other clauses are satisfied by any assignment from $X \times Y$). Hence H solves Search_ψ on $X \times Y$. Note that by definition of a good partition $|X| \geq 0.9 \cdot 2^{n-5\sqrt{n}} \geq 2^{n-6\sqrt{n}}$ and $|Y| \geq 2^{n-6\sqrt{n}}$.

Denote by $\mathcal{M} := \{R \subseteq X \times Y \mid R \text{ is a rectangle}, \exists C \in \mathcal{C}, \forall (x,y) \in R, C(x,y) \text{ is unsat}\}$ the collection of monochromatic rectangles of Search_ψ.

4.1.2 Formal Idea

Following the ideas of a bottleneck counting argument we define a partial map $\mu\colon X \cup Y \to H$ such that:

- $|\operatorname{Dom}(\mu)| \geq \frac{1}{4}\min(|X|,|Y|) \geq 2^{n-10\sqrt{n}}$;
- for all $h \in H$: $|\mu^{-1}(h)| \leq 2^{n-k\Delta\delta/4}$.

Hence size of the image of μ (that is the size of H) is as desired.

We start with a definition of a 2-**width** complexity measure $w\colon H \times (X \cup Y) \to \mathbb{N}$ that helps us to define the mapping μ. We define w as follows for all $h \in H$ and all $z \in X \cup Y$:

$$w(h,z) := \min(\mathcal{M}' \subseteq \mathcal{M} \mid \mathcal{M}' \text{ is a covering of } T_h^z).$$

See the Figure 2 for the example.

Remark 4.3

Measure w is **semi-additive** wrt second argument, i.e. for all $z \in X \cup Y$, $w(h,z) \leq w(h',z) + w(h'',z)$ where h', h'' are children of h.

Proof. Note that $T_h \subseteq T_{h'} \cup T_{h''}$ hence $T_h^z \subseteq T_{h'}^z \cup T_{h''}^z$. Thus union of monochromatic coverings of $T_{h'}^z$ and $T_{h''}^z$ are also covering of T_h^z and the observation follows. □

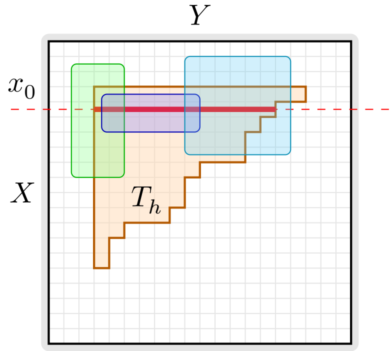

Figure 2: "Width"-measure. The covering witnesses the fact $w(h, x_0) \leq 3$

Now we describe the construction of μ, see Algorithm 1. Informally, for each vertex h in topological order starting from leaves we put all $z \in X \cup Y$ such that $w(h, z) > k$ and erase z from the universe and repeat this process.

Algorithm 1 Definition of μ

```
for h ∈ H in topological order starting from leaves do
    for x ∈ X do
        if w(h, x) > k then
            μ(x) := h.
            Erase the line {x} × Y from triangles T_h for all h ∈ H.
    for y ∈ Y do
        if w(h, y) > k then
            μ(y) := h.
            Erase the line X × {y} from triangles T_h for all h ∈ H.
```

Remark 4.4

1. In the beginning of the iteration of Algorithm 1 in node h: for all $z \in X \cup Y$ it holds that $w(h, z) \leq 2k$.
2. In the end of the iteration of Algorithm 1 in node h: for all $z \in X \cup Y$ it holds that $w(h, z) \leq k$.

Proof. The second observation follows from the description of Algorithm 1 and the fact that after erasing any point from triangle T_h the measure w may only decrease.

The first observation follows from the Remark 4.3 and the fact that we process the node h when we have already processed all the children of h. □

To conclude the proof we show the required properties of μ.

4.1.3 Size of Domain

In this section we show that $|\mathrm{Dom}(\mu)| \geq \frac{1}{2}\min(|X|, |Y|)$. For the sake of contradiction assume that $|\mathrm{Dom}(\mu)| < \frac{1}{2}\min(|X|, |Y|)$. In the remainder of the section we analyse the triangle protocol H after the application of Algorithm 1.

Note that after the application of Algorithm 1 in the root r of H we are left with a triangle (that is also a rectangle) $T_r = X_r \times Y_r \subseteq X \times Y$ that consists of pairs (x, y) such that $x \notin \mathrm{Dom}(\mu)$ and $y \notin \mathrm{Dom}(\mu)$. By our assumption $|X_r| > \frac{1}{2}|X|$ and $|Y_r| > \frac{1}{2}|Y|$. For any point $x_0 \in X_r$ it holds that $w(r, x_0) \leq k$ by Remark 4.4, and hence there are at most k monochromatic rectangles that cover the line $T_r^{x_0} = \{x_0\} \times Y_r$.

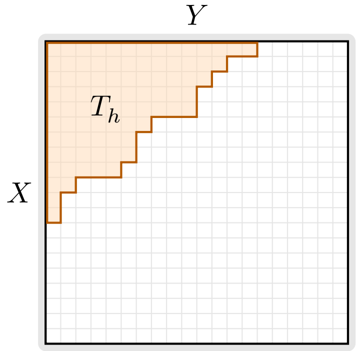

Figure 3: Reordering of rows and columns

We recall that monochromatic rectangles consist of points that violate some specific clause of ψ. Thus the fact that the line $T_r^{x_0}$ can be covered by at most k monochromatic rectangles implies that there is a set S of at most k clauses of the formula ψ such that any point $y \in Y_r$ does not satisfy at least one clause in S. At the same time, pick a point y from Y' uniformly at random:

$$\begin{aligned} \Pr_{y \sim Y'}[y \in Y_r] \leq \\ \Pr_{y \sim Y'}[y \text{ does not satisfy some clause } C \in S] \leq \\ \sum_{C \in S} \Pr_{y \sim Y'}[y \text{ does not satisfy } C] \leq \\ |S| \cdot \max_{C \in S} \Pr_{y \sim Y'}[y \text{ does not satisfy } C] \leq \\ k \cdot 2^{-\delta\Delta}, \end{aligned}$$

where the last inequality holds since each clause of ψ has at least $\delta\Delta$ of V_y variables. Hence

$$|Y_r| \leq k \cdot 2^{-\delta\Delta}|Y'| \leq \frac{10}{9} 2^{\delta\Delta/8} \cdot 2^{-\delta\Delta}|Y| \leq \frac{10}{9} 2^{-\frac{7}{8}\delta\Delta}|Y| \leq \frac{|Y|}{2}.$$

This is a contradiction with the assumption.

4.1.4 Size of Preimage

In this section we show that for all $h \in H$: $|\mu^{-1}(h)| \leq 2^{n-k\Delta\delta/4}$.

Pick some vertex $h \in H$. We consider a situation in the beginning of the iteration of Algorithm 1 in node h and fix this time moment for the rest of the section. By Remark 4.4 we know that for all $z \in X \cup Y$, $w(h, z) \leq 2k$. By using this fact we estimate the number of $x \in X$ that can be mapped into the vertex h by μ, or in other words, the number of $x \in X$ such that $w(h, x) > k$. In order to realize this we build a collection $\mathcal{P}$ of *potential* monochromatic coverings of size k and count the number of $x \in X$ such that T_h^x is not covered by any covering in our collection $\mathcal{P}$. By analogy the same counting holds for the set Y.

Let $X_0 \subseteq X$ be a set of points x such that T_h^x is not empty. The set $X \setminus X_0$ is not interesting for us since $w(h, x') = 0$ for all $x' \in X \setminus X_0$. Let $T_h = \{(x, y) \subseteq X \times Y \mid a_T(x) < b_T(y)\}$. We sort $x \in X$ wrt to increasing order of $a_T(x)$ and we sort $y \in Y$ wrt to decreasing order of $b_T(x)$, see the resulting triangle in Figure 3. Now we are ready to build the potential monochromatic coverings. During this process we create an auxiliary tree L whose vertices will be marked by subsets (subtriangles) of T_h and all edges are marked by monochromatic rectangles. The collection $\mathcal{P}$ will correspond to the set of root-leaf paths in this tree. We start with the formal construction and give a description after.

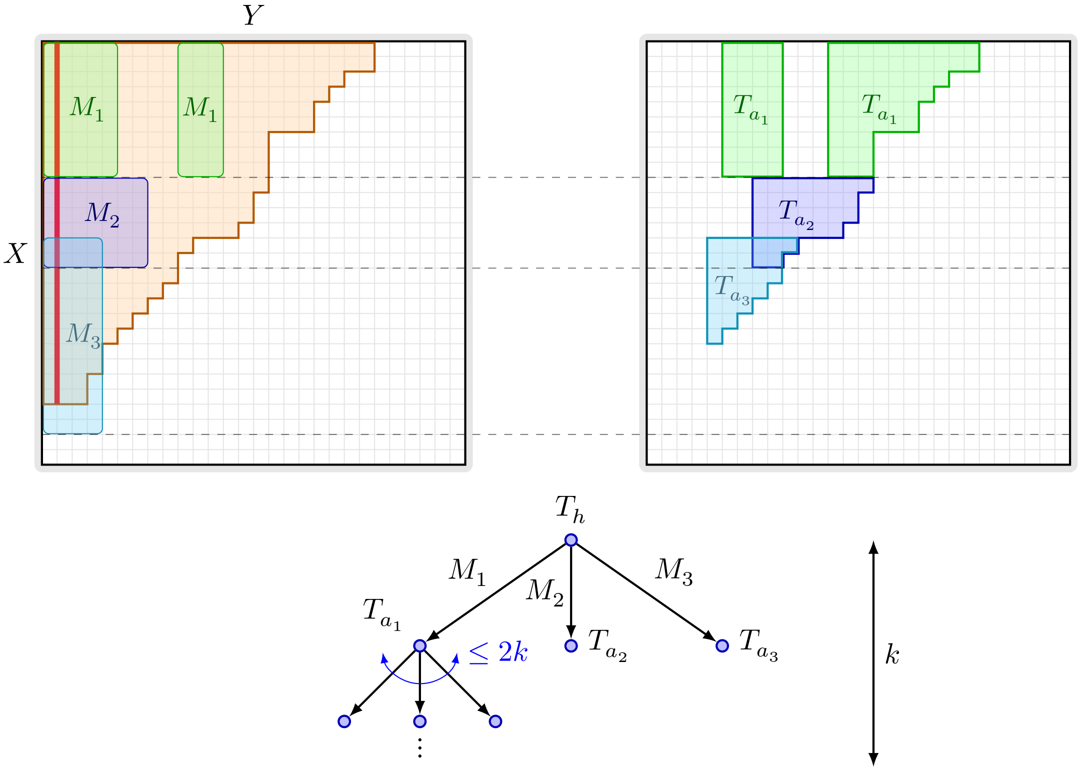

Figure 4: First iteration of Algorithm 2

Algorithm 2 Construction of $\mathcal{P}$

L consists of a single node r that is labelled by T_h.
$A := \{r\}$ set of active leaves of L.
while A is not empty **do**
  Pick $a \in A$. Erase a from A. Let $T \subseteq T_h$ be the label of a.
  Pick a first y such that $T^y \neq \emptyset$.
  Let $M_1, M_2, M_3, \dots, M_\ell \in \mathcal{M}$ be the smallest covering of T^y ▷ Note that $T^y \subseteq T_h^y$
  for $i \in [\ell]$ **do**
    Let $M_i = X_i \times Y_i$.
    Add a_i in L as a child of a. Mark the edge (a, a_i) by M_i.
    Mark a_i by $T_{a_i} := T \cap (X_i \times (Y \setminus Y_i))$.
    if the height of a_i in T is less than k and $T_{a_i} \neq \emptyset$ **then**
      Add a_i into A.

In this algorithm we start with a triangle T_h and first (in our order) $y \in Y$. We start building our collection $\mathcal{P}$ by considering the smallest monochromatic covering $M_1, M_2, \dots, M_\ell$ of the line T_h^y. For any $x \in X_0$ the line T_h^x intersects with at least one rectangle M_i, or in other words M_i covers part of T_h^x. We divide X_0 into ℓ parts (intersections are allowed) wrt which monochromatic rectangles intersects with T_h^x and deal with each part independently. This division induces division of T_h into ℓ subtriangles $T_{a_1}, \dots, T_{a_\ell}$ such that $T_{a_i} := T_h \cap (X_i \times (Y \setminus Y_i))$, here we also erase parts that are already covered by M_i. Note that $\ell \leq 2k$ since $w(h, y) \leq 2k$. See an example of the first iteration of the Algorithm 2 in Figure 4. We repeat this process for each T_{a_i} independently.

For each $x \in X_0$ we trace a path $P_x \subseteq L$ starting from the root in the natural way: suppose we reach node a with a label T,

- if $T^x = \emptyset$ then stop;

- if a is a leaf then stop;
- pick an edge (a, a') marked by $M = X_M \times Y_M$ such that $x \in X_M$ (if there are more than one edge pick any). Note that such an edge is always guaranteed to exists since the collection $\mathcal{M}$, as defined in Algorithm 2, covers T^y where y is the smallest element of Y according to our order. Hence point (x, y) is in T and it is also covered.

Note two properties of these constructed paths.

1. If an edge (a, a') is marked by M is in P_x then the monochromatic rectangle M intersects with T_h^x, or in other words x does not satisfy a clause of ψ that corresponds to M.
2. If the length of P_x is less than k then T_h^x is covered by the monochromatic rectangles from edges of P_x. Indeed, consider a node $a \in P_x$ with label T and it is child $a' \in P_x$ with label T'. By construction $T^x \setminus (T')^x$ is covered by monochromatic rectangle on the edge (a, a'), or say otherwise, while tracing P_x we step by step cut parts of T_h^x that are covered by monochromatic rectangles on edges of P_x. But note that $(T'')^x = \emptyset$ where T'' is the label of the last vertex of P_x since the length of P_x is less than k and the desired property follows.

Let $\mathcal{P}$ be the set of paths in L from root of size k. A straightforward corollary from the second property is that $w(x, h) \geq k$ implies $P_x \in \mathcal{P}$. Hence the number of $x \in X$ such that $w(x, h) \geq k$ is

$$\sum_{P \in \mathcal{P}} |\{x \in X \mid P \text{ is } P_x\}| \leq |\mathcal{P}| \cdot \max_{P \in \mathcal{P}} |\{x \in X \mid P \text{ is } P_x\}|$$

By Remark 4.4 for any $y \in Y$ there is a monochromatic covering of T_h^y of size at most $2k$. Hence the degree of the tree L is at most $2k$ and there are at most $(2k)^k$ different paths in $\mathcal{P}$. Fix some path $P \in \mathcal{P}$. By the first property there is a set S of size k of clauses of the formula ψ such that if $P = P_x$ for some x then x does not satisfy *any* clause from S. Note that there is at most one assignment with support $\mathrm{N}(S) \cap V_x$ that does not satisfy all clauses from S hence there are at most $2^{n-|\mathrm{N}(S) \cap V_x|}$ different points $x \in X$ that do not satisfy any clause in S.

$$\begin{aligned}
2^{n-|\mathrm{N}(S) \cap V_x|} &= 2^{n-|\mathrm{N}(S) \setminus V_y|} \\
&\leq 2^{n-(|\mathrm{N}(S)|-(1-\delta)\Delta|S|)} && \text{partition } V_x \sqcup V_y \text{ is } \delta\text{-good} \\
&\leq 2^{n-(1-\delta/2)\Delta|S|-(1-\delta)\Delta|S|} && \text{dependency graph is an expander} \\
&\leq 2^{n-|S|\Delta\delta/2}.
\end{aligned}$$

We can use expansion property since $|S| = k$ that is at most r. Hence

$$\max_{P \in \mathcal{P}} |\{x \in X \mid P \text{ is } P_x\}| \leq 2^{n-k\Delta\delta/2}.$$

Altogether, the number of $x \in X$ such that $w(x, h) \geq k$ is at most $2^{n-k\Delta\delta/2+k(\log k+1)} = 2^{n-k(\Delta\delta/2-(\log k+1))}$. An analoguous counting argument showes that the number of $y \in Y$ such that $w(y, h) \geq k$ is bounded by $2^{n-k(\Delta\delta/2-(\log k+1))}$. Hence $|\mu^{-1}(h)| \leq 2^{n-k(\Delta\delta/2-(\log k+1))+1} \leq 2^{n-k(\Delta\delta/2-2\log k)} \leq 2^{n-k\Delta\delta/4}$ as needed.

Acknowledgments

I would like to thank Kilian Risse, Paul Beame, Nikola Galesi and others who forced me to write down this paper. Thank Kilian Risse for error correction. Part of this work was done while the author was visiting the Simons Institute for the Theory of Computing. This work was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00026.

References

[ABE02] Albert Atserias, Maria Luisa Bonet, and Juan Luis Esteban. "Lower Bounds for the Weak Pigeonhole Principle and Random Formulas beyond Resolution." In: *Inf. Comput.* 176.2 (2002), pp. 136–152. DOI: 10.1006/inco.2002.3114. URL: https://doi.org/10.1006/inco.2002.3114.

[Ale+04] Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. "Pseudorandom Generators in Propositional Proof Complexity." In: *SIAM J. Comput.* 34.1 (2004), pp. 67–88. DOI: 10.1137/S0097539701389944. URL: https://doi.org/10.1137/S0097539701389944.

[Ale11] Michael Alekhnovich. "Lower Bounds for k-DNF Resolution on Random 3-CNFs." In: *Comput. Complex.* 20.4 (2011), pp. 597–614. DOI: 10.1007/s00037-011-0026-0. URL: https://doi.org/10.1007/s00037-011-0026-0.

[AR03] Michael Alekhnovich and Alexander A. Razborov. "Lower Bounds for Polynomial Calculus: Non-Binomial Case." In: *Proceedings of the Steklov Institute of Mathematics* 242 (2003). Available at http://people.cs.uchicago.edu/~razborov/files/misha.pdf. Preliminary version in *FOCS '01.*, pp. 18–35.

[Ats+18] Albert Atserias, Ilario Bonacina, Susanna F. de Rezende, Massimo Lauria, Jakob Nordström, and Alexander A. Razborov. "Clique is hard on average for regular resolution." In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018.* Ed. by Ilias Diakonikolas, David Kempe, and Monika Henzinger. ACM, 2018, pp. 866–877. DOI: 10.1145/3188745.3188856. URL: https://doi.org/10.1145/3188745.3188856.

[Bea+02] Paul Beame, Richard M. Karp, Toniann Pitassi, and Michael E. Saks. "The Efficiency of Resolution and Davis–Putnam Procedures." In: *SIAM J. Comput.* 31.4 (2002), pp. 1048–1075. DOI: 10.1137/S0097539700369156. URL: https://doi.org/10.1137/S0097539700369156.

[BI99] Eli Ben-Sasson and Russell Impagliazzo. "Random CNF's are Hard for the Polynomial Calculus." In: *40th Annual Symposium on Foundations of Computer Science, FOCS '99, 17-18 October, 1999, New York, NY, USA.* IEEE Computer Society, 1999, pp. 415–421. DOI: 10.1109/SFFCS.1999.814613. URL: https://doi.org/10.1109/SFFCS.1999.814613.

[CCT87] William J. Cook, Collette R. Coullard, and György Turán. "On the complexity of cutting-plane proofs." In: *Discret. Appl. Math.* 18.1 (1987), pp. 25–38. DOI: 10.1016/0166-218X(87)90039-4. URL: https://doi.org/10.1016/0166-218X(87)90039-4.

[CS88] Vašek Chvátal and Endre Szemerédi. "Many Hard Examples for Resolution." In: *J. ACM* 35.4 (Oct. 1988), pp. 759–768. ISSN: 0004-5411. DOI: 10.1145/48014.48016. URL: http://doi.acm.org/10.1145/48014.48016.

[Fei02] Uriel Feige. "Relations between Average Case Complexity and Approximation Complexity." In: *Proceedings of the 17th Annual IEEE Conference on Computational Complexity, Montréal, Québec, Canada, May 21-24, 2002.* IEEE Computer Society, 2002, p. 5. DOI: 10.1109/CCC.2002.10006. URL: http://doi.ieeecomputersociety.org/10.1109/CCC.2002.10006.

[FKO06] Uriel Feige, Jeong Han Kim, and Eran Ofek. "Witnesses for non-satisfiability of dense random 3CNF formulas." In: *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings.* IEEE Computer Society, 2006, pp. 497–508. DOI: 10.1109/FOCS.2006.78. URL: https://doi.org/10.1109/FOCS.2006.78.

[Fle+22] Noah Fleming, Denis Pankratov, Toniann Pitassi, and Robert Robere. "Random $\Theta(\log n)$-CNFs are Hard for Cutting Planes." In: *J. ACM* 69.3 (2022), 19:1–19:32. DOI: 10.1145/3486680. URL: https://doi.org/10.1145/3486680.

[Gar+18] Ankit Garg, Mika Göös, Pritish Kamath, and Dmitry Sokolov. "Monotone circuit lower bounds from resolution." In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*. Ed. by Ilias Diakonikolas, David Kempe, and Monika Henzinger. ACM, 2018, pp. 902–911. DOI: 10.1145/3188745.3188838. URL: https://doi.org/10.1145/3188745.3188838.

[Göö+20] Mika Göös, Sajin Koroth, Ian Mertz, and Toniann Pitassi. "Automating cutting planes is NP-hard." In: *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*. Ed. by Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy. ACM, 2020, pp. 68–77. DOI: 10.1145/3357713.3384248. URL: https://doi.org/10.1145/3357713.3384248.

[GP18] Mika Göös and Toniann Pitassi. "Communication Lower Bounds via Critical Block Sensitivity." In: *SIAM J. Comput.* 47.5 (2018), pp. 1778–1806. DOI: 10.1137/16M1082007. URL: https://doi.org/10.1137/16M1082007.

[Gri01] Dima Grigoriev. "Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity." In: *Theoretical Computer Science* 259.1 (2001), pp. 613–622. ISSN: 0304-3975. DOI: https://doi.org/10.1016/S0304-3975(00)00157-2. URL: http://www.sciencedirect.com/science/article/pii/S0304397500001572.

[HC99] Armin Haken and Stephen A. Cook. "An Exponential Lower Bound for the Size of Monotone Real Circuits." In: *Journal of Computer and System Sciences* 58.2 (1999), pp. 326–335. ISSN: 0022-0000. DOI: http://dx.doi.org/10.1006/jcss.1998.1617.

[HP17] Pavel Hrubeš and Pavel Pudlák. "Random Formulas, Monotone Circuits, and Interpolation." In: *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*. Ed. by Chris Umans. IEEE Computer Society, 2017, pp. 121–131. DOI: 10.1109/FOCS.2017.20. URL: https://doi.org/10.1109/FOCS.2017.20.

[HP18] Pavel Hrubeš and Pavel Pudlák. "A note on monotone real circuits." In: *Inf. Process. Lett.* 131 (2018), pp. 15–19. DOI: 10.1016/j.ipl.2017.11.002. URL: https://doi.org/10.1016/j.ipl.2017.11.002.

[Hru13] Pavel Hrubeš. "A note on semantic cutting planes." In: *Electronic Colloquium on Computational Complexity (ECCC)* 20 (2013), p. 128. URL: http://eccc.hpi-web.de/report/2013/128.

[Kra97] Jan Krajícek. "Lower Bounds for a Proof System with an Expentential Speed-up over Constant-Depth Frege Systems and over Polynomial Calculus." In: *Mathematical Foundations of Computer Science 1997, 22nd International Symposium, MFCS'97, Bratislava, Slovakia, August 25-29, 1997, Proceedings*. 1997, pp. 85–90. DOI: 10.1007/BFb0029951. URL: https://doi.org/10.1007/BFb0029951.

[Lov+22] Shachar Lovett, Raghu Meka, Ian Mertz, Toniann Pitassi, and Jiapeng Zhang. "Lifting with Sunflowers." In: *13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA*. Ed. by Mark Braverman. Vol. 215. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, 104:1–104:24. DOI: 10.4230/LIPIcs.ITCS.2022.104. URL: https://doi.org/10.4230/LIPIcs.ITCS.2022.104.

[MT14] Sebastian Müller and Iddo Tzameret. "Short propositional refutations for dense random 3CNF formulas." In: *Ann. Pure Appl. Log.* 165.12 (2014), pp. 1864–1918. DOI: 10.1016/j.apal.2014.08.001. URL: https://doi.org/10.1016/j.apal.2014.08.001.

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. ISBN: 978-0-521-83540-4. DOI: 10.1017/CBO9780511813603. URL: https://doi.org/10.1017/CBO9780511813603.

[Pud10] Pavel Pudlák. "On extracting computations from propositional proofs (a survey)." In: *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2010, December 15-18, 2010, Chennai, India.* 2010, pp. 30–41. DOI: 10.4230/LIPIcs.FSTTCS.2010.30. URL: http://dx.doi.org/10.4230/LIPIcs.FSTTCS.2010.30.

[Pud97] Pavel Pudlák. "Lower Bounds for Resolution and Cutting Plane Proofs and Monotone Computations." In: *J. Symb. Log.* 62.3 (1997), pp. 981–998. DOI: 10.2307/2275583. URL: https://doi.org/10.2307/2275583.

[Raz15] Alexander A. Razborov. "Pseudorandom generators hard for k-DNF resolution and polynomial calculus resolution." In: *Ann. of Math.* 181 (2 2015), pp. 415–472. DOI: https://doi.org/10.4007/annals.2015.181.2.1.

[Raz95] Alexander A. Razborov. "Bounded Arithmetic and Lower Bounds in Boolean Complexity." In: *Feasible Mathematics II.* Ed. by Peter Clote and Jeffrey B. Remmel. Boston, MA: Birkhäuser Boston, 1995, pp. 344–386. ISBN: 978-1-4612-2566-9.

[RGR22] Susanna F. de Rezende, Mika Göös, and Robert Robere. "Guest Column: Proofs, Circuits, and Communication." In: *SIGACT News* 53.1 (2022), pp. 59–82. DOI: 10.1145/3532737.3532746. URL: https://doi.org/10.1145/3532737.3532746.

[SBI04] Nathan Segerlind, Samuel R. Buss, and Russell Impagliazzo. "A Switching Lemma for Small Restrictions and Lower Bounds for k-DNF Resolution." In: *SIAM J. Comput.* 33.5 (2004), pp. 1171–1200. DOI: 10.1137/S0097539703428555. URL: https://doi.org/10.1137/S0097539703428555.

[Sok17] Dmitry Sokolov. "Dag-Like Communication and Its Applications." In: *Computer Science - Theory and Applications - 12th International Computer Science Symposium in Russia, CSR 2017, Kazan, Russia, June 8-12, 2017, Proceedings.* Ed. by Pascal Weil. Vol. 10304. Lecture Notes in Computer Science. Springer, 2017, pp. 294–307. DOI: 10.1007/978-3-319-58747-9_26. URL: https://doi.org/10.1007/978-3-319-58747-9%5C_26.

[Sok20] Dmitry Sokolov. "(Semi)Algebraic proofs over ± 1 variables." In: *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020.* Ed. by Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy. ACM, 2020, pp. 78–90. DOI: 10.1145/3357713.3384288. URL: https://doi.org/10.1145/3357713.3384288.

[SS22] Anastasia Sofronova and Dmitry Sokolov. "A Lower Bound for k-DNF Resolution on Random CNF Formulas via Expansion." In: *Electron. Colloquium Comput. Complex.* TR22-054 (2022). ECCC: TR22-054. URL: https://eccc.weizmann.ac.il/report/2022/054.

A Properties of Random Formulas

A.1 Probabilistic Tools

Theorem A.1 [Chernoff bound, [MU05]]

Suppose $Z_1, \dots, Z_n$ are independent random variables taking values in $\{0, 1\}$. Let X denote their sum and let $\mu = \mathrm{E}(X)$ denote the sum's expected value. Then for any $0 < \delta \leq 1$ we have:

$$\Pr[|X - \mu| \geq \delta\mu] \leq \exp\left(-\frac{\delta^2\mu}{3}\right).$$

A.2 Expansion of Random Formulas

For $m, n, \Delta \in \mathbb{N}$, we denote by $\mathfrak{G}(m, n, \Delta)$ the distribution over bipartite graphs with disjoint vertex sets $U := \{u_1, \dots, u_m\}$ and $V := \{v_1, \dots, v_n\}$ where the neighbourhood of a vertex $u \in U$ is chosen by sampling a subset of size Δ uniformly at random from V.

Theorem A.2

Let $\Delta := c\log n$, $m \le \alpha n 2^{\Delta}$, for some constants $\alpha > 0, c > 0$. For any constant $\varepsilon > 0$ there is a constant $\kappa > 0$ such that whp for $r := \kappa \cdot \frac{n}{\Delta}$ a randomly sampled graph $G \sim \mathfrak{G}(m, n, \Delta)$ is an $(r, \Delta, (1-\varepsilon)\Delta)$-expander.

Proof. Let $\varepsilon < 1/2$. We estimate the probability that G is not an $(r, \Delta, (1-\varepsilon)\Delta)$-expander for some parameter r. Let $G := (U, V, E)$. We first estimate the probability that a set $S \subseteq U$ of size at most r violates the expansion. For brevity, let us write $s = |S|$ and $d = (1-\varepsilon)\Delta$. The probability that S violates the expansion can be bounded by:

$$\begin{aligned}
\Pr[|\mathrm{N}\,(S)\,| < ds] &\le \binom{n}{ds} \cdot \left(\frac{\binom{ds}{\Delta}}{\binom{n}{\Delta}}\right)^{s} \\
&\le \binom{n}{ds} \cdot \left(\frac{ds}{n}\right)^{\Delta s} \\
&\le \left[\left(\frac{en}{ds}\right)^{d} \cdot \left(\frac{ds}{n}\right)^{\Delta}\right]^{s}
\end{aligned}$$

Hence, the probability that G is not an expander can be bounded by

$$\begin{aligned}
\Pr[G \text{ is not an expander}] &\le \sum_{s\in[r]} \binom{m}{s} \left[\left(\frac{en}{ds}\right)^{d} \cdot \left(\frac{ds}{n}\right)^{\Delta}\right]^{s} \\
&\le \sum_{s\in[r]} \left(\frac{me}{s}\right)^{s} \left[\left(\frac{en}{ds}\right)^{d} \cdot \left(\frac{ds}{n}\right)^{\Delta}\right]^{s} \\
&\le \sum_{s\in[r]} \left[\frac{me}{s}\left(\frac{en}{ds}\right)^{d} \cdot \left(\frac{ds}{n}\right)^{\Delta}\right]^{s} \\
&\le \sum_{s\in[r]} \left[e^{1+d}\frac{m}{s}\left(\frac{ds}{n}\right)^{\varepsilon\Delta}\right]^{s} \\
&\le \sum_{s\in[r]} \left[e^{1+d}\alpha n 2^{\Delta}\,(\kappa)^{\varepsilon\Delta}\right]^{s} \\
&\le \sum_{s\in[r]} \left[\alpha 2^{(2c+1)\log n}\,(2\kappa^{\varepsilon})^{c\log n}\right]^{s} \\
&\le \sum_{s\in[r]} \left[\alpha\left(2^{3c+1}\kappa^{c\varepsilon}\right)^{\log n}\right]^{s}.
\end{aligned}$$

And if $\kappa < 2^{-(3c+1)/(\varepsilon c)}$ this sum is $o(1)$. □

A.3 Proof of 3.2

Lemma 3.2

For every constants $\alpha > 0, c > 1$, if $\varphi \sim \mathfrak{F}(m, n, \Delta)$ where $m = \alpha n 2^{\Delta}$ and $\Delta \ge c\log n$, then there exists δ-good partition of variables $V_x \sqcup V_y$ of φ for any δ such that $c > (1 - \delta - \mathrm{H}\,(\delta))^{-1}$.

Proof. We add variables into V_x with probability $\frac{1}{2}$ uniformly at random, and put all other variables into V_y respectively. We show that with constant probability the partition is good.

Note that by Chernoff bound:

$$\Pr\left[\left||V_x| - \frac{n}{2}\right| \geq 5\sqrt{n}\right] \leq \exp\left(-\frac{(10/\sqrt{n})^2 \cdot (n/2)}{3}\right) \leq e^{-10}.$$

Hence $\left||V_x| - |V_y|\right| \leq 10\sqrt{n}$ with probability $1 - e^{-10}$.

Let $\psi_x \subseteq \varphi$ consists of all clauses that contain more than $(1-\delta)\Delta$ variables from V_x, and $\psi_y \subseteq \varphi$ is defined by analogy. To show the remaining properties we we show that size of ψ_x and ψ_y are not so big, and random assignment satisfy so small formulas with high probability. We analyze ψ_x and by analogy the same holds for ψ_y.

Consider some clause $C \in \varphi$:

$$\Pr[C \text{ contains at most } \delta\Delta \text{ variables from } V_x] = 2^{-\Delta}\sum_{i=0}^{\delta\Delta}\binom{\Delta}{i} \leq 2^{(\mathrm{H}(\delta)-1)\Delta}.$$

Hence by Markov inequality

$$\Pr\left[|\psi_x| \geq 3m2^{(\mathrm{H}(\delta)-1)\Delta}\right] = \Pr\left[|\psi_x| \geq 3\alpha n 2^{\mathrm{H}(\delta)\Delta}\right] \leq \frac{1}{3}.$$

Now assume that $|\psi_x| \leq 3m2^{(\mathrm{H}(\delta)-1)\Delta}$. Random assignment to variables V_x does not satisfy some clause $C \in \psi_x$ with probability at most $2^{-(1-\delta)\Delta}$ since it contains at least $(1-\delta)\Delta$ variables from V_x. Hence

$$\begin{aligned}\Pr_{a\in\{0,1\}^{|V_x|}}[\exists C \in \psi_x \colon C(a) \text{ is not satisfied}] &\leq 3\alpha n\frac{2^{\mathrm{H}(\delta)\Delta}}{2^{(1-\delta)\Delta}} \\ &= 3\alpha\frac{n}{2^{(1-\delta-\mathrm{H}(\delta))\Delta}} \\ &= 3\alpha\frac{n}{n^{(1-\delta-\mathrm{H}(\delta))c}} \\ &= o(1). && \text{by the choice of } \delta\end{aligned}$$

Hence with probability $\frac{2}{3} - e^{-10} - o(1)$ random assignment is δ-good. ☐

A.4 From Cutting Planes to Communication Protocols

Lemma A.3 [Sokolov, [Sok17]]

Let φ be an unsatisfiable CNF formula on variables $V_x \sqcup V_y$, X be a set of assignments to variables V_x and Y is a set of assignment to variables V_y. If there is a semantic CP-proof for φ of size S then there is a triangle-dag of size S for Search_φ.

Proof. Let graph H of the triangle dag be the graph of the semantic CP proof of the formula φ with inverted edges. Consider a vertex $h \in H$, there is a proof line $f(V_x) + g(V_y) \geq c$ that corresponds to h. We associate with the node h a traingle T_h defined by labelling functions: $a_{T_h}(x) = f(u) - c$ and $b_{T_h}(y) = -g(y)$. Note that $(x, y) \in T_h$ iff $a_{T_h}(x) < b_{T_h}(y)$, hence $f(x) + g(y) < c$, i.e. the inequality is falsified by the assignment (x, y).

The root r of H corresponds to the trivially false inequality $0 \geq 1$, hence the triangle $T_r = X \times Y$. If an assignment satisfies all inequalities in the children $h', h'' \in H$ of some vertex $h \in H$ then this assignment also satisfies the inequality in h. Thus, $T_h \subseteq T_{h'} \cup T_{h''}$.

By contruction in a leaf h we have a triangle that consists of points that violate some clause $C \in \varphi$, we mark the leaf h by C. ☐

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il