Electronic Colloquium on Computational Complexity, Report No. 182 (2022)

Criticality of AC^0-Formulae

Prahladh Harsha * Tulasimohan Molli * Ashutosh Shankar *

Abstract

Rossman [In *Proc. 34th Comput. Complexity Conf.*, 2019] introduced the notion of *criticality*. The criticality of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is the minimum $\lambda \geq 1$ such that for all positive integers t,

$$\Pr_{\rho \sim \mathcal{R}_p}\left[\mathsf{DT}_{\mathrm{depth}}(f|_\rho) \geq t\right] \leq (p\lambda)^t.$$

Håstad's celebrated switching lemma shows that the criticality of any k-DNF is at most $O(k)$. Subsequent improvements to correlation bounds of AC^0-circuits against parity showed that the criticality of any AC^0-*circuit* of size S and depth $d+1$ is at most $O(\log S)^d$ and any *regular* AC^0-*formula* of size S and depth $d+1$ is at most $O(\frac{1}{d}\cdot \log S)^d$. We strengthen these results by showing that the criticality of *any* AC^0-formula (not necessarily regular) of size S and depth $d+1$ is at most $O(\frac{\log S}{d})^d$, resolving a conjecture due to Rossman.

This result also implies Rossman's optimal lower bound on the size of any depth-d AC^0-formula computing parity [Comput. Complexity, 27(2):209–223, 2018.]. Our result implies tight correlation bounds against parity, tight Fourier concentration results and improved #SAT algorithm for AC^0-formulae.

1 Introduction

Understanding the power of various models of computation is the central goal of complexity theory. With respect to small-depth AND-OR circuits, the early works of Furst, Saxe and Sipser [FSS84], Sipser [Sip83], Ajtai [Ajt83], Yao [Yao85] and Håstad [Hås89] using *random restrictions* and Razborov [Raz87] and Smolensky [Smo87] using the *polynomial method* laid out a promising direction.

Furst, Saxe and Sipser [FSS84] and Ajtai [Ajt83] independently proved that the parity function requires super-polynomial sized constant depth AND-OR circuits to compute it. This was then

*Tata Institute of Fundamental Research, Mumbai, India. Email: prahladh@tifr.res.in, tulasimohanm@gmail.com, ashushankar98@gmail.com. Research supported by the Department of Atomic Energy, Government of India, under project 12-R&D-TFR-5.01-0500. Research of second author partially supported through the MATRICS grant MTR/2019/001226 of the Science and Engineering Research Board, Department of Science and Technology, Government of India.

ISSN 1433-8092

later improved by Yao [Yao85] and Håstad [Hås89] who proved that any depth-$(d+1)$, AND-OR circuit computing parity on n bits requires size $2^{n^{\Theta(1/d)}}$. The pièce de résistance of these results is the *switching lemma* method introduced by Furst, Saxe and Sipser [FSS84]. Informally stated, it states that any k-DNF[1] reduces (aka *switches*) to a low-width CNF with high probability, when acted upon by a p-random restriction. Very soon (in Håstad's paper itself [Hås89]), it was discovered that it was more convenient and useful to state the switching lemma in terms of the depth of decision trees. This leads us to Håstad's switching lemma, one of the most celebrated theorems in theoretical computer science. Let f be a k-DNF and $\mathcal{R}_p$ denote the distribution of p-random restrictions ($p \in [0,1]$) where each variable independently is left unrestricted with probability p and otherwise set uniformly to 0 or 1. Then,

$$\Pr_{\rho \sim \mathcal{R}_p} \left[\mathsf{DT}_{\text{depth}}(f|_\rho) \geq s \right] \leq (5pk)^s \,.$$

Despite the immense success of these methods in proving optimal lower bounds for small-depth AND-OR circuits and related models, they did not help in understanding limits of considerably stronger computational models. It was soon discovered that "other techniques" are needed to tackle these stronger models and this was made formal in the *natural proof* approach by Razborov and Rudich [RR97]. About a decade ago, interest in an *improved* switching lemma was revived while trying to understand optimal correlation bounds of small depth AND-OR circuits with the parity function. While the early results of Ajtai [Ajt83] were only able to show a correlation bound of $\exp(-\Omega(n^{1-\varepsilon}))$, Beame, Impagliazzo and Srinivasan [BIS12] proved a considerably smaller correlation bound of $\exp(-\Omega(n/2^{2d(\log S)^{4/5}}))$ for depth-d AND-OR circuits of size S with the parity function over n bits. This was then improved by Impagliazzo, Matthews and Paturi [IMP12] and Håstad [Hås14] who proved the optimal correlation bound of $\exp(-\Omega(n/(\log S)^d))$ for depth-$(d+1)$ AND-OR circuits of size S with the n-bit parity function. Håstad proved this optimal correlation bound by proving the *multi-switching lemma*, a significant strengthening of his earlier switching lemma. The multi-switching lemma is best described in terms of *criticality*, a notion introduced subsequently by Rossman [Ros19].

The *criticality* of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is the minimum $\lambda \geq 1$ such that

$$\Pr_{\rho \sim \mathcal{R}_p} \left[\mathsf{DT}_{\text{depth}}(f|_\rho) \geq s \right] \leq (p\lambda)^s.$$

Thus, Håstad's switching lemma in terms of criticality states that k-DNFs (and k-CNFs) have criticality $O(k)$. The Multi-switching lemma (in Rossman's reformulation in terms of criticality) states that a depth-$(d+1)$ AND-OR circuit of size S has criticality $O(\log S)^d$. In the same paper, Rossman [Ros19] showed a stronger result that depth $(d+1)$ *regular* AND-OR *formulae* of size S have criti-

[1] A k-DNF is a Boolean formula in disjunctive normal form (DNF) in which each term has at most k literals. A k-CNF is defined similarly.

cality $O\left(\frac{\log S}{d}\right)^d$, where regular means that all gates at the same height have equal fan-in. Parallel and independent of this line of work involving correlation bounds with parity, Rossman [Ros18] showed that any depth-$(d+1)$ AND-OR formula (not necessarily regular) that computes the n-bit parity requires size at least $2^{\Omega(d(n^{1/d}-1))}$. Our main result is a common strengthening (and unification) of all the above mentioned results, where we prove that any (not necessarily regular) depth $(d+1)$ AND-OR formula of size S has criticality $O\left(\frac{\log S}{d}\right)^d$. More precisely,

Theorem 1.1. *Let F be an AND-OR formula of depth $d+1$ and size at most S, then for any $p \in [0,1]$*

$$\Pr_{\rho\sim\mathcal{R}_p}\left[\mathsf{DT}_{depth}(F|_\rho) \geq s\right] \leq \left(p \cdot O\left(32^d\left(\frac{\log S}{d}+1\right)^d\right)\right)^s.$$

As an immediate corollary of the above criticality result and [Ros19, Theorem 14], we get the following results for general AND-OR formulae of depth $(d+1)$. Rossman had proved similar results for regular AND-OR formulae of depth $(d+1)$ [Ros19].

Corollary 1.2. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be computable by an AND-OR formula of depth $d+1$ and size at most S. Then*

1. *Decision tree size bounds:* $\mathsf{DT}_{size}(f) \leq O\left(2^{\left(1-1/O\left(\frac{1}{d}\log S\right)^d\right)n}\right)$,
2. *Correlation bound with parity:* $\mathrm{Cor}(f,\oplus_n) \leq O\left(2^{-n/O\left(\frac{1}{d}\log S\right)^d}\right)$,
3. *Degree bounds:* $\Pr_{\rho\sim\mathcal{R}_p}\left[\deg(F|_\rho) \geq s\right] \leq \left(p \cdot O\left(32^d\left(\frac{1}{d}\log S+1\right)^d\right)\right)^s$.
4. *ℓ_2-Fourier concentration (Linial-Mansour-Nisan [LMN93]):* $\sum_{S\subseteq[n]\colon |S|\geq k}\widehat{f}(S)^2 \leq 2e\cdot e^{-k/O\left(\frac{1}{d}\log S\right)^d}$,
5. *ℓ_1-Fourier concentration (Tal [Tal17]):* $\sum_{S\subseteq[n]\colon |S|=k}\left|\widehat{f}(S)\right| \leq O\left(\frac{1}{d}\log S\right)^{dk}$.

As indicated before, Theorem 1.1 unifies (and arguably simplifies) all previous results for AC^0 circuits and formulae in this context. It also yields satisfiability results for AC^0-formulae along the lines of the Impagliazzo, Matthews and Paturi result [IMP12] (see Section 6 for more details).

1.1 Proof Overview

The proof of Theorem 1.1 is an adaptation of the proof of the multi-switching lemma due to Håstad [Hås14] and Rossman's proof in the regular case [Ros19]. As is typical in all proofs of the switching lemma, we construct a canonical decision tree (CDT) for the depth-d formula. This CDT under a restriction is constructed in an inductive fashion by progressively refining the restriction. The

main theorem is proved via the following statement (see Lemma 5.5 for the exact statement), which we prove inductively. For any s-length bitstring $a \in \{0,1\}^s$,

$$\Pr_{\boldsymbol{\rho}}[\text{ There exists a path in } \mathrm{CDT}(F,\boldsymbol{\rho}) \text{ labelled by the instruction-set } a \mid \boldsymbol{\rho} \in \mathcal{T}] \leq (p \cdot \lambda(F))^s,$$

where $\mathcal{T}$ is any family of downward-closed set of restrictions and $\lambda(F)$ is the desired criticality bound that we wish to prove. The crucial difference from Rossman's proof in the regular setting is we prove the above statement subject to $\boldsymbol{\rho}$ belonging to any downward-closed set. As in Rossman's proof, the event "There exists a path ...." is broken into several subevents, $\mathcal{E}_t$ for t ranging over a polynomially large set and each of the events $\mathcal{E}_t$ is typically a conjunction of 3 events $\mathcal{A}_t, \mathcal{B}_t$ and $\mathcal{C}_t$. The required probability can then bounded by an expression as follows:

$$\sum_t \Pr[\mathcal{A}_t \cap \mathcal{B}_t \cap \mathcal{C}_t \mid \mathcal{T}] = \sum_t \Pr[\mathcal{A}_t \mid \mathcal{T}] \cdot \Pr[\mathcal{B}_t \mid \mathcal{A}_t \cap \mathcal{T}] \cdot \Pr[\mathcal{C}_t \mid \mathcal{A}_t \cap \mathcal{B}_t \cap \mathcal{T}].$$

The advantage of using conditioning is that some of these intermediate probabilites (c.f., $\Pr[\mathcal{C}_t \mid \mathcal{A}_t \cap \mathcal{B}_t \cap \mathcal{T}]$) can be bound using the inductive assumption provided the conditioned events are themselves downward-closed. The events $\mathcal{A}_t$ and $\mathcal{B}_t$ are chosen such that this is indeed the case. The sum over t is then handled via convexity. The use of downward-closed sets to prove the inductive claim is inspired from Håstad's use of downward-closed sets in his proof of the multi-switching lemma [Hås14]. However, the situation for depth-d formulae is considerably more involved than the DNF/CNF setting and both the choice of the events as well as bounding these conditional probabilities require considerable care and subtlety (see Section 4 and Claim 5.9). The use of downward-closed sets considerably simplifies the proof and yields an arguably simpler proof of the criticality bound, even in the regular setting [Ros19].

Organization

The rest of the paper is organized as follows. We begin with some preliminaries in Section 2, where we recall the standard notions of restrictions, decision trees and introduce variants of these notions such as restriction trees etc, which will be of use later. We then define canonical decision trees for depth-d formulae in Section 3, identical to the corresponding notion in the regular setting due to Rossman [Ros19]. We then demonstrate the downward-closedness of some properties related to CDTs in Section 4 and finally prove the main theorem in Section 5. In Section 6, we use the main lemma to give a randomized #SAT algorithm for arbitrary AC^0 formulae generalizing the corresponding algorithm due to Rossman in the regular setting [Ros19].

2 Preliminaries

For a positive integer $n \in \mathbb{N}$, $[n]$ refers to the set $\{1, 2, \ldots, n\}$. All logarithms in this paper are to base 2.

While studying distributions D over some finite set Σ, we will use bold letters (i.e., $\boldsymbol{\sigma}$) to distinguish a random sample according to D from a fixed element $\sigma \in \Sigma$. Given any distribution D on a finite set Σ, we let $\mu_D \colon S \to [0,1]$ denote the corresponding probability distibution (i.e., $\mu_D(\sigma) = \Pr_{\boldsymbol{\sigma} \sim D}[\boldsymbol{\sigma} = \sigma]$). We will drop the subscript D from μ_D typically.

We begin by recalling the definition of an AC^0-formula.

Definition 2.1 (AC^0-formulae)**.** *Let d be a non-negative integer. Let V be a set of variable indices. An AC^0-formula F of depth-d over variables V is (inductively) defined as follows: A depth-0 formula is the constant 0 or 1 or a literal x_v or $\neg x_v$ where v is a variable index. For $d \geq 1$, a depth-d AC^0-formula is either an OR-formula or an AND-formula which are defined below. A depth-d OR-forumula is of the form $F_1 \vee F_2 \vee \cdots \vee F_m$ where the F_i's are either depth-d' AND-formulae for some $1 \leq d' < d$ or depth-0 formulae. A depth-d AND-formula $F = F_1 \wedge F_2 \cdots \wedge F_m$ is defined similarly.*

Depth-1 AND-formulae and OR-formulae are usually referred to as terms and clauses respectively, while depth-2 AND-formulae and OR-formulae are called DNFs and CNFs respectively.

The size of a formula is given by the number of depth-1 subformulae[2]. More precisely, $\mathrm{size}(F)$ *is inductively defined as*

$$\mathrm{size}(F) := \begin{cases} 0 & \text{if } \mathrm{depth}(F) = 0 \\ 1 & \text{if } \mathrm{depth}(F) = 1 \\ \sum_{i=1}^{m} \mathrm{size}(F_i) & \text{if } F = F_1 \vee F_2 \vee \cdots \vee F_m \text{ or } F_1 \wedge F_2 \cdots \wedge F_m. \end{cases}$$

The variable index set V of a given formula F unless otherwise specified is always assumed to be $[n]$.

We will sometimes identify a formula F with the Boolean function it computes. We say "$F \equiv 1$" if this Boolean function is a tautology and "$F \equiv 0$" if it is a contradiction. ⌟

We will be chiefly concerned with restrictions.

Definition 2.2 (restriction)**.** *Given a variable index set V, a* restriction ρ *is a function $\rho \colon V \to \{0, 1, *\}$ or equivalently a partial function from V to $\{0,1\}$. We refer to the domain of this partial function as* $\mathrm{dom}(\rho)$ *and the remaining set of unrestricted variables, namely $V \setminus \mathrm{dom}(\rho)$, as* $\mathrm{stars}(\rho)$.

We say that two restrictions ρ_1 and ρ_2 are consistent if for every $v \in \mathrm{dom}(\rho_1) \cap \mathrm{dom}(\rho_2)$, we have $\rho(v_1) = \rho(v_2)$.

We can define a partial ordering among restrictions as follows: we say $\rho_1 \preccurlyeq \rho_2$ if (1) $\mathrm{stars}(\rho_1) \subseteq \mathrm{stars}(\rho_2)$ *and (2) ρ_1 and ρ_2 are consistent. In words, ρ_1 only "sets more variables" than ρ_2. Sometimes, we*

[2]Traditionally, the size is defined by the number of leaves or depth-0 formulas but for this paper, it would be more convenient to work with this definition.

will only be interested in this order with respect to a particular subset S of the variable index set V. In this case, we say

$$\rho_1 \preccurlyeq_S \rho_2 \text{ if (1) } \mathrm{stars}(\rho_1) \cap S \subset \mathrm{stars}(\rho_2) \cap S \text{ and (2) } \rho_1 \text{ and } \rho_2 \text{ are consistent.}$$

Given a formula F and a restriction ρ, the restricted formula $F|_\rho$ refers to the formula obtained by relabeling literals involving variables indices in $\mathrm{dom}(\rho)$ according to ρ (we make no further simplification to the formula). Given two consistent restrictions ρ_1, ρ_2, $F|_{\rho_1,\rho_2}$ refers to the formula $(F|_{\rho_1})|_{\rho_2}$ (which is identical to $(F|_{\rho_2})|_{\rho_1}$). ⌟

It will sometimes be convenient to consider an ordering among the variables in the domain of a restriction, especially when studying restrictions arising from decision trees.

Definition 2.3 (ordered restriction)**.** *An* ordered restriction *on a variable set V is a sequence of the form $\alpha = (x_{v_1} \to b_1, \ldots, x_{v_t} \to b_t)$ where $t \in \mathbb{N}$, $b_i \in \{0,1\}$, and $v_1, \ldots, v_t$ are distinct elements of V. We will use $\mathrm{dom}(\alpha)$ to refer to the set $\{v_1, \ldots, v_t\}$. (We will typically use α or β for ordered restrictions.)*

Any ordered restriction can be interpreted as a restriction ρ with $\mathrm{dom}(\rho) = \{v_1, \ldots, v_t\}$. Similarly, given a restriction ρ on V, and an ordering on $\mathrm{dom}(\rho)$, we have a natural representation of ρ as an ordered restriction on V. ⌟

The following is the standard definition of a decision tree except that we allow the internal nodes of the tree to have (out-)degree either 1 or 2.

Definition 2.4 (decision tree)**.** *A* decision tree *is a a finite rooted binary tree where*

- *each internal node is labelled by a variable, has one or two children and the edges to its children have distinct labels from the set $\{0,1\}$,*
- *the leaves are labelled by 0 or 1, and*
- *the variables appearing in any root-to-leaf path are distinct.*

For each node v (including leaf node), the root-to-node path in the decision tree naturally corresponds to an ordered restriction, which we denote by α_v^Γ (this restriction is non-trivial for every non-root node).

The depth *of a decision tree T, denoted by $\mathrm{depth}(T)$, is defined as the maximum number of degree-2 nodes along any root-to leaf path in T. Note that this may be shorter than the length of the corresponding ordered restriction, which includes the degree-1 nodes also.*

A decision tree is said to compute a Boolean function $F\colon \{0,1\}^V \to \{0,1\}$ under a restriction ρ if the following conditions hold

- *any internal vertex labelled by a variable index, say v, that is in $\mathrm{dom}(\rho)$ has degree one, with the edge to the only child labelled with $\rho(v)$,*
- *any internal vertex labelled by a variable index in $\mathrm{stars}(\rho)$ has degree two and*

- *for every leaf v, we have $F|_{\rho,\alpha_v} \equiv label(v)$.* ⌟

An "honest-to-god" decision tree (with all internal nodes having degree two) can be obtained from the above decision tree by contracting the degree-1 edges. However, we will find it convenient to keep this information about degree-1 nodes while constructing decision trees for functions under a restriction. Note that if a decision tree T computes a function F under the restriction ρ, then the contracted decision tree T' computes the function $F|_\rho$.

To prove the criticality bound for a given formula F, we construct a canonical decision tree (CDT) for F under a (random) restriction ρ. This CDT is constructed in an inductive fashion by constructing the CDT's for F's subformulae first and then using these CDT's to construct F's CDT. While doing so, we progressively refine the restriction so that the final restriction under which the CDT is constructed is the target restriction ρ. This naturally leads us to the notion of *restriction trees*, which is essentially a family of restrictions, one for each subformula of a given formula, such that the restrictions get refined as we move from child to parent in the formula tree.

Definition 2.5 (restriction tree)**.** *Let F be a formula on the variable index set V and T_F the set of all subformulae of F. The elements of T_F have a natural bijection with the underlying formula tree of F. A restriction tree for F, denoted by $\tilde{\rho}$, associates a restriction with each node in T_F, formally $\tilde{\rho}\colon T_F \to \{0,1,*\}^V$, such that for $G, H \in T_F$ where G is a subformula of H, we have $\tilde{\rho}(H) \preccurlyeq \tilde{\rho}(G)$. In other words, the sequence of restrictions on any leaf-to-root path sets increasingly more variables as we approach the root.*

For any subformula G of F, we let $\tilde{\rho}|_G$ denote the restriction of $\tilde{\rho}$ to the set T_G of subformulae of G. ⌟

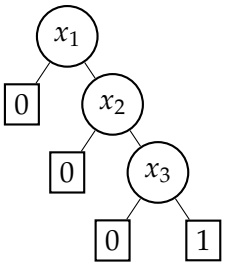
We will use the "tilde" notation to distinguish between restrictions ρ and restriction trees $\tilde{\rho}$. Observe that, by definition, every restriction ρ in a restriction tree $\tilde{\rho}$ corresponding to a formula F satisfies $\rho \preccurlyeq \tilde{\rho}(F)$ and are hence consistent with each other.

3 Canonical decision tree

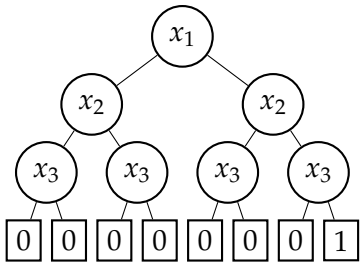
In this section, we construct a canonical decision tree (CDT) for a formula F. This definition is identical to Rossman's definition [Ros19, Definition 19] (except that Rossman defines it completely in terms of ordered restrictions while we define it using decision trees which have both degree-1 and degree-2 internal nodes).

Let us first recall the CDT construction for DNFs in the proof of Håstad's classical switching lemma [Bea94, Raz95, Hås14]. Let $F = T_1 \vee \cdots \vee T_m$ be a DNF and ρ a restriction on the variables of F. To construct $\mathrm{CDT}(F, \rho)$ we do the following:

1. Find the first term T (from left to right), not forced to 0 by ρ. If there is no such term, return the tree comprising of a single leaf node labelled 0.
2. If $T|_\rho \equiv 1$, return the tree comprising of a single leaf node labelled 1.



(a) Optimal DT for $x_1 \wedge x_2 \wedge x_3$



(b) Completed balanced DT for $x_1 \wedge x_2 \wedge x_3$

Figure 1: Illustration of $\mathsf{DT}(x_1 \wedge x_2 \wedge x_3)$ used in the CDT construction in the proof of Håstad's Switching Lemma.

3. Let Y be the set of ρ-unrestricted variables in T. Let Γ be the $\mathsf{CDT}(T, \rho)$ constructed from the complete balanced binary tree of depth $|Y|$ indexed by the variables of Y and labelling the $2^{|Y|}$ appropriately.

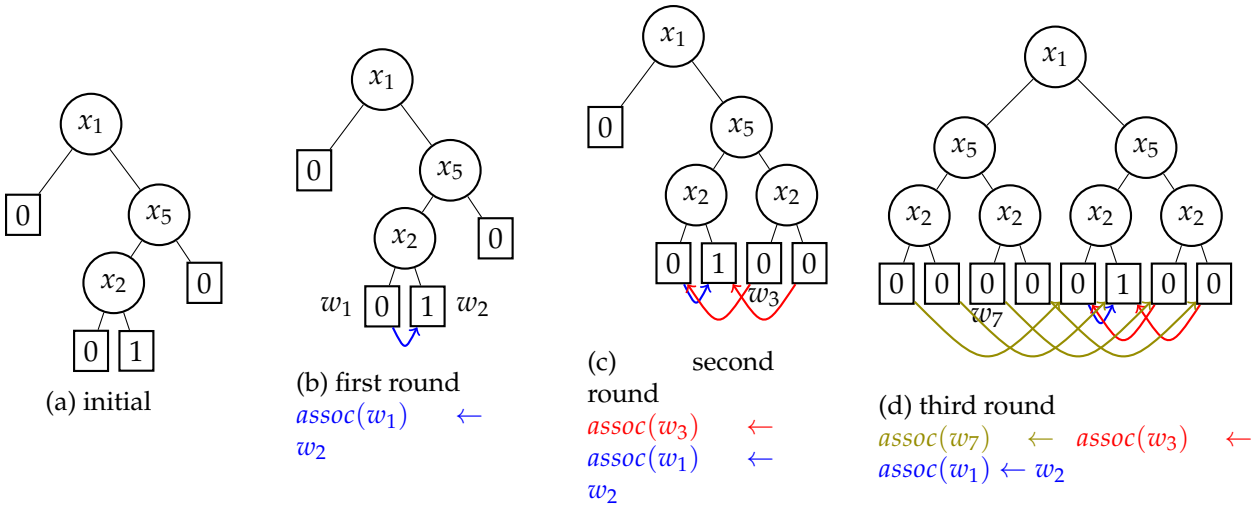
4. For each leaf v of Γ, inductively replace v with $\mathsf{CDT}(F|_{\alpha_v}, \rho)$ where α_v is the (ordered) restriction corresponding to leaf v.

The construction of CDTs for depth-d formulae will be inspired by the above CDT construction for DNFs. Note that in Step 3, we used a complete binary tree instead of the best decision tree for the term T (see Fig. 1). The rationale for doing this is because while proving the switching lemma, we wanted to attribute a 0-leaf in Γ to a 1-leaf which shares the same set of variables. We will need a similar property in our construction. To this end, we perform a balancing operation which ensures that every 0-leaf has a corresponding 1-leaf such that the two associated ordered restrictions share the same set of variables (this is the 0-balancing operation defined below. The 1-balancing operation is similar with the roles of 0 and 1 reversed).

3.1 0-Balancing and 1-Balancing

Given a decision tree Γ for a Boolean function F, the 0-balanced version Γ' is constructed as follows. We first pull-up the zeros, in other words, if there is any subtree all of whose leaves are labelled 0, we contract the entire subtree to a single leaf node labelled 0. The construction then proceeds in d rounds where d is the length of the longest root-to-leaf path in Γ (note this is not necessarily the depth of Γ due to the presence of degree-1 nodes). This process leaves the 1-leaves in Γ unaltered. As we proceed, we also construct a map assoc which associates each leaf (both 0 and 1 leaves) in Γ' with a 1-leaf in Γ'. To begin with, this map assoc associates each 1-leaf to itself (i.e, if u is a 1-leaf, then $\mathsf{assoc}(u) = u$).

In the i^{th} round, we consider all 0-leaves in Γ at distance $(d - i)$ from the root. Let u be one such 0-leaf and T_u the subtree rooted at the sibling of u. Observe that T_u necessarily has some leaf labelled 1, else the entire subtree rooted at the parent of u would have been contracted to a



(a) initial

(b) first round
$assoc(w_1) \leftarrow w_2$

(c) second round
$assoc(w_3) \leftarrow assoc(w_1) \leftarrow w_2$

(d) third round
$assoc(w_7) \leftarrow assoc(w_3) \leftarrow assoc(w_1) \leftarrow w_2$

Figure 2: Illustration of 0-balancing process

single leaf node labelled 0. We then mirror the entire subtree T_u at the leaf node u and relabel all the leaves of this mirrored subtree with 0. These are the 0-leaves of Γ'. For each such newly created 0-leaf w (in the mirrored subtree T_u), let w' be the corresponding leaf in the tree T_u. Set $\mathrm{assoc}(w) \leftarrow \mathrm{assoc}(w')$.

See Fig. 2 for an illustration of the 0-balancing process. Observe that if we 0-balance the best decision tree for a term, we obtain the complete balanced tree (see Fig. 1).

At the end of this process, observe that Γ is transformed into another decision tree Γ' such that the following hold.

- If Γ computes a function F under some restriction ρ, so does Γ'.
- The 1-leaves in Γ' are in 1-1 correspondence with the 1-leaves in Γ. Furthermore, the two 1-leaves (the one in Γ and its associated 1-leaf in Γ') correspond to identical ordered restrictions.
- Every 0-leaf w in Γ' has an associated 1-leaf in Γ' given by $\mathrm{assoc}(w)$. Furthermore, the corresponding ordered restrictions (namely $\alpha_w^{\Gamma'}$ and $\alpha_{\mathrm{assoc}(w)}^{\Gamma'}$) share the same set of variables which are queried in the same order along both these root-to-leaf paths.

Let us now try to understand what are the 0-leaves constructed in the 0-balancing process. Let w be any 0-leaf in Γ' and $w' = \mathrm{assoc}(w)$ be the corresponding 1-leaf. Furthermore, let $\alpha := \alpha_w^{\Gamma'} = (v_1 \mapsto c_1, \ldots, v_t \mapsto c_t)$ and $\beta := \alpha_{w'}^{\Gamma'} = (v_1 \mapsto d_1, \ldots, v_t \mapsto d_t)$. First, we must have that $\mathrm{dom}(\alpha) = \mathrm{dom}(\beta)$ and that the variables in this common domain must be queried in the same order. Furthermore, whenever α differs from β, the ordered restriction formed by following β upto the step prior to this particular point of disagreement and then taking a step according to α must cause the formula $F|_\rho$ to evaluate to 0. This occurs as the mirroring operation is performed only

at such nodes. More precisely, let α_i denote the ordered restriction $(v_1 \mapsto d_1, v_2 \mapsto d_2, \ldots, v_{i-1} \mapsto d_{i-1}, v_i \mapsto c_i)$. Note, α_i is the ordered restriction of length i which is identical to β in the first $i-1$ variables and then is similar to α on the i^{th} variable. The ordered restrictions α and β satisfy the following:

$$\forall i \in [t], c_i \neq d_i \implies F|_{\rho,\alpha_i} \equiv 0.$$

Furthermore, the converse also holds. That is, let β corresponds to an ordered restriction of some 1-leaf in Γ', then the ordered restriction α (with the same domain and same order of querying) corresponds to a 0-node in Γ' only if the above condition holds.

Since this is an important point, we summarize the above discussion in the following definition and claim.

Definition 3.1. *Let F be a Boolean function, ρ a restriction and $\alpha = (v_1 \mapsto c_1, \ldots, v_t \mapsto c_t), \beta = (v_1 \mapsto d_1, \ldots, v_t \mapsto d_t)$ be two ordered restrictions (on the same domain and order of querying). We say $\alpha \in \text{ASSOC}_0(F, \rho, \beta)$ iff*

$$\forall i \in [t], c_i \neq d_i \implies F|_{\rho,\alpha_i} \equiv 0$$

where α_i refers to the ordered restriction $(v_1 \mapsto d_1, v_2 \mapsto d_2, \ldots, v_{i-1} \mapsto d_{i-1}, v_i \mapsto c_i)$. ⌟

Claim 3.2. *Let Γ compute the formula F under the restriction ρ and Γ' be the 0-balanced version of Γ. Let w be a 1-leaf in Γ (and hence also Γ') and β be the corresponding ordered restriction. Then α is an ordered restriction corresponding to a 0-leaf w' with $\text{assoc}(w') = w$ iff $\alpha \in \text{ASSOC}_0(F, \rho, \beta)$.*

1-balancing is defined similarly with the roles of 0 and 1 reversed

3.2 CDT Definition

We are now ready to define the canonical decision tree (CDT). As indicated before, this definition is identical to [Ros19, Definition 19].

Definition 3.3. *Given a formula F on variable set V and associated restriction tree $\tilde{\rho}\colon T_F \to \{0,1,*\}^V$, we define the canonical decision tree, denoted by $\text{CDT}(F, \tilde{\rho})$, inductively (on depth and the number of variables) as follows:*

1. *If F is a constant 0 or 1, then $\text{CDT}(F, \tilde{\rho})$ is the unique tree with a single leaf node labelled by the appropriate constant.*
2. *If F is a literal x or $\neg x$, then*
    - *if x is set by $\tilde{\rho}(F)$ to a constant, then $\text{CDT}(F, \tilde{\rho})$ is the unique tree with a single node labelled by the appropriate constant.*

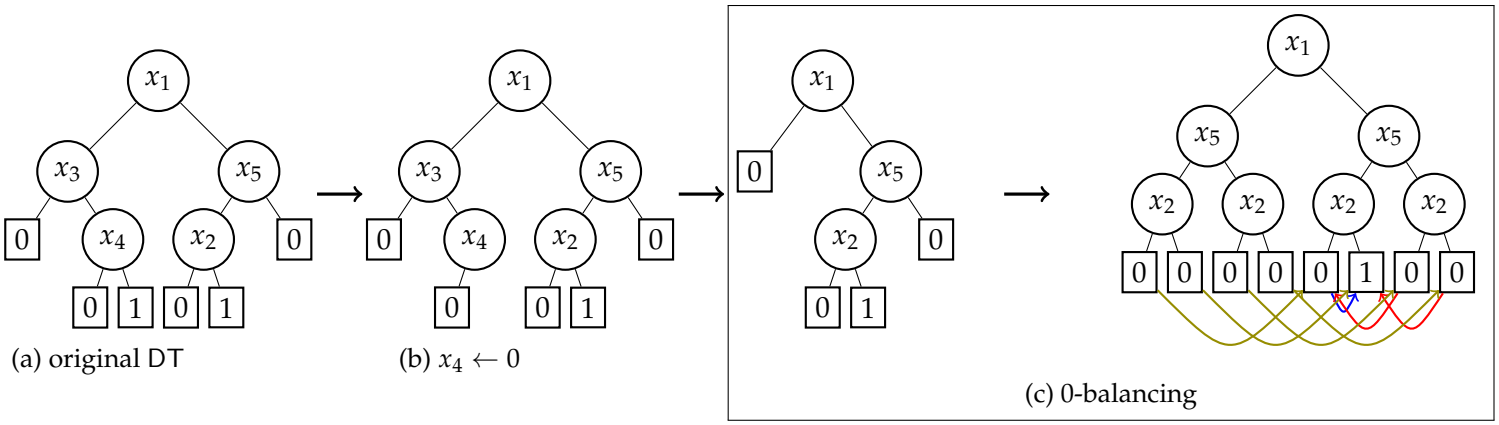


Figure 3: Illustration of Items 3a to 3c in CDT construction

- *Otherwise if x is unset by $\tilde{\rho}(F)$, then $\mathrm{CDT}(F, \tilde{\rho})$ is the tree with 3 nodes where the root is labelled by x and the two children are labelled appropriately by 0 or 1.*

3. *If $F = F_1 \vee \cdots \vee F_m$, then*

   - *If $F_1|_{\tilde{\rho}(F)} \equiv F_2|_{\tilde{\rho}(F)} \equiv \cdots \equiv F_m|_{\tilde{\rho}(F)} \equiv 0$, then $\mathrm{CDT}(F, \tilde{\rho})$ is the unique tree with a single leaf node labelled 0.*
   - *Else, there is some $1 \leq \ell \leq m$ such that $F_1|_{\tilde{\rho}(F)} \equiv \cdots \equiv F_{\ell-1}|_{\tilde{\rho}(F)} \equiv 0$ and $F_\ell|_{\tilde{\rho}(F)} \not\equiv 0$.*
   - *If $F_\ell|_{\tilde{\rho}(F)} \equiv 1$, then $\mathrm{CDT}(F, \tilde{\rho})$ is the unique tree node with a single leaf node labelled 1.*
   - *If $F_\ell|_{\tilde{\rho}(F)} \not\equiv$ constant, then do the following steps to construct $\mathrm{CDT}(F, \tilde{\rho})$*
     (a) *Let Γ be $\mathrm{CDT}(F_\ell, \tilde{\rho}|_{F_\ell})$ constructed inductively (since $\mathrm{depth}(F_\ell) < \mathrm{depth}(F)$).*
     (b) *Apply the restriction $\tilde{\rho}(F)$ to Γ to get Γ' and remove all the sub-trees which are inconsistent with $\tilde{\rho}(F)$[3].*
     (c) *0-balance Γ' to get Γ''.*
     (d) *For each 0-leaf u of Γ'', replace u by $\mathrm{CDT}(F|_{\alpha_u}, \tilde{\rho})$ where α_u is the ordered restriction corresponding to u in Γ''.*

   The case when $F = F_1 \wedge \cdots \wedge F_m$ is a conjunction of sub-formulas is handled similarly (with the roles of 0 and 1 reversed). ⌟

Given any s-long bitstring $a = (a_1, a_2, \ldots, a_s)$, we can walk along the CDT using a as an "instruction set". In other words, we walk from the root to a node of the tree by using a to make choices at the degree-2 nodes and otherwise following the degree one-edges. If this walks ends at a node w (possibly leaf node) of the CDT, we denote the corresponding ordered restriction α_w by $\mathrm{CDT}^{(a)}(F, \tilde{\rho})$, else $\mathrm{CDT}^{(a)}(F, \tilde{\rho})$ is undefined. When this node is a leaf node, then it is labelled either 0 or 1. In this case, we further enhance this definition as follows.

[3]This step introduces degree 1 nodes in the decision tree.

Definition 3.4. *Let F be a formula and $\tilde{\rho}$ an associated restriction tree. For any bitstring $a = (a_1, \ldots, a_s)$ and $b \in \{0,1\}$, define*

$$\mathrm{CDT}_b^{(a)}(F, \tilde{\rho}) = \begin{cases} \alpha_w & \text{if the walk according to instruction set "a" ends on a leaf } w \text{ labelled } b, \\ \perp & \text{otherwise.} \quad \lrcorner \end{cases}$$

3.3 Unpacking the CDT

Fix a formula F on n variables and an associated restriction tree $\tilde{\rho}\colon T_F \to \{0,1,*\}^n$. Let $a = (a_1, \ldots, a_s)$ be an s-bitstring with $s \geq 1$. Let us assume $F = F_1 \vee F_2 \vee \cdots \vee F_m$ is a disjunction. In this section, we try to understand when $\mathrm{CDT}_0^{(a)}(F, \tilde{\rho})$ exists.

Suppose $\mathrm{CDT}_0^{(a)}(F, \tilde{\rho})$ exists and is the ordered restriction α. Then, the following must be true.

- There exists a unique $\ell \in [m]$ such that for all $\ell' < \ell$, we have $F_{\ell'}|_{\tilde{\rho}(F)} \equiv 0$ and $F_\ell|_{\tilde{\rho}(F)} \not\equiv$ constant.
- Let Γ be $\mathrm{CDT}(F_\ell, \tilde{\rho}|_{F_\ell})$. Let Γ' be the tree obtained by restricting Γ by $\tilde{\rho}(F)$ and Γ'' be the 0-balancing of Γ'. There must be some $1 \leq r \leq s$ such that, a walk to a leaf of Γ'' using instruction set $a_{\leq r}$ leads us to a leaf in Γ''. Let α' be the corresponding ordered restriction (which ought to be a prefix of α).
- $\mathrm{CDT}_0^{(a_{>r})}(F|_{\alpha'}, \tilde{\rho})$ exists, and is α'' say. In that case, $\alpha = (\alpha', \alpha'')$.

Let us peer deeper into the balancing operation. Since Γ'' is the 0-balancing of Γ', we must have that $\mathrm{assoc}(\alpha') = \beta$ for some 1-leaf β of Γ'' and this β is a 1-leaf of $\Gamma = \mathrm{CDT}(F_\ell, \tilde{\rho}|_{F_\ell})$ as well. In fact, β must be consistent with $\tilde{\rho}(F)$ as this path in $\mathrm{CDT}(F_\ell, \tilde{\rho}|_{F_\ell}) = \Gamma$ survived in Γ' as well. Thus, there is some instruction set $b \in \{0,1\}^t$, for some $t \geq r$, such that $\mathrm{CDT}_1^{(b)}(F_\ell, \tilde{\rho}|_{F_\ell}) = \beta$.

Let us focus on the differences between the ordered restriction α' and β. We know there were t degree-2 nodes on the path to β in Γ'', and there were r degree-2 nodes on the path to α_1 in $\mathrm{CDT}(F_\ell, \tilde{\rho}_\ell)$. Thus, among the t degree-2 nodes on the path to β, we must have that $t - r$ of them belong to $\mathrm{dom}(\tilde{\rho}(F))$ (with β being consistent with $\tilde{\rho}(F)$) and the path to α_1 uses $a_{\leq r}$ as instructions for the other r nodes (instead of whatever route was taken by the path to β).

We summarize this discussion in the following lemma. We will be using this lemma for a *random* restriction tree $\tilde{\boldsymbol{\rho}}$ (chosen according to a suitable distribution). To distinguish the quantites that depend on this random variable from the rest, we use bold font to indicate all the quantities (including $\tilde{\boldsymbol{\rho}}$ itself) that are functions of $\tilde{\boldsymbol{\rho}}$.

Lemma 3.5 (Unpacking $\mathrm{CDT}_0^{(a)}(F, \tilde{\boldsymbol{\rho}})$)**.** *Let $F = F_1 \vee \cdots \vee F_m$ be a formula and $\tilde{\boldsymbol{\rho}}\colon T_F \to \{0,1,*\}^n$ an associated restriction tree. Let $s \geq 1$ and $a \in \{0,1\}^s$.*

Then $\mathrm{CDT}_0^{(a)}(F, \tilde{\boldsymbol{\rho}})$ exists if and only if there exist

- *non-negative integer* $r \in [s]$,
- $\ell \in [m]$
- *non-negative integer* $t \geq r$
- *a bitstring* $b \in \{0,1\}^t$ *and*
- $Q \in \binom{[t]}{r}$

such that the following three conditions $\mathcal{A}, \mathcal{B}, \mathcal{C}$ *are met.*

$\mathcal{A}(\ell, t, b)$**:** *(i).* $F_{\ell'}|_{\tilde{\boldsymbol{\rho}}(F)} \equiv 0$ *for all* $\ell' < \ell$,

(ii). $\mathrm{CDT}_1^{(b)}(F_\ell, \tilde{\boldsymbol{\rho}}|_{F_\ell})$ *exists (and is* $\boldsymbol{\beta}$ *say).*

(iii). $\boldsymbol{\beta}$ *is consistent with* $\tilde{\boldsymbol{\rho}}(F)$,

(iv). $\boldsymbol{\alpha'} \in \mathrm{ASSOC}_0(F_\ell, \tilde{\boldsymbol{\rho}}(F), \boldsymbol{\beta})$ *where* $\boldsymbol{\alpha'}$ *is the ordered restriction obtained by walking along* $\boldsymbol{\Gamma''}$ *using the instruction set a till one reaches a leaf where* $\boldsymbol{\Gamma''}$ *defined as follows.* $\boldsymbol{\Gamma} \leftarrow \mathrm{CDT}(F_\ell, \tilde{\boldsymbol{\rho}}_{F_\ell})$, $\boldsymbol{\Gamma'}$ *is obtained by applying* $\tilde{\boldsymbol{\rho}}(F)$ *on* $\boldsymbol{\Gamma}$ *and* $\boldsymbol{\Gamma''}$ *is the 0-balanced version of* $\boldsymbol{\Gamma'}$.

$\mathcal{B}(\ell, t, b, r, Q)$**:** Q *identifies* $\mathrm{stars}(\tilde{\boldsymbol{\rho}}(F))$ *within* $\mathrm{dom}(\boldsymbol{\beta}) \cap \mathrm{stars}(\tilde{\boldsymbol{\rho}}(F_\ell))$.

$\mathcal{C}(\ell, r)$**:** $\mathrm{CDT}_0^{(a_{>r})}(F|_{\boldsymbol{\alpha'}}, \tilde{\boldsymbol{\rho}})$ *exists (is* $\boldsymbol{\alpha''}$ *say).*

Furthermore, when $\mathrm{CDT}_0^{(a)}(F, \tilde{\boldsymbol{\rho}})$ *exists, we have* $\mathrm{CDT}_0^{(a)}(F, \tilde{\boldsymbol{\rho}}) = (\boldsymbol{\alpha'}, \boldsymbol{\alpha''})$.

When clear from context, we will drop the arguments ℓ, t, b, r, Q from the properties $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$.

4 Downward closure property

Let ρ, ρ' be two restrictions on the variable set V and let $S \subseteq V$ any subset of the variables. Recall that we say that $\rho' \preccurlyeq_S \rho$ if (1) $\mathrm{stars}(\rho') \cap S \subseteq \mathrm{stars}(\rho) \cap S$ and (2) ρ_1 and ρ_2 are consistent. We say that a set $\mathcal{F}$ of restrictions is downward-closed with respect to the set of variables S if the following holds:

$$\rho \in \mathcal{F} \implies (\forall \rho' \in \mathcal{F},\ \rho' \preccurlyeq_S \rho \implies \rho' \in \mathcal{F}).$$

We now extend this defintion of downward-closed sets to restriction trees.

Definition 4.1 (downward-closed set of restriction trees)**.** *Let* F *be a formula on the variable set* V *and* $\tilde{\rho}, \tilde{\rho}' \colon T_F \to \{0,1,*\}^V$ *be two associated restriction trees. Let* $S \subseteq V$*. We say* $\tilde{\rho}' \preccurlyeq_S \tilde{\rho}$ *iff for all* $G \in T_F$*, we have* $\tilde{\rho}'(G) \preccurlyeq_S \tilde{\rho}(G)$.

We call a set $\mathcal{T}$ of restriction trees downward-closed with respect to the variable set S *iff*

$$\tilde{\rho} \in \mathcal{T} \implies (\forall \tilde{\rho}' \in \mathcal{T},\ \tilde{\rho}' \preccurlyeq_S \tilde{\rho} \implies \tilde{\rho}' \in \mathcal{T}).$$

If $S = [n]$, then we drop the subscript S in the above definitions. ⌟

It is evident that if $\mathcal{T}$ and $\mathcal{T}'$ are two downward-closed set of restriction trees with respect to a variable set, so is their intersection. The key property that enables our proof of the main lemma is the following downward-closure property.

Lemma 4.2. *Let $F = F_1 \vee F_2 \vee \cdots \vee F_m$ be a formula on variable set V and $\tilde{\rho}\colon T_F \to \{0,1,*\}^{|V|}$ be an associated restriction tree. Let $s \in \mathbb{Z}_{>0}$, $a \in \{0,1\}^s$ and α be an ordered restriction such that*

$$\mathrm{CDT}_0^{(a)}(F, \tilde{\rho}) = \alpha.$$

Suppose $\tilde{\rho}'\colon T_F \to \{0,1,*\}^{|V|}$ is another restriction tree satisfying

- *$\tilde{\rho}' \preccurlyeq \tilde{\rho}$ and*
- *$\tilde{\rho}'(G)|_{\mathrm{dom}(\alpha)} = \tilde{\rho}(G)|_{\mathrm{dom}(\alpha)}$ for all $G \in T_F$,*

then $\mathrm{CDT}_0^{(a)}(F, \tilde{\rho}') = \alpha$.

Similarly, when $F = F_1 \wedge F_2 \wedge \cdots \wedge F_m$, the same holds for "$\mathrm{CDT}_1^{(a)}(F, \tilde{\rho}) = \alpha$".

Note that the lemma implies that the set $\mathcal{T}_{F,a,\alpha} := \{\tilde{\rho}\colon \mathrm{CDT}_0^{(a)}(F, \tilde{\rho}) = \alpha\}$ is downward-closed with respect to the variable set $V \setminus \mathrm{dom}(\alpha)$.

Proof. We are given that $\tilde{\rho}'$ and $\tilde{\rho}'$ behave identically on $\mathrm{dom}(\alpha)$, and $\tilde{\rho}'$ only sets more variables (all of them outside of $\mathrm{dom}(\alpha)$) than $\tilde{\rho}$. The proof is by induction on the depth and number of variables in the formula.

Base case: The base case is when $F|_{\tilde{\rho}(F)}$ is a literal or a constant. The lemma is clearly true in this case as $\tilde{\rho}'$ only sets more variables than $\tilde{\rho}$ and does not change the variables in $\mathrm{dom}(\alpha)$.

Induction step: Let F be a formula of depth d on the variable set $[n]$. Assume the lemma is true for all formulae of either depth less than d or involving less than n variables.

By the Unpacking lemma (Lemma 3.5), we have that $\mathrm{CDT}_0^{(a)}(F, \tilde{\rho}) = \alpha$ if and only if there exist ℓ, r, t, b, Q and ordered restrictions α', α'', β such that the following are true.

(i). $F_{\ell'}|_{\tilde{\rho}(F)} \equiv 0$ for all $\ell' < \ell$,

(ii). $\mathrm{CDT}_1^{(b)}(F_\ell, \tilde{\rho}|_{F_\ell}) = \beta$,

(iii). β is consistent with $\tilde{\rho}(F)$,

(iv). $\alpha' \in \mathrm{ASSOC}_0(F_\ell, \tilde{\rho}(F), \beta)$ where α' is the ordered restriction obtained by walking along Γ'' using the instruction set a till one reaches a leaf where Γ'' is defined.

$$\Gamma = \mathrm{CDT}(F_\ell, \tilde{\rho}_{F_\ell}) \overset{\text{Apply } \tilde{\rho}(F)}{\rightsquigarrow} \Gamma' \overset{\text{0-balance}}{\rightsquigarrow} \Gamma''.$$

(v). Q identifies $\mathrm{stars}(\tilde{\rho}(F))$ within $\mathrm{dom}(\beta) \cap \mathrm{stars}(\tilde{\rho}(F_\ell))$.

(vi). $\mathrm{CDT}_0^{(a_{>r})}(F|_{\alpha'}, \tilde{\rho}) = \alpha''$.

(vii). $\alpha = (\alpha', \alpha'')$.

We will demonstrate that for the same ℓ, r, t, b, Q and ordered restrictions α', α'', β all the above conditions continue to hold good when $\tilde{\rho}$ is replaced by $\tilde{\rho}'$. This will prove that $\mathrm{CDT}_0^{(a)}(F, \tilde{\rho}') = \alpha$.

Item (vii) is trivially true as this is independent of $\tilde{\rho}$ or $\tilde{\rho}'$. The other conditions are met for the following reasons.

- Since $\alpha' \in \mathrm{ASSOC}_0(F_\ell, \tilde{\rho}(F), \beta)$, we have that $\mathrm{dom}(\beta) = \mathrm{dom}(\alpha')$. Furthermore, we also have $\mathrm{dom}(\alpha'), \mathrm{dom}(\alpha'') \subset \mathrm{dom}(\alpha)$. The restriction tree $\tilde{\rho}'$ only sets more variables than $\tilde{\rho}$ and is unaltered on $\mathrm{dom}(\alpha)$. These imply that Items (i), (iii) and (v) hold when $\tilde{\rho}$ is replaced by $\tilde{\rho}'$.
- Items (ii) and (vi) are also true when $\tilde{\rho}$ is replaced by $\tilde{\rho}'$ due to the inductive assumption (since F_ℓ is a formula of smaller depth, and $F|_{\alpha'}$ is a formula on fewer variables).
- As for Item (iv), let $\tilde{\Gamma}$ be $\mathrm{CDT}(F_\ell, \tilde{\rho}'_{F_\ell})$, let $\tilde{\Gamma}'$ be obtained by applying $\tilde{\rho}'(F)$ on $\tilde{\Gamma}$ and $\tilde{\Gamma}'$ is the 0-balanced version of $\tilde{\Gamma}'$. Since $\beta = \mathrm{CDT}_1^{(b)}(F_\ell, \tilde{\rho}'|_{F_\ell})$ and β is consistent with $\tilde{\rho}'(F)$, the ordered restriction continues to be a root-leaf path in $\tilde{\Gamma}'$. Since $\tilde{\rho}'(F)$ does not alter any of the variables in $\mathrm{dom}(\beta)$ and is otherwise sets only more variables than $\tilde{\rho}(F)$, we have $\alpha' \in \mathrm{ASSOC}_0(F_\ell, \tilde{\rho}'(F), \beta)$ and hence α' is a 0-leaf in $\tilde{\Gamma}'$. Since $\tilde{\rho}'$ and $\tilde{\rho}$ behave identically on $\mathrm{dom}(\alpha) \supset \mathrm{dom}(\alpha')$, we have that α' is obtained by walking along $\tilde{\Gamma}'$ using the instruction set a till one reaches a leaf.

Thus, we have proved the claim. □

5 Bounds on criticality

In this section, we prove Theorem 1.1 (the criticality result for AC^0 formulae). To this end, we first define $\lambda(F)$, the bound on criticality that we eventually prove. We then define a sampling procedure to sample random restriction trees $\tilde{\boldsymbol{\rho}}$ for a given formula F such that the marginal distribution $\tilde{\boldsymbol{\rho}}(F)$ (i.e, the distribution of the restriction corresponding to the entire formula) is the standard p-random restriction. Finally, we state and prove the main inductive lemma (Lemma 5.5) that proves Theorem 1.1.

We begin by defining $\lambda(F)$ for any AC^0-formula.

Definition 5.1 (lambda)**.** *For a positive integer $S \in \mathbb{Z}_{>0}$ and non-negative integer $d \in \mathbb{Z}_{\geq 0}$, define*

$$\lambda_{S,d} := 32^{d+1}\left(\frac{\log S}{d}+1\right)^d = 32^{d+1}\left(\frac{\log(2^d \cdot S)}{d}\right)^d.$$

Given an AC^0 formula F of depth $d+1$ and size S, define $\lambda(F) := \lambda_{S,d+1}$. ⌟

Note, that the above expression simplifies to 32 for depth-1 formulae (i.e., terms and clauses), where we have used the convention that $\frac{0}{0} = 1$.

Claim 5.2. $8\lambda_{S,d} \leq \lambda_{S,d+1}$

Proof. $8\lambda_{S,d} \leq 8 \cdot 32^{d+1}\left(\frac{\log 2^{d+1}S}{d}\right)^d = \frac{\lambda_{S,d+1}}{4\cdot \log 2^{d+1}S}\frac{(d+1)^{d+1}}{d^d} \leq \frac{\lambda_{S,d+1}}{4}\frac{e(d+1)}{d+1+\log S} \leq \lambda_{S,d+1}$. □

5.1 Sampling restriction trees

We begin by recalling the definition of the classical $\mathcal{R}_p$ distribution over restrictions.

Definition 5.3 (p-random restriction)**.** *For $p \in [0,1]$ and a variable set V, $\mathcal{R}_p([n])$ is the distribution on restrictions $\boldsymbol{\rho}\colon V \to \{0,1,*\}^n$ defined as follows: independently for each $v \in V$,*

$$\boldsymbol{\rho}(v) \leftarrow \begin{cases} * & \text{with probability } p, \\ 0 & \text{with probability } \frac{1-p}{2}, \text{and} \\ 1 & \text{with probability } \frac{1-p}{2}. \end{cases}$$

An equivalent way of sampling this distibution is by choosing independently a uniformly random string $\boldsymbol{\sigma} \in \{0,1\}^V$ and a uniformly random point $\boldsymbol{\tau} \in [0,1]^V$. Then set for each $v \in V$

$$\boldsymbol{\rho}(v) \leftarrow \begin{cases} * & \text{if } \boldsymbol{\tau}_v \leq p \\ \boldsymbol{\sigma}_v & \text{otherwise.} \end{cases}$$ ⌟

We now extend this definition to distribution over restriction trees. Given a formula F, we say that $\tilde{p}\colon T_F \to [0,1]$ is a valid set of probabilities if whenever G is a subformula of H, we have $\tilde{p}(G) \geq \tilde{p}(H)$.

Definition 5.4 ($\tilde{\mathcal{R}}_p$-distribution)**.** *Let F be a formula on the variable set V and $\tilde{p}\colon T_F \to [0,1]$ be a valid set of probabilities. A random restriction tree $\tilde{\boldsymbol{\rho}}$ from the distribution $\tilde{\mathcal{R}}_{\tilde{p}}(F)$ is sampled as follows: Choose independently a uniformly random string in $\boldsymbol{\sigma} \in \{0,1\}^V$ and a uniformly random point $\boldsymbol{\tau} \in [0,1]^V$. For*

each $G \in T_F$ and $i \in V$, set

$$\tilde{\boldsymbol{\rho}}(G)(v) \leftarrow \begin{cases} * & \text{if } \boldsymbol{\tau}_v \leq \tilde{p}(G) \\ \boldsymbol{\sigma}_v & \text{otherwise.} \end{cases}$$

For any $p \in [0, 1/\lambda(F)]$, let $\tilde{\mathcal{R}}_p(F)$ denote the distribution $\tilde{\mathcal{R}}_{\tilde{p}}(F)$ where $\tilde{p}$ is defined as follows $\tilde{p}(F) = p$ and for all $G \in T_F$ other than F, we have $\tilde{p}(G) = 1/8\lambda(G)$.[4] ⌟

It trivially follows that the marginal distribution $\tilde{\boldsymbol{\rho}}(G)$ on any subformula G is distributed exactly according $\mathcal{R}_{\tilde{p}(G)}$.

5.2 Main Lemma

We are now ready to state the main lemma of the paper.

Lemma 5.5. *Let $d \geq 0$ and $F = F_1 \vee F_2 \vee \cdots \vee F_m$ be an AC^0 formula of size S and depth $d+1$ on n variables. Let $\mathcal{T}$ be any set of downward-closed set of restriction trees with respect to the variables of the formula F, then for all integers $s \geq 1$ and $a \in \{0,1\}^s$,*

$$\Pr_{\tilde{\boldsymbol{\rho}} \sim \tilde{\mathcal{R}}_p(F)} \left[\mathrm{CDT}_0^{(a)}(F, \tilde{\boldsymbol{\rho}}) \text{ exists} \mid \tilde{\boldsymbol{\rho}} \in \mathcal{T} \right] \leq (p \cdot \lambda(F))^s.$$

The statement for conjunctions $F = F_1 \wedge F_2 \wedge \cdots \wedge F_m$ is identical with $\mathrm{CDT}_0^{(a)}$ replaced by $\mathrm{CDT}_1^{(a)}$.

Theorem 1.1 stated in the introduction clearly follows from the above lemma. The above lemma is stronger than what is needed for Theorem 1.1 as it proves the statement even when conditioned under any downward-closed set of restriction trees. This stronger statement is needed for the inductive proof to go through.

Proof. The proof is by induction on the depth d of the formula and the number of variables in the formula F.

Let us begin with the base case (depth-1 AC^0-formulae). This base case will also serve as a warmup to one of the key claims (Claim 5.9) in the proof of the induction step. The proof of this is similar to the proof of [Hås14, Lemma 3.4]

Base case: The base case is when F is a depth-1 formula and we need to bound the probability by $(32p)^s$ since in this case $\lambda(F) = 32$. A depth-1 formula is a term or a clause. Without loss of generality let's assume that F is a clause of the form $x_1 \vee \cdots \vee x_m$, where the x_i's are distinct variables.

[4] For this to be well-defined, we need $\lambda(F) \geq 8\lambda(G)$ for any subformula G of F. This follows from Claim 5.2

For a given $a \in \{0,1\}^s$, let $\mathcal{E}_a = \left\{\tilde{\rho}\colon \mathrm{CDT}_0^{(a)}(F,\tilde{\rho}) \text{ exists}\right\}$. We need to bound the quantity $\mu(\mathcal{E}_a \cap \mathcal{T})/\mu(\mathcal{T})$. For every $\tilde{\rho} \in \mathcal{E}_a \cap \mathcal{T}$, define the set $N(\tilde{\rho})$ as outlined below. The fact that $\tilde{\rho} \in \mathcal{E}_a$ implies that there exists a unique subset of variables $Q \subset [m]$ of size s such that $\mathrm{stars}(\tilde{\rho}(F)) \cap [m] = Q$ and for all variables $i \in [m] \setminus Q$, we have $\tilde{\rho}(F)(x_i) = 0$. $N(\tilde{\rho})$ is the set of all restriction trees $\tilde{\rho}'$ which are identical to $\tilde{\rho}$ everywhere except for how $\tilde{\rho}(F)$ behaves on the variables specified by the set Q where we allow it to be any restriction consistent with 1^Q (i.e., either let the variable continue to be unrestricted or set it to 1). Note that every $\tilde{\rho}' \in N(\tilde{\rho})$ is in $\mathcal{T}$ as $\mathcal{T}$ is downward closed, however exactly one element in $N(\tilde{\rho})$, namely $\tilde{\rho}$ is in $\mathcal{E}_a$. It is easy to verify that

$$\frac{\mu(\tilde{\rho})}{p^s} = \frac{\mu(N(\tilde{\rho}))}{\left(p + \frac{1-p}{2}\right)^s}.$$

Since for distinct $\tilde{\rho}$, the corresponding $N(\tilde{\rho})$ are disjoint, we have the following bound on the probability that we wish to bound.

$$\begin{aligned}\Pr_{\tilde{\rho}}\left[\mathcal{E}_a \mid \mathcal{T}\right] &= \frac{\mu(\mathcal{E}_a \cap \mathcal{T})}{\mu(\mathcal{T})} = \frac{\sum_{\tilde{\rho}\in\mathcal{E}_a\cap\mathcal{T}} \mu(\tilde{\rho})}{\sum_{\tilde{\rho}\in\mathcal{E}_a\cap\mathcal{T}} \mu(N(\tilde{\rho})) + \mu\left(\mathcal{T} \setminus \bigcup_{\tilde{\rho}\in\mathcal{E}_a\cap\mathcal{T}} N(\tilde{\rho})\right)} \\ &\leq \frac{\sum_{\tilde{\rho}\in\mathcal{E}_a\cap\mathcal{T}} \mu(\tilde{\rho})}{\sum_{\tilde{\rho}\in\mathcal{E}_a\cap\mathcal{T}} \mu(N(\tilde{\rho}))} = \left(\frac{2p}{1+p}\right)^s \leq (2p)^s.\end{aligned}$$

Induction step: Let us assume without loss of generality that $F = F_1 \vee \cdots \vee F_m$ and the main lemma holds for all formulae of smaller depth (in particular the F_i's) and all formulae with smaller number of variables (in particular $F|_\beta$ for any non-trivial restriction β). By the Unpacking Lemma (Lemma 3.5) and a union bound we have that

$$\Pr_{\tilde{\boldsymbol{\rho}}}\left[\mathrm{CDT}_0^{(a)}(F,\tilde{\boldsymbol{\rho}}) \text{ exists} \mid \tilde{\boldsymbol{\rho}} \in \mathcal{T}\right] \leq \sum_{r\in[s]} \sum_{\ell\in[m]} \sum_{t\colon t\geq r} \sum_{Q\in\binom{[t]}{r}} \sum_{b\in\{0,1\}^t} \Pr\left[\mathcal{A}\cap\mathcal{B}\cap\mathcal{C} \mid \mathcal{T}\right]. \tag{5.6}$$

For each fixed choice r, ℓ, t, b, Q, the summand in the above expression can be factorized as

$$\Pr\left[\mathcal{A} \mid \mathcal{T}\right] \cdot \Pr\left[\mathcal{B} \mid \mathcal{A}\cap\mathcal{T}\right] \cdot \Pr\left[\mathcal{C} \mid \mathcal{A}\cap\mathcal{B}\cap\mathcal{T}\right]. \tag{5.7}$$

The following three (3) claims bound each of the terms in the above product.

Claim 5.8. *For a fixed a, ℓ, r, t, b and Q, we have*

$$\Pr_{\tilde{\boldsymbol{\rho}}\sim\tilde{\mathcal{R}}_p(F)}\left[\mathrm{CDT}_0^{(a_{>r})}(F|_{\boldsymbol{\alpha}'},\tilde{\boldsymbol{\rho}}) \text{ exists} \mid \mathcal{A}(\ell,t,b)\cap\mathcal{B}(\ell,t,b,r,Q)\cap\mathcal{T}\right] \leq (p\cdot\lambda(F))^{s-r}.$$

Proof. We first note that the formula being considered in the above expression, namely $F|_{\boldsymbol{\alpha}'}$, is itself random since the ordered restriction $\boldsymbol{\alpha}'$ is random. To deal with this, we prove the above

bound for each fixing of $\boldsymbol{\alpha'}$. More precisely, we rewrite the above expression as follows (here we not only fix $\boldsymbol{\alpha}$, but also $\boldsymbol{\beta}$).

$$\mathbb{E}_{\alpha',\beta}\left[\underbrace{\Pr_{\tilde{\rho}}\left[\mathrm{CDT}_0^{(a_{>r})}(F|_{\alpha'},\tilde{\boldsymbol{\rho}}) \text{ exists} \mid \mathcal{A}\cap\mathcal{B}\cap\mathcal{T}\cap\mathcal{E}_{\alpha',\beta}\right]}\right],$$

where $\mathcal{E}_{\alpha',\beta}$ is the set of restriction trees $\tilde{\boldsymbol{\rho}}$ that satisfy $\boldsymbol{\alpha'}=\alpha'$ and $\boldsymbol{\beta}=\beta$. We will prove that for any fixing of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the indicated quantity in the above expression is at most $(p\cdot\lambda(F))^{s-r}$, which would imply the claim.

Consider any fixing (α',β) of $(\boldsymbol{\alpha'},\boldsymbol{\beta})$. We first observe that since α' is a non-trivial ordered restriction (which is true since $r\geq 1$), the variable set of the formula $F|_{\alpha'}$ is less than that of F and hence we can apply the inductive assumption provided the set of restriction trees $\mathcal{A}\cap\mathcal{B}\cap\mathcal{T}\cap\mathcal{E}_{\alpha',\beta}$ is downward closed with respect to the variables of $F|_{\alpha'}$. Below, we verify that this is indeed the case.

$\mathcal{A}(\ell,t,b)\cap\mathcal{E}_{\alpha',\beta}$**:** We will show that each of the 4 components of $\mathcal{A}$ are downward-closed.

(i). We first observe that since α' and β are fixed, $\mathcal{A}$(iv) $\cap\,\mathcal{E}_{\alpha',\beta}$ is a deterministic event and nothing needs to be checked.

(ii). $\mathcal{A}$(i) is clearly downward-closed.

(iii). $\mathcal{A}$(ii) $\cap\,\mathcal{E}_{\alpha',\beta}$ is the event that "$\mathrm{CDT}_1^{(b)}(F_\ell,\tilde{\boldsymbol{\rho}}|_{F_\ell})=\beta$". This is downward-closed on the variable set $[n]\setminus\mathrm{dom}(\beta)=[n]\setminus\mathrm{dom}(\alpha')$ by Lemma 4.2.

(iv). $\mathcal{A}$(iii) $\cap\,\mathcal{E}_{\alpha',\beta}$ is downward-closed on the variable set $[n]\setminus\mathrm{dom}(\beta)=[n]\setminus\mathrm{dom}(\alpha')$ (note this is not necessarily downward-closed on the entire set of variables).

$\mathcal{B}(\ell,t,b,r,Q)\cap\mathcal{E}_{\alpha',\beta}$**:** This is a condition concerning variables in $\mathrm{dom}(\beta)$ and hence is downward-closed for the complementary set.

Combined with the fact that $\mathcal{T}$ is downward-closed, we have $\mathcal{A}\cap\mathcal{B}\cap\mathcal{T}\cap\mathcal{E}_{\alpha',\beta}$ is downward-closed on $[n]\setminus\mathrm{dom}(\alpha')$ and hence by the inductive assumption, we have the required bound. □

Claim 5.9. *For fixed a,ℓ,r,t,b and Q, we have* $\Pr[\mathcal{B}\mid\mathcal{A}\cap\mathcal{T}]\leq(16\cdot p\cdot\lambda(F_\ell))^r$.

As indicated earlier, the proof of this claim is similar in spirit to the proof of the base case, which in turn is similar to the proof of [Hås14, Lemma 3.4]. Things are however considerably more involved here and one has to do a careful conditioning argument to obtain the bound.

Proof. We need to bound the following quantity for a fixed choice of a,ℓ,r,t,b and Q:

$$\Pr_{\tilde{\rho}\sim\tilde{\mathcal{R}}_p(F)}[Q \text{ identifies } \tilde{\boldsymbol{\rho}}(F) \text{ within } \mathrm{dom}(\boldsymbol{\beta})\cap\mathrm{stars}(\tilde{\boldsymbol{\rho}}(F_\ell)) \mid \mathcal{A}(\ell,t,b)\cap\mathcal{T}].$$

The random restriction tree $\tilde{\boldsymbol{\rho}}$ contains within it the random restriction tree $\tilde{\boldsymbol{\rho}}|_{F_\ell}$ (which we will refer to as $\tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}$ for short) and the random restriction $\tilde{\boldsymbol{\rho}}(F)$. So, we can first sample $(\tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}, \tilde{\boldsymbol{\rho}}(F))$ first and then $\tilde{\boldsymbol{\rho}}$ from the corresponding conditional distribution. Furthermore, given the random restriction tree $\tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}$, the random restriction $\tilde{\boldsymbol{\rho}}(F)$ can be obtained by sampling a random restriction $\boldsymbol{\rho}$ from $\mathcal{R}_{p/\tilde{p}(F_\ell)}(\mathrm{stars}(\tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}(F_\ell)))$. We make this formal in Algorithm 1. Here $q = p/\tilde{p}(F_\ell) = 8p \cdot \lambda(F_\ell)$.

Algorithm 1: Sampling $\tilde{\boldsymbol{\rho}}$ from $\tilde{\mathcal{R}}_p(F)$

1 Sample $\tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}$ from $\tilde{\mathcal{R}}_{\tilde{p}(F_\ell)}(F_\ell)$.
2 Set $\boldsymbol{\rho}' \leftarrow \tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}(F_\ell)$.
3 Sample $\boldsymbol{\rho}$ from $\mathcal{R}_q(\mathrm{stars}(\boldsymbol{\rho}'))$.
4 Set $\boldsymbol{\rho}'' \leftarrow \boldsymbol{\rho}' \cdot \boldsymbol{\rho}$.
5 Sample $\tilde{\boldsymbol{\rho}}$ from the conditional distribution $\tilde{\mathcal{R}}_p(F)|_{\tilde{\boldsymbol{\rho}}|_{F_\ell}=\tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}, \tilde{\boldsymbol{\rho}}(F)=\boldsymbol{\rho}''}$.
6 Output $\tilde{\boldsymbol{\rho}}$.

The probability which we wish to bound can be rewritten as follows:

$$\mathbb{E}_{\tilde{\rho}_\ell}\left[\underbrace{\Pr_{\rho,\tilde{\rho}}\left[Q \text{ identifies } \tilde{\boldsymbol{\rho}}(F) \text{ within } \mathrm{dom}(\boldsymbol{\beta}) \cap \mathrm{stars}(\tilde{\boldsymbol{\rho}}(F_\ell)) \mid \mathcal{A}(\ell,t,b) \cap \mathcal{T}\right]}\right].$$

We will show that for every $\tilde{\rho}_\ell$, the quantity indicated (using the underbrace) is bounded by $(2q)^r$. If the restriction tree $\tilde{\rho}_\ell$ is fixed, so is the ordered restriction β (since $\beta = \mathrm{CDT}_1^{(b)}(F_\ell, \tilde{\rho}_\ell)$). Consider any pair $(\rho, \tilde{\rho})$ that satisifies the three properties (1) $\mathcal{B}(\ell,t,b,r,Q)$, (2) $\mathcal{A}(\ell,t,b)$ and (3) $\mathcal{T}$. Recall that $\mathcal{B}$ refers to the condition "Q identifies $\tilde{\boldsymbol{\rho}}(F)$ within $\mathrm{dom}(\beta) \cap \mathrm{stars}(\tilde{\rho}_\ell(F_\ell))$. Recall that ρ is a sample from the distribution $\mathcal{R}_q(\mathrm{stars}(\tilde{\rho}_\ell(F_\ell)))$ and $\tilde{\rho}$ from the appropriate conditional distribution. Any such ρ (and all the other restrictions in $\tilde{\rho}$) leave the variables specified by Q unrestricted. For every such pair $(\rho, \tilde{\rho})$, define the set $N(\rho, \tilde{\rho})$ as the set of all pairs $(\rho_1, \tilde{\rho}_1)$ which are identical to $(\rho, \tilde{\rho})$ everywhere except for how ρ_1 behaves on the variables specified by the set Q where we allow it to be any restriction consistent with β (i.e., either let the variable continue to be unrestricted or set them according to β). Note that every pair $(\rho_1, \tilde{\rho}_1) \in N(\rho, \tilde{\rho})$ satisfies the latter two properties, namely $\mathcal{A}(\ell,t,b)$ and $\mathcal{T}$, however exactly one element in $N(\rho, \tilde{\rho})$, namely $(\rho, \tilde{\rho})$ satisfies all the 3 properties. Furthermore, since ρ is sampled from $\mathcal{R}_q$, we have

$$\frac{\mu(\rho, \tilde{\rho})}{q^r} = \frac{\mu(N(\rho, \tilde{\rho}))}{\left(q + \frac{1-q}{2}\right)^r}.$$

Since for distinct $(\rho, \tilde{\rho})$, the corresponding $N(\rho, \tilde{\rho})$ are disjoint, we have the following bound on

the indicated probability (in underbraces) for every fixing of $\tilde{\rho}_\ell$.

$$\begin{aligned}&\Pr_{\boldsymbol{\rho},\tilde{\boldsymbol{\rho}}}\left[Q \text{ identifies } \tilde{\boldsymbol{\rho}}(F) \text{ within } \operatorname{dom}(\boldsymbol{\beta}) \cap \operatorname{stars}(\tilde{\boldsymbol{\rho}}(F_\ell)) \mid \mathcal{A}(\ell,t,b) \cap \mathcal{T}\right] \\ &\leq \frac{\sum_{(\rho,\tilde{\rho})} \mu(\rho,\tilde{\rho})}{\sum_{(\rho,\tilde{\rho})} \mu(N(\rho,\tilde{\rho}))} = \left(\frac{2q}{1+q}\right)^r \leq (2q)^r = (16 \cdot p \cdot \lambda(F_\ell))^r,\end{aligned}$$

where the summation (in both the numerator and denominator) in the second step above is over all $(\rho,\tilde{\rho})$ that satisfy all three properties. This completes the proof of the claim. □

Claim 5.10. *For fixed a, ℓ, t and b, we have*

$$\eta(\ell,t,b) := \Pr\left[\mathcal{A}(\ell,t,b) \mid \mathcal{T}\right] \leq \left(\frac{1}{8}\right)^t.$$

Proof.

$$\begin{aligned}\eta(\ell,t,b) &= \Pr_{\tilde{\boldsymbol{\rho}} \sim \tilde{\mathcal{R}}_p(F)}\left[\mathcal{A}(\ell,t,b) \mid \mathcal{T}\right] \leq \Pr_{\tilde{\boldsymbol{\rho}} \sim \tilde{\mathcal{R}}_p(F)}\left[\mathrm{CDT}_1^{(b)}(F_\ell, \tilde{\boldsymbol{\rho}}|_{F_\ell}) \text{ exists} \mid \mathcal{T}\right] \\ &= \Pr_{\tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}} \sim \tilde{\mathcal{R}}_{\tilde{p}(F_\ell)}(F_\ell)}\left[\mathrm{CDT}_1^{(b)}(F_\ell, \tilde{\boldsymbol{\rho}}_{\boldsymbol{\ell}}) \text{ exists} \mid \mathcal{T}\right] \\ &\leq (\tilde{p}(F_\ell) \cdot \lambda(F_\ell))^t = \left(\frac{1}{8\lambda(F_\ell)} \cdot \lambda(F_\ell)\right)^t = \left(\frac{1}{8}\right)^t.\end{aligned}$$

The last inequality follows from the induction assumption since F_ℓ has depth strictly smaller than that of F. □

Plugging the results of these claims back into the the expression in (5.7), we have

$$\eta(\ell,t,b) \cdot (16 \cdot p \cdot \lambda(F_\ell))^r \cdot (p \cdot \lambda(F))^{s-r} \leq \left(\frac{1}{8}\right)^t \cdot (16 \cdot p \cdot \lambda(F_\ell))^r \cdot (p \cdot \lambda(F))^{s-r}$$

We need to bound the sum of this expression when summed over all r, ℓ, t, b, Q as given by (5.6). However, even if we just over all possible ℓ the sum turns out to be prohibitively expensive. To keep this sum over ℓ (and also r, t, b, Q) under control, we further observe that the events $\mathcal{A}(\ell,t,b)$ are mutually disjoint over disjoint ℓ, t, b. This lets us conclude the following claim.

Claim 5.11.

$$\sum_{\ell,t,b} \eta(\ell,t,b) = \sum_{\ell,t,b} \Pr\left[\mathcal{A}(\ell,t,b) \mid \mathcal{T}\right] \leq 1.$$

We now have all the ingredients to bound the quantity of concern. The rest of the proof is a roller-coaster ride along the Jensen highway. We now bound the quantity in (5.6) as follows:

$$\Pr_{\tilde{\rho}}\left[\mathrm{CDT}_0^{(a)}(F,\tilde{\boldsymbol{\rho}}) \text{ exists } \mid \tilde{\boldsymbol{\rho}} \in \mathcal{T}\right] \leq \sum_{r,\ell,t,b,Q} \eta(\ell,t,b)\cdot(16\cdot p\cdot\lambda(F_\ell))^r\cdot(p\cdot\lambda(F))^{s-r}$$

$$= (p\cdot\lambda(F))^s\cdot\sum_{r,\ell}\left[\left(\frac{16\cdot\lambda(F_\ell)}{\lambda(F)}\right)^r\cdot\nu(\ell)\cdot\underbrace{\sum_{t,b}\frac{\eta(\ell,t,b)}{\nu(\ell)}\cdot t^r}\right] \tag{5.12}$$

where $\nu(\ell) := \sum_{t,b}\eta(\ell,t,b)$. We only sum over those ℓ that satisfy $\nu(\ell) > 0$. Observe that $\sum_\ell \nu(\ell) = \sum_{\ell,t,b}\eta(\ell,t,b) \leq 1$. We first bound the quantity indicated (using underbraces) in the above expression using Jensen's inequality and Claim 5.10 as follows.

Subclaim 5.13.

$$\sum_{t,b}\frac{\eta(\ell,t,b)}{\nu(\ell)}\cdot t^r \leq \left(\log\left(\frac{1}{\nu(\ell)}\right)\right)^r.$$

Proof. Rewriting t^r as $(\log 2^t)^r$, the lefthand side can be written as $\sum_{t,b}\frac{\eta(\ell,t,b)}{\nu(\ell)}\cdot(\log 2^t)^r$. Since $\sum_{t,b}\frac{\eta(\ell,t,b)}{\nu(\ell)} = 1$, we can apply Jensen's inequality to the concave function $x \mapsto (\log x)^r$ to obtain

$$\begin{aligned}
\sum_{t,b}\frac{\eta(\ell,t,b)}{\nu(\ell)}\cdot\left[\log 2^t\right]^r &\leq \left[\log\left(\sum_{t,b}\frac{\eta(\ell,t,b)}{\nu(\ell)}\cdot 2^t\right)\right]^r \\
&\leq \left[\log\left(\sum_{t,b}\frac{1}{\nu(\ell)\cdot 4^t}\right)\right]^r && \text{[Since } \eta(\ell,t,b)\leq 8^{-t} \text{ from Claim 5.10]} \\
&\leq \left[\log\left(\frac{1}{\nu(\ell)}\sum_t\frac{1}{2^t}\right)\right]^r && \text{[Since there are at most } 2^t \text{ } b\text{'s]} \\
&\leq \left[\log\frac{1}{\nu(\ell)}\right]^r && \square
\end{aligned}$$

Substituting this bound back into the expression (5.12) above, we obtain

$$\Pr\left[\mathrm{CDT}_0^{(a)}(F,\tilde{\boldsymbol{\rho}}) \text{ exists} \mid \tilde{\boldsymbol{\rho}} \in \mathcal{T}\right] \leq (p\cdot\lambda(F))^s\cdot\sum_r\left[\left(\frac{16}{\lambda(F)}\right)^r\cdot\sum_\ell\nu(\ell)\cdot\left(\underbrace{\lambda(F_\ell)\cdot\log\left(\frac{1}{\nu(\ell)}\right)}\right)^r\right]$$

We now apply AM-GM inequality and the definition of $\lambda(F_\ell)$ to bound the indicated quantity.

Subclaim 5.14. *Let $S_\ell := \mathrm{size}(F_\ell)$. Then,*

$$\lambda(F_\ell) \cdot \log\left(\frac{1}{\nu(\ell)}\right) \leq 32^d \left[\frac{\log\left(2^{d-1} \cdot S_\ell / \nu(\ell)\right)}{d}\right]^d.$$

Proof. If $d = 1$, then $S_\ell = 1$ (recall the 'size' defined in Definition 2.1), $\lambda(F_\ell) = 32$ (recall Definition 5.1). Thus, both sides of the above claim simplify to $32 \cdot \log(1/\nu(\ell))$.

For larger d,

$$\begin{aligned}
\lambda(F_\ell) \cdot \log\left(\frac{1}{\nu(\ell)}\right) &= 32^d \left(\frac{\log 2^{d-1} \cdot S_\ell}{d-1}\right)^{d-1} \cdot \log\left(\frac{1}{\nu(\ell)}\right) && \text{[Since } F_\ell \in \mathsf{AC}^0[S_\ell, d]\text{]} \\
&\leq 32^d \left[\frac{\log(2^{d-1} \cdot S_\ell) + \log\left(\frac{1}{\nu(\ell)}\right)}{d}\right]^d && \text{[Applying AM-GM inequality]} \\
&= 32^d \left[\frac{\log\left(2^{d-1} \cdot S_\ell / \nu(\ell)\right)}{d}\right]^d.
\end{aligned}$$

□

Plugging this bound back into our expression, we have

$$\Pr\left[\mathrm{CDT}_0^{(a)}(F, \tilde{\boldsymbol{\rho}}) \text{ exists} \mid \tilde{\boldsymbol{\rho}} \in \mathcal{T}\right] \leq (p \cdot \lambda(F))^s \cdot \sum_r \left[\left(\frac{16 \cdot 32^d}{\lambda(F)}\right)^r \cdot \underbrace{\sum_\ell \nu(\ell) \cdot \left(\frac{\log\left(\frac{2^{d-1} \cdot S_\ell}{\nu(\ell)}\right)}{d}\right)^{dr}}\right]$$

We bound the indicated quantity using yet another application of Jensen's inequality using the fact that $\sum_\ell \nu(\ell) \leq 1$ as follows.

Subclaim 5.15.

$$\sum_\ell \nu(\ell) \cdot \left(\frac{\log\left(\frac{2^{d-1} \cdot S_\ell}{\nu(\ell)}\right)}{d}\right)^{dr} \leq \left[\frac{\log\left(2^d \cdot S\right)}{d}\right]^{dr}.$$

Proof. Recall that $\sum_\ell \nu(\ell) \leq 1$. Consider the random variable Y defined as follows:

$$Y \leftarrow \begin{cases} \frac{2^{d-1} \cdot S_\ell}{\nu(\ell)} & \text{with probability } \nu(\ell) \text{ for each } \ell \text{ such that } \nu(\ell) \neq 0, \\ 1 & \text{with probability } 1 - \sum_\ell \nu(\ell). \end{cases}$$

and the concave function $x \xmapsto{f} \left(\frac{\log x}{d}\right)^{dr}$. Applying Jensen's inequality, we obtain

$$\begin{aligned}
\mathbb{E}[f(\boldsymbol{Y})] \leq f(\mathbb{E}\,\boldsymbol{Y}) &= \left[\frac{\log\left(\left(\sum_\ell 2^{d-1}\cdot S_\ell\right) + \left(1-\sum_\ell \nu(\ell)\right)\right)}{d}\right]^{dr} \\
&\leq \left[\frac{\log\left(2^{d-1}\cdot S + 1\right)}{d}\right]^{dr} && \left[\text{Since } S = \sum_\ell S_\ell\right] \\
&\leq \left[\frac{\log\left(2^{d}\cdot S\right)}{d}\right]^{dr} && [\text{Since } S \geq 1]
\end{aligned}$$

□

Plugging this bound back into our expression and recalling that $\lambda(F) = 32^{d+1}\cdot\left(\frac{\log 2^d\cdot S}{d}\right)^d$, we obtain

$$\Pr\left[\mathrm{CDT}_0^{(a)}(F,\tilde{\boldsymbol{\rho}}) \text{ exists} \mid \tilde{\boldsymbol{\rho}} \in \mathcal{T}\right] \leq (p\cdot\lambda(F))^s \cdot \sum_r \frac{1}{2^r} \leq (p\cdot\lambda(F))^s,$$

which completes the proof of our main lemma. □

6 Satisfiablity algorithms

In this section, we a give a randomized #SAT algorithm for general AC^0 formulae, matching the Impagliazzo-Matthews-Paturi result for AC^0 circuits. Rossman [Ros19] had obtained a similar result for regular formulae. The proof below is a verbatim adaptation of Rossman's corresponding result [Ros19, Theorem 30] for regular formulae to the general setting.

Theorem 6.1. *There is a randomized, zero-error algorithm which, given an* AC^0 *formula* F *of depth* $d+1$ *and size* S *on* n *variables, outputs a decision tree for* F *of size* $O\left(Sn\cdot 2^{(1-\varepsilon)n}\right)$ *where* $\varepsilon = 1/O\left(\left(\frac{1}{d}\log S\right)^d\right)$. *This algorithm also solves the #*SAT *problem, that is, it counts the number of satisfying assignments for* F.

Proof. Given any depth d formula, and restriction tree $\tilde{\rho}$, the decision tree algorithm from Definition 3.3 computes $\mathrm{CDT}(F,\tilde{\rho})$ in time $O(n)\cdot\sum_{G\in T_F}\mathrm{size}(\mathrm{CDT}(G,\tilde{\rho}|_G))$. Given an AC^0 formula, consider the following tree of subsets $\tilde{D}\colon T_F \to [n]$ such that for each $G, H \in T_F$, such that G is a parent of H, $\tilde{D}(H) \subseteq \tilde{D}(G)$. For each such $\tilde{D}$, we get a decision tree for F as follows: We first construct a decision tree Γ by querying all the variables in $\tilde{D}(F)$ and labelling each leaf with the corresponding restriction on $\tilde{D}(F)$. For each such leaf σ (i.e, for each choice $\sigma\colon \tilde{D}(F) \to \{0,1\}$), we get a corresponding restriction tree $\tilde{\rho}_{\tilde{D},\sigma}$ in the natural manner. For each such σ, construct $\mathrm{CDT}(F,\tilde{\rho}_{\tilde{D},\sigma})$ and plug it in instead of the leaf corresponding to σ in the complete binary tree Γ. Clearly, this resultant tree $\Gamma_{\tilde{D}}$ is a decision tree for F.

We construct a (random) $\mathbf{\Gamma}_{\tilde{\boldsymbol{D}}}$ by sampling a $\tilde{\boldsymbol{D}}$ as follows: randomly choose a $\boldsymbol{\tau} \in_R [0,1]^n$ and set $\tilde{\boldsymbol{D}}(G) := \{i \colon \boldsymbol{\tau}_i \leq 1 - 1/8\lambda(G)\}$ for each $G \in T_F$. Therefore the expected running time of the algorithm which computes the decision tree for F is $O(n) \cdot \sum_{G \in T_F} \mathbb{E}_{\boldsymbol{\tau}} \left[\sum_{\sigma \colon \tilde{\boldsymbol{D}}(F) \to \{0,1\}} \text{size}(\text{CDT}(G, \tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma}|_G)) \right]$ while the expected size of the decision tree is $\mathbb{E}_{\boldsymbol{\tau}} \left[\sum_{\sigma \colon \tilde{\boldsymbol{D}}(F) \to \{0,1\}} \text{size}(\text{CDT}(F, \tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma})) \right]$.

We bound these expression as follows. For each $G \in T_F$, size of the decision tree $\text{CDT}(G, \tilde{\rho})$ is given by the expression,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\tau}} \left[\sum_{\sigma \colon \{0,1\}^{\tilde{\boldsymbol{D}}(F)}} \text{size}(\text{CDT}(F, \tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma})) \right] &= \mathbb{E}_{\boldsymbol{\tau}} \left[2^{|\tilde{\boldsymbol{D}}(F)|} \cdot \mathbb{E}_{\sigma} \left[\text{size}(\text{CDT}(F, \tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma})) \right] \right] \\
&\leq 2^{n(1-1/16\lambda(F))} \cdot \underbrace{\mathbb{E}_{\boldsymbol{\tau},\sigma} \left[\text{size}(\text{CDT}(F, \tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma})) \right]} + 2^n \cdot e^{-(\frac{1}{2})^2 \cdot \frac{1}{2} \cdot \frac{n}{8\lambda(G)}}
\end{aligned}$$

where in the last expression, we have used the Chernoff Bound $\Pr[\sum X_i \leq (1-\delta)\mu] \leq e^{-\delta^2\mu/2}$ to bound the probability $\Pr[|[n] \setminus \tilde{\boldsymbol{D}}(F)| \leq (1 - 1/2)\mu]$ where $\mu = \mathbb{E}[|[n] \setminus \tilde{\boldsymbol{D}}(F)|] = n/8\lambda(F)$. We can now further simplify the expression indicated in the underbraces as follows:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\tau},\sigma} \left[\text{size}(\text{CDT}(F, \tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma})) \right] &= \sum_{t \geq 0} \sum_{a \in \{0,1\}^t} \sum_{b \in \{0,1\}} \Pr_{\tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma}} \left[\text{CDT}_b^{(a)}(F, \tilde{\boldsymbol{\rho}}_{\tilde{\boldsymbol{D}},\sigma}) \text{ exists} \right] \\
&\leq 1 + \sum_{t=1}^{\infty} 2^t \left(\frac{1}{8} \right)^t = \frac{4}{3}.
\end{aligned}$$

We thus conclude that the expected size of the decision tree is at most $2^{n(1-\frac{1}{C\lambda})}$ for a suitably large constant C. □

Acknowledgements

The first and the second authors spent several years thinking about this problem and we are indebted to several people along the way. First and foremost, we thank Jaikumar Radhakrishnan and Ramprasad Saptharishi for spending innumerable hours in the early and latter stages of this project respectively going over various parts of the proof and giving us very helpful feedback. In addition, we also thank Srikanth Srinivasan, Siddharth Bhandari, Ben Rossman, Yuval Filmus and Mrinal Kumar for their comments and feedback during the various stages of this project.

References

[Ajt83] MIKLÓS AJTAI. *Σ_1^1-formulae on finite structures*. Ann. Pure Appl. Logic, 24(1):1–48, July 1983. 1, 2

[Bea94] PAUL BEAME. *A switching lemma primer*. Technical Report UW-CSE-95-07-01, University of Washington, 1994. 7

[BIS12] PAUL BEAME, RUSSELL IMPAGLIAZZO, and SRIKANTH SRINIVASAN. *Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#AC^0$-SAT*. In VENKATESAN GURUSWAMI, ed., *Proc. 27th IEEE Conf. on Comput. Complexity*, pages 117–125. 2012. 2

[FSS84] MERRICK L. FURST, JAMES B. SAXE, and MICHAEL SIPSER. *Parity, circuits, and the polynomial-time hierarchy*. Mathematical Systems Theory, 17(1):13–27, 1984. (Preliminary version in 22*nd FOCS*, 1981). 1, 2

[Hås89] JOHAN HÅSTAD. *Almost optimal lower bounds for small depth circuits*. In SILVIO MICALI, ed., *Randomness and Computation*, volume 5 of *Advances in Computing Research*, pages 143–170. JAI Press, Greenwich, Connecticut, 1989. (Preliminary version in *18th STOC* 1986). 1, 2

[Hås14] ———. *On the correlation of parity and small-depth circuits*. SIAM J. Comput., 43(5):1699–1708, 2014. eccc:2012/TR12-137. 2, 3, 4, 7, 17, 19

[IMP12] RUSSELL IMPAGLIAZZO, WILLIAM MATTHEWS, and RAMAMOHAN PATURI. *A satisfiability algorithm for AC^0*. In YUVAL RABANI, ed., *Proc. 23rd Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 961–972. 2012. arXiv:1107.3127. 2, 3

[LMN93] NATHAN LINIAL, YISHAY MANSOUR, and NOAM NISAN. *Constant depth circuits, Fourier transform, and learnability*. J. ACM, 40(3):607–620, 1993. (Preliminary version in *30th FOCS*, 1989). 3

[Raz87] ALEXANDER A. RAZBOROV. Нжние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического сложения *(Russian) [Lower bounds on the size of bounded depth circuits over a complete basis with logical addition]*. Mathematicheskie Zametki, 41(4):598–607, 1987. (English translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987). doi:10.1007/BF01137685. 1

[Raz95] ———. *Bounded arithmetic and lower bounds in Boolean complexity*. In PETER CLOTE and JEFFREY B. REMMEL, eds., *Feasible Mathematics II*, volume 13 of *Progress in Computer Science and Applied Logic*, pages 344–387. Birkhäuser, 1995. 7

[Ros18] BENJAMIN ROSSMAN. *The average sensitivity of bounded-depth formulas*. Comput. Complexity, 27(2):209–223, 2018. (Preliminary version in *56th FOCS*, 2015). arXiv:1508.07677, eccc:2015/TR15-143. 3

[Ros19] ———. *Criticality of regular formulas*. In AMIR SHPILKA, ed., *Proc. 34th Comput. Complexity Conf.*, volume 137 of *LIPIcs*, pages 1:1–1:28. Schloss Dagstuhl, 2019. 2, 3, 4, 7, 10, 24

[RR97] ALEXANDER A. RAZBOROV and STEVEN RUDICH. *Natural proofs*. J. Comput. Syst. Sci., 55(1):24–35, 1997. (Preliminary version in *26th STOC*, 1991). 2

[Sip83] MICHAEL SIPSER. *Borel sets and circuit complexity*. In DAVID S. JOHNSON, RONALD FAGIN, MICHAEL L. FREDMAN, DAVID HAREL, RICHARD M. KARP, NANCY A. LYNCH, CHRISTOS H. PAPADIMITRIOU, RONALD L. RIVEST, WALTER L. RUZZO, and JOEL I. SEIFERAS, eds., *Proc. 15th ACM Symp. on Theory of Computing (STOC)*, pages 61–69. 1983. 1

[Smo87] ROMAN SMOLENSKY. *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*. In ALFRED V. AHO, ed., *Proc. 19th ACM Symp. on Theory of Computing (STOC)*, pages 77–82. 1987. 1

[Tal17] AVISHAY TAL. *Tight bounds on the Fourier Spectrum of AC^0*. In RYAN O'DONNELL, ed., *Proc. 32nd Comput. Complexity Conf.*, volume 79 of *LIPIcs*, pages 15:1–15:31. Schloss Dagstuhl, 2017. eccc:2014/TR14-174. 3

[Yao85] ANDREW CHI-CHIH YAO. *Separating the polynomial-time hierarchy by oracles (preliminary version)*. In MANUEL BLUM, JOHN HOPCROFT, JEFF LAGARIAS, TOM LEIGHTON, CHARLES RACKOFF, LARRY RUZZO, LARRY STOCKMEYER, ROBERT TARJAN, and FRANCES YAO, eds., *Proc. 26th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 1–10. 1985. 1, 2

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il