Criticality and Decision Tree Size of AC^0

Benjamin Rossman*
University of Toronto

June 21, 2017

Abstract

We show that the decision tree size of a function $f : \{0,1\}^n \to \{0,1\}$ computed by an AC^0 circuit of depth d and size s is at most $O(2^{(1-p)n})$ where $p = 1/O(\log s)^{d-1}$. This result follows from an exponential tail bound on the decision tree depth of the randomly restricted function $f{\upharpoonright}\mathbf{R}_p$: for all $t \geq 0$, we show that

$$\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p) \geq t\ ] \leq \exp(-t).$$

For $t \leq \log s$, this is a straightforward consequence of Håstad's Switching Lemma (1986); for $t \geq \log s$, we combine Håstad's Multi-Switching Lemma (2014) with a shrinkage lemma for decision trees.

Qualitatively, our bound on $\mathsf{DT}_{\mathsf{size}}(f)$ improves a similar bound on the subcube partition number of f due to Impagliazzo, Matthews and Paturi (2011). Our bound on $\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p)$ improves a similar bound of Tal (2014) on the degree of $f{\upharpoonright}\mathbf{R}_p$ as a real polynomial.

1 Introduction

For every Boolean function f, there is a sufficiently small value of $p > 0$ such that the random variable $\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p)$ (i.e. the decision tree depth of f under the random restriction $\mathbf{R}_p$) obeys an exponential tail bound. To give a name to this phenomenon, let's say that f is *p-critical* if for all $t \geq 0$,

$$\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p) \geq t\ ] \leq \exp(-t).$$

We observe that criticality implies an upper bound on decision tree size. Namely, we show that if $f : \{0,1\}^n \to \{0,1\}$ is p-critical, then $\mathsf{DT}_{\mathsf{size}}(f) \leq O(2^{(1-p)n})$ (Proposition 5.2).

The main result in this note (Theorem 6.1) shows every function f computable by an AC^0 circuit of depth d and size s is p-critical for $p = 1/O(\log s)^{d-1}$. The argument is based on a combination of Håstad's Switching Lemma [2] (for $t \leq \log s$) and Multi-Switching Lemma [3] (for $t \geq \log s$).[1] As a corollary, we get a bound of $O(2^{(1-p)n})$ on the decision tree size of f. Qualitatively, this improves a similar bound of Impagliazzo, Matthews and Paturi [4] on the subcube partition number

*Supported by NSERC.

[1]The argument is a slight extension of Håstad's methods. The main technical novelty is a tweak in the application of the Multi-Switching Lemma (see Lemma 4.3).

of f.[2] Our bound on $\mathsf{DT_{size}}(f)$ is moreover effective in the following sense: there is a randomized algorithm which, given the circuit computing f, outputs a decision tree of size $O(2^{(1-p)n})$ in time $\mathrm{poly}(s) \cdot O(2^{(1-p)n})$. We omit the details of this algorithm, but note that it gives an AC^0-SAT algorithm that essentially matches the running time of the algorithm of [4].

Our result on the criticality of AC^0 circuits also qualitatively strengthens recent results of Tal [7], who considers a different property of Boolean functions f that we will call *p-degree-criticality*: for all $t \geq 0$,

$$\mathbb{P}[\ \mathsf{deg}(f{\restriction}\mathbf{R}_p) \geq t\ ] \leq \exp(-t).$$

(Note that p-criticality implies p-degree-criticality since $\mathsf{deg}(\cdot) \leq \mathsf{DT_{depth}}(\cdot)$.) By a similar combination of the Switching and Multi-Switching Lemmas, Tal shows that every function f computable by a depth-d size-s AC^0 circuit is p-degree-critical for $p = 1/O(\log s)^{d-1}$. Tal further shows that p-degree-criticality (called the "switching lemma type property" in [7]) implies tight bounds on the Fourier tails of AC^0 functions, improving the bounds given by the LMN Theorem [5].

2 Preliminaries

$\log(\cdot)$ denotes the base-2 logarithm.

$\mathsf{DT_{depth}}(f)$ and $\mathsf{DT_{size}}(f)$ denote the decision tree depth and decision tree size of a Boolean function f (i.e. the minimum depth/size of a decision tree that computes f).

A *restriction* w.r.t. a Boolean function f is a function ϱ from the variables of f (w.l.o.g. the set $\{1,\dots,n\}$) to $\{0,1,\star\}$. The restricted Boolean function is denoted $f{\restriction}\varrho : \{0,1\}^{\varrho^{-1}(\star)} \to \{0,1\}$.

For $p \in [0,1]$, $\mathbf{R}_p$ denotes the random restriction which maps each variable independently to $\star$ with probability p and to 0 and 1 with probability $(1-p)/2$.

Circuit refers to single-output AC^0 circuits. *Depth* of a circuit is the maximum number of AND and OR gates on any input-to-output path. *Size* of a circuit is the total number of gates. Under this definition, depth-0 circuits have size 0 and depth-1 circuits have size 1.

Depth of a decision tree is the maximum number of variables queried on a branch. *Size* of a decision tree is the number of branches (i.e. the number of leaves).

3 Decision Trees

For a decision tree T, random variable $\mathbf{W}(T)$ is the number of variables read on a uniform random input (in other words, the length of a random walk down T). This random variable has density function

$$\mathbb{P}[\ \mathbf{W}(T) = \ell\ ] = 2^{-\ell} \cdot \#\{\text{leaves of } T \text{ at distance } \ell \text{ from the root}\}.$$

For a decision tree T and a restriction ϱ, let $T{\restriction}\varrho$ be the syntactically restricted decision tree (defined in the obvious way).

[2] Impagliazzo, Matthews and Paturi show that there is a partition of $\{0,1\}^n$ into at most $2^{(1-\varepsilon)n}$ subcubes on which f is constant, where $\varepsilon = 1/O(\log(s/n) + d\log d)^{d-1}$. Our bound on decision tree size implies an equivalent $O(2^{(1-p)n})$ bound on the subcube partition number of f. Note that p and ε are within a factor of $O(1)^{d-1}$ when $s \geq n^{1+\Omega(1)}$.

Lemma 3.1 (Syntactic Decision Tree Shrinkage Lemma). *If T is a depth-k decision tree, then*

$$\mathbb{P}[\ T{\restriction}\mathbf{R}_p \textit{ has depth} \geq \ell\ ] \leq (2epk/\ell)^\ell.$$

Proof. Without loss of generality, assume that no variable is queries more than once on any branch of T. Observe that random variables $\mathbf{W}(T{\restriction}\mathbf{R}_p)$ and $\mathbf{Bin}(\mathbf{W}(T), p)$ are identically distributed. Using this observation, we have

$$\begin{aligned}
\mathbb{P}[\ T{\restriction}\mathbf{R}_p \text{ has depth} \geq \ell\ ] &= \mathbb{P}_{\varrho\sim\mathbf{R}_p}[\ \mathbb{P}[\ \mathbf{W}(T{\restriction}\varrho) \geq \ell\ ] \geq 2^{-\ell}\ ] \\
&\leq 2^\ell\, \mathbb{P}[\ \mathbf{W}(T{\restriction}\mathbf{R}_p) \geq \ell\ ] \\
&= 2^\ell\, \mathbb{P}[\ \mathbf{Bin}(\mathbf{W}(T), p) \geq \ell\ ] \\
&\leq 2^\ell\, \mathbb{P}[\ \mathbf{Bin}(k, p) \geq \ell\ ] \\
&\leq (2p)^\ell \binom{k}{\ell} \\
&\leq (2epk/\ell)^\ell.
\end{aligned}$$

□

Corollary 3.2 (Semantic Decision Tree Shrinkage Lemma). *If $\mathsf{DT}_{\mathsf{depth}}(f) \leq k$, then*

$$\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\restriction}\mathbf{R}_p) \geq \ell\ ] \leq (2epk/\ell)^\ell.$$

4 Switching Lemmas

We consider the following classes.

- $\mathcal{DT}(k)$ is the class of [Boolean functions computed by] depth-k decision trees.
- $\mathcal{CKT}(d, s)$ is the class of single-output depth-d size-s circuits. $\mathcal{CKT}(d; s_1, \dots, s_d)$ is the subclass of circuits in $\mathcal{CKT}(d, s)$ which have s_i depth-i subcircuits for all $i \in \{1, \dots, d\}$ (where $s_1 + \cdots + s_d = s$ and $s_d = 1$).
- $\mathcal{CKT}(d, s) \circ \mathcal{DT}(k)$ is the class of circuits in $\mathcal{CKT}(d, s)$ whose inputs are labeled by decision trees in $\mathcal{DT}(t)$ over a common set of variables.
- $\mathcal{DT}(t) \circ \mathcal{CKT}(d, s) \circ \mathcal{DT}(k)$ is the class of depth-t decision trees with leaves labeled by circuits in $\mathcal{CKT}(d, s) \circ \mathcal{DT}(k)$ over a common set of variables.

Note that $\mathcal{CKT}(d, s) = \mathcal{CKT}(d, s) \circ \mathcal{DT}(1) = \mathcal{DT}(0) \circ \mathcal{CKT}(d, s) \circ \mathcal{DT}(1)$.

4.1 Håstad's Switching and Multi-Switching Lemmas

We state the Swiching Lemma of Håstad (1986) and Multi-Swiching Lemma of Håstad (2014) in the form we will use. (The original statements speak about k-DNF/CNFs. Lemma 4.1, below, includes a union bound over depth-1 subcircuits.)

Lemma 4.1 (Switching Lemma [2]). *If $f \in \mathcal{CKT}(d; s_1, \dots, s_d) \circ \mathcal{DT}(k)$, then*

$$\mathbb{P}[\ f{\restriction}\mathbf{R}_p \notin \mathcal{CKT}(d-1; s_2, \dots, s_d) \circ \mathcal{DT}(t-1)\ ] \leq s_1(5pk)^t.$$

Lemma 4.2 (Multi-Switching Lemma [3]). *If $f \in \mathcal{CKT}(d; s_1, \dots, s_d) \circ \mathcal{DT}(k)$ and $\ell \geq \log s_1 + 1$, then*

$$\mathbb{P}[\ f{\restriction}\mathbf{R}_p \notin \mathcal{DT}(t-1) \circ \mathcal{CKT}(d-1; s_2, \dots, s_d) \circ \mathcal{DT}(\ell)\ ] \leq s_1(50pk)^t.$$

4.2 Combined Multi-Switching Lemma

Key to our main result is the following lemma, which combines the Multi-Switching Lemma 4.2 with the Syntactic Decision Tree Shrinkage Lemma 3.1.

Lemma 4.3 (Combined Multi-Switching Lemma)**.** *If* $f \in \mathcal{DT}(t-1) \circ \mathcal{CKT}(d; s_1, \dots, s_d) \circ \mathcal{DT}(k)$ *and* $\ell \geq \log s_1 + 1$*, then*

$$\mathbb{P}[\ f{\restriction}\mathbf{R}_p \notin \mathcal{DT}(t-1) \circ \mathcal{CKT}(d-1; s_2, \dots, s_d) \circ \mathcal{DT}(\ell)\ ] \leq s_1(200pk)^{t/2}.$$

Note that Lemma 4.3 has a stronger hypothesis than Lemma 4.2, but gives a weaker bound of $s_1(200pk)^{t/2}$ compared to $s_1(50pk)^t$; these bounds are comparable when $t \geq \log s_1$ and $pk \ll 1$.

Proof. Suppose f is computed by a depth $t-1$ decision tree T in which each leaf λ is labeled by a circuit $C_\lambda \in \mathcal{CKT}(d, s, m) \circ \mathcal{DT}(k)$. Consider events

$$\mathcal{A} \overset{\text{def}}{\iff} T{\restriction}\mathbf{R}_p \text{ has depth} \leq \lceil t/2 \rceil - 1,$$
$$\mathcal{B} \overset{\text{def}}{\iff} C_\lambda{\restriction}\mathbf{R}_p \in \mathcal{DT}(\lceil t/2 \rceil - 1) \circ \mathcal{CKT}(d-1; s_2, \dots, s_d) \circ \mathcal{DT}(\ell) \text{ for every leaf } \lambda \text{ of } T.$$

Observe that

$$\mathcal{A} \wedge \mathcal{B} \implies f{\restriction}\mathbf{R}_p \in \mathcal{DT}(t-1) \circ \mathcal{CKT}(d-1; s_2, \dots, s_d) \circ \mathcal{DT}(\log s + 1).$$

By the Syntactic Decision Tree Shrinkage Lemma 3.1, we have

$$\begin{aligned} \mathbb{P}[\ \neg\mathcal{A}\ ] &= \mathbb{P}[\ T{\restriction}\mathbf{R}_p \text{ has depth} \geq \lceil t/2 \rceil\ ] \\ &\leq (2ep(t-1)/\lceil t/2 \rceil)^{\lceil t/2 \rceil} \\ &\leq (4ep)^{t/2}. \end{aligned}$$

By a union bound and the Multi-Switching Lemma 4.2, we have

$$\begin{aligned} \mathbb{P}[\ \neg\mathcal{B}\ ] &\leq \sum_\lambda \mathbb{P}[\ C_\lambda{\restriction}\mathbf{R}_p \notin \mathcal{DT}(\lceil t/2 \rceil - 1) \circ \mathcal{CKT}(d-1; s_2, \dots, s_d) \circ \mathcal{DT}(\ell)\ ] \\ &\leq \sum_\lambda s_1(50pk)^{\lceil t/2 \rceil} \\ &\leq 2^{t-1} s_1(50pk)^{\lceil t/2 \rceil}. \end{aligned}$$

Putting things together, we have

$$\begin{aligned} \mathbb{P}[\ f{\restriction}\mathbf{R}_p \notin \mathcal{DT}(t-1) \circ \mathcal{CKT}(d-1; s_2, \dots, s_d) \circ \mathcal{DT}(\log s + 1)\ ] &\leq \mathbb{P}[\ \neg\mathcal{A}\ ] + \mathbb{P}[\ \neg\mathcal{B}\ ] \\ &\leq (4ep)^{t/2} + 2^{t-1} s_1(50pk)^{t/2} \\ &\leq \tfrac{1}{2}(16ep)^{t/2} + \tfrac{1}{2} s_1(200pk)^{t/2} \\ &\leq s_1(200pk)^{t/2}. \end{aligned}$$

□

5 Criticality

Definition 5.1. We say that a Boolean function f is *p-critical* if for all $t \geq 0$,

$$\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p) \geq t\ ] \leq \exp(-t).$$

Observe that every n-variable Boolean function $f : \{0,1\}^n \to \{0,1\}$ is $1/en$-critical, since

$$\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_{1/en}) \geq t\ ] \leq \mathbb{P}[\ \mathbf{Bin}(n, 1/en) \geq t\ ] \leq \exp(-t).$$

Every Boolean function is thus p-critical for some $p > 0$. (Note that the original Switching Lemma of [2] implies that every k-DNF (or k-CNF) is $1/5ek$-critical.)

A key property of criticality is that it implies an bound on decision tree size.

Proposition 5.2. *Every p-critical function $\{0,1\}^n \to \{0,1\}$ has a decision tree of size $O(2^{(1-p)n})$.*

Proof. Suppose $f : \{0,1\}^n \to \{0,1\}$ is p-critical. Let $\mathbf{I}$ be a random $(1-p)$-binomial random subset of $[n]$ (with density function $\mathbb{P}[\ \mathbf{I} = I\ ] = (1-p)^{|I|}p^{n-|I|}$ for every $I \subseteq [n]$). Let $\boldsymbol{\varrho} : \mathbf{I} \to \{0,1\}$ be a uniform random function. Note that $\boldsymbol{\varrho}$ has distribution $\mathbf{R}_p$ when viewed on its own as a random restriction $[n] \to \{0,1,\star\}$.

We obtain a decision tree for f by querying all variables in $\mathbf{I}$ and considering the decision tree depth of $f{\upharpoonright}\varrho$ for each $\varrho : \mathbf{I} \to \{0,1\}$. We show that the resulting decision tree has size at most $20 \cdot 2^{(1-p)n}$ with probability < 1. Therefore, f has a decision tree of this size (by the magic of the probabilistic method).

First, we observe that, for any fixed $I \subseteq [n]$,

$$\mathsf{DT}_{\mathsf{size}}(f) \leq \sum_{\varrho : I \to \{0,1\}} 2^{\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\varrho)} = 2^{|I|} \mathop{\mathbb{E}}_{\boldsymbol{\varrho} : I \to \{0,1\}}[\ 2^{\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\boldsymbol{\varrho})}\ ]. \tag{1}$$

Using the fact that every median of $\mathbf{Bin}(n,p)$ is at least $\lfloor pn \rfloor$, we have

$$\mathbb{P}\Big[\ |\mathbf{I}| > \lceil (1-p)n \rceil\ \Big] = \mathbb{P}\Big[\ \mathbf{Bin}(n, 1-p) > n - \lfloor pn \rfloor\ \Big] = \mathbb{P}\Big[\ \mathbf{Bin}(n,p) < \lfloor pn \rfloor\ \Big] \leq \frac{1}{2}. \tag{2}$$

Putting things together, we have

$$\begin{aligned}
\mathop{\mathbb{P}}_{\mathbf{I}}[\ \mathsf{DT}_{\mathsf{size}}(f) > 20 \cdot 2^{(1-p)n}\ ] &\leq \mathop{\mathbb{P}}_{\mathbf{I}}\Big[\ 2^{|\mathbf{I}|} \mathop{\mathbb{E}}_{\boldsymbol{\varrho}:\mathbf{I}\to\{0,1\}}[\ 2^{\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\boldsymbol{\varrho})}\ ] > 20 \cdot 2^{(1-p)n}\ \Big] \quad \text{(by (1))}\\
&\leq \mathop{\mathbb{P}}_{\mathbf{I}}\Big[\ \Big(2^{|\mathbf{I}|} > 2^{(1-p)n+1}\Big) \vee \Big(\mathop{\mathbb{E}}_{\boldsymbol{\varrho}:\mathbf{I}\to\{0,1\}}[\ 2^{\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\boldsymbol{\varrho})}\ ] > 10\Big)\ \Big]\\
&\leq \mathop{\mathbb{P}}_{\mathbf{I}}\Big[\ |\mathbf{I}| > \lceil (1-p)n \rceil\ \Big] + \mathop{\mathbb{P}}_{\mathbf{I}}\Big[\ \mathop{\mathbb{E}}_{\boldsymbol{\varrho}:\mathbf{I}\to\{0,1\}}[\ 2^{\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\boldsymbol{\varrho})}\ ] > 10\ \Big]\\
&\leq \frac{1}{2} + \frac{1}{10}\mathbb{E}[\ 2^{\mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p)}\ ] \quad \text{(by (2) and Markov's inequality)}\\
&= \frac{1}{2} + \frac{1}{10}\sum_{t=0}^{\infty} 2^t \cdot \underbrace{\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p) = t\ ]}_{\leq \exp(-t) \text{ by } p\text{-crit. of } f}\\
&= \frac{1}{2} + \frac{1}{10} \cdot \frac{1}{1-(2/e)}\\
&< 1.
\end{aligned}$$

Therefore, $\mathsf{DT}_{\text{size}}(f) \leq 20 \cdot 2^{(1-p)n}$. ∎

Although not needed in this paper, we state another property of p-criticality.

Proposition 5.3. *If f is a p-critical Boolean function, then for all $0 \le q \le p$ and $t \ge 0$,*

$$\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\restriction}\mathbf{R}_q) \ge t\ ] \le O(q/p)^t.$$

Proof. Since the bound is trivial if $q \ge p$, we assume that $q \le p$. Generate $\mathbf{R}_q$ as the composition of a random restriction $\boldsymbol{\varrho}_1 \sim \mathbf{R}_p$ (over the variables of f) and $\boldsymbol{\varrho}_2 \sim \mathbf{R}_{q/p}$ (over the variables of $f{\restriction}\boldsymbol{\varrho}_1$). We have

$$\begin{aligned}
&\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\restriction}\mathbf{R}_q) \ge t\ ]\\
&\quad = \mathop{\mathbb{E}}_{\boldsymbol{\varrho}_1}\Big[\ \mathop{\mathbb{P}}_{\boldsymbol{\varrho}_2}[\ \mathsf{DT}_{\mathsf{depth}}((f{\restriction}\boldsymbol{\varrho}_1){\restriction}\boldsymbol{\varrho}_2) \ge t\ ]\ \Big]\\
&\quad = \sum_{k=t}^{\infty} \underbrace{\mathop{\mathbb{P}}_{\boldsymbol{\varrho}_1}[\ \mathsf{DT}_{\mathsf{depth}}(f{\restriction}\boldsymbol{\varrho}_1) = k\ ]}_{\le \exp(-k) \text{ by } p\text{-crit. of } f} \cdot \mathop{\mathbb{E}}_{\boldsymbol{\varrho}_1}\Big[\ \underbrace{\mathop{\mathbb{P}}_{\boldsymbol{\varrho}_2}[\ \mathsf{DT}_{\mathsf{depth}}((f{\restriction}\boldsymbol{\varrho}_1){\restriction}\boldsymbol{\varrho}_2) \ge t\ ]}_{\le (2eqk/pt)^t \text{ by Cor. 3.2}}\ \Big|\ \mathsf{DT}_{\mathsf{depth}}(f{\restriction}\boldsymbol{\varrho}_1) = k\ \Big]\\
&\quad \le \sum_{k=t}^{\infty} \exp(-k) \cdot (2eqk/pt)^t\\
&\quad = (4eq/p)^t \cdot \sum_{i=0}^{\infty} \exp(-t-i) \cdot \big(\underbrace{(t+i)/2t}_{\le \exp(i/2t)}\big)^t\\
&\quad \le (4q/p)^t \cdot \sum_{i=0}^{\infty} \exp(-i/2)\\
&\quad < 3(4q/p)^t.
\end{aligned}$$

□

6 Criticality of AC^0 Circuits

Theorem 6.1. *Every Boolean function computed by an AC^0 circuit of depth d and size s is p-critical for $p = 1/O(\log s)^{d-1}$.*

Proof. Let C be a circuit of depth d and size s, which computes a Boolean function f. Let $s = s_1 + \cdots + s_d$ where s_i is the number of gates at depth i. Note that $s_d = 1$, corresponding to the output gate of C.

Let $\ell = \lceil \log s \rceil + 1$ and let $p = 1/12800^{d+1}\ell^{d-1}$. For $i \in \{1, \dots, d\}$, let $p_i = 1/12800^i \ell^{d-1}$. Note that $p_1 = p/p_d = 1/12800$ and $p_i/p_{i-1} = 1/12800\ell$ for all $i \in \{2, \dots, d\}$.

We wish to show that

$$\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\restriction}\mathbf{R}_p) \ge t\ ] \le \exp(-t)$$

for all $t \ge 1$. For the case $t \le \log s$, we show this using Håstad's Switching Lemma 4.1 in the completely standard way. For the case $t \ge \log s$, we show this using the Combined Multi-Switching Lemma 4.3.

Case 1: $t \le \log s$**.**

For $i \in \{1, \dots, d-1\}$, let $\mathcal{A}_i$ denote the event that $\mathsf{DT}_{\mathsf{depth}}(g \restriction \mathbf{R}_{p_i}) \le \ell$ for every function g computed by a depth-i subcircuit of C. By the Switching Lemma 4.1, we have

$$\mathbb{P}[\ \neg \mathcal{A}_1\ ] \le s_1(5p_1)^{\ell} = s_1(1/2560)^{\ell}.$$

Again by the Switching Lemma 4.1, we have

$$\mathbb{P}[\ \neg \mathcal{A}_2 \mid \mathcal{A}_1\ ] \le s_2(5(p_2/p_1)\ell)^{\ell} = s_2(1/2560)^{\ell}.$$

Here we view $\mathbf{R}_{p_2}$ as the composition of $\mathbf{R}_{p_1}$ (over the variables of f) and $\mathbf{R}_{p_2/p_1}$ (over the free variables of $\mathbf{R}_{p_1}$).

Similarly, we have

$$\mathbb{P}[\ \neg \mathcal{A}_i \mid \mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_{i-1}\ ] \le s_i(1/2560)^{\ell}$$

for $i = 3, \dots, d-1$. Therefore,

$$\begin{aligned}
\mathbb{P}[\ \neg \mathcal{A}_{d-1}\ ] &\le \textstyle\sum_{i=1}^{d-1} \mathbb{P}[\ \neg \mathcal{A}_i \mid \mathcal{A}_1 \wedge \cdots \wedge \mathcal{A}_{i-1}\ ] \\
&\le (s_1 + \cdots + s_{d-1})(1/2560)^{\ell} \\
&= (s-1)(1/2560)^{\ell} \\
&\le (1/1280)^{\ell} \quad \text{(since } \ell > \log s\text{)} \\
&\le (1/1280)^{t} \quad \text{(since } \ell > t\text{).}
\end{aligned}$$

By a final application of the Switching Lemma, we have

$$\begin{aligned}
\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f \restriction \mathbf{R}_p) \ge t \mid \mathcal{A}_{d-1}\ ] &\le (5(p/p_{d-1})\ell)^{t} \\
&= (1/32768000)^{t}.
\end{aligned}$$

Therefore, we get a final bound

$$\begin{aligned}
\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f \restriction \mathbf{R}_p) \ge t\ ] &\le \mathbb{P}[\ \neg \mathcal{A}_{d-1}\ ] + \mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f \restriction \mathbf{R}_p) \ge t \mid \mathcal{A}_{d-1}\ ] \\
&\le (1/1280)^{t} + (1/32768000)^{t} \\
&\le \exp(-t).
\end{aligned}$$

Case 2: $t \ge \log s$**.**

Initially, we have $f \in \mathcal{CKT}(d; s_1, \dots, s_d) \circ \mathcal{DT}(1)$.

For $i \in \{1, \dots, d\}$, let $\mathcal{B}_i$ be the event

$$\mathcal{B}_i \stackrel{\text{def}}{\iff} f \restriction \mathbf{R}_{p_i} \in \mathcal{DT}(t-1) \circ \mathcal{CKT}(d-i; s_{i+1}, \dots, s_d) \circ \mathcal{DT}(\ell).$$

In particular, note that

$$\mathcal{B}_d \iff f \restriction \mathbf{R}_{p_d} \in \mathcal{DT}(t + \ell - 1)$$

since $\mathcal{DT}(t-1) \circ \mathcal{CKT}(0,0) \circ \mathcal{DT}(\ell) = \mathcal{DT}(t + \ell - 1)$.

By the Multi-Switching Lemma 4.2, we have

$$\mathbb{P}[\ \neg\mathcal{B}_1\ ] \leq s_1(50p_1)^t = s_1(1/256)^t.$$

Next, for all $i = 2, \dots, d$, by the Combined Multi-Switching Lemma 4.3 we have

$$\mathbb{P}[\ \neg\mathcal{B}_i \mid \mathcal{B}_1 \wedge \cdots \wedge \mathcal{B}_{i-1}\ ] \leq s_i(200(p_i/p_{i-1})\ell)^{t/2} = s_i(1/64)^{t/2} = s_i(1/8)^t.$$

Therefore,

$$\begin{aligned}
\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_{p_d}) \geq t+\ell\ ] &= \mathbb{P}[\ \neg\mathcal{B}_d\ ] \\
&\leq \textstyle\sum_{i=1}^d \mathbb{P}[\ \neg\mathcal{B}_i \mid \mathcal{B}_1 \wedge \cdots \wedge \mathcal{B}_{i-1}\ ] \\
&\leq s_1(1/256)^t + (s_2 + \cdots + s_d)(1/8)^t \\
&\leq s(1/8)^t \\
&\leq s(1/8)^{\frac{1}{3}\log s + \frac{2}{3}t} \qquad \text{(since } t \geq \log s) \\
&= (1/4)^t.
\end{aligned}$$

As a last step, we apply the Semantic Decision Tree Shrinkage Lemma 3.2:

$$\begin{aligned}
\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p) \geq t \mid \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_{p_d}) \leq t+\ell-1\ ] &\leq (2e(p/p_{d+1})t/(t+\ell-1))^t \\
&\leq (e/3200)^t
\end{aligned}$$

using $p/p_{d+1} = 1/12800$ and $t + \ell - 1 = t + \lceil \log s \rceil \geq 2t$.

Putting things together, we have

$$\begin{aligned}
\mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p) \geq t\ ] &\leq \mathbb{P}[\ f{\upharpoonright}\mathbf{R}_{p_d} \geq t+\ell\ ] + \mathbb{P}[\ \mathsf{DT}_{\mathsf{depth}}(f{\upharpoonright}\mathbf{R}_p) \geq t \mid f{\upharpoonright}\mathbf{R}_{p_d} \leq t+\ell-1\ ] \\
&\leq (1/4)^t + (e/3200)^t \\
&\leq \exp(-t).
\end{aligned}$$

□

Acknowledgements

This note emerged from conversations with Prahladh Harsha, Rahul Santhanam, Srikanth Srinivasan and Avishay Tal. I thank Ian Mertz and Toni Pitassi as well for helpful discussions.

References

[1] Paul Beame, Russell Impagliazzo, and Srikanth Srinivasan. Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#AC^0$-SAT. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 117–125. IEEE, 2012.

[2] Johan Hastad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, pages 6–20. ACM, 1986.

[3] Johan Hastad. On the correlation of parity and small-depth circuits. *SIAM Journal on Computing*, 43(5):1699–1708, 2014.

[4] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 961–972. SIAM, 2012.

[5] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.

[6] Avishay Tal. Properties and applications of boolean function composition. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 441–454. ACM, 2013.

[7] Avishay Tal. Tight bounds on the Fourier spectrum of AC^0. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 21, page 174, 2014.