**COMS 6998: Unconditional Lower Bounds & Derandomization** **Spring 2024**

# Lecture 4: February 6, 2024

*Lecturer: Rocco Servedio* *Scribe: Ryan Anselm*

## Overview

Last time, we talked about:

- The KRW conjecture
- Lower bounds for full-basis formulas
- Constant-depth circuits: start $2^{\Omega(n^{1/(d-1)})}$ size lower bound for depth-$d$ circuits for computing the parity function; intuition for and statement of the Switching Lemma

Today, we will:

- Demonstrate how to use Håstad's Switching Lemma (HSL) to get a $2^{\Omega(n^{1/(d-1)})}$ size lower bound for depth-$d$ circuits for parity (PAR)
- Prove a weak version of the Switching Lemma
- Prove the full version of Håstad's Switching Lemma

## 1 HSL $\Longrightarrow$ Lower Bound for Constant Depth Circuits

Recall that $R_p$ is the distribution over restrictions $\rho : [n] \to \{0, 1, *\}$ such that each element $\boldsymbol{\rho}(i)$ of a restriction $\boldsymbol{\rho} \sim R_p$ is an i.i.d. random variable where $\Pr[\boldsymbol{\rho}(i) = *] = p$ and $\Pr[\boldsymbol{\rho}(i) = 0] = \Pr[\boldsymbol{\rho}(i) = 1] = \frac{1-p}{2}$.

**Theorem 1.** *(Håstad's Switching Lemma)*
*Let $f(x_1, \ldots, x_n)$ be computed by a width-$w$ DNF (or CNF). Let DT-depth($f$) be the decision tree depth of $f$. Let $f\restriction_{\boldsymbol{\rho}}$ be the restriction of $f$ by $\boldsymbol{\rho}$. Then for $t \geq 1$, $0 < p < 1$,*

$$\Pr_{\boldsymbol{\rho} \sim R_p}[DT\text{-}depth(f\restriction_{\boldsymbol{\rho}}) \geq t] \leq (7 \cdot p \cdot w)^t.$$

Assume for now that Håstad's Switching Lemma is true. Let $C$ be a depth-$d$, size-$M$ circuit that computes $\mathrm{PAR}_n$. We'll argue using HSL that $M \geq 2^{\Omega(n^{1/(d-1)})}$, unconditionally. Let's say $C$ is a circuit that has alternating layers of the form AND-OR-...-AND-OR.[1] We will now collapse this circuit by hitting it with random restrictions in successive stages.

**Stage 0:** Perform an initial "trim" to reduce the bottom fan-in. Hit C with $\boldsymbol{\rho} \sim R_{\frac{1}{100}}$: $*$-prob. $= \frac{1}{100}$, 1-prob. $= \frac{49.5}{100}$, 0-prob. $= \frac{49.5}{100}$.

Observation: If some bottom-level OR gate has fan-in $> 10 \log M$, then it will simplify to a constant 1 as long as there is at least one literal assigned to 1 beneath it (each variable is assigned to 1 w.p. 0.495). So for $p = \frac{1}{100}$

$$\Pr_{\boldsymbol{\rho} \sim R_p}[\text{output of the OR gate is fixed to 1 and vanishes}] \geq 1 - (.505)^{10 \log M} \gg 1 - \frac{1}{M^5}.$$

By upper bounding over all ($\leq M$) bottom level gates, we get that:

1. w.p. $> \frac{1}{2}$ (actually much higher), $\boldsymbol{\rho} \sim R_{\frac{1}{100}}$ kills all gates of fan-in $> 10 \log M$.
2. Also w.p. $> \frac{1}{2}$, at least $\frac{n}{200}$ variables survive $\boldsymbol{\rho} \sim R_{\frac{1}{100}}$.

Since both events occur w.p. $> \frac{1}{2}$, there exists some $\boldsymbol{\rho}$ where both occur simultaneously. Fix that $\boldsymbol{\rho}$ and call $C_0$ the circuit $C \restriction_{\boldsymbol{\rho}}$.

Any restriction of a circuit computing the parity function will compute either a parity function or its negation on the subset of variables that are left unfixed by the restriction. Therefore $C_0$ computes $\mathrm{PAR}_k$ or its negation on the remaining $k \geq \frac{n}{200}$ variables and has bottom fan-in $\leq 10 \log M$.

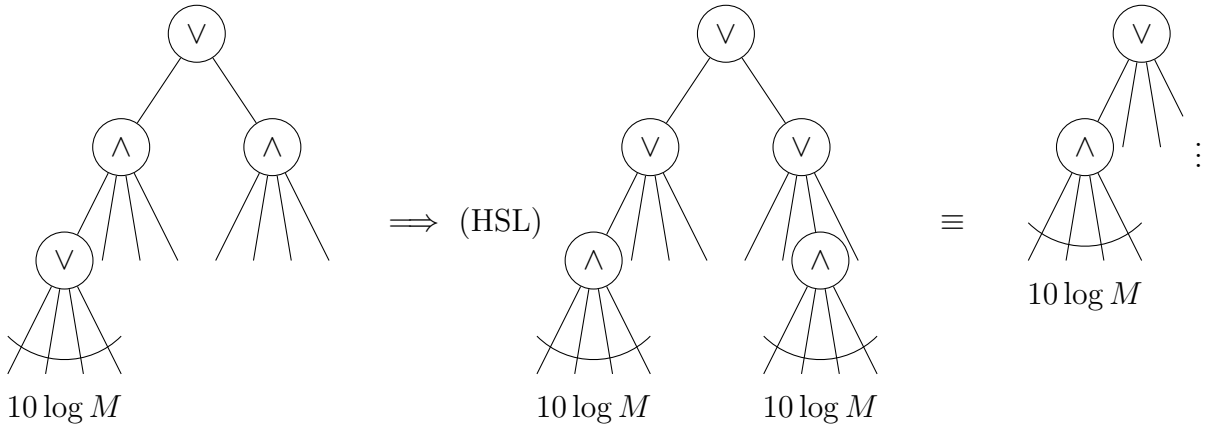
**Stage 1:** Hit $C_0$ with $\boldsymbol{\rho} \sim R_{\frac{1}{100 \log M}}$.

Each depth-2 circuit at the bottom of $C_0$ is an at most $(10 \log M)$-width CNF, so we can fruitfully apply HSL to it! Take $t = 10 \log M$. HSL $\implies$ this depth-2 circuit collapses to a $\leq (10 \log M)$-depth decision tree except with failure probability $(7 \cdot p \cdot w)^{t=10 \log M} = (7 \cdot (\frac{1}{100 \log M}) \cdot (10 \log M))^{10 \log M} = (0.7)^{10 \log M} \approx \frac{1}{M^5}$.

Upper-bounding over all ($\leq M$) depth-2 circuits at the bottom level of $C_0$, we get that:

[1] We can safely assume this because two consecutive ANDs or two consecutive ORs could be collapsed down to a single layer.

1. w.p. $> \frac{1}{2}$ (again, actually much higher), each subcircuit becomes a decision tree of depth at most $10 \log M$, which can be rewritten as a DNF of width at most $10 \log M$. Observe that since the original bottom-level depth-2 subcircuits were CNFs, the last three layers of the original circuit were of the form ...-OR-AND-OR. Upon conversion of the bottom-level depth-2 subcircuits from width-$(10 \log M)$ CNFs to width-$(10 \log M)$ DNFs, the last three layers become of the form ...-OR-OR-AND, which after combining the adjacent OR layers can be collapsed into ...-OR-AND, reducing the overall depth of the circuit by 1. See the figure below for a visual depiction of depth reduction.




Depth reduction at bottom level of circuit

2. w.p. $> \frac{1}{2}$, the number of surviving variables $\geq \frac{n}{200} \cdot \frac{1}{200 \log M}$

Again, there exists a $\boldsymbol{\rho}$ satisfying both these properties which we fix, and then we define $C_1 = C_0 \restriction_{\boldsymbol{\rho}}$. $C_1$ computes PAR or its negation on $\geq \frac{n}{200} \cdot \frac{1}{200 \log M}$ variables and is a depth-$(d-1)$ circuit with $\leq M$ gates at distance 2 from the bottom.

**Stages $\mathbf{2}, \mathbf{3}, \ldots, \mathbf{d-3}, \mathbf{d-2}$:** In each successive stage, we hit $C_{i-1}$ with $\boldsymbol{\rho} \sim R_{\frac{1}{100 \log M}}$ to obtain $C_i = C_{i-1} \restriction_{\boldsymbol{\rho}}$, as we did in Stage 1. We can tabulate the properties of the circuits obtained.

| Current circuit | Depth | max bottom fan-in | # variables in play |
|---|---|---|---|
| $C$ | $d$ | $n$ | $n$ |
| $C_0$ | $d$ | $10 \log M$ | $\frac{n}{200}$ |
| $C_1$ | $d-1$ | $10 \log M$ | $\frac{n}{200 \cdot (200 \log M)}$ |
| $C_2$ | $d-2$ | $10 \log M$ | $\frac{n}{200 \cdot (200 \log M)^2}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $C_{d-2}$ | 2 | $10 \log M$ | $\frac{n}{200 \cdot (200 \log M)^{d-2}}$ |

$C_{d-2}$ computes PAR or its negation on at least $\frac{n}{200 \cdot (200 \log M)^{d-2}}$ variables and it is a width-$(10 \log M)$ CNF. By the lower bound on the width of a CNF computing PAR, $C_{d-2}$ must satisfy

$$\frac{n}{200^{d-1} \cdot (\log M)^{d-2}} \leq 10 \log M$$

i.e. that the width of the circuit exceeds the number of unfixed variables, which is met only when $M \geq 2^{\Omega(n^{1/(d-1)})}$. $\square$

# 2 Proof of Weak Switching Lemma

We'll now prove a weaker version of Håstad's Switching Lemma.

**Theorem 2.** *(Weak Switching Lemma) Let $f(x_1, \ldots, x_n)$ be computed by a width-$w$ DNF (or CNF). Let DT-depth($f$) be the decision tree depth of $f$. Let $f\restriction_{\boldsymbol{\rho}}$ be the restriction of $f$ by $\boldsymbol{\rho}$. Then for $t \geq 1$, $0 < p < 1$,*

$$\Pr_{\boldsymbol{\rho} \sim R_p}[DT\text{-}depth(f\restriction_{\boldsymbol{\rho}}) \geq t] \leq (40 \cdot p \cdot w \cdot 2^w)^t.$$

There is a major difference in the structure of the proofs:

- In the Håstad's Switching Lemma proof, we first restrict $f$ by $\boldsymbol{\rho}$, then analyze the decision tree for $f\restriction_{\boldsymbol{\rho}}$.
- In the Weak Switching Lemma proof, we first construct a decision tree $T$ for $f$, then analyze $T\restriction_{\boldsymbol{\rho}}$.

The Weak Switching Lemma proof is easier because a decision tree $T$ hit by $\boldsymbol{\rho}$ is easy to understand! If

$$\mathrm{T} = \begin{array}{ccc} & x_i & \\ \swarrow & & \searrow \\ T_0 & & T_1 \end{array}$$

is hit by $\boldsymbol{\rho}$, then

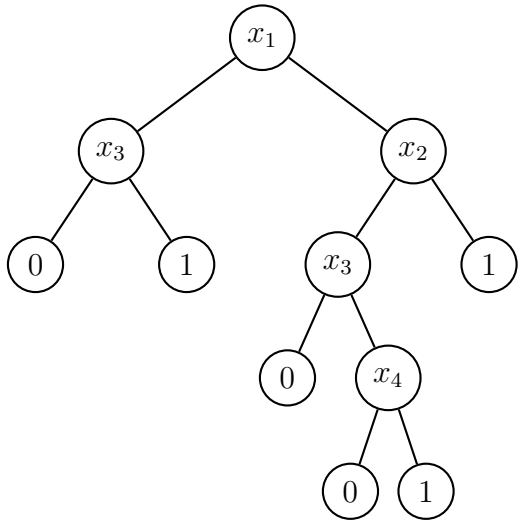
$$T\restriction_{\boldsymbol{\rho}} = \begin{cases} \begin{array}{ccc} & x_i & \\ \swarrow & & \searrow \\ T_0\restriction_{\boldsymbol{\rho}} & & T_1\restriction_{\boldsymbol{\rho}} \end{array} & \text{if } x_i \leftarrow * \\ T_0\restriction_{\boldsymbol{\rho}} & \text{if } x_i \leftarrow 0 \\ T_1\restriction_{\boldsymbol{\rho}} & \text{if } x_i \leftarrow 1 \end{cases}$$

**Definition 3.** *A decision tree $T$ is w-clipped if every node in $T$ has $\geq$ one leaf at a distance of $\leq w$ below it.*

**Lemma 4.** *Any width-$w$ DNF (or CNF) is equivalent to a $w$-clipped decision tree.*

*Proof.* We can build decision tree $T$ by querying variables in each successive term. Read/query variables in the current term one by one, adding a 1-leaf under the path that satisfies the current term, and moving on to the next term on paths where the current term is falsified. When we have satisfied all terms, put down 0-leaves on all remaining variables to finish the tree. This decision tree will be $w$-clipped because at any internal node, the branch satisfying the current term ends in a 1-leaf at a distance $\leq w$. ■

**Example:** Suppose we want to construct the 2-clipped decision tree for the width-2 DNF $(x_1x_2) \vee (\overline{x_1}x_3) \vee (x_3x_4)$. By following the path that satisfies the term currently being queried, there will always be at least one 1-leaf no more than 2 levels away from any node. For this DNF the decision tree will look as follows:




The Weak Switching Lemma follows from the following lemma:

**Theorem 5.** *(w-clipped Switching Lemma)*
*Let $f(x_1, \ldots, x_n)$ be computed by a $w$-clipped decision tree. Then for $t \geq 1$, $0 < p < 1$,*

$$\Pr_{\boldsymbol{\rho} \sim R_p} [DT\text{-}depth(f \restriction_{\boldsymbol{\rho}}) \geq t] \leq (40 \cdot p \cdot w \cdot 2^w)^t.$$

Define Branches($T$) to be the set of all root-to-leaf paths in $T$. Consider a random walk down $T$. Let $W(T)$ be the distribution over branches corresponding to a uniform random walk down from the root. A path $\pi \in$ Branches($T$) will have probability mass $2^{-|\pi|}$ under $W(T)$ where $|\pi|$ = length of $\pi$.

**Lemma 6.** *Let $T$ be a proper decision tree (no variable occurs twice on any branch). The following distributions are equivalent:*

- $\mathcal{D}_1(T)$*: draw* $\boldsymbol{\rho} \sim R_p$*, consider* $T\restriction_{\boldsymbol{\rho}}$*, output* $\sigma \sim W(T\restriction_{\boldsymbol{\rho}})$*.*
- $\mathcal{D}_2(T)$*: draw* $\boldsymbol{\pi} = (\boldsymbol{\pi_1}, \boldsymbol{\pi_2}, \dots) \sim W(T)$*, and output sublist* $\boldsymbol{\sigma}$ *obtained by going through* $\boldsymbol{\pi_1}, \boldsymbol{\pi_2}, \dots$ *and including each one with probability* $p$*. (*$|\boldsymbol{\sigma}| \sim Bin(|\boldsymbol{\pi}|, p)$*)*

*Proof:* Official Homework Problem

Now we present the proof of the $w$-clipped Switching Lemma.

*Proof.* ($w$-clipped Switching Lemma) Observe that for any $\boldsymbol{\rho} \sim R_p$, we have that

$$\text{depth}(T\restriction_{\boldsymbol{\rho}}) \geq t \implies \Pr_{\boldsymbol{\sigma} \sim W(T\restriction_{\boldsymbol{\rho}})}[|\boldsymbol{\sigma}| \geq t] \geq 2^{-t}.$$

By some basic rearrangement, Markov's inequality states that if $Z \geq 0$ is any random variable, we have that $\Pr[Z \geq 2^{-t}] \leq 2^t \cdot \mathbb{E}[Z]$. So we have

$$\begin{aligned}
\Pr_{\boldsymbol{\rho} \sim R_p}[\text{depth}(T\restriction_{\boldsymbol{\rho}} \geq t] &\leq \Pr_{\boldsymbol{\rho} \sim R_p}\left[\Pr_{\boldsymbol{\sigma} \sim W(T\restriction_{\boldsymbol{\rho}})}[|\boldsymbol{\sigma}| \geq t] \geq 2^{-t}\right] \\
&\leq 2^t \cdot \mathbb{E}_{\boldsymbol{\rho} \sim R_p}\left[\Pr_{\boldsymbol{\sigma} \sim W(T\restriction_{\boldsymbol{\rho}})}[|\boldsymbol{\sigma}| \geq t]\right] \quad \text{(Markov)} \\
&= 2^t \cdot \mathbb{E}_{\boldsymbol{\pi} \sim W(T)}\left[\Pr_{y \in \text{Bin}(|\boldsymbol{\pi}|, p)}[y \geq t]\right] \quad \text{(Equivalence lemma)} \\
&\leq 2^t \cdot \mathbb{E}_{\boldsymbol{\pi} \sim W(T)}\left[p^t \binom{|\boldsymbol{\pi}|}{t}\right] \\
&= (2p)^t \cdot \mathbb{E}_{\boldsymbol{\pi} \sim W(T)}\left[\binom{|\boldsymbol{\pi}|}{t}\right] \\
&\leq (40 \cdot p \cdot w \cdot 2^w)^t \quad \text{(from the following claim)}
\end{aligned}$$

■

**Claim 7.** *If $T$ is $w$-clipped, then* $\mathbb{E}_{\boldsymbol{\pi}\sim W(T)}[\binom{|\boldsymbol{\pi}|}{t}] \leq (20 \cdot w \cdot 2^w)^t$

*Proof.* Let $\boldsymbol{X}$ = the number of tosses that have occured when it is the first time that a fair coin lands on $w$ consecutive heads. ($\boldsymbol{X} \in \{w, w+1, w+2, \dots\}$). This r.v. *stochastically dominates* $|\boldsymbol{\pi}|$ for $\boldsymbol{\pi} \sim W(T)$:

$$\forall \alpha : \Pr[\boldsymbol{X} \geq \alpha] \geq \Pr[|\boldsymbol{\pi}| \geq \alpha].$$

For any monotonically increasing function $g$, we have that $\mathbb{E}[g(\boldsymbol{X})] \geq \mathbb{E}[g(|\boldsymbol{\pi}|)]$, so it is sufficient to show that $\mathbb{E}[\binom{\boldsymbol{X}}{t}] \leq (20 \cdot w \cdot 2^w)^t$.

We can show that $\mathbb{E}[\boldsymbol{X}^t] \leq (7 \cdot w \cdot t \cdot 2^w)^t$ (Official Homework Problem).

So $\mathbb{E}[\binom{\boldsymbol{X}}{t}] \leq \mathbb{E}[(\frac{e\boldsymbol{X}}{t})^t] \leq (\frac{e}{t})^t \cdot (7 \cdot w \cdot t \cdot 2^w)^t < (20 \cdot w \cdot 2^w)^t$. ■

# 3 Proof of Håstad's Switching Lemma

First, we will change the coefficient in HSL from 7 to a 10 for this version of the proof.

**Theorem 8.** *(Håstad's Switching Lemma, slightly modified)*
*Let $f(x_1, \dots, x_n)$ be computed by a width-$w$ DNF (or CNF). Let DT-depth($f$) be the decision tree depth of $f$. Let $f\restriction_{\boldsymbol{\rho}}$ be the restriction of $f$ by $\boldsymbol{\rho}$. Then for $t \geq 1$, $0 < p < 1$,*

$$\Pr_{\boldsymbol{\rho}\sim R_p}[\textit{DT-depth}(f\restriction_{\boldsymbol{\rho}}) \geq t] \leq (10 \cdot p \cdot w)^t.$$

Note: If $p \geq \frac{1}{10w}$ then $(10 \cdot p \cdot w) \geq 1$ and the lemma is trivially true, so the lemma is only relevant when $p$ is small ($p < \frac{1}{10w}$), meaning that $*$'s are unlikely.

Let $F = T_1 \vee T_2 \vee \dots$ be a width-$w$ DNF (fixing order of terms). What happens to <u>one</u> term e.g. $T_i = x_1 \wedge \overline{x_3} \wedge x_4$ under $\boldsymbol{\rho} \sim R_p$?

- Some literal could get set to 0; $T_i\restriction_{\boldsymbol{\rho}} \equiv 0$. i.e. $T_i$ vanishes from $F\restriction_{\boldsymbol{\rho}}$. If <u>every</u> $T_i$ has this happen, $F\restriction_{\boldsymbol{\rho}} \equiv 0$.
- Could have each of $T_i$'s $w$ literals get set to 1. If this happens for <u>any</u> $T_i$, $F\restriction_{\boldsymbol{\rho}} \equiv 1$.
- Otherwise, maybe some literals in $T_i$ get $*$, other literals in $T_i$ get 1. This means $T_i$ (potentially) "shrinks", but survives.

So overall, F could simplify to 1, simplify to 0, or some $T_i$ vanish to 0, while others may shrink.

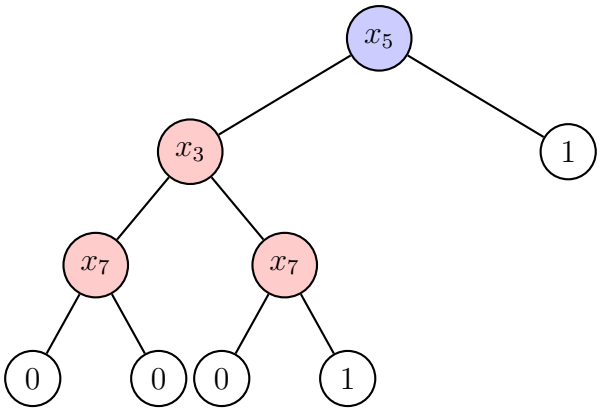
Goal: show that w.h.p. with respect to $\boldsymbol{\rho} \sim R_p$, some decision tree for $F\restriction_{\boldsymbol{\rho}}$ is shallow. We'll argue that w.h.p., a particular way of writing a decision tree for $F\restriction_{\boldsymbol{\rho}}$ is shallow. Denote this $\mathrm{CDT}(F, \boldsymbol{\rho})$ the "Canonical Decision Tree for $F$ after $\boldsymbol{\rho}$".

**Definition 9.** *Let $F = T_1 \vee T_2 \vee \dots$. $CDT(F, \boldsymbol{\rho})$ is defined as follows.*

1. *If $F\restriction_{\boldsymbol{\rho}}$ is killed or set to 0 or 1, that's $CDT(F, \boldsymbol{\rho})$. Otherwise, go to next step.*
2. *Find first $T_i$ not killed by $\boldsymbol{\rho}$. $CDT(F, \boldsymbol{\rho})$ obliviously queries all surviving variables in that term.*
3. *Recurse at each leaf: for each leaf, do steps (1), (2) under the augmented restriction corresponding to that leaf.*

Note: When $\mathrm{CDT}(F, \boldsymbol{\rho})$ queries a block of variables, some unique assignment/path satisfies all literals in that restricted term, and we get a 1-leaf in $\mathrm{CDT}(F, \boldsymbol{\rho})$ there.

**Example:** Suppose $F = (x_1 \wedge \overline{x_2} \wedge x_4) \vee (x_2 \wedge x_5 \wedge \overline{x_6}) \vee (x_3 \wedge \overline{x_5} \wedge x_7)$.
Suppose $\boldsymbol{\rho}$ is chosen as $x_1 = *; x_2 = 1; x_3 = *; x_4 = 1; x_5 = *; x_6 = 0; x_7 = *$. Then, $T_1\restriction_{\boldsymbol{\rho}} = 0$, $T_2\restriction_{\boldsymbol{\rho}} = x_5$, $T_3\restriction_{\boldsymbol{\rho}} = x_3 \wedge \overline{x_5} \wedge x_7$, so $\boxed{F\restriction_{\boldsymbol{\rho}} = (x_5) \vee (x_3 \wedge \overline{x_5} \wedge x_7)}$. The $\mathrm{CDT}(F, \boldsymbol{\rho})$ would be

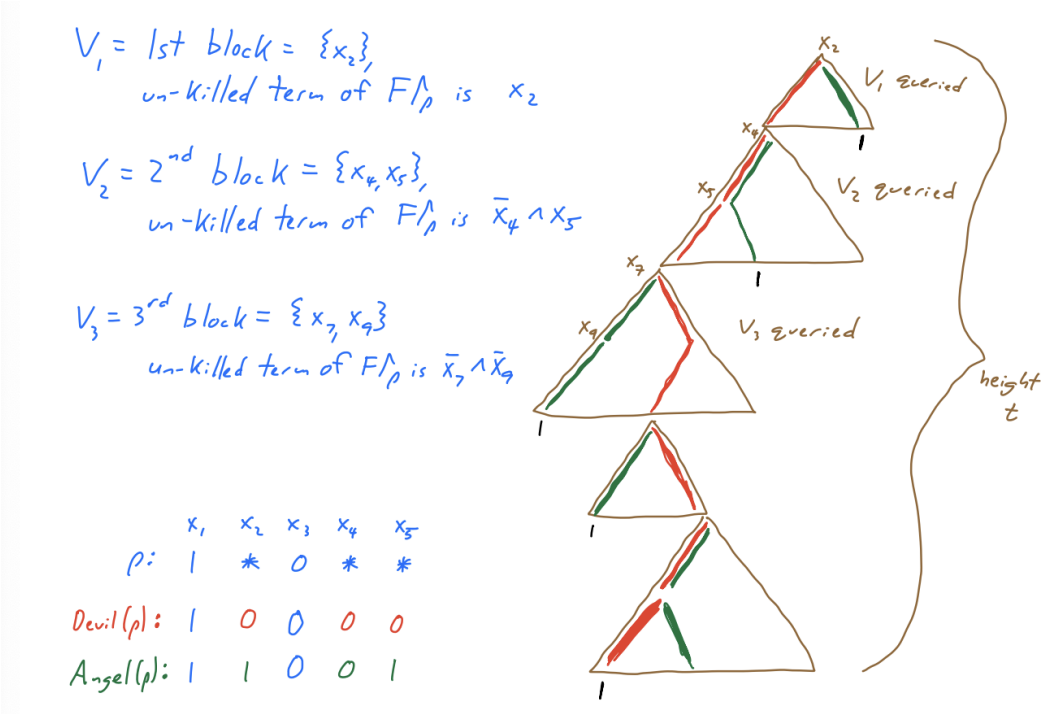



Blue labels correspond to the first block and red labels correspond to the second block.

Define BAD = set of restrictions $\boldsymbol{\rho}$ such that $\mathrm{CDT}(F, \boldsymbol{\rho})$ has depth $\geq t$. We want to show that

$$\Pr_{\boldsymbol{\rho} \sim R_p}[\boldsymbol{\rho} \in \mathrm{BAD}] \leq (10 \cdot p \cdot w)^t.$$

Fix a $\boldsymbol{\rho} \in \text{BAD}$. $\text{CDT}(F, \boldsymbol{\rho})$ has depth $\geq t$, so at least one path of depth $\geq t$ exists. Let $P$ be the variable assignments corresponding to following the leftmost path of depth $\geq t$ in $\text{CDT}(F, \boldsymbol{\rho})$. For simplicity, assume that the depth of $P$ is $t$. So $\boldsymbol{\rho} + P$ is a new restriction that fixes $t$ more variables than $\boldsymbol{\rho}$ did, which we denote as $\text{Devil}(\boldsymbol{\rho})$. Observe that $\text{Devil}(\boldsymbol{\rho})$ has "segments" based on blocks of $\text{CDT}(F, \boldsymbol{\rho})$ that are queried along the path induced by $P$. Let $V_1$ be the first block queried along $P$, $V_2$ the second block queried along $P$, and so on.




Let $\delta_1$ be the restriction fixing the variables of $V_1$ as they are fixed in $\text{Devil}(\boldsymbol{\rho})$. Similarly, let $\delta_2$ be the restriction fixing the variables of $V_2$ as they are fixed in $\text{Devil}(\boldsymbol{\rho})$. So $\text{Devil}(\boldsymbol{\rho})$ is $\boldsymbol{\rho} \circ \delta_1 \circ \delta_2 \circ \delta_3 \circ \ldots$. $\text{Devil}(\boldsymbol{\rho})$ is kind of a "worst case" path down $\text{CDT}(F, \boldsymbol{\rho})$ because it avoids reaching satisfying assignments until $t$ variables have been queried.

Consider a different restriction called $\text{Angel}(\boldsymbol{\rho})$, which is a set of disconnected path segments that tells us the best path we can follow *from the start of each block* to reach the 1-leaf of the current block, given that up to this point we have received only bad query results taken from $\text{Devil}(\boldsymbol{\rho})$. Like $\text{Devil}(\boldsymbol{\rho})$, $\text{Angel}(\boldsymbol{\rho})$ fixes $t$ additional variables beyond $\boldsymbol{\rho}$. However, $\text{Angel}(\boldsymbol{\rho})$ fixes variables in $V_i$ to reach the 1-leaf of the block $V_i$, so within each block $\text{Angel}(\boldsymbol{\rho})$ and $\text{Devil}(\boldsymbol{\rho})$ disagree.

Q: Is Angel($\boldsymbol{\rho}$) or $\boldsymbol{\rho}$ more likely under $R_p$?
A: Angel($\boldsymbol{\rho}$) is *much* more likely: it has more fixed bits (0 or 1), and fixed bits are much more likely to be drawn than $*$'s.

In more detail: if we fix any $\boldsymbol{\rho} \in \{0, 1, *\}^n$ with $k$ many $*$'s

$$\Pr_{R_p}[\boldsymbol{\rho}] = \left(\frac{1-p}{2}\right)^{n-k} \cdot p^k$$

So for any bad $\boldsymbol{\rho} \sim R_p$ with $k$ $*$'s

$$\frac{\Pr_{R_p}[\boldsymbol{\rho}]}{\Pr_{R_p}[\text{Angel}(\boldsymbol{\rho})]} = \frac{\left(\frac{1-p}{2}\right)^{n-k} \cdot p^k}{\left(\frac{1-p}{2}\right)^{n-(k-t)} \cdot p^{k-t}} = \left(\frac{2p}{1-p}\right)^t \leq (2.5p)^t$$

Note that the last inequality holds because $p$ is small ($p < 0.2$). For any bad $\boldsymbol{\rho}$, we have that

$$\Pr_{R_p}[\boldsymbol{\rho}] \leq (2.5p)^t \cdot \Pr_{R_p}[\text{Angel}(\boldsymbol{\rho})] \quad (!)$$

**Key Fact** (to be proven next time): *Any restriction $\sigma$ is Angel($\boldsymbol{\rho}$) for $\leq (4w)^t$ many bad $\boldsymbol{\rho}$'s.*

Finally we can apply all of the inequalities we've collected:

$$\begin{aligned}
\Pr_{\boldsymbol{\rho} \sim R_p}[\text{BAD}] &= \sum_{\boldsymbol{\rho} \in \text{BAD}} \Pr_{R_p}[\boldsymbol{\rho}] \\
&\leq \sum_{\boldsymbol{\rho} \in \text{BAD}} (2.5p)^t \cdot \Pr_{R_p}[\text{Angel}(\boldsymbol{\rho})] \quad (!) \\
&\leq (2.5p)^t \cdot (4w)^t \cdot \underbrace{\sum_{\sigma} \Pr_{\boldsymbol{\rho} \sim R_p}[\sigma]}_{=1} \quad \text{(Key Fact)} \\
&= (10 \cdot p \cdot w)^t. \ \square
\end{aligned}$$

# Next Time

- Prove "Key Fact" to the proof of HSL
- average case $\text{AC}^0$ lower bounds

- average case lower bound for DNF/CNFs via "random projections"
- $\mathbb{F}_2$ polynomials