Today we want to prove what we just claimed last lecture, namely that

1. Using random restrictions of the right type, we can collapse d-DNF formulas to decision trees of small height with good probability (this is known as a SWITCHING LEMMA).

2. If we have a d-DNF resolution refutation of \( F \) \( \Pi = (H_1, H_2, \ldots, H_{k-1}, H_k = 1) \) where all lines \( H_i \) can be represented by decision trees of small height, then \( F \) can be refuted in (standard) resolution in small width.

In future lectures, we will combine 1 and 2 with resolution width lower bounds to prove d-DNF resolution length lower bounds.

Let us start by proving 2, which is easier.

We need to refresh our memories about the concepts:

(a) Strong implication
(b) Strong soundness
(c) Strong representation
In what follows $G_1, \ldots, G_s, H$ are d-DNF formulas.

$G_1, \ldots, G_s$ imply $H$ if for any (total) truth value assignment $\tau$ s.t. $\tau(G_i) \neq 1 \forall i$ it also holds that $\tau(H) = 1$

Notation $G_1 \land \ldots \land G_s \vdash H$

$G_1, \ldots, G_s$ strongly imply $H$ if for any terms $t_i \in G_i$, i.e. that are mutually consistent (i.e., $t_i$ satisfiable) there is a $t \in H$ s.t. $\bigwedge_{i=1}^s t_i \vdash t$.

A proof system operating with DNF formulas is strongly sound if whenever

$$G_1, \ldots, G_s \vdash H$$

is a derivation rule, it holds that $G_1, \ldots, G_s$ strongly imply $H$

As noted last time, d-DNF resolution is strongly sound, and this is all we need to know about the derivation rules today.
DECISION TREE for $H$

Binary tree; internal nodes labelled by queried variables; leaves by 0/1

Every leaf $v$ defines partial assignment $g_v$
Every total assignment $x$ leads to node $v_x$

$T$ represents $H$ if for any total $x$

$x(H) = \text{label}(v_x)$

Let $b_T(H) = \{ g_v | T \text{ labelled by } g_v \}$

$T$ strongly represents $H$ if

(a) $\forall g \in b_T(H) \forall t \in H \quad t \neg g = 0$

(b) $\forall g \in b_T(H) \exists t \in H \quad t g = 1$

The representation height $h(H)$ of $H$ is the minimum height of any decision tree strongly representing $H$. 
THEOREM 1 Let $F$ be a $h$-CNF formula and suppose that $\overline{\pi} = (H_1, H_2, \ldots, H_k)$ is a $d$-DNF resolution refutation of $F$ such that for every line $H_i$, it holds that $h(H_i) \leq h$. Then there is a resolution refutation of $F$ in width $W(\overline{\pi}) \leq 2h$.

For convenience, let us use resolution augmented with the subsumption or weakening rule saying that from a clause $C$ we can derive $C \lor D$ for any $D$. All use of weakening can be eliminated from a resolution refutation without increasing the length (in fact, the length will decrease).

For every line $H$ in $\overline{\pi}$, fix a decision tree $T_H$ of minimal height that strongly represents $H$.

For a clause $C = a_1 \lor \cdots \lor a_h \in F$ we fix $T_C = 0 \leftarrow 1 \rightarrow a_1 \rightarrow 0 \rightarrow a_2 \rightarrow 1 \rightarrow \cdots \rightarrow a_h \rightarrow 1$.

For the final line $H_k$, we have $H_k = 1$ we have $T_{H_k} = 0$. 
For any partial assignment \( \delta \), let us define
\[
C_\neg \delta = \bigvee_{a \in \delta} \neg a
\]
1.e., \( C_\neg \delta \) is the unique maximal clause falsified by \( \delta \).

By induction over \( \Pi = (H_1, H_2, \ldots, H_L) \),
we will derive the set of clauses
\[
\{ C_\neg \delta \mid \delta \in \Gamma_0 (T_{H_i}) \}
\]
in width at most \( 2h \).

**Base case** For a clause \( C = a_1 \lor \cdots \lor a_h \in F \)
we have \( \Gamma_0 (T_C) = \{ \neg a_1, \neg a_2, \ldots, \neg a_h \} \)
and hence
\[
\{ C_\neg \delta \mid \delta \in \Gamma_0 (T_C) \} = \{ a_1, a_2, \ldots, a_h \}
\]
which is exactly what we want.

So if we can just do the inductive step,
then we are done.
Suppose that \( H \) was derived from \( H_1 \) and \( H_2 \) (\( H \) derived from just \( H_1 \) is similar).

By our inductive hypothesis, we have derived all \( C_{i_2} \) for \( i_2 \in \mathcal{B}_0(T_{H_i}) \), \( i = 1, 2 \).

We want to derive \( C_{i_2} \) for all \( i_2 \in \mathcal{B}_0(T_H) \).

To guide our inductive proof, we build a decision tree \( T \) for \( H_1 \wedge H_2 \) in the following way:

1. Take \( T_{H_1} \) and keep all internal nodes and 0-leaves, say correspondingly to \( i_2 \).
2. For every 1-leaf \( i_2 \) replace this leaf by a (separate) copy of \( T_{H_2} \).

(See example picture on next page)

The height of \( T \) is clearly at most \( 2h \).

In \( \mathcal{L} \cup \mathcal{M}_2 \), the clauses \( \{ C_{i_2} \mid i_2 \in \mathcal{B}_0(T) \} \) are derivable from \( \{ C_{i_2} \mid i_2 \in \mathcal{B}_0(T_{H_1}) \} \) or

\[ \{ C_{i_2} \mid i_2 \in \mathcal{B}_0(T_{H_2}) \} \]

by weakening.

Proof. By construction every \( i_2 \in \mathcal{B}_0(T) \) contains either some \( i_1 \in \mathcal{B}_0(T_{H_1}) \) or some \( i_2 \in \mathcal{B}_0(T_{H_2}) \), and for this \( i_2 \), it then holds that \( C_{i_2} \leq C_{i_2} \). \( \square \)
\[ C_1 = x \lor y \lor z \]
\[ C_2 = x \lor y \lor w \]

\[ T_{C_1} \]

\[ T_{C_2} \]

\[ T_{(C_1 \lor C_2)} \]
LEMMA 3. With notation as above, fix any \( \delta \in \delta_0 \left( \mathcal{T}_H \right) \). Then for all \( \varphi \in \delta_0 \left( \mathcal{T} \right) \cup \delta_2 \left( \mathcal{T} \right) \), it holds that if \( \varphi \) and \( \delta \) are consistent, then \( \varphi \in \delta_0 \left( \mathcal{T} \right) \).

**Proof.** Suppose that \( \varphi \in \delta_2 \left( \mathcal{T} \right) \). We claim that there exists a term \( t \in \mathcal{H} \) such that \( \varphi \) satisfies \( t \) (i.e., \( t / \varphi = 1 \)). By construction of \( \mathcal{T} \) there are \( \varphi_1 \in \delta_2 \left( \mathcal{T}_H \right) \) and \( \varphi_2 \in \delta_2 \left( \mathcal{T}_H \right) \) such that \( \varphi_1 \cup \varphi_2 = \varphi \).

Since \( \mathcal{T}_1 \) and \( \mathcal{T}_2 \) strongly represent \( \mathcal{H}, \) and \( \mathcal{H}_2 \), there are \( t_1 \in \mathcal{H}, \) and \( t_2 \in \mathcal{H}_2 \) such that \( \varphi = \varphi_1 \cup \varphi_2 \) satisfy \( t_1 \lor t_2 \). By the strong soundness of \( \mathcal{H}, \) \( \mathcal{D} \)-DNF resolution, there exists a term \( t \in \mathcal{H} \) such that \( t / \varphi = 1 \).

Suppose now that \( \delta \in \delta_0 \left( \mathcal{T}_H \right) \). Since \( \mathcal{T}_H \) strongly represents \( \mathcal{H}, \) \( \delta \) falsifies all terms \( t \in \mathcal{H} \). Since \( \varphi \in \delta_2 \left( \mathcal{T} \right) \) satisfies some term in \( \mathcal{H}, \) \( \delta \) and \( \varphi \) cannot be consistent.
Why does this lemma help us? Recall that we have \( \{ C_{10} \mid g \in \delta_0 (T) \} \) by our inductive hypothesis.

We want to derive \( C_{10} \) for \( \sigma \in \delta_0 (T_0) \). Let us consider what happens when we walk in \( T \) according to \( \sigma \). A priori 3 cases

1. Hit a leaf labelled 1. Let \( g \) correspondingly be assigned. Then \( g_\sigma \in \delta_{10} (T) \) and \( g_\sigma \) and \( \sigma \) consistent.

   This doesn't happen — the lemma we just proved rules this out.

2. Hit a leaf labelled 0. Then \( g_\sigma \in \delta_0 (T) \) and \( g_\sigma \leq \sigma_0 \).

   But then \( C_{10} \geq C_{g_\sigma} \in \{ C_{10} \mid g \in \delta_0 (T) \} \),

   which is a clause we already have, so \( C_{10} \) can be derived by weakening.

3. Walk some way down in \( T \) but don't hit a leaf. Then look at all \( g \in \delta_0 (T) \), \( g \leq \sigma \), corresponding to leaves in \( T \) below where we walk according to \( \sigma \) stopped. We claim that we can resolve these clauses to derive \( C_{10} \). If we can do this, then we are done with the proof of the theorem.
But let us think a little bit more...

Let $V_0$ be the node reached in $T$ when walking according to $\sigma \in B_0(T_H)$. Then there is no path to a 1-labelled leaf in $T$ going through $V_0$.

**Proof:** Let $w$ be such a leaf and consider $\sigma_w \in B_1(T)$. By construction $\sigma \leq \sigma_w$ and so $\sigma$ and $\sigma_w$ are consistent. But that contradicts the lemma.

So, pictorially, $V_0$ is the root of a subtree containing only 0-labelled leaves.

We have all clauses $C_{\neg\sigma_w}$ for the leaves $w$ in this subtree. It is intuitively kind of clear that we should just systematically resolve away the variables in the subtree bottom-up to derive $C_{\neg\sigma}$. So, for any $s$, let $V_0$ be node in $T$ reached when walking according to $s$. 
Formally, we prove the following by induction.

Lemma 5: With notation as above, let \( \sigma \in B_0(T_r) \) and let \( v_2 \in T \) be the node reached when walking from the root of \( T \) according to \( \sigma \). Let \( v_3 \) be any node in the subtree rooted at \( v_2 \), i.e., such that \( j = 0 \). Then \( C_{v_3} \) is derivable in width at most \( 2h \) from \( \delta = \{ C_{v_2} \mid \sigma \in B_0(T_r), j = 0 \} \).

Proof: All clauses we will see will have width-depth of \( T \leq 2h \), so let us not worry about that.

Base case: \( j \in B_0(T_r) \) corresponds to a leaf. Then \( \delta \) we know:

Then \( C_{v_3} \in \delta \) and there is nothing to prove.

Inductive step: \( j \notin B_0(T_r) \), so there are nodes in \( T \) labelled by \( y \in \exists x \exists \beta \) and \( y \in \forall \exists x \beta \) for the variable \( x \) queried at \( v_3 \).

By induction, can derive

\[
C_{v_3}(y \exists x \beta) = C_{v_2} \exists x \beta \quad \text{and} \quad C_{v_3}(y \forall \exists x \beta) = C_{v_2} \forall \exists x \beta
\]

Resolve to get \( C_{v_2} \). Done.

Note that width of clauses only shrink as we work our way up the tree.

Now we have established Thm 1.
SWITCHING LEMMA

Q1 When can an OR of ANDs be represented as an AND of ORs?
A1 Always. Both CNFs and ONFs can express any Boolean function.

Q2 When can the conversion from DNF to CNF (and vice versa) be efficient?
A2 Not always.

Consider 2-DNF

\[(x_1 \land x_2) \lor (x_3 \land x_4) \lor \ldots \lor (x_{2m-1} \land x_{2m})\]

CNF becomes

\[(x_1 \lor x_3 \lor \ldots \lor x_{2m-3} \lor x_{2m-1}) \land (x_1 \lor x_3 \lor \ldots \lor x_{2m-2} \lor x_{2m}) \land (x_1 \lor x_3 \lor \ldots \lor x_{2m-2} \lor x_{2m}) \land \ldots \]

et cetera.

Switching lemmas say that when we first hit DNF with suitable random restrictions, then conversion likely to be efficient.
In our case, want to convert DNF to shallow decision tree (and a decision tree of height $h$ can always be represented as a \( h \)-CNF formula).

We stated last time:

**Theorem 2 [SB1’04]**

Let $d, s, t \in \mathbb{N}^+$, let $\gamma, \delta \in (0, 1)$, and let $D$ be a distribution on partial truth value assignments such that for every $d$-DNF formula $G$ it holds that

$$\Pr_{D \leftarrow D} \big[ G / \gamma + 1 \big] \leq t \cdot 2^{-5 \log d} \delta$$

Then for every $d$-CNF formula $H$ it holds that

$$\Pr_{D \leftarrow D} \big[ h(H / \gamma) > 2s \big] \leq d t \cdot 2^{-\delta' \log d}$$

for $\delta' = 2(\delta/4)^d$ and $\gamma' = \gamma d$. 
For \( H \) \( d \)-DNF formula \( S \subseteq \text{Vars}(H) \) is a \underline{cover} of \( H \) if every term \( t \in H \) contains a variable from \( S \).

The covering number \( \text{cov}(H) \) of \( H \) is the minimum size of any cover of \( H \).

Will actually prove more general theorem from which Thm 1 follows as corollary. Take a deep breath...

\textbf{THEOREM 2} \ [5B1 '07]

Let \( d \in \mathbb{N}^+ \); let \( s_i \in \mathbb{N}^+ \), \( i \in \{0, d-1\} \) and \( p_i \in (0, 1) \), \( i \in \{1, d\} \), and be sequences of positive numbers, and let \( D \) be a distribution on partial assignments such that for every \( i \leq d \) and every \( i \)-DNF formula \( G \), it holds that if \( \text{cov}(G) > s_{i-1} \), then

\[
P[D_n \leq \lfloor G \rfloor_{s_i + 1}] \leq p_i
\]

Then for every \( d \)-DNF formula \( H \) it holds that

\[
P[D_n \leq h(H|G) \leq \sum_{i=0}^{d-1} s_i] \leq \sum_{i=1}^{d} p_i \cdot 2 \sum_{i=0}^{d-1} s_i.
\]
Wait... Aren't there minus signs missing somewhere? $\sum_{j=i}^{d-1} s_j > 1$, so $2 \sum_{j=i}^{d-1} s_j$ is kind of large, no?

Yes, so the $p_i$ had better be small...

For what remains of today's lecture we will prove this theorem. The proof is by induction on $d$, the size of the terms.

Base case [d = 1, i.e., H is a clause]

(6) If $\text{core}(H) \leq s_0$, this means that the width of the clause $H$ is $W(H) \leq s_0$, or expressed differently that $H$ contains at most $s_0$ variables.

Build decision tree of height $\leq s_0$ that queries all variables in $H$. Since all $g \in \overline{b_0(T)} \cup b_2(T)$ assign all variables in $H$, $T$ obviously strongly represents $H$.

(4) If $\text{core}(H) > s_0$, then by assumption:

$$\Pr[\text{geo} \left[ H/\gamma \neq 1 \right] \leq p_1, \text{ and if } H/\gamma = 1 \text{ then the decision tree } 2 \text{ of height 0 would}$$
So,
\[ P_{g \geq 0} \left( h(H|g) > s_0 \right) \leq P_{g} \left( h(H|g) \neq 1 \right) \leq P_1 = P_1 \cdot 2 \sum_{j=1}^{d} s_j = P_1 \cdot 2^d. \]

**Inductive step**

Theorem holds for \( d \)-DNF formulas.

Let \( H \) be a \((d+1)\)-DNF formula.

Let \( s_0, \ldots, s_d, P_0, \ldots, P_{d+1} \) be sequences as in the statement of the theorem.

**Case (a) \( \cov(H) > s_d \)**

Similar to base case:
\[ P_{g \geq 0} \left( h(H|g) > \sum_{i=0}^{d} s_i \right) \leq P_{g \geq 0} \left( H|g \neq 1 \right) \leq P_{d+1} \leq \sum_{i=1}^{d+1} P_i \cdot 2^d s_i. \]

**Case (b) \( \cov(H) \leq s_d \)**

Let \( S \leq \vars(H) \) cover of size \( |S| \leq s_d \).

For any restriction \( \gamma : S \to \{0,1\}^S \),
\( H^\gamma \) is a \( d \)-DNF formula (since \( \gamma \) hits at least one literal in every term \( t \in H \)).
What to prove that \( \text{h} \) is very likely to have shallow decision trees.

By induction, this is true for

\[
(h \land \varphi) \land \psi
\]

First sample \( g \sim D \).

For any \( \varphi \) consistent with \( g \) we have \( (h \land \varphi) \land \gamma = (h \land \varphi) \land \varphi \) and hence by the induction hypothesis:

\[
Pr \left[ h((h \land \varphi) \land \gamma) \right] \leq \sum_{i=0}^{d-1} \pi_i \cdot 2^{\sum_{j=i+1}^{d} s_j}
\]

(1)

There are at most \( 2 |S| \leq 2^d \) restrictions \( T : S \Rightarrow \{0,1\} \) that are consistent with \( g \). Hence, by the union bound, we get

\[
Pr \left[ \exists \varphi \text{ consistent with } g \text{ s.t. } h((h \land \varphi) \land \gamma) > \sum_{i=0}^{d-1} s_i \right] \leq
\]

\[
2^d \left( \sum_{i=1}^{d} \pi_i \cdot 2^{\sum_{j=i+1}^{d} s_j} \right)
\]

\[
= \sum_{i=1}^{d} \pi_i \cdot 2^i \sum_{j=1}^{d} s_j < \sum_{i=1}^{d+1} \pi_i \cdot 2^{i-1} \sum_{j=1}^{d} s_j
\]

(2)
Note that (2) is exactly the probability bound that we are shooting for.

Hence, if there is a $\tau$ s.t. $h((H|S)\Lambda \tau)$ is too large, we can afford to just give up.

But if for all $\tau$ consistent with $p$ it holds that

$$h((H|S)\Lambda \tau) \leq \sum_{i=0}^{d-1} s_i; \quad (3)$$

we proceed to construct a decision tree strongly representing $H|S$.

First query all variables in $S$ left unset by $p$.

Formally, let $\beta = T \setminus S$ (where again we think of $S = \{s_i\}$, each set to true by $p$).

Build complete binary tree (if necessary) for all assignments $\beta : (\text{dom}(\tau) \setminus \text{dom}(S)) \to \{0,1\}$.

Let this be the "upper part" of $\tau$.

At each leaf, the residual formula is

$$(H|S)\Lambda \beta = (H|S)\Lambda \tau$$

at the corresponding leaf.

For each $\beta$, plug in a decision tree of minimum height that strongly represents

$$(H|S)\Lambda \beta,$$

which has height

$$\leq \sum_{i=0}^{d-1} s_i \quad \text{by (3)}.$$
Let this be the "lower part" of the tree. Let $T$ be the tree constructed in this way. We claim two properties of $T$:

(1) $T$ has height $\leq \sum_{i=0}^{d} s_i$.

(2) $T$ strongly represents $H_p$.

Note that if we can establish these two claims, then the inductive step is finished and Thm2 follows by the induction principle.

Claim 1: This is the easy part. The upper part has height

$$|\text{dom}(13)| = |\text{dom}(S) \setminus \text{dom}(g)|$$

$$\leq |\text{dom}(N)|$$

$$\leq |S|$$

$$\leq sd \quad (4)$$

As already noted, the lower part has height at most

$$\sum_{i=0}^{d-1} s_i \quad (5)$$

Just sum up (4) & (5). Done.
Let \( \psi \in \delta_0(T) \cup \delta_2(T) \)
Can write
\[ \psi = \beta \cup \sigma \]
for
\[ \beta : S \setminus \text{dom} \sigma \rightarrow \{0,1\} \]
and
\[ \sigma \in \delta_0(T(\beta)) \cup \delta_2(T(T(\beta))) \]
where \( T(\beta) \) is a tree that strongly represents \( (H/\beta) \beta \)

Suppose \( \psi \in \delta_2(T) \). Then there is a term \( t' \in (H/\beta) \beta \) such that
\[ t' \sigma = 1. \]

This \( t' \) comes from \( t \in T \) such that
\[ t' = t \sigma_{\beta} = (t \beta) \beta \]
for \( t \beta \in H/\beta \). For this term
\[ (t \beta) \psi = (t \beta) \beta_{\psi} \equiv 1 \]
Suppose $\psi \in \mathcal{B}_0 (T)$.

As before, write $\psi = \beta \cup \delta$.

We get two cases:

(a) $(H/\beta) \setminus \beta = 0$

(b) $(H/\beta) \setminus \beta \neq 0$

In case (a), $\beta$ already falsifies every term in $H/\beta$, and externally $\beta \cup \delta$ does not change this.

In case (b), since subface underneath $\beta$ strongly represents $(H/\beta) \setminus \beta$ and $\delta \in \mathcal{B}_0 (T(\beta))$ for every term $t'' \in (H/\beta) \setminus \beta$ it holds that $t'' \setminus \delta = 0$.

Thus, for every $t' \in H/\beta$, it holds that $t' \setminus (\beta \cup \delta) = t' \setminus \psi = 0$;

either because $t'$ is inconsistent with $\beta$, or if not, then with $\delta$.

This shows that $T$ strongly represents $H/\beta$, and the theorem follows.