In this lecture we want to use the machinery set up in the last two lectures to prove a lower bound for the formula PHP_n^{cn} where c is constant. That is we will prove the following result:

> **Theorem**: for each $c>1$ there exists $\varepsilon>0$ s.t. for all n sufficiently large, if $k < \sqrt{\frac{\log n}{\log\log n}}$ then every $Res(k)$ refutation of PHP_n^{cn} require size $\geq 2^{n^\varepsilon}$.

- Recall that PHP_n^{cn} is the following CNF formula resulting from the conjunction of the following clauses
  - (pigeon clauses) for each $i \in [cn]$ we have in PHP_n^{cn} the clause
    $$P_i = x_{i1} \vee x_{i2} \vee \cdots \vee x_{in}$$
  - (hole clauses) for each $i, i' \in [cn]$ $i \neq i$ and for each $j \in [n]$ we have the clause: $H_j^{i,i'} = \overline{x}_{ij} \vee \overline{x}_{i'j}$.

- The lines in $Res(k)$ are k-DNF formulas and as set of inference rules we use the following:

$$\frac{A}{A \vee \ell} \qquad \frac{A \vee \ell_1 \quad \cdots \quad A \vee \ell_k}{A \vee (\ell_1 \wedge \cdots \wedge \ell_k)} \qquad \frac{A \vee (\ell_1 \wedge \cdots \wedge \ell_j)}{A \vee \ell_i} \; \ell_i \in \{\ell_1, \ldots, \ell_j\},\ j \leq k$$

$$\frac{A \vee (\ell_1 \wedge \cdots \wedge \ell_j) \quad B \vee \overline{\ell}_1 \vee \cdots \vee \overline{\ell}_j}{A \vee B}$$

In all such rules A, B are k-DNFs and $\ell_1, \ldots, \ell_j$ are literals $(\overline{\ell}_i = \neg \ell_i)$.

Sanity check: control that $Res(k)$ with those rules is p-equivalent to any other definition you saw before.

Regarding the Thm we are going to prove:

- the proof we present follows closely the one from [Segerlind-Buss-Impagliazzo '04]
- the bound we show is not optimal in κ: [Razborov '15] pushed the technique to work till $\kappa < \frac{\log n}{\log\log n}$ getting a l.b. for the size of PHP_n^{cn} of the form $2^{n/(\log n)^{O(\kappa)}}$.
- PHP_n^{cn} for $c \geq 2$ has quasi-polynomial $Res(\text{poly}\log n)$ refutations, i.e. refutations of size $\leq 2^{(\log n)^{O(1)}}$.
- The result we show is actually stronger in the sense that it holds for a variation of the PHP_n^{cn} formula called the "graph PHP", $PHP(G)$.

Def: ($PHP(G)$) Let G be a bipartite graph with bipartition $P = [cn]$ and $H = [n]$ and let E be the set of edges of G. $PHP(G)$ is the CNF formula obtained restricting the PHP_n^{cn} formula with the partial restriction α setting $x_{ij} = 0$ iff $\{i,j\} \notin E$.

obs: Every size l.b. we are able to prove for $PHP(G)$ will also hold for the corresponding PHP_n^{cn}. (why?) (The size l.b. are intended in "good" proof systems such as Res, Res(k) ...)

Theorem 2: If $G = (P \dot\cup H, E)$ is an (r,δ)-boundary expander then every resolution refutation of $PHP(G)$ will contain (at least) a clause of width $\frac{cr}{2}$. We denote this fact with the notation $\text{width}(PHP(G) \vdash) \geq \frac{cr}{2}$.

- A graph $G = (U \dot\cup V, E)$ is an (r,δ)-boundary expander iff for each $A \subseteq U$ $|A| \leq r \longrightarrow |\partial A| \geq \delta |A|$, where ∂A denotes here the "unique neighbors" of A: $\partial A = \{v \in V : \exists! u \in A \ \{u,v\} \in E\}$

proof of Thm 2: Let $\pi = (C_1, \ldots, C_\ell)$ be a resolution proof of PHP(G).

for any clause consider the measure μ defined as follows:

$$\mu(C) = \min\left\{ |I| : I \subseteq P \ \&\ \bigwedge_{i \in I} P_i \wedge \bigwedge_{\substack{i,i' \in P \\ i \neq i' \\ j \in H}} H_j^{i,i'} \models C \right\}$$

1. if C is in PHP(G) then $\mu(C) \leq 1$
2. $\mu(\perp) \geq r$ because G is an (r,c)-boundary expander and hence any subset of P of size $\leq r$ admit a matching in G (why?)
3. μ is sub additive, i.e. $\mu(C \vee D) \leq \mu(C \vee x) + \mu(D \vee \bar{x})$ for every C, D clauses and x variable

(1)+(2)+(3) $\Rightarrow \exists C \in \pi$ s.t. $\frac{r}{2} \leq \mu(C) \leq r$.

Fix such C and let I_C be the subset of P realizing $\mu(C)$.

claim: for each $v \in \partial I_C$ there exist some $u \in P$ s.t. $x_{uv} \in C$.

(then claim $\Rightarrow |C| \geq |\partial I_C| \geq c|I_C| \geq \frac{cr}{2}$ - since $|I_C| \leq r$)

proof of claim: suppose none of the variables x_{uv} appears in C. let $\tilde{u}$ be the unique neighbor of v in I_C

by minimality of I_C,

$$\bigwedge_{i \in I_C \setminus \{\tilde{u}\}} P_i \wedge \bigwedge H_j^{i,i'} \not\models C$$

take an assignment α that satisfies $\bigwedge_{i \in I_C \setminus \{\tilde{u}\}} P_i \wedge \bigwedge H_j^{i,i'}$ but falsifies C

extend it to α' in the following way:

$$\alpha' = \alpha \cup \{x_{\tilde{u}v} = 1\} \cup \{x_{uv} = 0\}_{u \neq \tilde{u}}$$

- $\alpha'(C) = \alpha(C) = \text{false}$ but α' satisfies $\bigwedge_{i \in I_C} P_i \wedge \bigwedge H_j^{i,i'}$.

Contradicting the choice of I_C

▨

The argument to prove Thm. 1 then goes as follows: suppose Π is a Res(k) refutation of PHP(G) of "small" size, then we want to find some partial assignment ρ s.t.

(i) every k-DNF ψ in Π is s.t. $h(\psi|_\rho) \leq \frac{W}{k}$ for some parameter W

- we will be able to do this using a Switching Lemma.
- $\Rightarrow$ $PHP(G)|_\rho$ will have a resolution proof of width $\leq \frac{W}{k} \cdot k$.

(ii) $PHP(G)|_\rho$ require resolution proofs of width $> W$.

Contradiction & this will imply that there can't be a Res(k) of "small" size. ☺

~ Random Restrictions ~

Let $G = (P \cup H, E)$ be a bipartite graph and let $p \in [0,1]$.

The distribution over partial assignment $\mathcal{M}_p(G)$ is defined as follows:

- for each $v \in H$ choose with probability $1-p$ that some pigeon must fly to it; and with prob. p leave it un-matched.
- If $v \in H$ is matched, uniformly select $u \in N(v)$, set $x_{uv} = 1$ and $x_{u'v} = 0$ for all the other $u' \neq u$ in $N(v)$.

Lemma 1: Let $p \in [0,1]$. $G = (P \cup H, E)$ bipartite and $d = \deg G = \max_{v \in P \cup H} \deg(v)$.

Let ψ be a k-DNF in pigeon-normal-form.

$$\Pr_{\alpha \in \mathcal{M}_p(G)} [\psi|_\alpha \neq 1] \leq e^{-\frac{(1-p)^k \operatorname{cov}(\psi)}{k\, d^{k+1}}},$$

where $\operatorname{cov}(\psi)$ = min size of a set S of vars s.t. each term in ψ contain a var. from S

Wait but what is a k-DNF in pigeon-normal-form ??

A k-DNF ψ is in pigeon-normal-form if it does not contain a term immediately contradicting a hole axiom of PHP(G), that is no term in ψ may contain $x_{uv} \wedge x_{u'v}$ for $u \neq u' \in N(v)$.

OBS: If PHP(G) has a Res(k) proof of size S then there is another proof of PHP(G) in Res(k) with all lines in pigeon-normal-form and size $\leq 2S$.

proof: exercise.

The previous observation tells us that restricting to proofs in pigeon-normal-form is not a big assumption and the previous lemma is of course in the direction of using a Switching Lemma: we will use the following particular version of the Switching Lemma proved last time.

Switching Lemma: Let K, s be positive numbers and consider the distribution $\mathcal{M}_p(G)$ defined above. Then if $\Pr_{\alpha \in \mathcal{M}_p(G)}[\psi\restriction_\alpha \neq 1] \leq 2^{-\delta \cdot \mathrm{cov}(\psi)}$ (for every K-DNF ψ in pigeon-normal-form) then

for every K-DNF ψ in pigeon-normal-form $\quad \Pr_{\alpha \in \mathcal{M}_p(G)}[h(\psi\restriction_\alpha) > 2s] \leq K 2^{-\delta' \cdot s}$ where $\delta' = 2\left(\frac{\delta}{4}\right)^K$

Proof of Lemma 1: The proof is just a long chain of inequalities:

$$\Pr_{\alpha \in \mathcal{M}_p(G)}[\psi\restriction_\alpha \neq 1] = \Pr_{\alpha \in \mathcal{M}_p(G)}\left[\bigwedge_{t \in \psi}(t\restriction_\alpha \neq 1)\right] \leq \Pr_{\alpha \in \mathcal{M}_p(G)}\left[\bigwedge_{t \in \psi'}(t\restriction_\alpha \neq 1)\right], \quad (*)$$

where $\psi' \subseteq \psi$ consists only of hole-disjoint terms, i.e. such that

$$\{v \in H \mid x_{uv} \in t \text{ or } \overline{x}_{uv} \in t\} \cap \{v \in H \mid x_{uv} \in t' \text{ or } \overline{x}_{uv} \in t'\} = \emptyset.$$

But then

$$(*) \leq \prod_{t \in \psi'}\left(1 - \Pr_\alpha[t\restriction_\alpha = 1]\right) \leq \prod_{t \in \psi'}\left(1 - \left(\frac{1-p}{d}\right)^K\right) \quad (\dagger)$$

(the events of satisf. hole-disj. terms are indep.)

$\Pr_\alpha[t\restriction_\alpha = 1] \geq (1-p)^K \cdot \frac{1}{d^K}$ since all holes in t must be matched and they must be matched in the right way.

N.B.: Here we used the hyp. that ψ is in pigeon-normal-form

$$(\dagger) \leq \left(1 - \left(\frac{1-p}{d}\right)^K\right)^{\frac{\mathrm{cov}(\psi)}{Kd}}$$

in ψ there are always $\geq \frac{\mathrm{cov}(\psi)}{K}$ variable-disj. terms and each hole appears in $\leq d$ variables so there must be $\geq \frac{\mathrm{cov}(\psi)}{Kd}$ hole-disj. terms in ψ.

□

To prove Thm 1 we need some G and some p s.t.

$\text{width}(\text{PHP}(G)\restriction_\alpha \vdash) \geq W$ for some large W $(= \Omega(n))$ and $\alpha \in \mathcal{M}_p(G)$.

We have the following result.

Theorem 3: For all $c > 1$ there exist $c' > 0$ such that for n sufficiently large there exists a graph $G = (P \cup H, E)$ with $P = [cn]$, $H = [n]$, $\deg(G) \leq c' \log n$ and such that

$$\Pr_{\alpha \in \mathcal{M}_{\frac{3}{4}}(G)} \left[\text{width}(\text{PHP}(G)\restriction_\alpha \vdash) \geq \frac{n}{24} \right] \geq \frac{1}{2}.$$

We will not prove this result but we mention that the graph G whose existance is provided by the theorem above result from the following construction:

let $\mathcal{G}_{m,n,q}$ be the distribution over bipartite graphs with vertex sets $[m]$ and $[n]$ in which every edge is included with indip. probability q.

the graph in theorem 3 is picked according to the previous construction with $q = \frac{48c \ln cn}{cn}$. "Intuitively" such random graph will have good expansion properties and hence will require large width in Res to refute PHP(G).

So, let's put all pieces together and (finally) prove Theorem 1

Proof (of Thm. 1):

Let Π be a Res(k) refutation of PHP(G) where G is the graph coming from Thm 3, so in particular $d = \deg(G) \le c' \log w$ and

$$\Pr_{\alpha \in M_{3/4}(G)} \left[\text{width}(\text{PHP}(G)\restriction_\alpha \vdash) \ge \frac{w}{24} \right] \ge \frac{1}{2} \quad - \qquad (**)$$

W.l.o.g. we can assume Π is pigeon-normal-form (why?)

On the other hand we can use the switching lemma with parameters: $p = \frac{3}{4}$ $\quad \delta = \frac{\text{cov}(\Psi)}{k\,(d^{k+1}} \cdot \left(1 - \frac{3}{4}\right)^k = \frac{\text{cov}(\Psi)}{k \cdot 4^k \cdot d^{k+1}}$ -

It is remained to choose s: with a momentary act of faith we set it to

$$s = \frac{1}{2k}\left(\frac{w}{24} - 1\right) -$$

So we have that

$$\Pr_{\alpha \in M_{3/4}(G)} \left[\exists \psi \in \Pi \;\; h(\psi\restriction_\alpha) > 2s \right] \le |\Pi| \cdot k \cdot 2^{-\delta' s}, \qquad (*)$$

where $\delta' = 2\left(\frac{\delta}{4}\right)^k$ - Suppose by contradiction that $|\Pi| < \frac{1}{2k} \cdot 2^{\delta' s}$,

then the probability in (*) is $< \frac{1}{2}$, so the complement of the event estimated in (*) and the one in (**) cannot be disjoint, i.e. there exists $\alpha \in M_{3/4}(G)$ s.t.

- $\text{width}(\text{PHP}(G)\restriction_\alpha \vdash) \ge \frac{w}{24}$ and
- each $\psi \in \Pi\restriction_\alpha$ has height $\le 2s = \frac{1}{k}\left(\frac{w}{24} - 1\right)$ which implies by a thm saw in the last lecture that $\text{width}(\text{PHP}(G)\restriction_\alpha \vdash) \le k \cdot \frac{1}{k}\left(\frac{w}{24} - 1\right)$

So it must be that $|\Pi| \ge \frac{1}{2k} 2^{\delta' s}$ - How good is this bound? $\Leftarrow$

$$|\Pi| \ge 2^{\frac{n}{d^{O(k^2)}}} \overset{\text{here we use the bound on } d \le c' \log n}{=} 2^{\frac{n}{(\log n)^{O(k^2)}}}$$

- So as long as $k < \sqrt{\frac{\log n}{\log \log n}}$ this bound is of the form 2^{n^ε} for some $\varepsilon > 0$ -

□