Function \( f \) mapping CSPs to CSPs is a complete linear-blowup (CL-) reduction if:

(i) \( f \) is polynomial-time computable
(ii) \( \text{val}(\varphi) = 1 \implies \text{val}(f(\varphi)) = 1 \)
(iii) If \( \varphi \) has \( m \) constraints, then \( f(\varphi) \) has at most \( Cm \) constraints and alphabet size \( W \). \( C \) and \( W \) do not depend on \# constraints or \# variables in \( \varphi \).

PCP theorem follows from 2 lemmas:

**Lemma 22.5 (Gap Amplification)**

\[ \exists \epsilon, \varphi \in \mathbb{N}^+, \exists W \in \mathbb{N}^+, \epsilon_0 \in (0,1), \text{CL-reduction } g \text{ such that} \]

(i) \( g_{\epsilon, \varphi} \) maps \( \text{CSP}_2 \) instances to \( \text{CSP}_W \) instances
(ii) \( \text{val}(\varphi) \leq 1 - \epsilon \) for \( \epsilon < \epsilon_0 \implies \text{val}(g_{\epsilon, \varphi}(\varphi)) \leq 1 - \epsilon^2 \)

That is, gap between satisfiable and unsatisfiable instances blows up by factor \( C \). But alphabet size also increases (a lot).

**Lemma 22.6 (Alphabet Reduction)**

\[ \exists \, g_0 \in \mathbb{N}^+ \text{ and CL-reduction } h \text{ such that} \]

(i) \( h \) maps \( \text{CSP}_W \) instances to \( g_0 \text{CSP}_2 \) instances
(ii) \( \text{val}(\varphi) \leq 1 - \epsilon \implies \text{val}(h(\varphi)) \leq 1 - \epsilon/3 \)
A d-regular graph \( G = (V, E) \) over \( n \) vertices (possibly with multi-edges and self-loops) is an \((n, d, \lambda)\)-spectral expander if the second largest eigenvalue in absolute value of the normalized adjacency matrix \( A_G \) of \( G \) is at most \( \lambda \).

\[ \lambda < 1 \] if \( G \) connected and non-bipartite
\[ \lambda \geq 2/\sqrt{d} \] always.

The smaller the \( \lambda \), the better the connectivity properties of the graph.

**CONSTRANT** GRPH4 \( G(\varphi) \) for CSP\( \omega \) instance \( \varphi \)

\( V(\varphi) = [n] \leftrightarrow \) identity with variables \( u_1, \ldots, u_n \)

Add edge \( (i, j) \) for every constraint over variables \( u_i, u_j \)

say that a qCSP\( \omega \) instance is "nice" if

1. \( q = 2 \)
2. \( G(\varphi) \) is d-regular for some (even) constant \( d \); at every vertex half of edges are self-loops
3. \( G(\varphi) \) is an \((n, d, \lambda)\)-expander

for, say, \( \lambda = 0.9 \)

In Lemma 22.5, can preprocess qCSP\( _2 \) instance to CSP\( \omega \) instance \( \varphi' \) such that

\( G(\varphi') \) is nice. So Lemma 22.5 follows

if we can prove the following:

We skipped the details. Not too hard, but not enlightening.
**Lemma 22.9 (Powering)**

There is an algorithm $\mathcal{A}$ that given

- a nice 2CSP instance $\Psi$
- a parameter $t < W^4$

produces 2CSP instance $\overline{A}(\Psi, t) = \Psi^t$ such that:

1. $W^4 < W^5$ 
2. $d$ degree of $\overline{G}(\Psi)$
   $\Psi^t$ has $\frac{n}{2}$ constraints

1. $\mathrm{val}(\Psi) = 1 \Rightarrow \mathrm{val}(\Psi^t) = 1$

2. If $\mathrm{val}(\Psi) \leq 1 - \varepsilon$, $\varepsilon < \frac{1}{dW^4}$, then
   $\mathrm{val}(\Psi^t) < 1 - \varepsilon'$, $\varepsilon' = \frac{\sqrt{1 - \varepsilon}}{105dW^4}$

4. $\mathcal{A}$ runs in time polynomial in $n$, $m$ and $Wd^t$.

Recall that unless stated otherwise:

- $n$ = number of variables in CSP instance
- $m$ = number of constraints in CSP instance

**Construction**

"New variable" $y_i$ encodes value of $u_i$ plus all $u_j$ at distance $\leq t + \sqrt{t}$ from $u_i$ in $G(\Psi)$

$y_i$ "claims" all these variables "has an opinion" about their value.
Graph terminology

Walk \textit{sequence of vertices} \((v_1, v_2, v_3, \ldots)\) such that \(v_i \text{ (} v_{i-1}, v_{i+1} \text{) edge} (\text{sometimes called a path})\)

Path \textit{walk without repeating vertices} (sometimes called a simple path)

Another seems to use walk and path as synonyms.

One constant \(C_p\) for every walk \(w\) with \(2t+1\) steps in \(G^*(\Psi) \setminus \{i_1, i_2, \ldots, i_{2t+2}\}\)

\[i = i_1\]
\[j = i_{2t+2}\]

\[\Psi_s (u_{ik}, u_{ik+1})\]

\(C_p\) checks all constraints in \(\Psi\) such that
1) \(y_i\) has an opinion about \(u_{ik}\), say \(v_i\) (i.e., \(u_{ik}\) within distance \(t + \frac{1}{T_1}\) from \(y_i\))
2) \(y_j\) has opinion about \(u_{ik+1}\), say \(v_j\) (i.e., \(u_{ik+1}\) within distance \(t + \frac{1}{T_2}\) from \(y_j\))

\(C_p\) accepts iff \(\Psi_s (v_i, v_j) \text{ satisfied for all such } \Psi_s\).
We skip proof of properties 21 & 24.
22 (completeness) is straightforward.
If \( y \) is satisfiable, then for some satisfying assignment to \( U_i \)-variables
let every \( y_i \) encode this assignment to variables in its neighborhood.

The hard part is 24 - the soundness/gap.

Proof idea: Take assignment to \( y_i \)'s of \( y \)
violating not too many constraints. "Decode"
to assignment of \( U_i \)'s of \( y \) violating very few constraints.

Or in the other direction: We know any assignment
to \( y \) violates \( \Omega \)-fraction of constraints
Every constraint \( G_o \) in \( y \) has two nodes \( y, y' \)
each checks \( \Omega(\sqrt{r}) \) constraints \( y, y' \) about which
both \( y_i \) and \( y' \) have opinions. In the best of
words, should have probability \( \Omega(\sqrt{r} \epsilon) \) of picking
up at least one violated constraint. But \( y \), of course
need not assign values to \( U_i \)'s consistently.

**Plurality Assignment**

Fix an arbitrary assignment \( y = y_1, \ldots, y_n \)
For every \( U_i \), define random variable \( L_i \) as follows:

1. Start at vertex \( i \).
2. Take \( \epsilon \)-step random walk in \( G(y) \), ending at, say \( j \).
3. Output the value that \( y_j \) claims for \( U_i \).
For every \( i \), let \( z_i = \arg\max_{z \in [0,1]} \Pr\left[ Z_i = z \right] \) (splitting ties arbitrarily) \( \Pr\left[ Z_i = z_i \right] \geq \frac{1}{w} \)

Planarity assignment to \( u = (u_1, \ldots, u_n) \)

\[ i's = (z_1, \ldots, z_n) \]

Any assignment to \( \psi \) violates \( E_{\psi} = E_{\text{viol}} \geq z \) constraints.

Exist a set \( F \) of at least \( E_n \) violated constraints violated by \( z = (z_1, \ldots, z_n) \) edges in \( G(\psi) \)

Use this set to show at least \( \epsilon' \) fraction of constraints \( C_p \) of \( \psi^c \) are violated by assignment \( y = (y_1, \ldots, y_n) \), \( \epsilon' = \Omega(\sqrt{\epsilon}) \)

Pick a random walk \( (i_2, \ldots, i_{2t+1}) \) in \( G \) of length \( 2t+1 \), corresponding to a constraint \( C_p \) (i.e., uniformly randomly chosen)

For \( j \in [2t+1] \), say that \( j \)-th edge in \( P(i_j, i_{j+1}) \) is \( \text{TRUE/FALSE} \) if

- \( y_{i_j} \) claims \( u_{i_j} = z_{i_j} \)
- \( y_{i_{2t+2}} \) claims \( u_{i_{2t+2}} = z_{i_{2t+2}} \)

i.e., new variables for end points of \( P \) claim that \( u_{i_j} \) and \( u_{i_{j+1}} \) agree with the planarity values.
OBSERVATION (SIMPLE BUT CRUCIAL)
If the walk $p$ has a truthful edge that lies in $F$, then $C_p$ is violated.

Proof
By parsing the definitions.

So, we want to prove that each fraction of constraints $C_p$'s walks $p$ contains a truthful edge in $F$.

Game plan (which we won't have time to carry out):

1. Fix any position $j$ in walk. Then $j$th edge of random walk $p$ is a uniformly random edge (not hard).

2. Midpoint edge of walk (distance $t$ from both endpooms $(i_{t-1,1}, i_{t+1,2})$ has $y_{i_{t-1,2}}$ taking part in plurality vote for $U_{i_{t+1,2}}$.

$y_{i_{t-1,2}} - 11 - U_{i_{t+1,2}}$

$\Rightarrow$ Edge is truthfull with probability $\geq \frac{1}{W} \cdot \frac{1}{W}$. (pretty much by definition)

3. For edges in interval $[t - \sqrt{t}, t + \sqrt{t}]$ both $y_{i_1}$ and $y_{i_{t+1}}$ have opinions about endpoints. "middle edges" if we are sufficiently close to midpoint — at most $\delta \sqrt{t}$ away for small $\delta$ — then
options likely to agree with opinions about midpoints of walls \(\Rightarrow\)
edges in interval \([t - \delta v, t + \delta v]\) are truthful with probability
\[\Omega\left(\frac{1}{W^2}\right)\]

(4) By (1) and (3) for any edge \(e\) and any fixed \(j \in [t - \delta v, t + \delta v]\) it holds that for a random walk \(p\) that
\[\Pr [e \text{ truthful} \mid e_{j \text{th edge}}] = \Omega\left(\frac{1}{W^2}\right)\]
\[\Pr [e \text{ is } j \text{th edge}] = \frac{1}{|E|}\]

Let \(V = V(p) = \# \{\text{middle edges both truthful and in } F\}\)
By linearity of expectation we have
\[E[V] = \sum_{j = t - \delta v}^{t + \delta v} \Pr [j \text{th edge truthful and in } E]\]
\[= \sum_j \sum_{e \in E} \Pr [e \text{ is } j \text{th edge and is truthful}]\]
\[= 2\delta v + 1 |F| \cdot \frac{1}{|E|} \cdot \Omega\left(\frac{1}{W^2}\right)\]
\[= \Omega\left(\frac{\sqrt{v} \cdot |e|}{W^2}\right)\]

This looks quite similar to the kind of expression we are shooting for! Except...
Except we need to bound
Pr[\text{Cp violated}] \geq Pr[\text{V > 0}]
not \ \mathbb{E}[V] !

In general, \mathbb{E}[V] says nothing about Pr[V > 0]. Unless we can prove that V is likely to be concentrated around its expectation.

Follows if variance \text{Var}(V) small
Follows if second moment \mathbb{E}[V^2] not too large

How could \mathbb{E}[V] be far off from Pr[V > 0]? Most of the time no truthful middle edges in F, but when there are, then there are tons of them! Want to rule this out.

Look at V' = \# middle edges in F
Clearly V' \geq V
So sufficient to upper-bound
\mathbb{E}[(V')^2]

This is where we are going to use that G(p) is an expander graph.
Recall from last lecture the claim that for an \((n, d, \lambda)\)-spectral expander \(G = (V, E)\) and any \(S \subseteq V\), \(|S| \leq n/2\):

\[
\Pr_{(u,v) \in E} \left[ u \in S \land v \in S \right] \leq \frac{|S|}{n} \cdot \frac{1 + \lambda}{2}
\]

If we instead sample length-\(r\) walks with endpoints \(u, v\), we get corollary:

\[
\Pr_{\text{length-r walk}} \left[ u \in S \land v \in S \right] \leq \frac{|S|}{n} \cdot \frac{1 + \lambda^r}{2}
\]

Because \(G^r\) is an \((n, d^r, \lambda^r)\)-spectral expander. And by assumption, \(F\) is a very small set of edges, covering very few vertices.

Since \(G(F)\) is an expander, it is very unlikely that a random walk will cluster inside vertex set touched by \(F\).

Doing the calculations, we get:

\[
E[V^2] \leq E[(V)^2] = O(\epsilon d V + d) (**)
\]

For a non-negative random variable \(V\) it holds that:

\[
\Pr[V > 0] \geq \frac{E[V]^2}{E[V^2]} \quad (***)
\]

[Useful exercise to prove this]
Plugging (*) and (**) into (***) we get

$$\Pr[Q \text{ violated}] \Rightarrow \Pr[ V > 0 ] \Rightarrow \Omega \left( \frac{\sqrt{\varepsilon}}{cl^{1/4}} \right)$$
as required.

We don't have time to do the details, but given the intuition above you could hopefully follow the proofs on pages 468 - 470 in Azar-Barak (and forgive the details if and when needed).

So, modulo all the details skipped, this establishes Lemma 22.9, which modulo even more details skipped, establishes the gap amplification Lemma 22.5.

[But to be fair, we did cover all the important and interesting ingredients of the proof.]
Lemma 22.6 (Alphabet Reduction)

For $g_0 \in \mathbb{N}^+$ and $\epsilon$-reduction $h$ such that

(i) $h$ maps $2\text{CSP}_n$ instances to $g_0 \cdot \text{CSP}_2$ instances

(ii) $\text{val}(\phi) \leq 1 - \epsilon \Rightarrow \text{val}(h(\phi)) \leq 1 - \epsilon/3$

We will use the PCP verifier in our proof of $\text{QUADFA} \in \text{PCP}_{1,\frac{1}{2}}(\text{poly}(n), 1)$

Proposition (poly-time)

There is a reduction from CIRCUIT-SAT to QUADFA such that if $C$ is a circuit of size $k$ with $k$ inputs, then $f(C)$ is a quadratic equation with $O(k^2)$ equations over $k$ variables. Furthermore, in a witness for $f(C)$ the first $k$ variables encode a satisfying input for $C$. 

Proof: Exercise — not hard.

We also need to remember correspondence:
Start with \( CSP \) instance

Variables \( u_1, \ldots, u_n \) taking values in \([W]\)

Constraints \( C_1, \ldots, C_m \)

Encode variables as binary strings in \( \{0,1\}^k \) for \( k = \log W \)

Then \( C_i \) is function \( C_i : \{0,1\}^{2k} \rightarrow \{0,1\} \)

Can be encoded as circuit of size \( \leq 2^{2k} = W^2 \)

Each such circuit has \( PCP \) verifier reading \( O(1) \) bits (universal constant) and using \( \text{poly}(W) \) randomness.

Proof is Walsh-Hadamard encoding of string of length \( \leq W^2 \), first \( k \) bits of which is satisfying assignment.

Idea:

Require proofs for all circuits/variables \( C_1, \ldots, C_m \)

Write down separately Walsh-Hadamard encodings of \( u_1, \ldots, u_n \in \{0,1\}^{\log W} \)

1) Pick \( 5 \) \( (u_{i_1}, u_{i_2}) \) randomly
2) check proof for circuit \( C_5 \) using \( PCP \) verifier from a couple of lectures ago
3) Run concentration test to check
\[ T_s = WH(U) \text{ for some } U \text{ such that} \]
\[ U = U_{i_{1,1}} \circ U_{i_{1,2}} \circ Z \quad (\dagger) \]

1) \((\dagger)\) is exactly as in our proof of
\[ \text{QuadEq} \in \text{PCP}_{1/2} (\text{poly}(n), 1) \]

For 3), we have (thanks to previous lemma)
\[ f = WH(U_{i_{1,1}}) \]
\[ g = WH(U_{i_{1,2}}) \]
\[ h = WH(U) \]

Pick random \(x, y \in \{0, 1\}^k\)

check if
\[ h(x \oplus y \oplus 0^2) = f(x) + g(y) \]

This detects violation of \((\dagger)\) with
probability \(1/2\) by the random subset principle.

If we fix circuit \(C_5\) and look at proofs \(f, g, h\) above, the argument yields the following theorem (or, rather, corollary of \(NP \leq \text{PCP}_{1/2} (\text{poly}(n), 1)\) theorem)
COROLLARY 22.13 (PCP of proximity)

There is a verifier $V$ that given any circuit $C_S$ of size $L$ with $2k$ inputs has the following properties:

1) If $u_1, u_2 \in \{0,1\}^k$ are such that $u_1 \circ u_2$ is a satisfying assignment for $C_S$, then $\exists \Pi_3$ of size $2 \text{poly}(L)$ such that $V$ accepts $WH(u_1) \circ WH(u_2) \circ \Pi_3$ with probability $1$.

2) For every triple of strings $\Pi_1, \Pi_2, \Pi_3 \in \{0,1\}^k$, if $V$ accepts $\Pi_1 \circ \Pi_2 \circ \Pi_3$ with probability $\frac{1}{2}$, then $\Pi_1, \Pi_2$ are $0.01$-close to $WH(u_1), WH(u_2)$ for $u_1, u_2 \in \{0,1\}^k$ s.t. $u_1 \circ u_2$ satisfies $C_S$.

3) $V$ runs in $\text{poly}(L)$ time, uses $\text{poly}(L)$ random bits, and queries only $O(1)$ bits.

Consider now a verifier that asks for $WH(u_1), \; i \in [n]$, as well as $\Pi_3$ for all $C_S$.

It then picks a random $S$ and runs verifier above and decides accordingly.
The tests for a given $C_s$ is a $q\text{CSP}_2$ instance $C_s$. The full verifier is a $q\text{CSP}_2$ instance $\forall = \bigvee_{s=1}^{n} C_s$

We claim that if $\bigvee_{s=1}^{n} C_s$ has assignment satisfying $(1-\delta)$-fraction of constraints then we can construct assignment to original $2\text{CSP}$ instance $\forall$ satisfying $(1-2\delta)$ constraints.

Here we get factor 2 while time bound get 3 for some reason.

Suppose we have assignment to $\forall$ falsifying only $\delta$ constraints.

Clearly, if at most $2\delta s' \in [m]$ s.t. 50% of constraints in $C_s$ falsified.

Look at all other $C_s$. Since acceptance probability $\geq 50\%$, we have Walsh-Hadamard encodings of assignments $a_{is,1}, a_{is,2} \in \{0,1\}^{\log W}$ such that $C_s(a_{is,1}, a_{is,2}) = 1$ (unique).

For all $c_i$ that do not get values from such $C_s$, let $c_i = \text{arbitrary value in } [W]$.
Than the assignment \((a_1, \ldots, a_n) \in [W]^n\) satisfies all but a fraction \(2\delta\) of the constraints in the original \(2CSPW\) instance \(\Phi\).

The Algebraic Reduction Lemma 22.6 follows.

[And for this lemma we didn't really cheat in the proof]

AND NOW THE COURSE IS OVER - WRAP-UP

- Basic computational complexity theory
- Selection of "advanced" topics:
  - distributed algorithms
  - proof complexity
  - PCPs and hardness of approximation
- Hope that course gave you
  - general overview of complexity theory (though by necessity biased)
  - opportunity to practise exposition skills
  - mental gymnastics!
- We have (what we think are) interesting topics for MSc theses
- We are always interested in sharp PhD students (though competition is stiff)

THANKS FOR TAKING THE COURSE — HOPE YOU ENJOYED IT