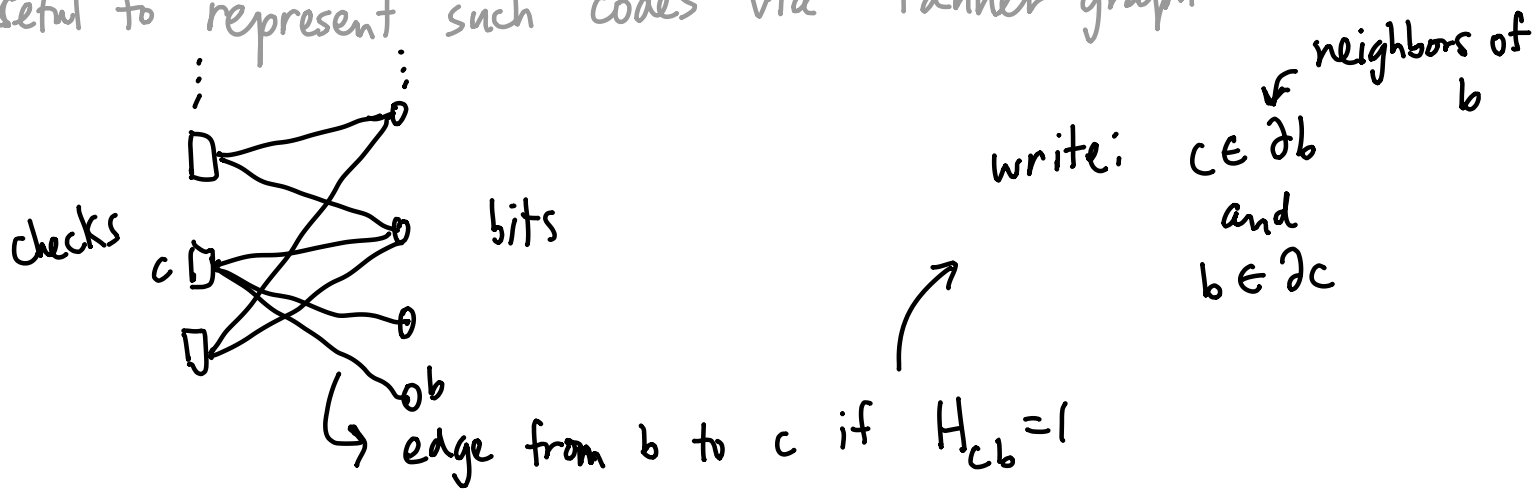


So far we studied codes that were either "small" n or local in space. What if we relaxed both of those restrictions? We'll start by thinking about classical codes

Classical (Δ_B, Δ_C) LDPC code:

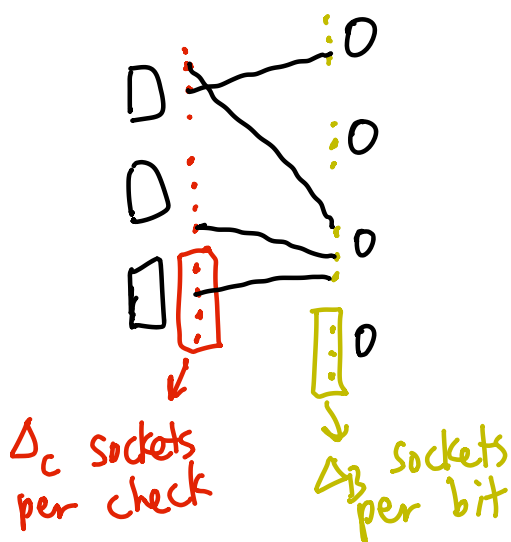
H is sparse: each row (check) has Δ_C bits
each column (bit) in Δ_B checks

Useful to represent such codes via Tanner graph



Idea: let's try to make a random code:

Def: Random (Δ_B, Δ_C) -LDPC code chosen by "configuration model":



① Pick integer n_0 divisible by Δ_B & Δ_C .

Set $n = \frac{n_0}{\Delta_B}$ and $n_c = \frac{n_0}{\Delta_C}$
of bits # of checks

② Pair up bit/check sockets via a random permutation

③ Propose Tanner graph by connecting bit to check when sockets connected.
Reject if any bit-check pair connected more than once.

Thm: The probability of success stays $O(1)$ as $n \rightarrow \infty$.

Proof sketch: let's ask how likely we are to connect a given bit to one check twice:

$$P[\text{bit } b \text{ connects to same check twice}] \approx \binom{\Delta_B}{2} \frac{\Delta_C - 1}{n_0 - 1}$$

which pair of sockets
wire to same check

prob second b socket
connects to same check:
 $\frac{\# \text{ of remaining } c \text{ socket}}{\# \text{ of possible choices}}$

$$\rightarrow \approx \frac{\Delta_B(\Delta_B - 1)(\Delta_C - 1)}{2n\Delta_B}$$

Now as $n \rightarrow \infty$ the chances of this happening are roughly independent for each b , so

$$P[\text{no double connection}] \approx \left[1 - \frac{(\Delta_B - 1)(\Delta_C - 1)}{2n} \right]^n \approx e^{-(\Delta_B - 1)(\Delta_C - 1)/2} = O(1)$$

Prop: Unlikely events in random configurations unlikely in random LDPC:

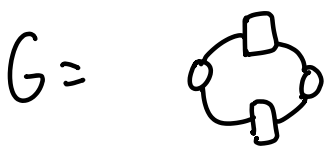
$$P[X \text{ happens in LDPC}] \leq e^{-(\Delta_B - 1)(\Delta_C - 1)/2} P[X \text{ happens in config}]$$

Proof: $P[X \text{ in config}] = P[X \text{ in LDPC}] P[\text{LDPC} | \text{config}] + P[X \text{ in not LDPC}] P[\text{not LDPC} | \text{config}] \geq P[X \text{ in LDPC}] \cdot e^{-(\Delta_B - 1)(\Delta_C - 1)/2}$

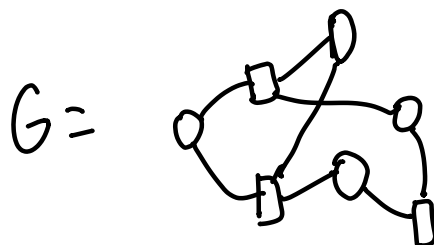
What would the Tanner graph of such a random code look like?

Thm: near a typical bit, Tanner graph of random code looks locally tree like:

Proof sketch: What's probability a subgraph of Tanner graph isomorphic to G ?



or



etc.

take any G w/ m_B bits, m_C checks, m_E edges.

$$P[\text{subgraph iso. to } G] \leq [\text{\# of choices of bit/check}] P[\text{all required edges exist}]$$

$$\leq n^{m_B} n^{m_C} = \left(\frac{\Delta_B}{\Delta_C}\right)^{m_C} n^{m_B+m_C} \leq \left(\frac{\Delta_B \Delta_C}{n_0}\right)^{m_E}$$

of candidate bits # of candidate checks $\approx$ prob. any specific bit/check connected

NB: exact combinatorics requires care (e.g. what are symmetries of G ?)

$$P[\text{subgraph iso. to } G] \lesssim n^{m_B+m_C-m_E}$$

Lemma: (Euler's formula for graphs):
 $m_B + m_C - m_E = 1 - (\text{\# of independent cycles in } G)$

Lemma: Number of distinct G is at most $2^{m_B m_C}$.

Proof: For each bit/check pair, there's either an edge or not;
 independently counting up options (neglecting disconnected / equivalent G)
 gives the claimed bound

So: $P[\text{any } G \text{ of size } \leq l \text{ w/ } \geq 2 \text{ loops}] \leq 2^{l^2} n^{-1} \rightarrow 0$
 for fixed l .

By same argument as before:

$$P[\text{no } G \text{ w/ one loop } \leq l] \lesssim \left(1 - \frac{O(1)}{n}\right)^n \rightarrow e^{-O(1)}$$

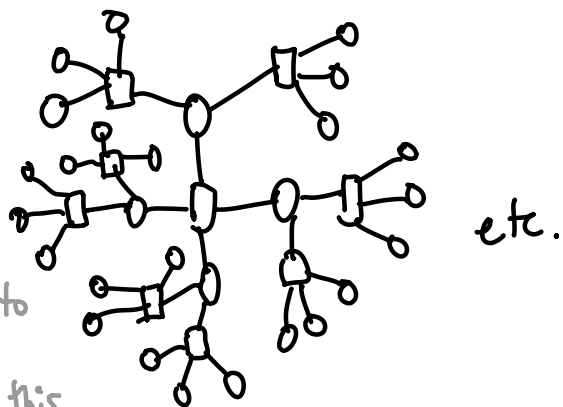
if l, l' are $O(1)$

$\hookrightarrow$ we expect typically $O(1)$ # of short loops.

$$\text{so } P[\text{a loop of length } \leq l \text{ w/in distance } l' \text{ of bit } b] \lesssim \frac{O(1)}{n}$$

by our earlier proposition

Hence we deduce that the Tanner graph typically looks like a tree:



note to self:
draw this
before class!!

Remains to determine code parameters k & d for these random codes...

Prop: Pick $\Delta_B < \Delta_C$. Then $k \geq n(1 - \frac{\Delta_B}{\Delta_C})$.

Proof: H has $n_c = \frac{\Delta_B}{\Delta_C}$ rows. If $\Delta_B < \Delta_C$,

$$n > n_c \geq \text{rank}(H) = n - k \quad [\text{equality if all checks independent}]$$

Thm: If $\Delta_B \geq 2$, w.h.p. $k = n(1 - \frac{\Delta_B}{\Delta_C}) + \begin{cases} 1 & \Delta_B \text{ even} \\ 0 & \Delta_B \text{ odd} \end{cases}$

because $\sum_c H_{cb} = 0$ for all b
 $\Rightarrow$ at least 1 redundant check

Proof uses similar tools to what we'll develop for the rest of the lecture. The next goal is to prove the following:

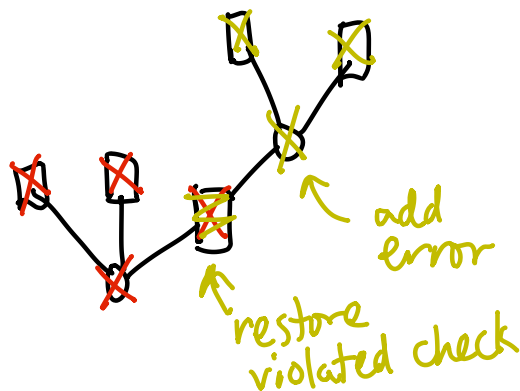
Thm: w.h.p. code distance $d = \Theta(n)$ [scales as n]

In order to prove this result we'll need to make following intuition precise:

Prop: on a perfect tree we have:

$$|H_e| \geq (\Delta_B - 2)|e|$$

Proof: every error violates
 $\geq (\Delta_B - 1)$ new checks,
removes ≤ 1 violated check.



This is such a useful property that we'll give it a name:

Def: Linear confinement if for every $|e| \leq L$,
 $|He| \geq \eta |e|$ for some $0 < \eta$

Lemma: If linear confinement for $|e| \leq L$, code distance $d > L$

This is not a sharp bound but it is satisfactory for our purposes...
 Proof: codeword $x \in \ker(H)$ has $|Hx| = 0 \Rightarrow |x| > L$ if $x \neq 0$.

Lemma: w.h.p. random code has η -linear confinement for $L \leq \alpha n$
 if $\alpha < \alpha_*$ and $0 < \eta < \Delta_B/2 - 1$,
 where $0 = (\Delta_B - 1)h(\alpha_*) - \Delta_C \alpha_* \left(\eta + \frac{\Delta_B}{2} \right) h\left(\frac{\Delta_B}{\Delta_C \left(\eta + \frac{\Delta_B}{2} \right)} \right) - \frac{\Delta_B}{\Delta_C} h\left(\frac{\alpha \Delta_C}{\Delta_B} \left(\frac{\Delta_B}{2} + \eta \right) \right)$,
 where $h(x) = -x \log x - (1-x) \log(1-x)$.
 ← take first positive zero of this equation to be α_* .

Proof: Let S be subset of bits; $|\partial S| = \#$ of checks connected to some bit in S .

convert to error

$$|He| \geq [\# \text{ of checks connected to 1 bit}] \geq |\partial S| - [\# \text{ connected to } > 1 \text{ bit}]$$

$$\geq |\partial S| - \frac{\Delta_B}{2} |S|, \quad \text{Goal: } |\partial S| \geq \left(\eta + \frac{\Delta_B}{2} \right) |S| =: \kappa |S|$$

Use exactly same configuration model counting as before:

$$\mathbb{P}[|\partial S| \leq \kappa |S|] \leq O(1) \cdot [\# \text{ of } S \text{ choices}] [\# \text{ of } C \text{ choices}] \mathbb{P}[\partial S \subseteq C]$$

for some $\hookrightarrow$ call S

$\binom{n}{S} \rightarrow e^{n h(\frac{S}{n})}$

$\binom{\frac{\Delta_B}{\Delta_C} n}{\kappa S} \rightarrow e^{\frac{\Delta_B n}{\Delta_C} h(\frac{\kappa \Delta_C S}{\Delta_B n})}$

$\leftarrow \kappa S$

$$\mathbb{P}[\partial S \subseteq C] = \frac{[\# \text{ of subsets of check sockets in } C \text{ of size } \Delta_B S]}{[\# \text{ of possible check sockets for edges from } S]}$$

$$= \begin{pmatrix} n\Delta_B \\ s\Delta_B \end{pmatrix}^{-1} \begin{pmatrix} \Delta_C \kappa s \\ s\Delta_B \end{pmatrix} \rightarrow e^{-n\Delta_B h\left(\frac{s}{n}\right) + \Delta_C \kappa s h\left(\frac{\Delta_B}{\Delta_C \kappa}\right)}$$

Now suppose $s := \alpha_0 n$; we get: (eliding a few steps...)

$$\mathbb{P}[|\mathcal{S}| \leq \kappa |\mathcal{S}| \text{ for some } |\mathcal{S}|] \leq O(1) \cdot e^{-n \cdot A} \quad \text{where}$$

$$A(\alpha_0) := (\Delta_B - 1)h(\alpha_0) - \Delta_C \kappa \alpha_0 h\left(\frac{\Delta_B}{\Delta_C \kappa}\right) - \frac{\Delta_B}{\Delta_C} h\left(\frac{\kappa \alpha_0 \Delta_C}{\Delta_B}\right)$$

Check that $A > 0$ for sufficiently small α_0 :

$$A \rightarrow \alpha_0 \log \frac{1}{\alpha_0} \cdot [\Delta_B - 1 - \kappa] + O(\alpha_0) \Rightarrow \text{need } \underline{\kappa < \Delta_B - 1}.$$

Now need to sum over all $s = \alpha_0 n$ until $s = \alpha n \dots$

$$\begin{aligned} \sum_{s=1}^{\alpha n} e^{-nA(s/n)} &= \sum_{s=1}^{\sqrt{n}} e^{-nA(s/n)} + \sum_{s=\sqrt{n}+1}^{\alpha n} e^{-nA(s/n)} \\ &\leq \sum_{s=1}^{\sqrt{n}} \left(\frac{c_1}{\sqrt{n}}\right)^{(\Delta_B - 1 - \kappa)s} + \sum_{s=\sqrt{n}+1}^{\alpha n} e^{-nA(s/n)} \\ &\quad \text{as } n \rightarrow \infty: \leq 2 \left(\frac{c_1}{\sqrt{n}}\right)^{\Delta_B - 1 - \kappa} + n e^{-c_2 \sqrt{n}} \end{aligned}$$

$O(1) : c_2 \leq \frac{A(x)}{x} \text{ in interval}$

So probability of any set violating linear confinement vanishes as $n \rightarrow \infty$

Thm: almost every random (Δ_B, Δ_C) code is, if $3 \leq \Delta_B < \Delta_C$,
asymptotically good: $k = \Theta(n)$ and $d = \Theta(n)$ as $n \rightarrow \infty$

Proof: combine previous Thm w/ earlier Lemma
linear confinement $\Rightarrow$ linear distance!