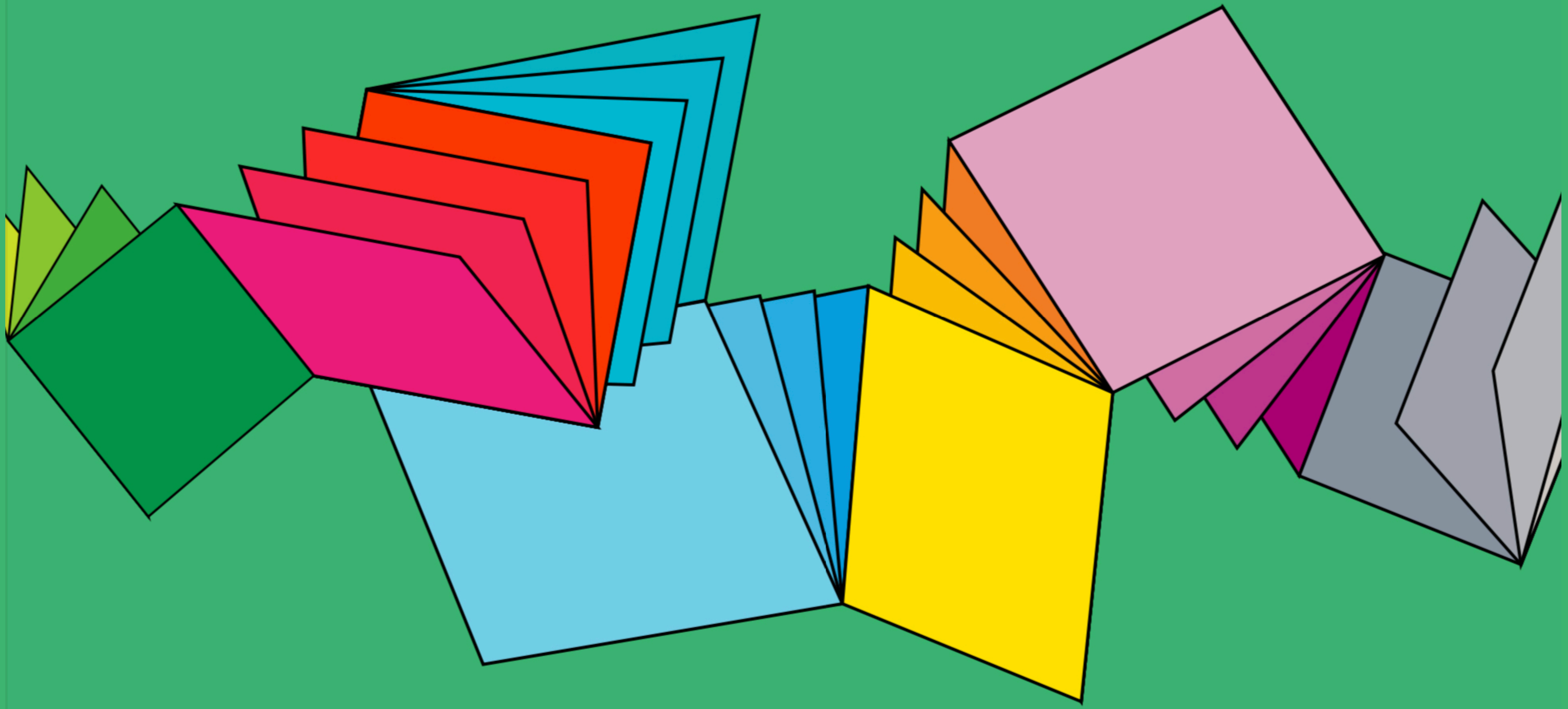
Locally testable codes with constant rate, distance, and locality

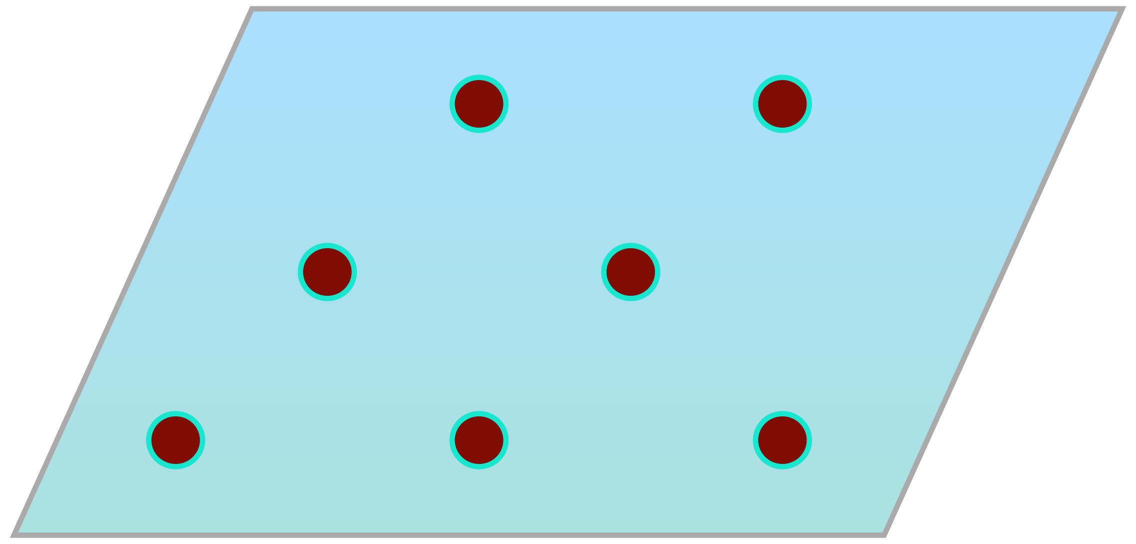


Irit Dinur

based on work with
Shai Evra
Ron Livne
Alexander Lubotzky
Shahar Mozes

Error Correcting Codes

A linear error-correcting code is a linear subspace $C \subseteq \{0,1\}^n$



The strings in this subspace are called **codewords**

Imagine sending a codeword over a noisy channel which flips a few bits

If the codewords are far apart, there is a unique way to recover the codeword

Distance = $min_{x \neq y \in C} \frac{|\{i : x_i \neq y_i\}|}{n}$

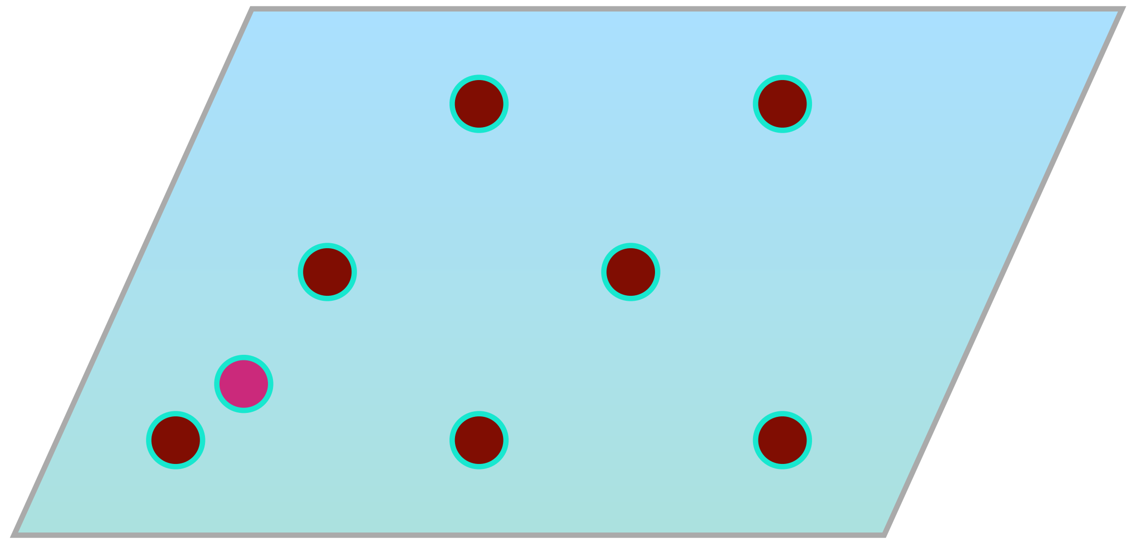
Of course, we want to be able to encode many messages using n bits

Rate = $\frac{dim(C)}{n}$,

Local Testability

Given a string $w \in \{0,1\}^n$, is $w \in C$?

Local testability - decide this question by reading a tiny (but random) part of w



Definition: A code C is locally testable with q queries if there is a tester T that has query access to a given word w, reads q randomized bits from w and accepts / rejects, such that

- If $w \in C$ then Pr[T accepts] $= 1$
- If $w \notin C$ then Pr[T rejects] $\geq const \cdot dist(w, C)$

q = the locality of the tester

such codes are called "LTCs"

Two measures of distance

Suppose $C \subseteq \{0,1\}^n$ is defined by linear constraints $t_1, t_2, t_3, \ldots, t_m$ ("local tests")

Given a string $w \in \{0,1\}^n$, there are two measures of how "far" w is from C

1. **Hamming distance**: how many bits do we need to flip to put $w \in C$

2. **T-distance**: how many linear constraints ("tests") are not satisfied by w

Hamming distance is natural - but may be hard to compute

T-distance is easy to compute / estimate (by selecting a few random constraints)

LTC (alternative definition): a code where the T-distance is local and gives a good estimate for Hamming-distance.

Potentially useful: we can test if many errors happened, and ask for retransmission

Historical background

- LTCs were studied implicitly in early works on PCPs (probabilistically checkable proofs) [BlumLubyRubinfeld 1990, BabaiFortnowLund 1990, ..]
- A systematic study initiated by Goldreich and Sudan in 2002.
  "what is the highest possible rate of an LTC?"
- Sequence of works (BenSasson-Sudan-Vadhan-Wigderson2003, BenSasson-Goldreich-Harsha-Sudan-Vadhan2004, Ben-Sasson-Sudan2005, Dinur2005, Kopparty-Meir-RonZewi-Saraf2017, Gopi-Kopparty-OliveiraRonZewi-Saraf2018) achieved rate = 1/polylog & constant locality+distance
- Are there "c^3 LTCs" (constant rate, constant distance, constant locality) ? experts doubt existence. Restricted lower bounds are shown [BenSasson-Harsha-Rashkhodnikova2005, Babai-Shpilka-Stefankovic2005, BenSasson-Guruswami-Kaufman-Sudan-Viderman2010, D.-Kaufman2011]
- High dimensional expansion: local to global features [Garland 1973, Kaufman-Lubotzky 2013, Kaufman-Kazhdan-Lubotzky 2014, Evra-Kaufman 2016, Oppenheim 2017, D.-Kaufman 2017, D.-Harsha-Kaufman-LivniNavon-TaShma 2019, Dikstein-D.-Harsha-Kaufman-RonZewi 2019, Anari-Liu-OveisGharan-Vinzant2019]

Main Result

For every $0 < r < 1$ there exist $\delta > 0$ and $q \in \mathbb{N}$ and an explicit construction of an infinite family of error-correcting codes $\{C_n\}_n$ with rate $\geq r$, distance $\geq \delta$ and locally testable with q queries.

(in fact, relying on previous reductions, r,δ —> "GV bound")

Panteleev & Kalachev [very recently]:

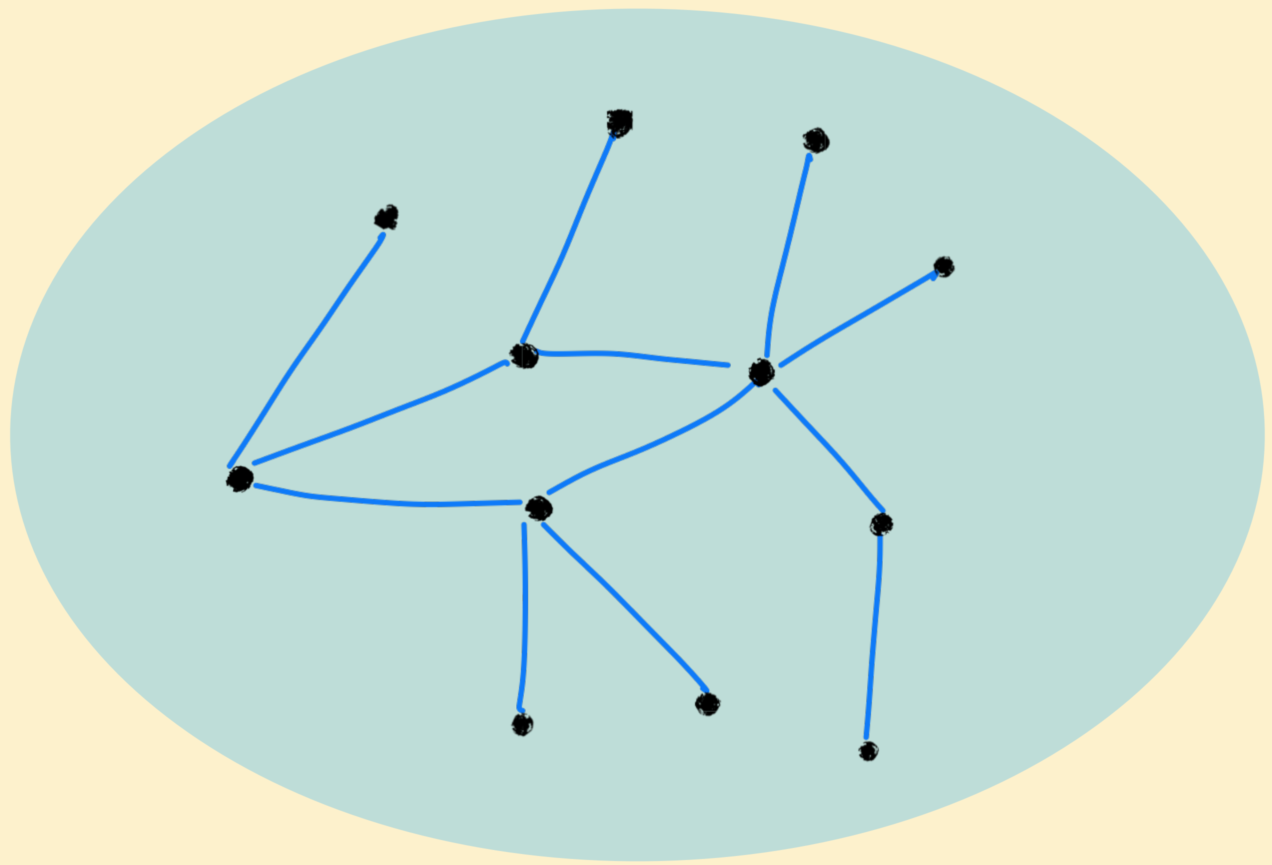
Similar result, almost identical construction, (+quantum LDPC codes!)

based on "balanced product" of Breuckmann & Eberhardt

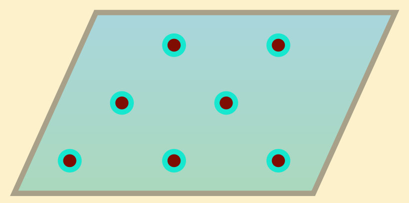
Plan of talk

1. Expander codes
2. New: left-right Cayley complex, “a graph-with-squares”
3. Define the code on the complex / graph-with-squares
4. Properties of the code

Expander Codes [Sipser & Spielman 1996]



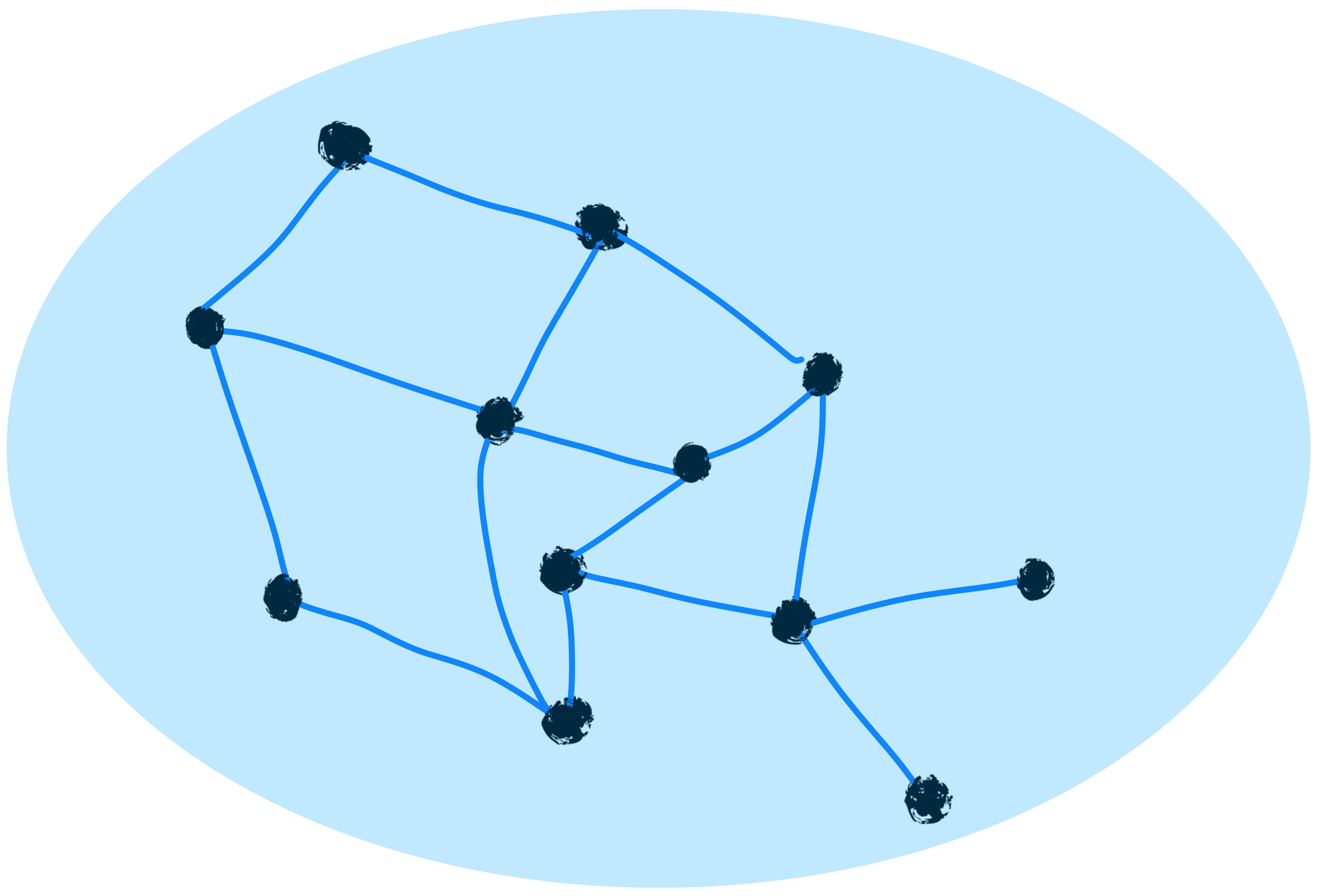
d-regular expander graph $G = (V, E)$

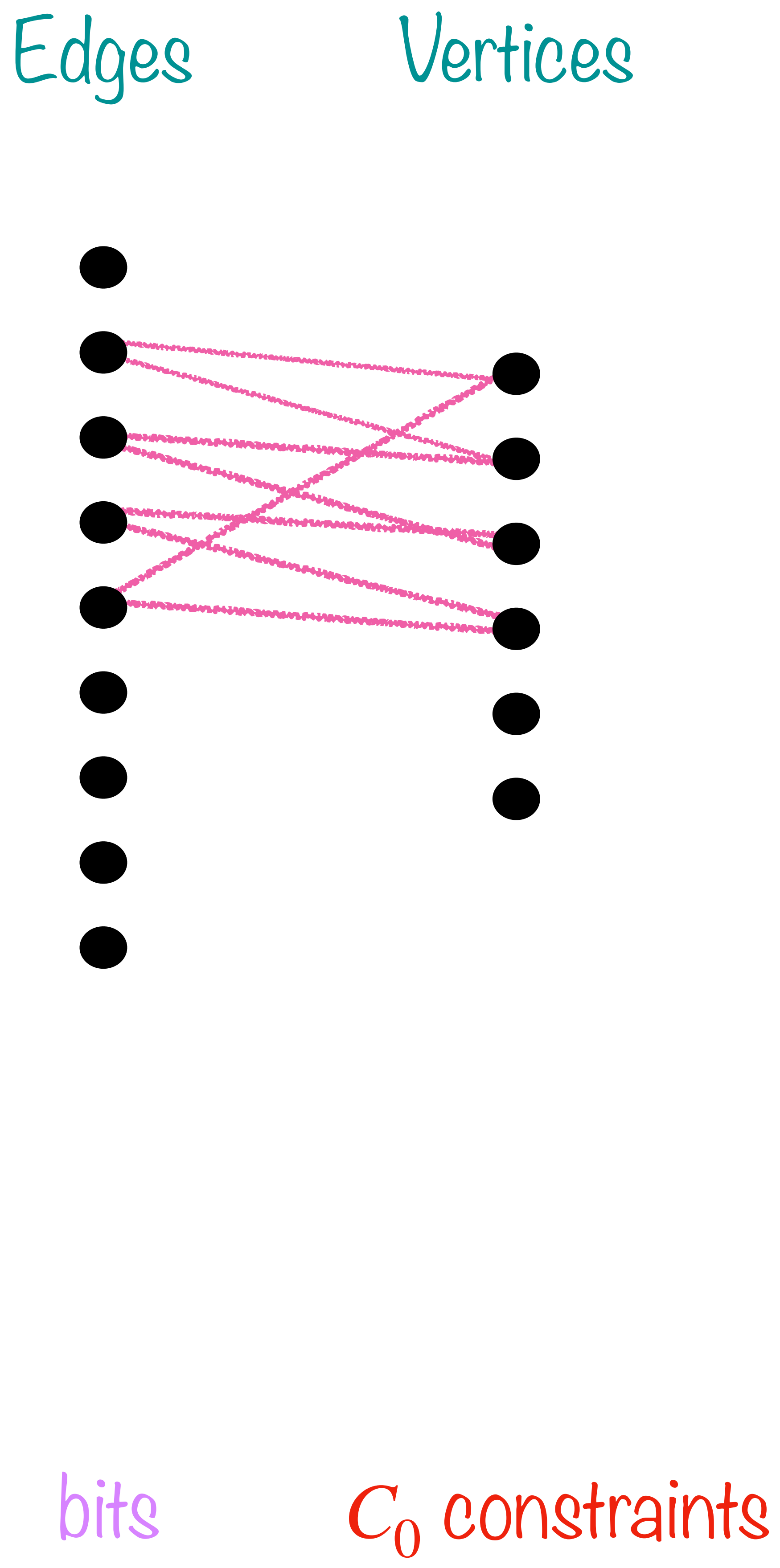


good base code C_0
on d bits

$$C[G, C_0] = \{ f : E \to \{0,1\} : \ f|_{edges(v)} \in C_0 \ \ \forall v \}$$

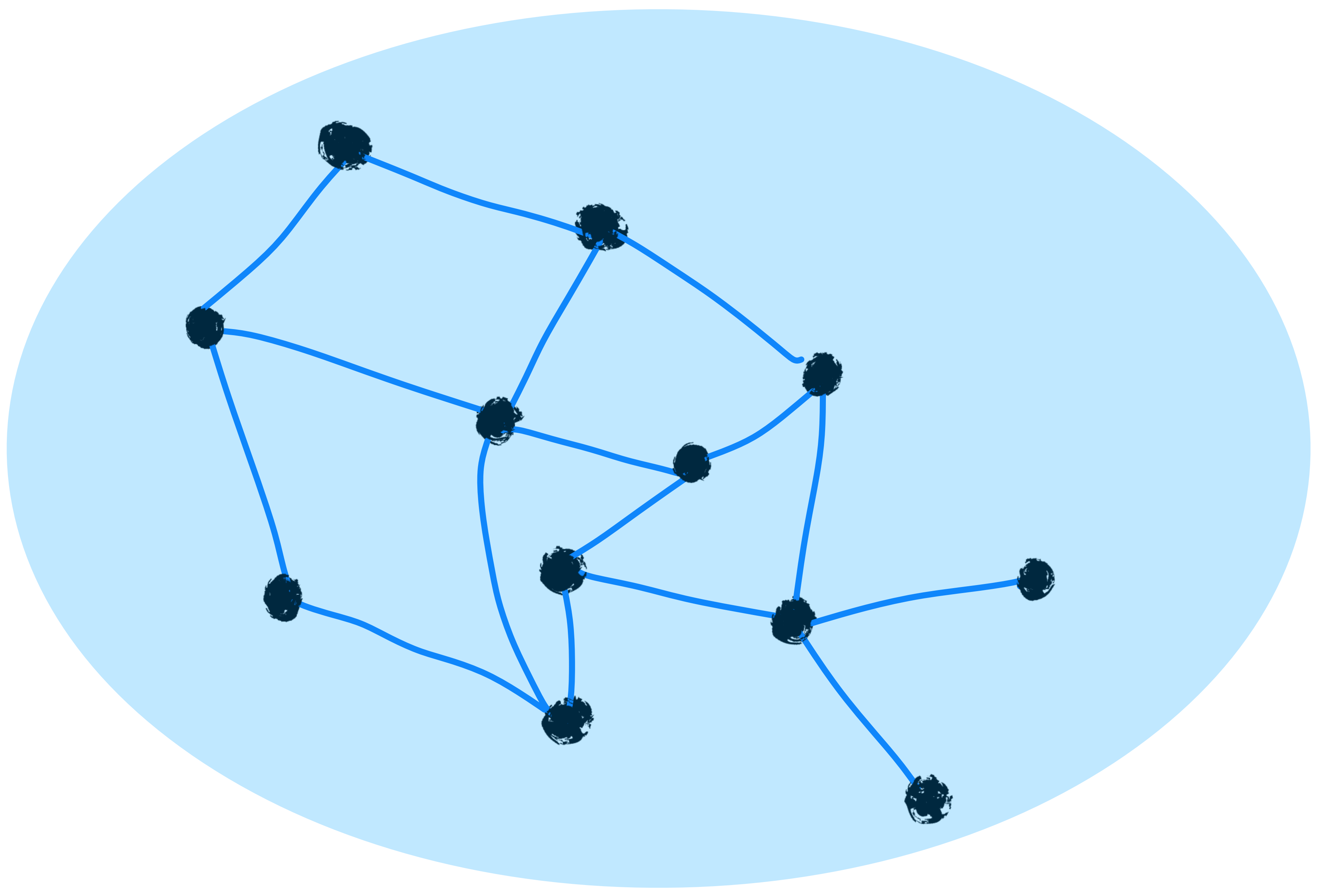
Expander Codes as Tanner Codes

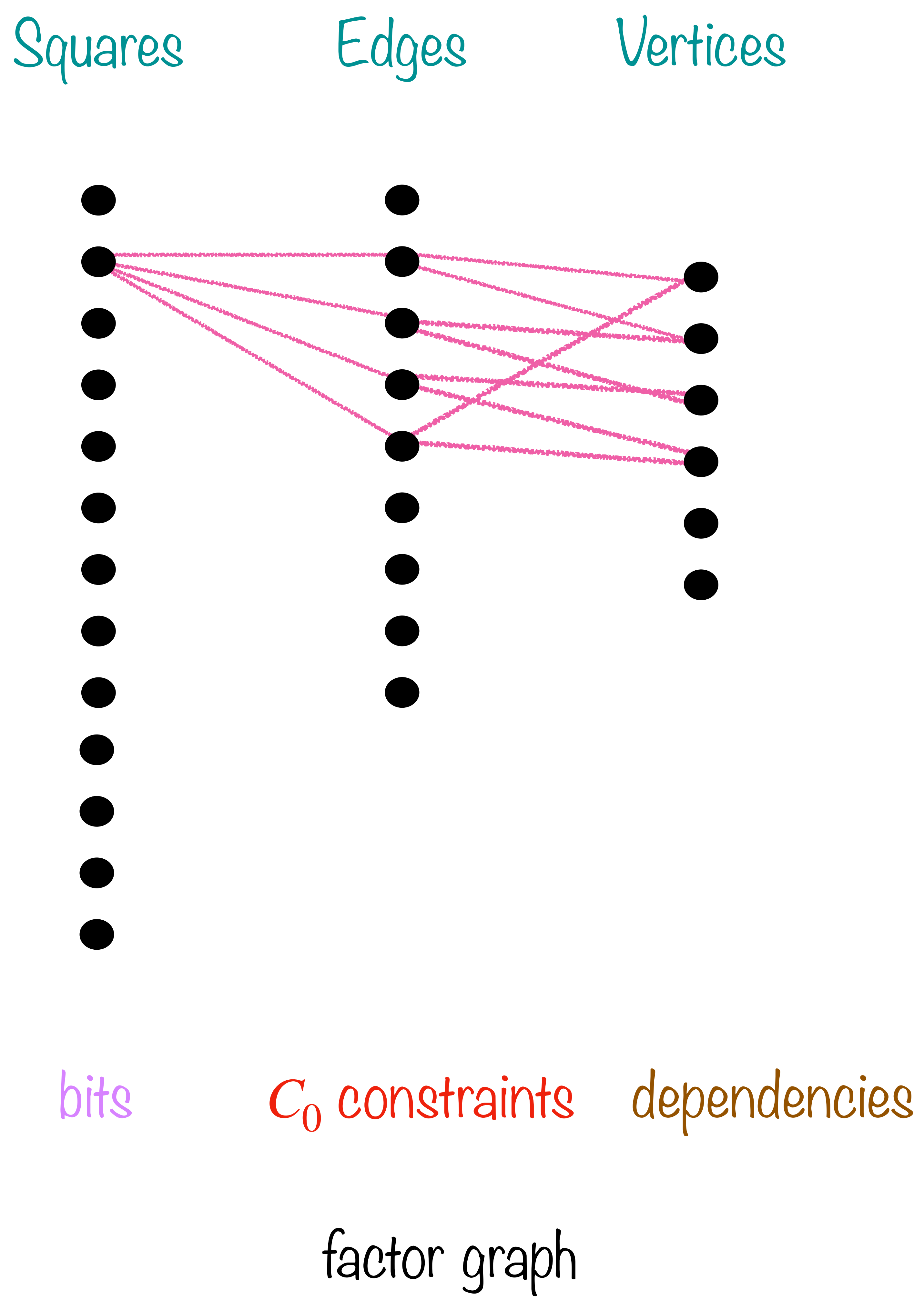




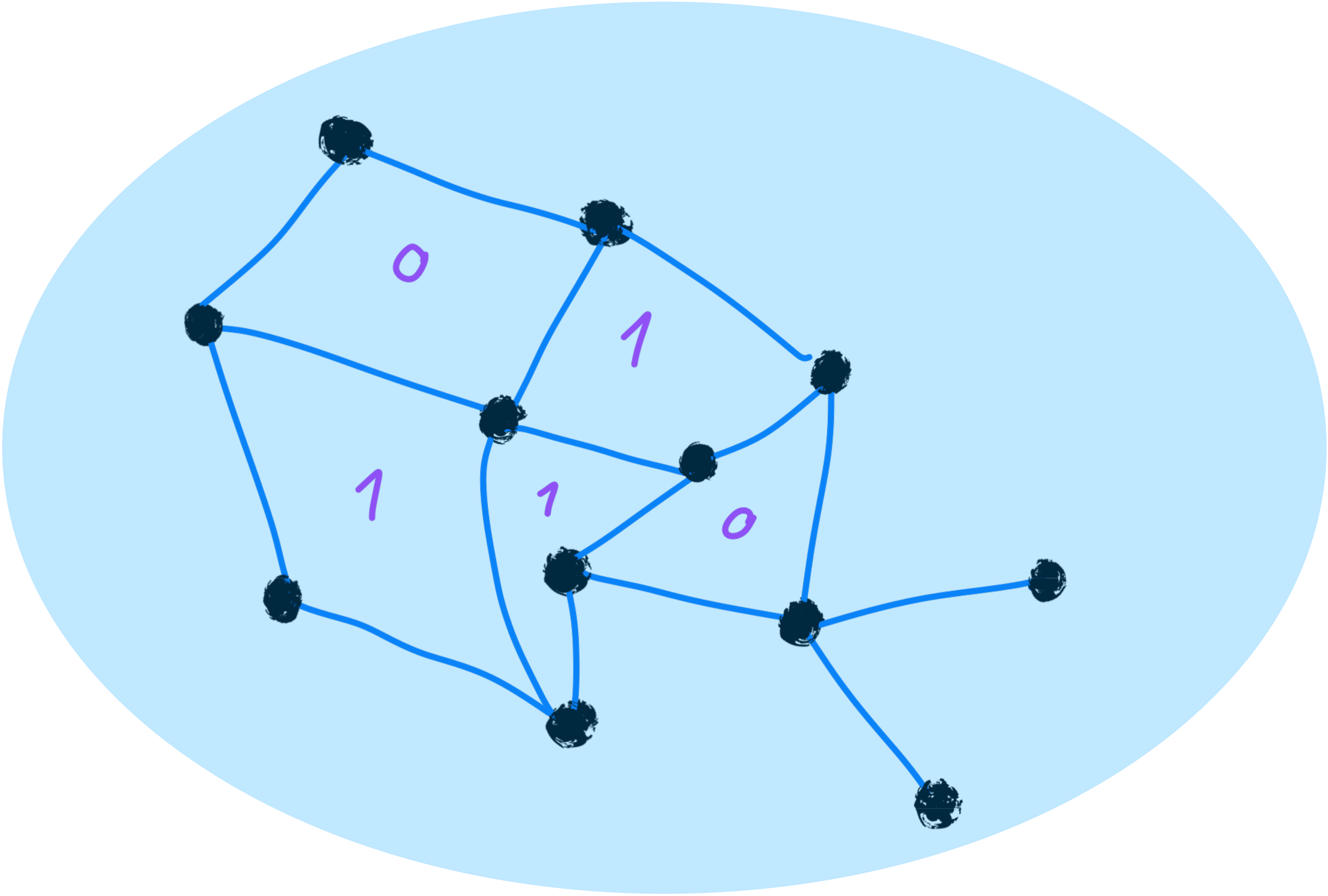
factor graph

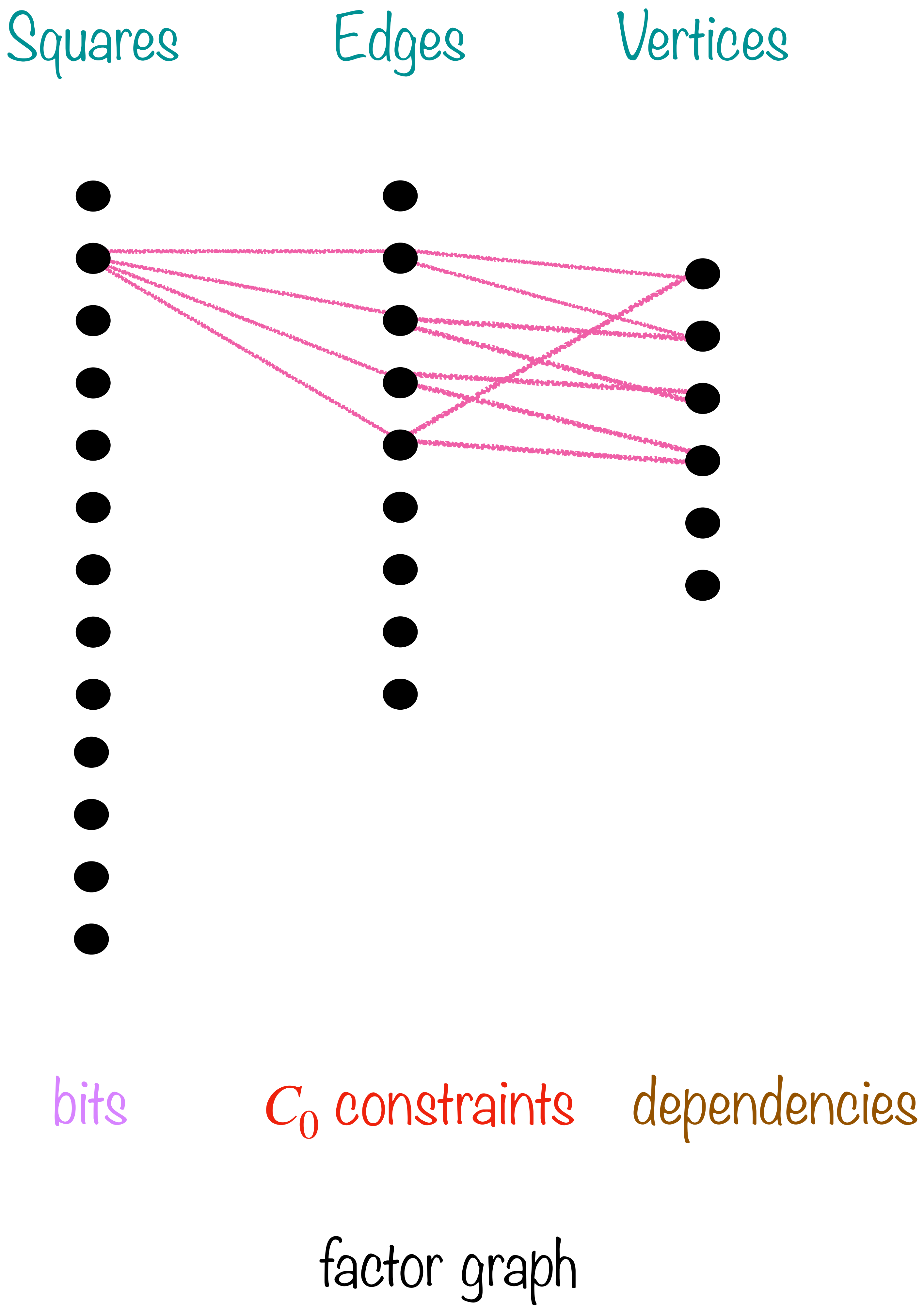
Expander Codes, one level up





Expander Codes, one level up





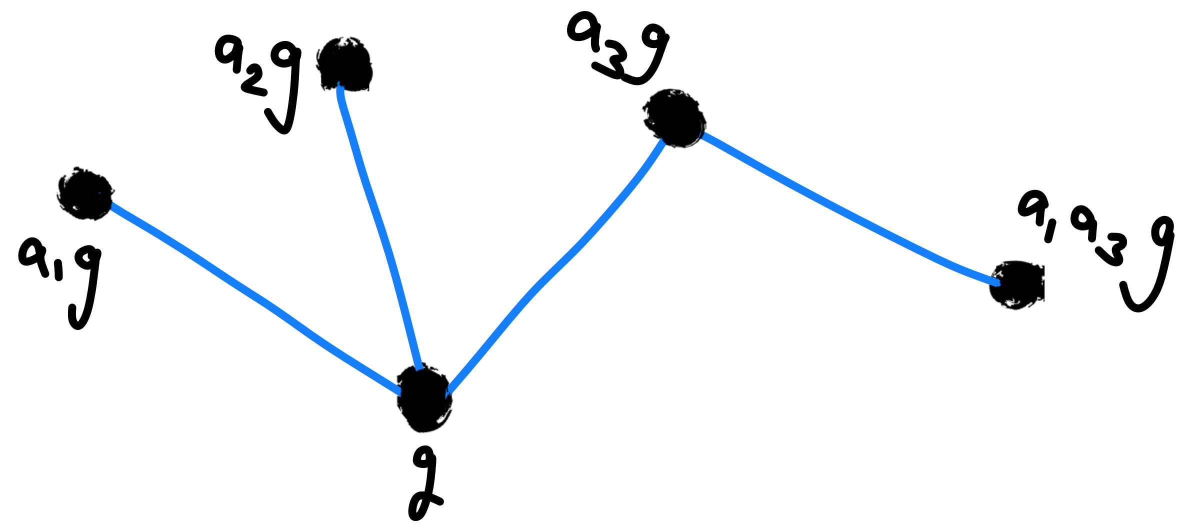
Left-right Cayley Complex

"a graph with squares"

Let G be a finite group,

Let $A \subset G$ be closed under taking inverses, i.e. such that $a \in A \rightarrow a^{-1} \in A$

Cay(G,A) is a graph with vertices G, and edges $E_A = \{\{g, ag\} : g \in G, a \in A\}$

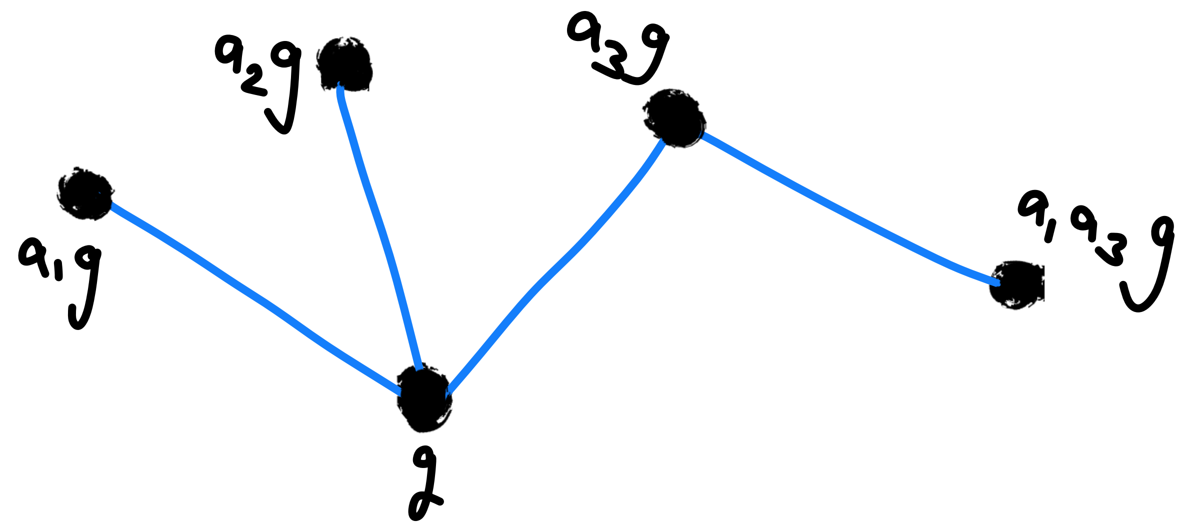


Left-right Cayley Complex

"a graph with squares"

Let G be a finite group,

Let $A, B \subset G$ be closed under taking inverses



Left-right Cayley Complex

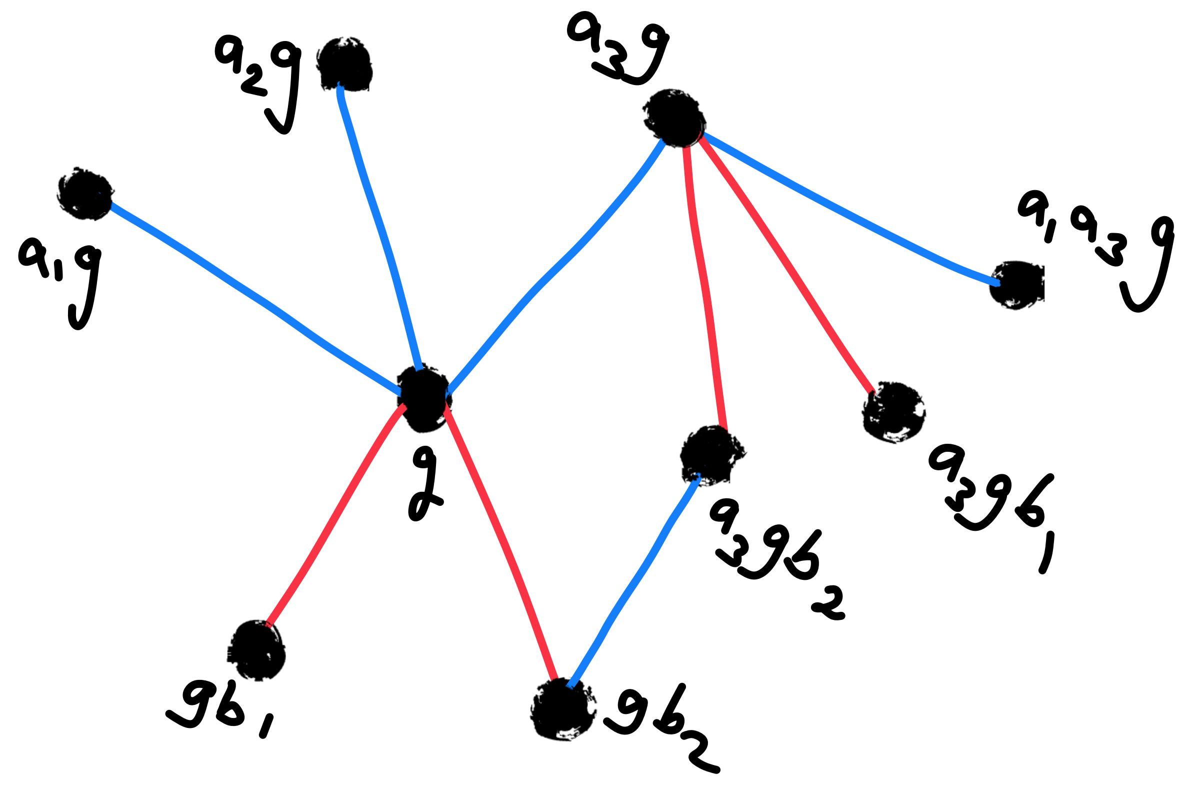
"a graph with squares"

Let G be a finite group,

Let $A, B \subset G$ be closed under taking inverses

Cay(G,A) is a graph with vertices G, and edges $E_A = \{\{g, ag\} : g \in G, a \in A\}$ (left *)

Cay(G,B) is a graph with vertices G, and edges $E_B = \{\{g, gb\} : g \in G, b \in B\}$ (right *)



Left-right Cayley Complex

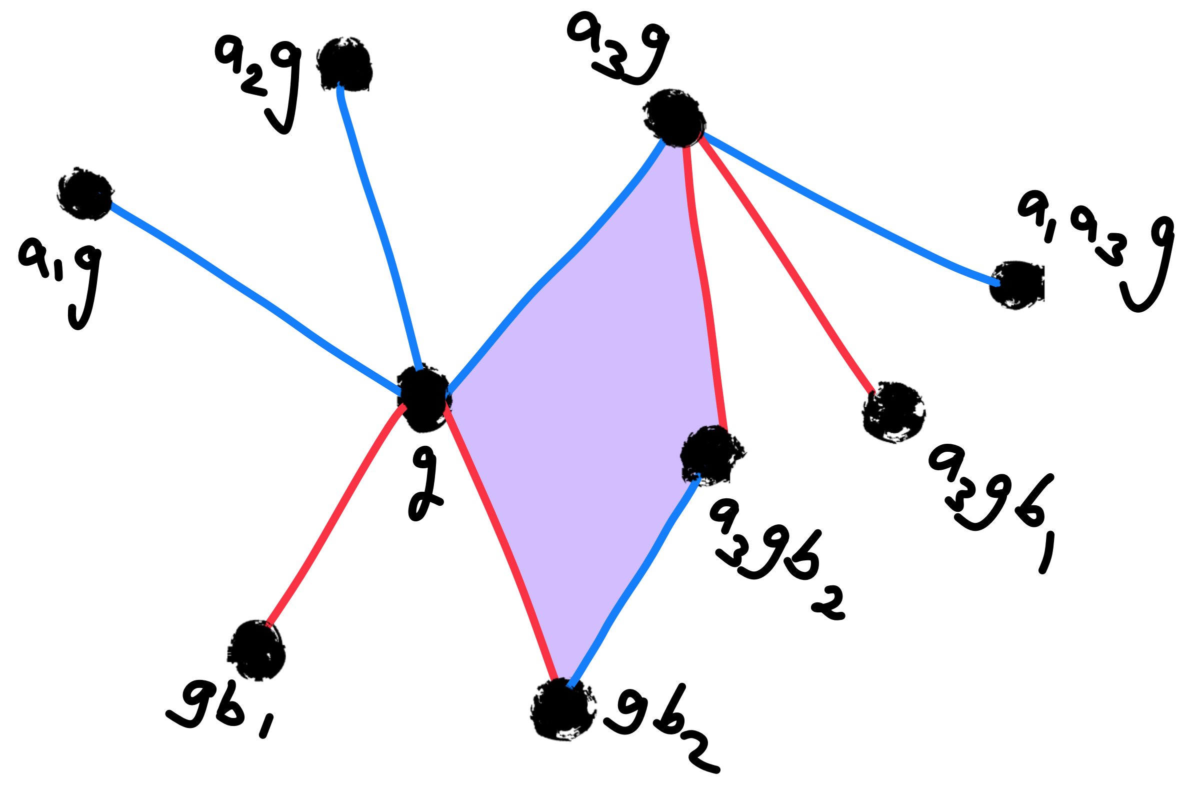
"a graph with squares"

Let G be a finite group,

Let $A, B \subset G$ be closed under taking inverses

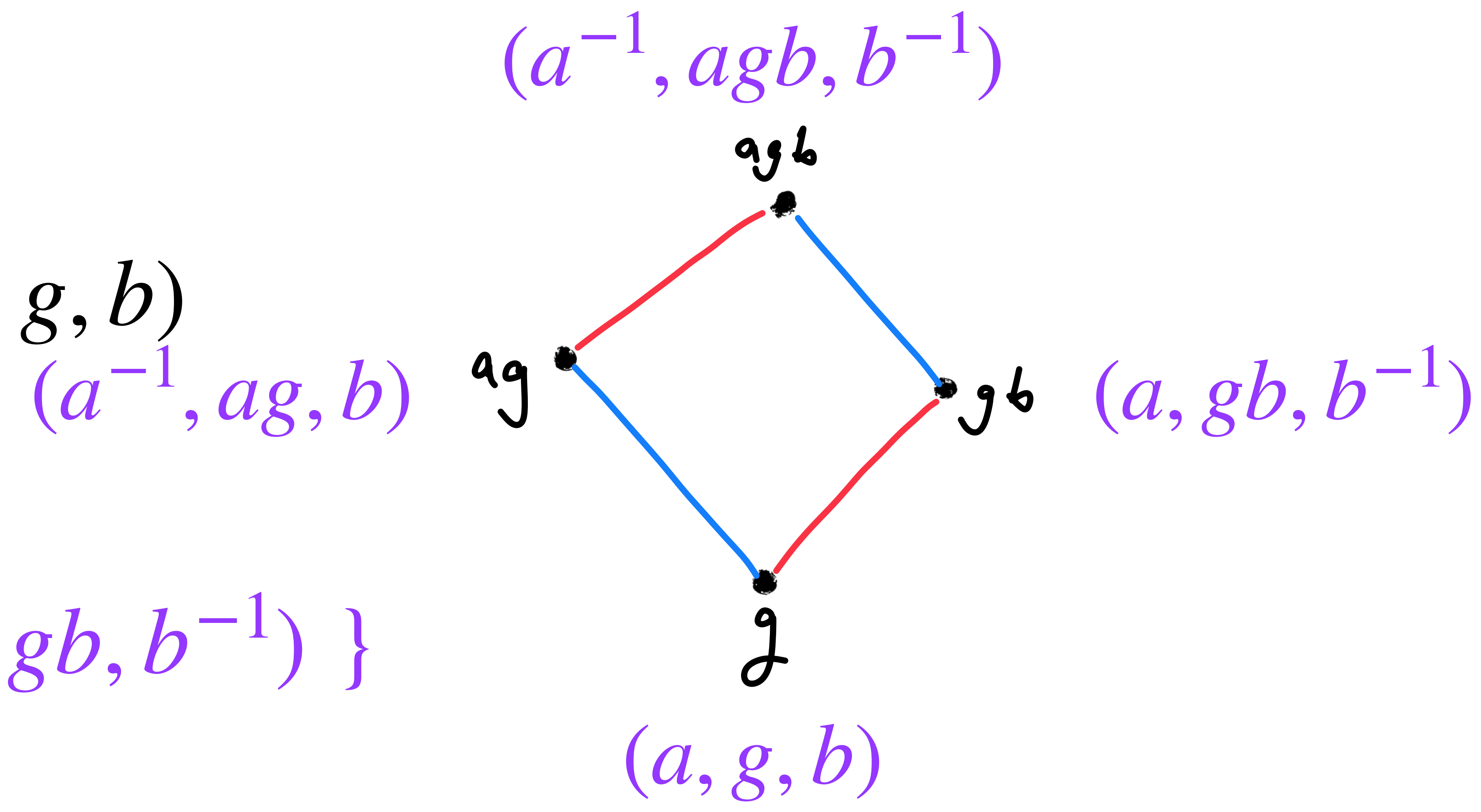
Cay(G,A) is a graph with vertices G, and edges $E_A = \{\{g, ag\} : g \in G, a \in A\}$ (left *)

Cay(G,B) is a graph with vertices G, and edges $E_B = \{\{g, gb\} : g \in G, b \in B\}$ (right *)



Left-right Cayley Complex

"a graph with squares"



Each triple $a \in A, g \in G, b \in B$ define a <u>rooted square</u> (a, g, b)

Each square can have 4 roots,

$[a, g, b] = \{\ (a, g, b),\quad (a^{-1}, ag, b),\quad (a^{-1}, agb, b^{-1}),\quad (a, gb, b^{-1})\ \}$

This square naturally contains

- The edges {g,ag}, {g,gb}, {gb,agb}, {ag,agb},
- The vertices g,ag,gb,agb

The set of squares is $X(2) = \{[a, g, b] : g \in G, a \in A, b \in B\} \quad = \quad A \times G \times B \,/ \sim$

Left-right Cayley Complex Cay²(A,G,B)

Let G be a finite group, and let $A, B \subset G$ be closed under taking inverses.

The left-right Cayley complex Cay²(A,G,B) has

- Vertices G
- Edges $E_A \cup E_B$

  $$E_A = \{\{g, ag\} : g \in G, a \in A\}, \quad E_B = \{\{g, gb\} : g \in G, b \in B\}$$

- Squares A x G x B / ~

We say that Cay²(A,G,B) is a λ-expander if Cay(G,A) and Cay(G,B) are λ-expanders.

Lemma: For every $\lambda > 0$ there are explicit infinite families of bounded-degree left-right Cayley complexes that are λ-expanders.

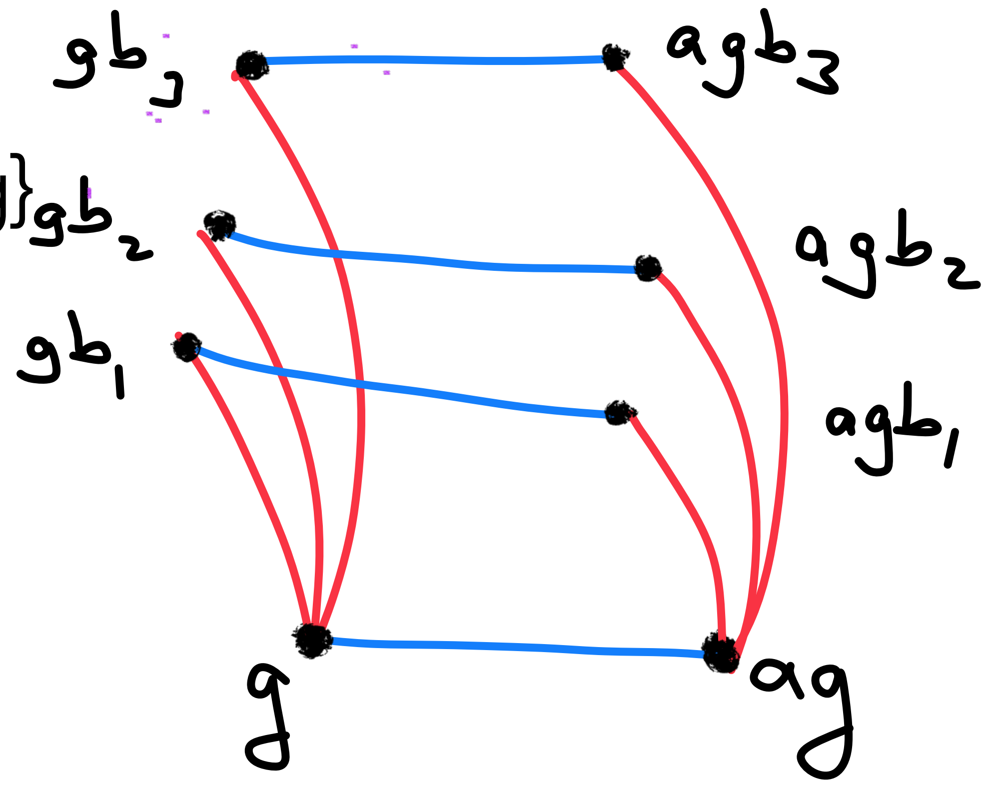
Left-right Cayley Complex

"a graph with squares"

Squares touching the edge {g,ag}

are naturally identified with B

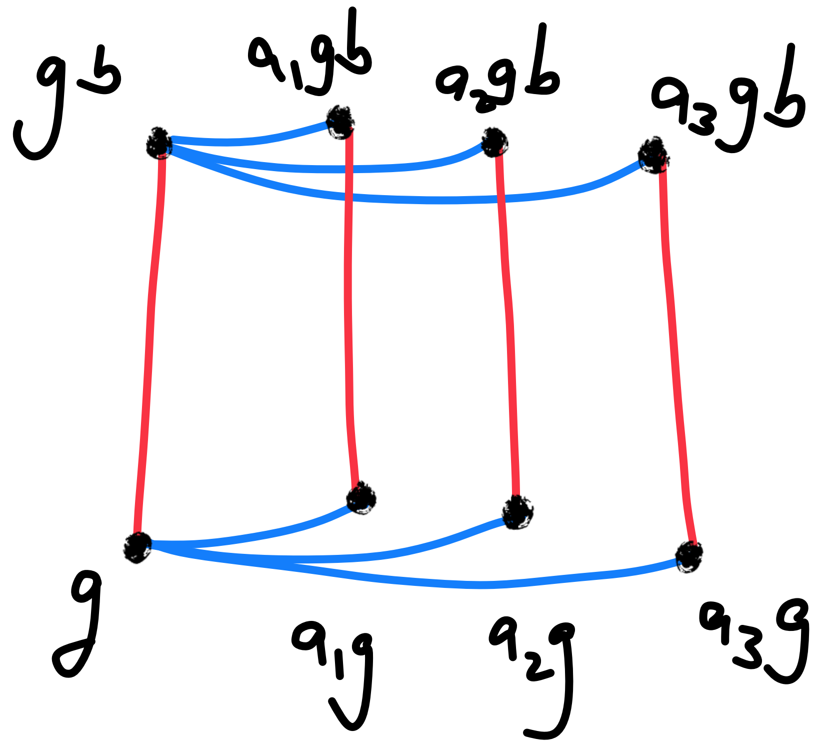
$b \mapsto [a, g, b]$



Squares touching the edge {g,gb}

are naturally identified with A

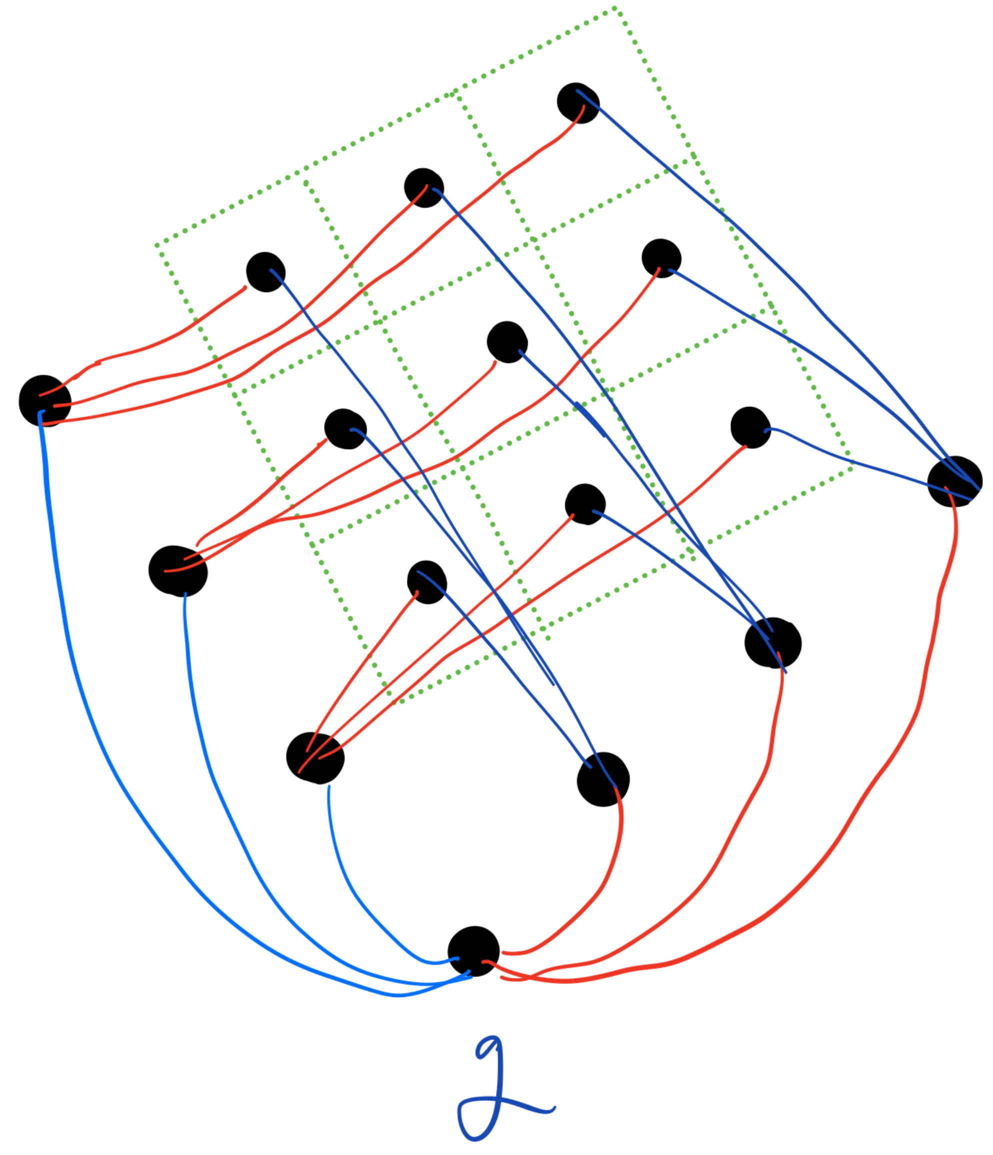
$a \mapsto [a, g, b]$



A vertex g has |A| + |B| neighbors

For each $a \in A, b \in B$ there is one square touching g,

so there is a natural bijection* $(a, b) \mapsto [a, g, b]$



* it is a bijection assuming $\forall a, b, g, \quad g^{-1}ag \neq b$

Left-right Cayley Complex

"a graph with squares"

Squares touching the edge {g,ag}

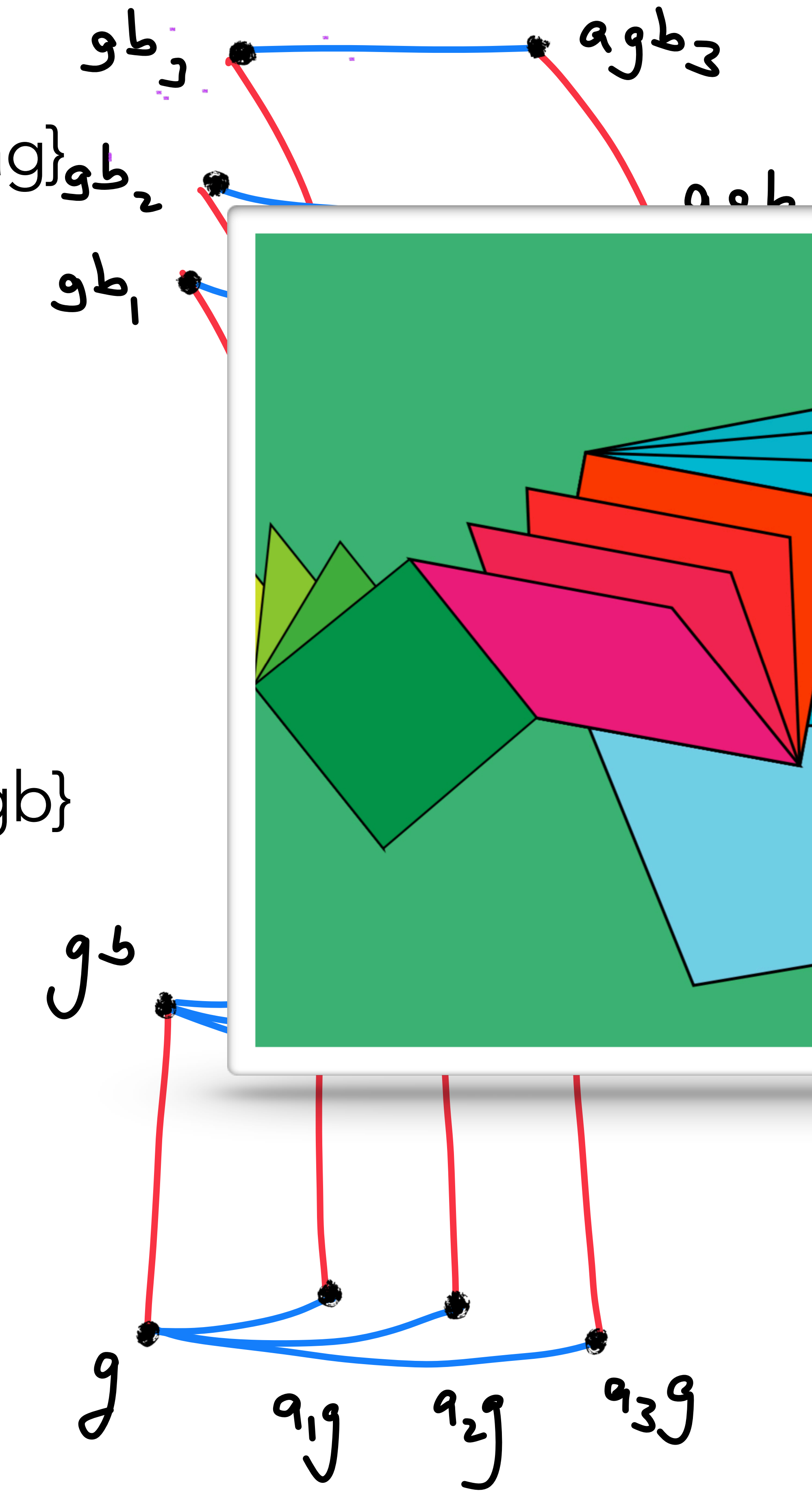
are naturally identified with B

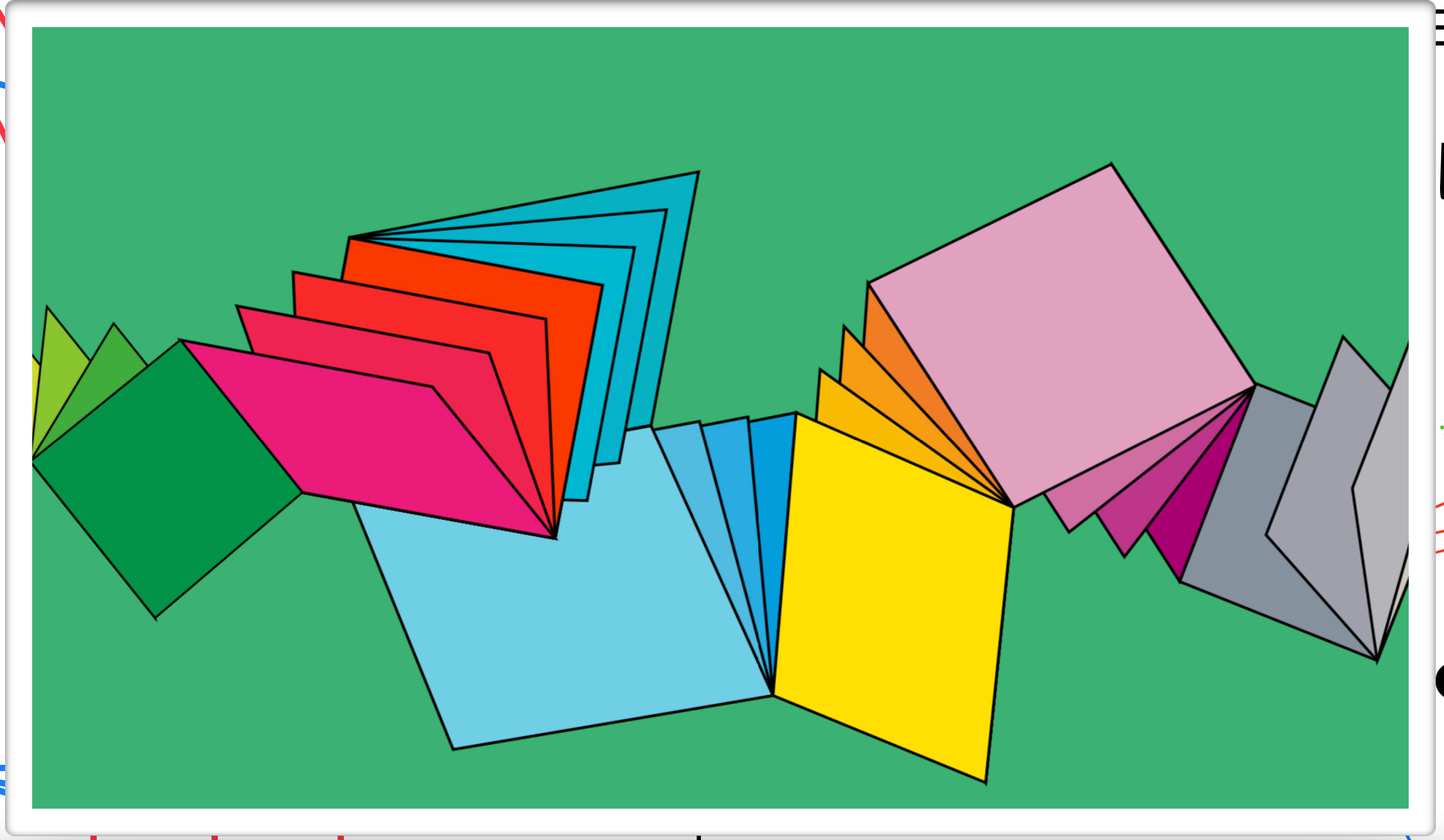
$$b \mapsto [a,g,b]$$

Squares touching the edge {g,gb}

are naturally identified with A

$$a \mapsto [a,g,b]$$

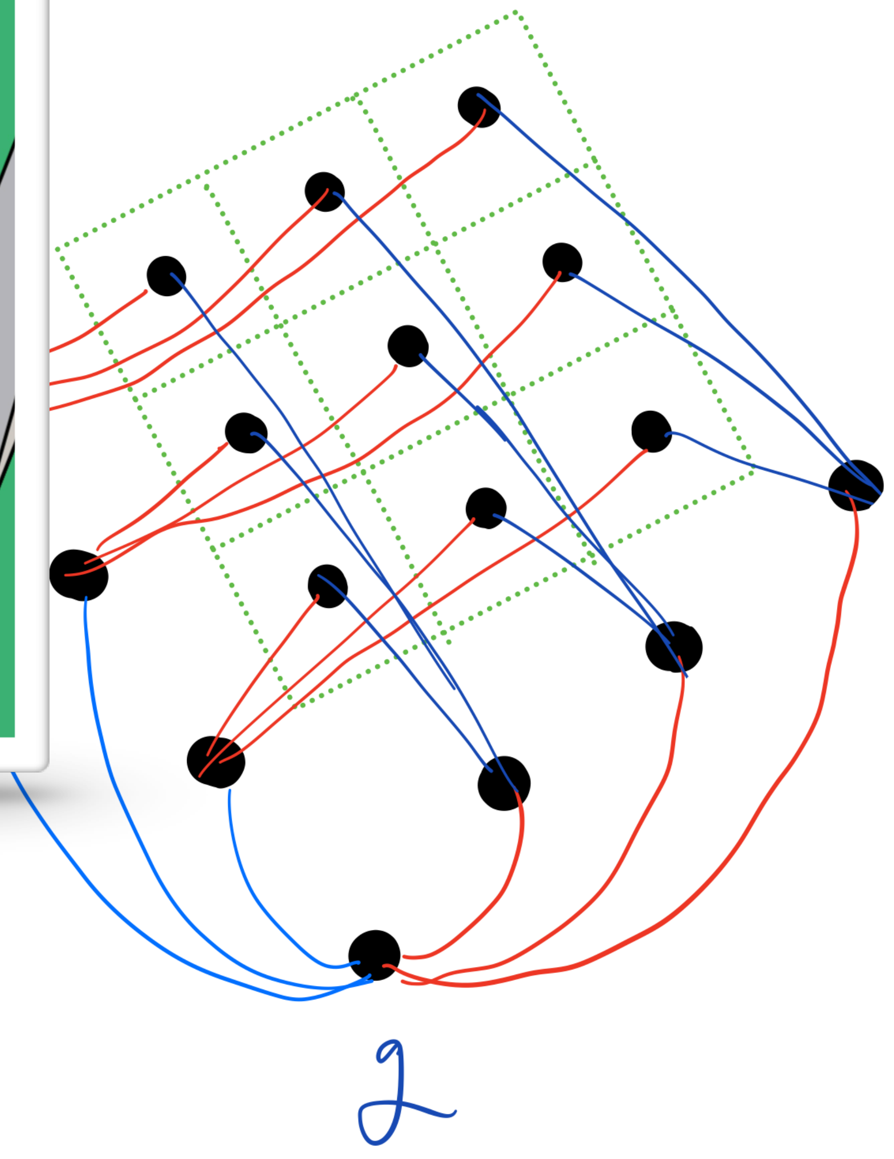




A vertex g has |A| + |B| neighbors

F... B there is one square touching g,

... bijection* $(a,b) \mapsto [a,g,b]$

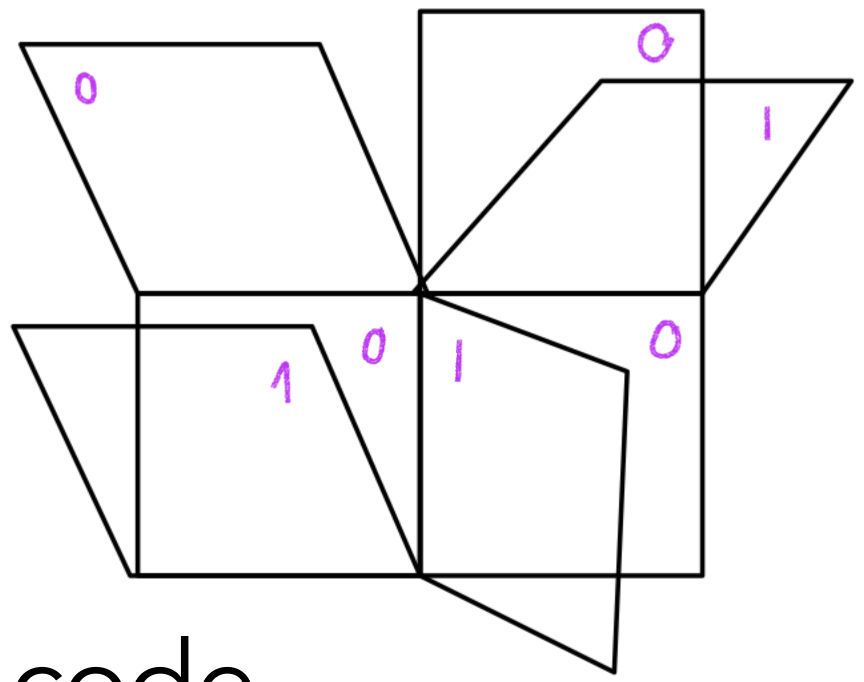


* it is a bijection assuming $\forall a,b,g, \quad g^{-1}ag \neq b$

The Code

Let Cay²(A,G,B) be a left-right Cayley complex.

Fix base codes $C_A \subseteq \{0,1\}^A, C_B \subseteq \{0,1\}^B$ (assuming |A| = |B| = d we can take one base code $C_0 \subseteq \{0,1\}^d$ and let $C_A, C_B \simeq C_0$)



Define a code CODE = $C[G, A, B, C_A, C_B]$:

- The codeword bits are placed on the squares
- Each edge requires that the bits on the squares around it are in the base code

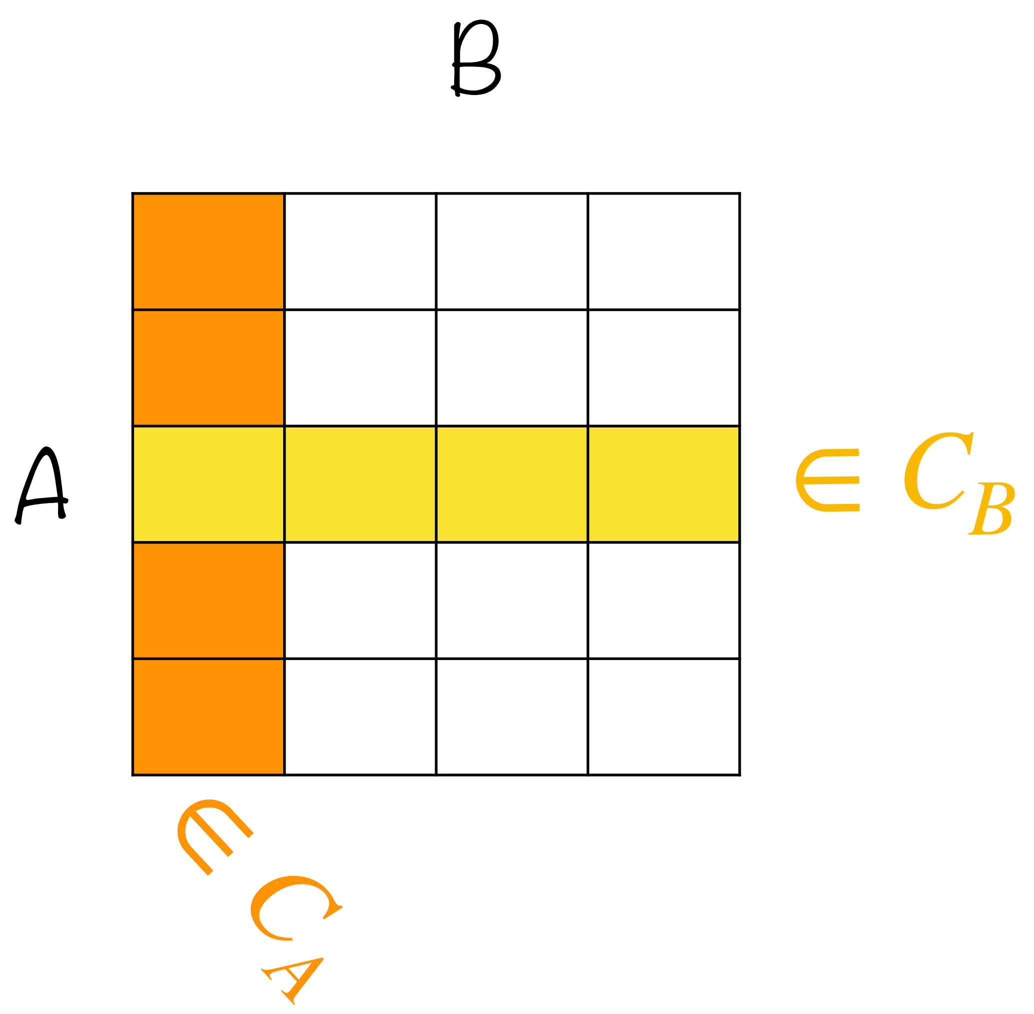
$$\boxed{\text{CODE} = \{f : Squares \to \{0,1\} : \forall a, g, b, \quad f([\,\cdot\,, g, b]) \in C_A,\ f([a, g, \cdot\,]) \in C_B\}}$$

Rate: $\geq 4r_0 - 3$ [calc: #squares - #constraints]

Distance: $\geq \delta_0^2(\delta_0 - \lambda)$ [easy propagation argument]

Local views are tensor codes

Claim: Fix f∈CODE. For each $g \in G$, $f([\cdot, g, \cdot]) \in C_A \otimes C_B$



Theorem: Assume Cay²(A,G,B) is a λ-expander, and $C_A \otimes C_B$ is ρ-robustly testable. If $\lambda < \delta_0\rho/5$, then $C[G, A, B, C_A, C_B]$ is locally testable.

The tester is as follows:

1. Select a vertex g uniformly,
2. Read f on all $|A| \cdot |B|$ squares touching g, namely $f([\cdot, g, \cdot])$.
3. Accept iff this belongs to $C_A \otimes C_B$

Then $\Pr_{g \in G}[f([\cdot, g, \cdot]) \notin C_A \otimes C_B) \geq const \cdot dist(f, C[G, A, B, C_A, C_B])$

CODE = $\{f : Squares \to \{0,1\} : \forall a, g, b, \quad f([\cdot, g, b]) \in C_A, f([a, g, \cdot]) \in C_B\}$

Proof of local-testability

Start with $f : Squares \to \{0,1\}$ and find $f' \in C,\ \ rej(f) \geq dist(f, f') \cdot const$

ALG "self-correct":

1. Init: Each $g \in G$ finds $T_g \in C_A \otimes C_B$ closest to $f([\cdot, g, \cdot])$

   [define a progress measure Φ = # dispute edges]

2. Loop: If g can change T_g and reduce Φ then do it

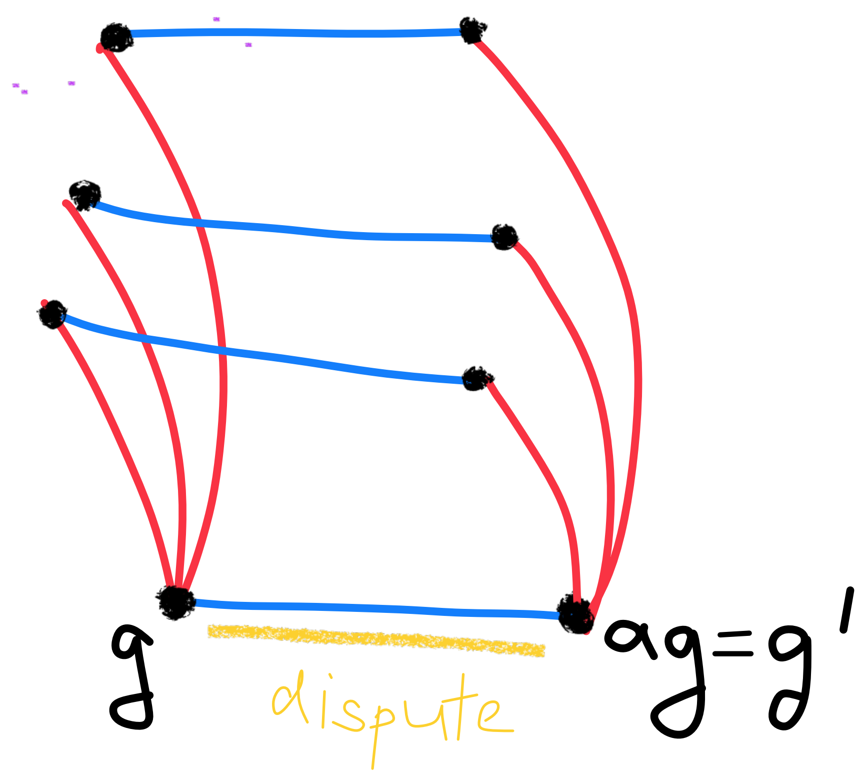
3. End: If $\Phi = 0$ let $f'([a,g,b]) = T_g(a,b)$ and output f',
   If $\Phi > 0$ quit

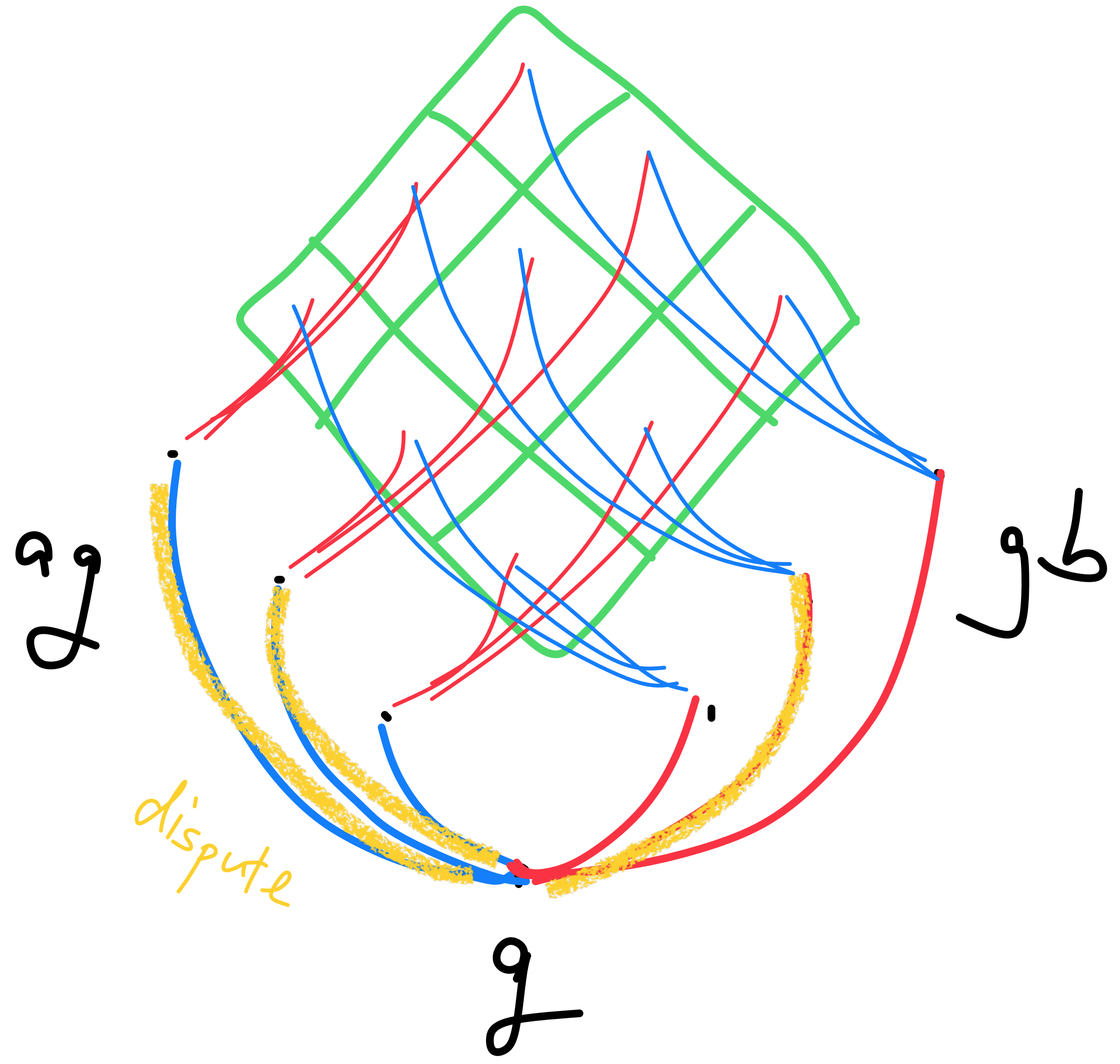
- steps $\leq \Phi\ \approx$ rej(f)
- If $\Phi = 0$ then $rej(f) \geq dist(f, f') \cdot const$
- **If $\Phi > 0$ then $\Phi > 0.1$** so $rej(f) \geq dist(f, f') \cdot 0.1$

Proof of local-testability

If ALG "self-correct" is stuck then rej (f) > 0.1

- If g,g′ are in dispute, there must be many squares on {g,g′} with further dispute edges
- Can try to propagate, but, they all might be clumped around g
- But then g's neighbors all agree, so there must have been a better choice for T_g (using the LTCness of tensor codes)
- Random walk **edge—>square—>edge** + expansion ==> dispute set is large





High dimensional expansion

The idea of using a higher-dimensional complex instead of a graph for LTCs has been circulating a number of years.

HDXs exhibit local-to-global features: prove something locally and then use expansion to globablize

[Garland 1973, Kaufman-Kazhdan-Lubotzky2014, Evra-Kaufman2016, Oppenheim2017, D.-Kaufman2017, D.-Harsha-Kaufman-LivniNavon-TaShma2018, Anari-Liu-OveisGharan-Vinzant2019]

Dikstein-D.-Harsha-RonZewi2019 - Locally testable codes on HDX can "theoretically" work

How to"instantiate" this? …we worked on the Lubotzky-Samuels-Vishne complexes (quotients of BT buildings), and have conjectured base codes, but no proof of local LTCness

…questions

- Can such ideas be used for constructing PCPs?
- Can these codes be made practical?