I. Approaches to bounding the exponent of matrix multiplication

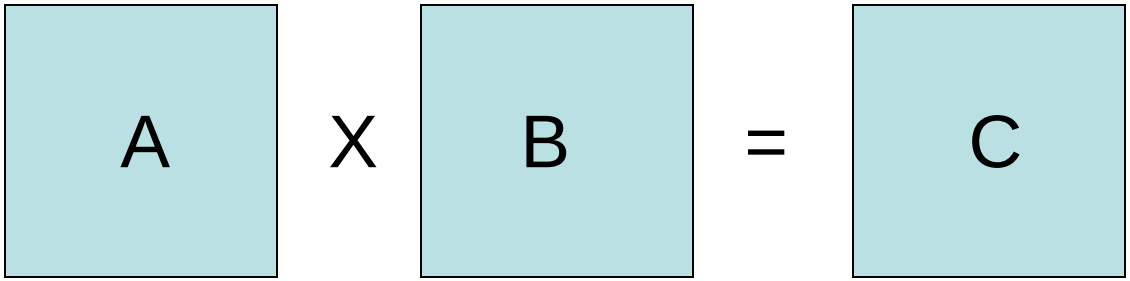
Chris Umans

Caltech

Based on joint work with Noga Alon, Henry Cohn, Bobby Kleinberg, Amir Shpilka, Balazs Szegedy

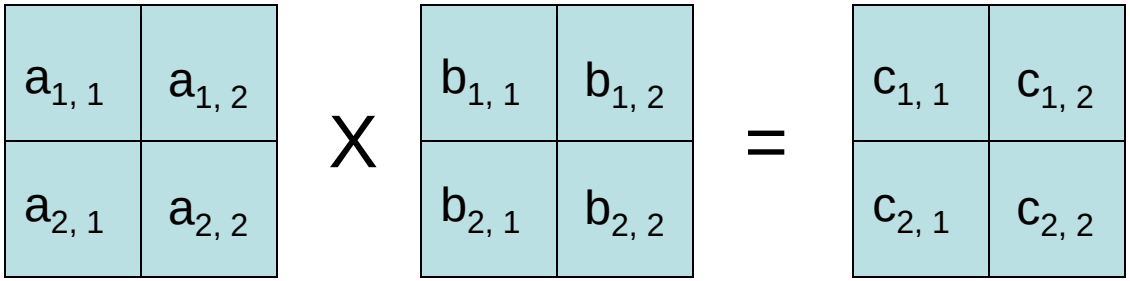
Modern Applications of Representation Theory, IMA, Chicago July 2014

Introduction



- Standard method: $O(n^3)$ operations
- Strassen (1969): $O(n^{2.81})$ operations

Strassen's Algorithm



- linear combos of A, B entries
- 7 mults
- linear combos of results

$P_1 = (a_{1,1} - a_{1,2}) \times (b_{2,1} + b_{2,2})$

$P_2 = (a_{1,1} + a_{2,2}) \times (b_{1,1} + b_{2,2})$

$P_3 = (a_{1,1} - a_{2,1}) \times (b_{1,1} + b_{1,2})$

$P_4 = (a_{1,1} + a_{1,2}) \times (b_{2,2})$

$P_5 = (a_{1,1}) \times (b_{1,2} + b_{2,2})$

$P_6 = (a_{2,2}) \times (b_{2,1} - b_{1,1})$

$P_7 = (a_{2,1} - a_{2,2}) \times (b_{1,1})$

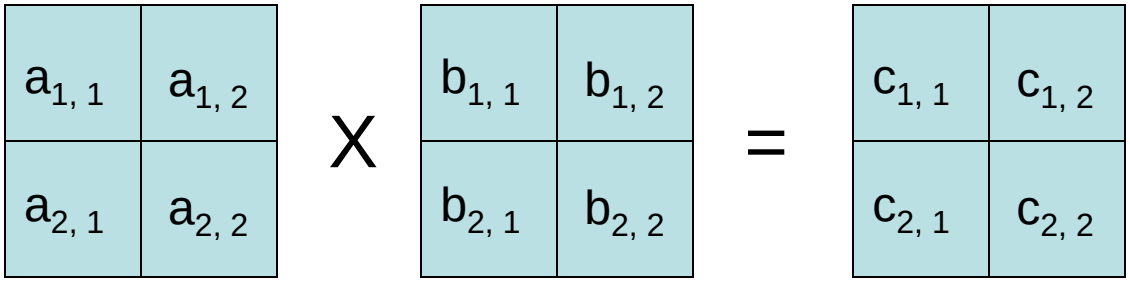
$c_{1,1} = P_1 + P_2 - P_4 + P_6$

$c_{1,2} = P_4 + P_5$

$c_{2,1} = P_6 + P_7$

$c_{2,2} = P_2 - P_3 + P_5 - P_7$

Strassen's Algorithm



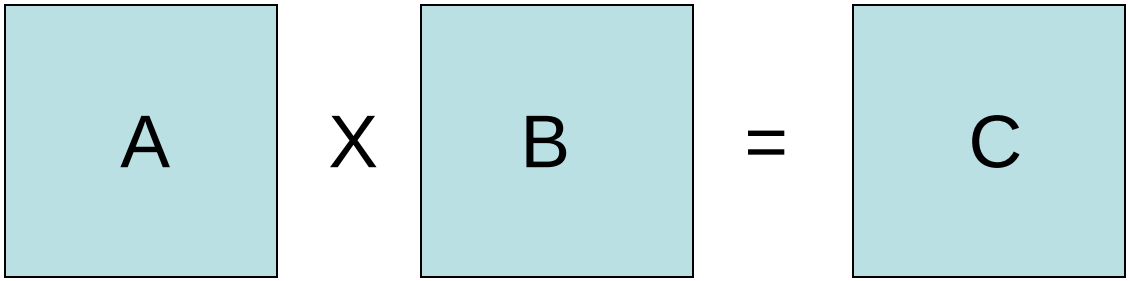
- linear combos of A, B entries
- 7 mults
- linear combos of results

T(n) = # operations to multiply n x n matrices:

$$T(n) \leq 7T(n/2) + O(n^2)$$

$$T(n) \leq O(n^{\log_2 7})$$

Introduction



- Standard method: $O(n^3)$ operations
- Strassen (1969): $O(n^{2.81})$ operations

The exponent of matrix multiplication:
smallest number ω such that for all $\varepsilon>0$
$O(n^{\omega + \varepsilon})$ operations suffice

History

• Standard algorithm	$\omega \leq 3$
• Strassen (1969)	$\omega < 2.81$
• Pan (1978)	$\omega < 2.79$
• Bini; Bini et al. (1979)	$\omega < 2.78$
• Schönhage (1981)	$\omega < 2.55$
• Pan; Romani; Coppersmith + Winograd (1981-1982)	$\omega < 2.50$
• Strassen (1987)	$\omega < 2.48$
• Coppersmith + Winograd (1987)	$\omega < 2.375$
• Stothers (2010)	$\omega < 2.3737$
• Williams (2011)	$\omega < 2.3729$
• Le Gall (2014)	$\omega < 2.37286$

Introduction

- Other important problems have same algorithmic complexity as matrix multiplication (via reductions)
  - determinants
  - LUP decompositions
  - matrix inversion
  - solving systems of linear equations
  - basic graph problems …

Outline

1. crash course on main ideas from Strassen 1969 through Le Gall 2014
2. conjectures implying ! = 2

Two more lectures:

II. group-theoretic approach

III. extending to coherent configurations

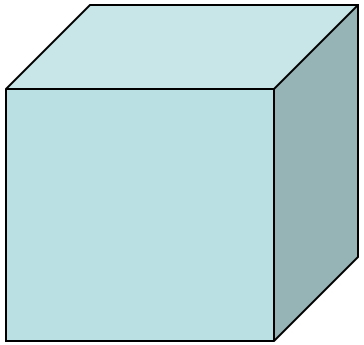
Bilinear algorithms & tensor rank

- Bilinear computation of complexity m:
  - m products of form (L.C. of A) x (L.C. of B)
  - result matrix entries = L.C.'s of these products
- equivalent: rank of matrix multiplication tensor "<n,n,n>" is at most m
- Strassen: "bilinear w.l.o.g."

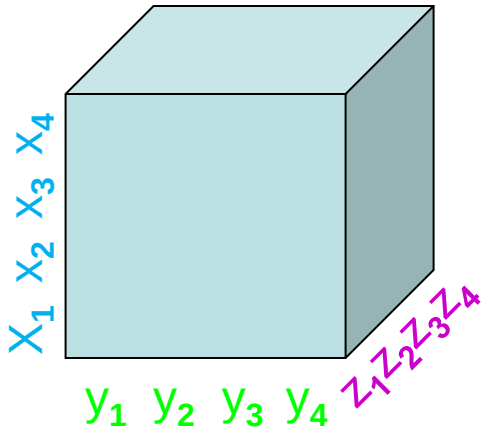
! = inf {¿ : n x n matmult in O(n¿) size}
= inf {¿ : rank(<n,n,n>) · O(n¿)}

Tensors

- tensor: 3-D array of complex numbers



- rank 1 tensor: (i,j,k) entry = $x_i y_j z_k$

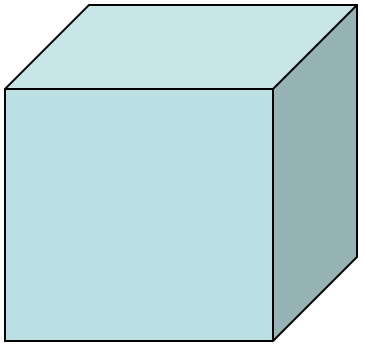


- rank r tensor: sum of r rank 1 tensors

equivalent: slices all L.C.'s of r rank 1 matrices

Tensors

- tensor: 3-D array of complex numbers



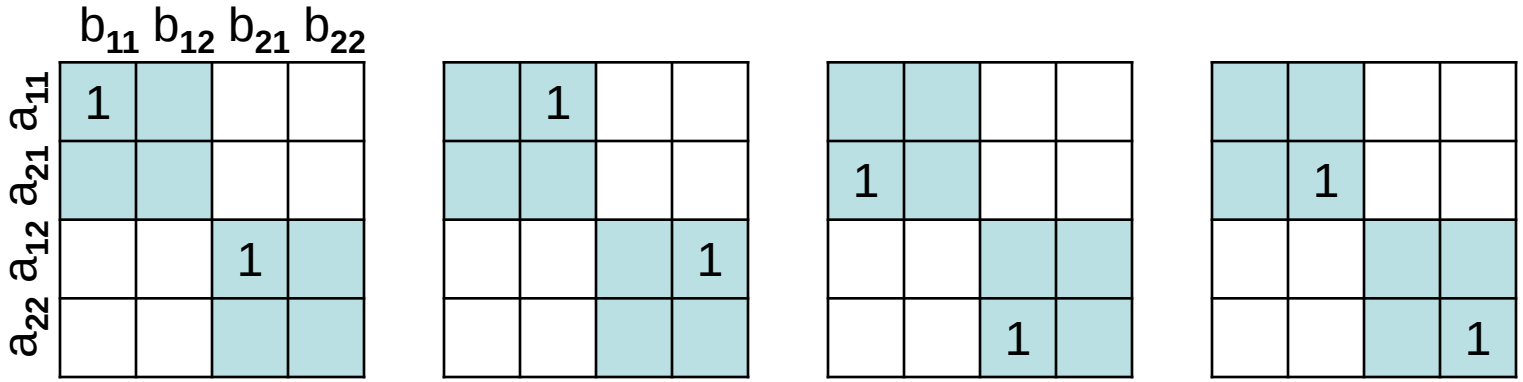
- tensor T with entries $T_{i,j,k}$
  - equivalent: trilinear form

$$\sum_{i,j,k} T_{i,j,k} X_i Y_j Z_k$$

The matrix multiplication tensor

<n,n,n> is a n^2 x n^2 x n^2 tensor described by trilinear form $\sum_{i,j,k} X_{i,j} Y_{j,k} Z_{k,i}$

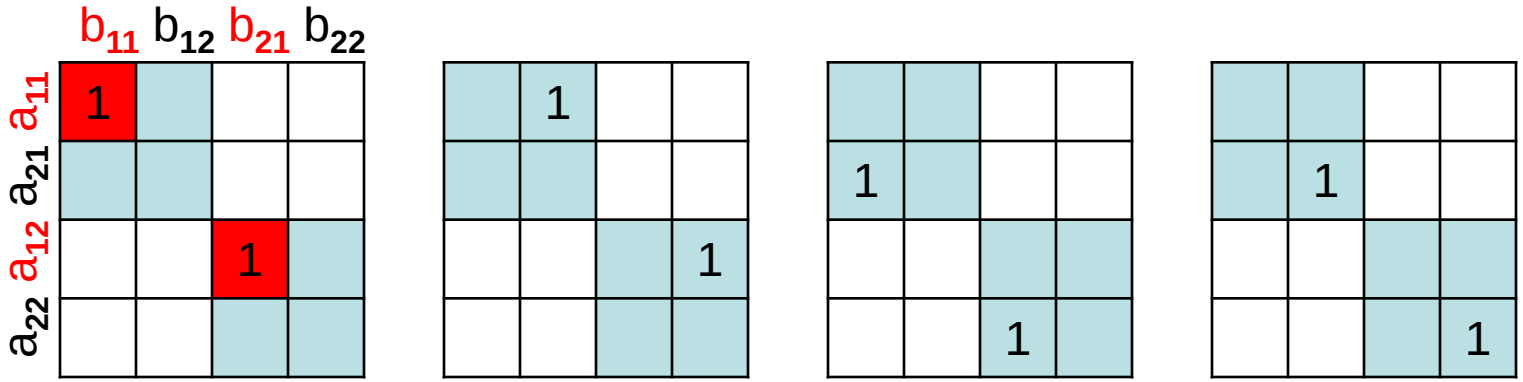
a_{11}	a_{12}	x	b_{11}	b_{12}	=	c_{11}	c_{12}
a_{21}	a_{22}		b_{21}	b_{22}		c_{21}	c_{22}



The matrix multiplication tensor

<n,n,n> is a $n^2 \times n^2 \times n^2$ tensor described by trilinear form $\sum_{i,j,k} X_{i,j} Y_{j,k} Z_{k,i}$

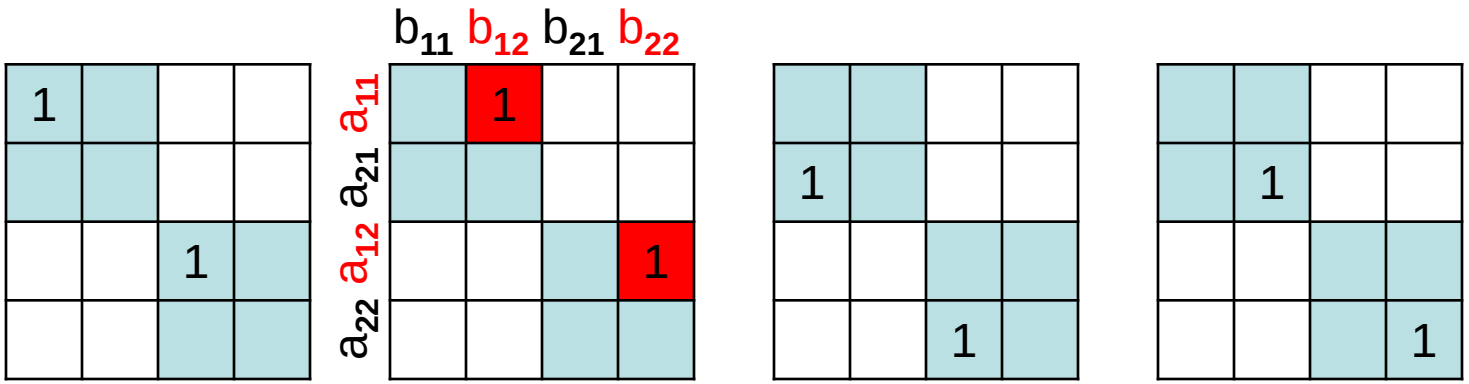
a_{11}	a_{12}	x	b_{11}	b_{12}	=	c_{11}	c_{12}
a_{21}	a_{22}		b_{21}	b_{22}		c_{21}	c_{22}



The matrix multiplication tensor

<n,n,n> is a n^2 x n^2 x n^2 tensor described by trilinear form $\sum_{i,j,k} X_{i,j} Y_{j,k} Z_{k,i}$

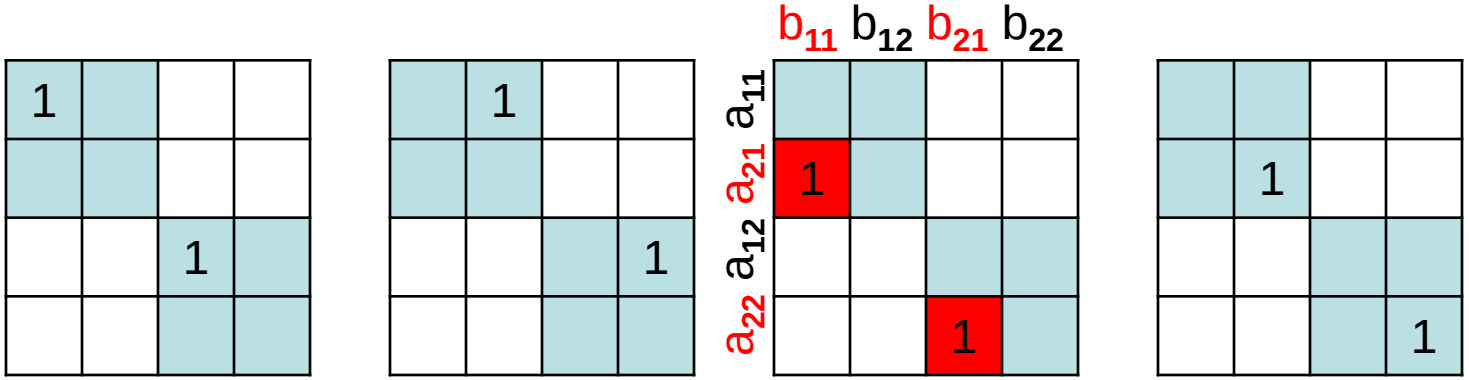
a_{11}	a_{12}	x	b_{11}	b_{12}	=	c_{11}	c_{12}
a_{21}	a_{22}		b_{21}	b_{22}		c_{21}	c_{22}



The matrix multiplication tensor

<n,n,n> is a n^2 x n^2 x n^2 tensor described by trilinear form $\sum_{i,j,k} X_{i,j} Y_{j,k} Z_{k,i}$

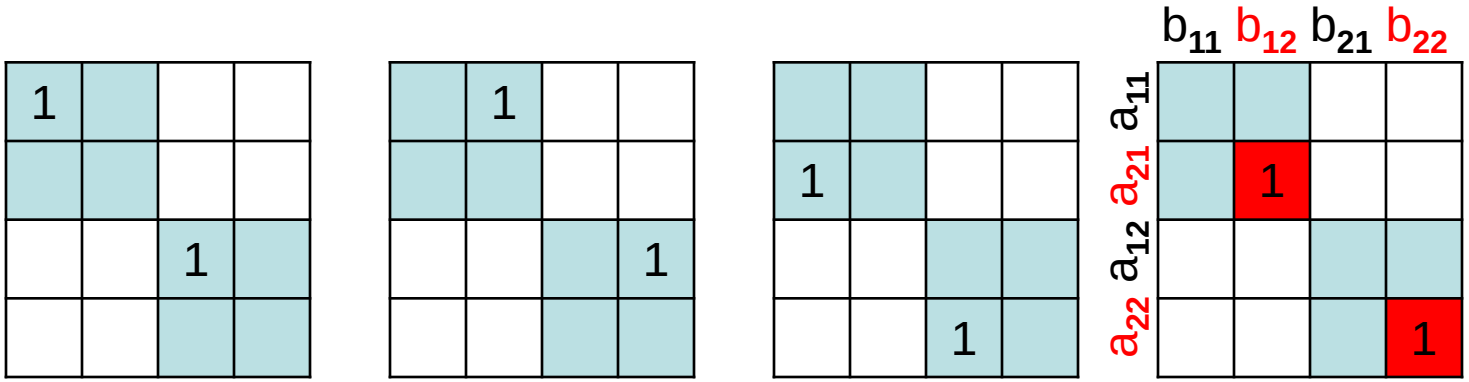
a_{11}	a_{12}	x	b_{11}	b_{12}	=	c_{11}	c_{12}
a_{21}	a_{22}		b_{21}	b_{22}		c_{21}	c_{22}



The matrix multiplication tensor

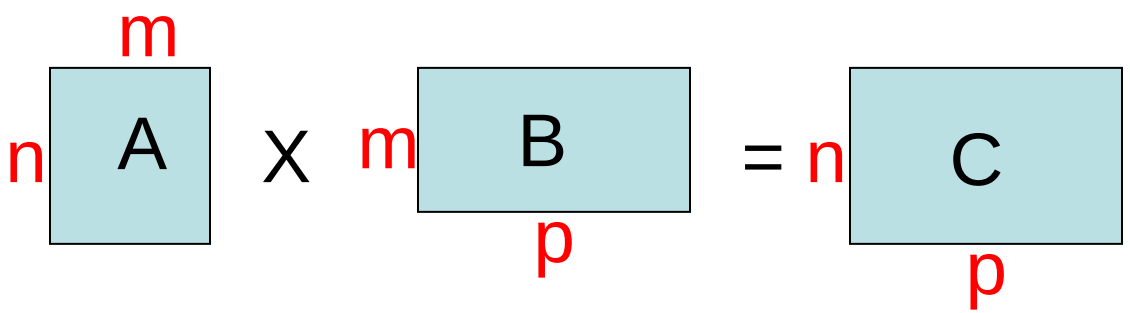
<n,n,n> is a n^2 x n^2 x n^2 tensor described by trilinear form $\sum_{i,j,k} X_{i,j} Y_{j,k} Z_{k,i}$

a_{11}	a_{12}	x	b_{11}	b_{12}	=	c_{11}	c_{12}
a_{21}	a_{22}		b_{21}	b_{22}		c_{21}	c_{22}

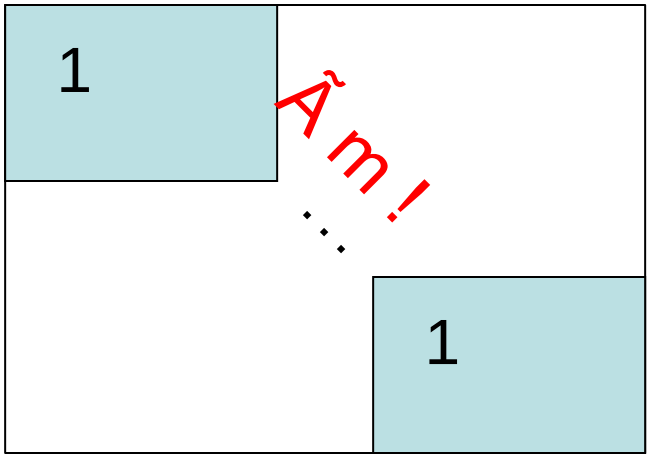


The matrix multiplication tensor

<n,m,p> is a nm £ mp £ pn tensor described by trilinear form $\sum_{i,j,k} X_{i,j} Y_{j,k} Z_{k,i}$



Each of np slices of <n,m,p>:



Strategies
for upper bounding the rank
of the
matrix multiplication tensor

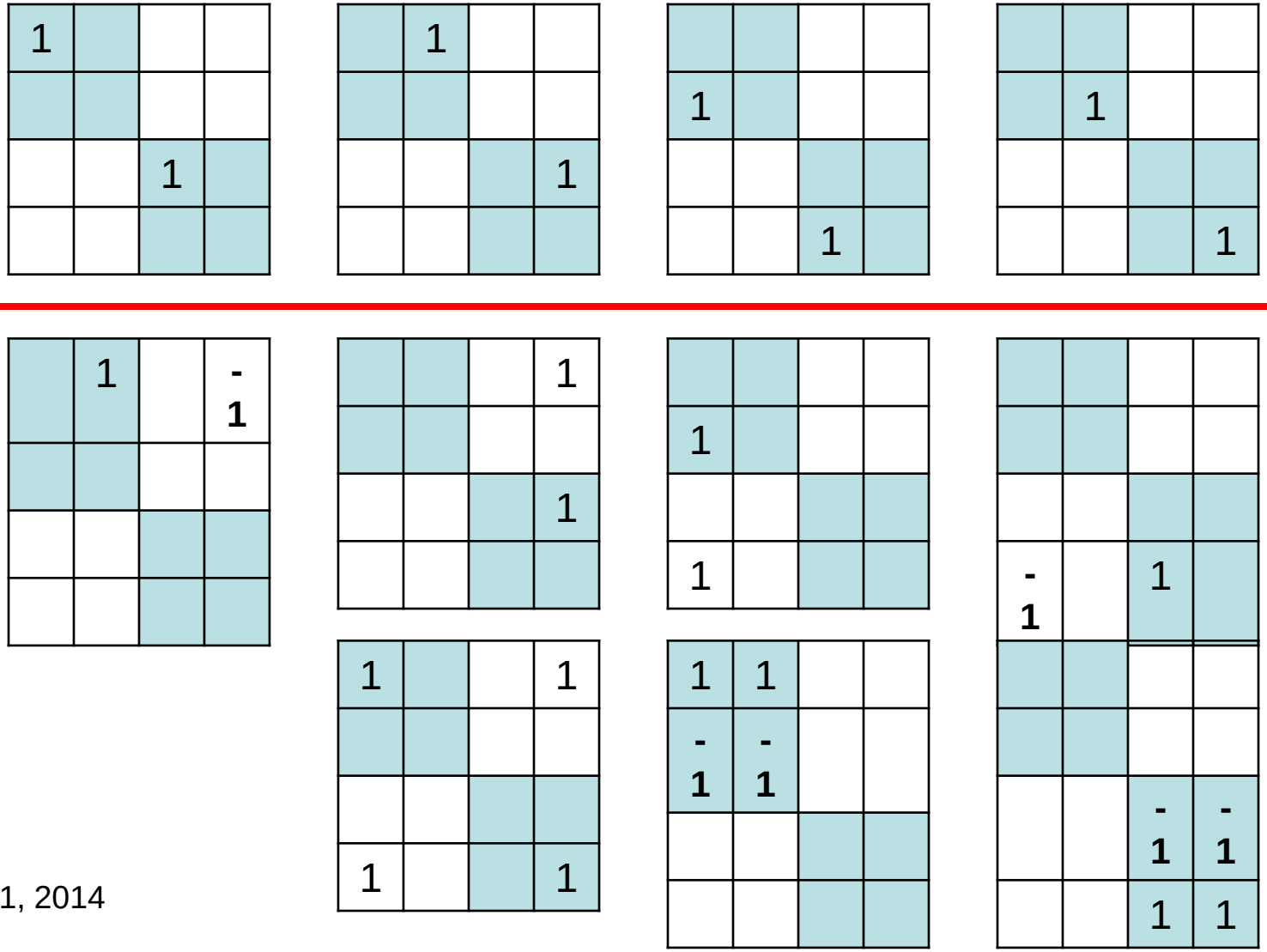
Upper bounds on rank

- Observation: $\langle n,n,n\rangle^i = \langle n^i, n^i, n^i\rangle$
) $R(\langle n^i, n^i, n^i\rangle) \cdot R(\langle n,n,n\rangle)^i$

- Strategy I: bound rank for small n by hand
  - $R(\langle 2,2,2\rangle) = 7$! < 2.81
  - $R(\langle 3,3,3\rangle)$ 2 [19..23] (worse bound)

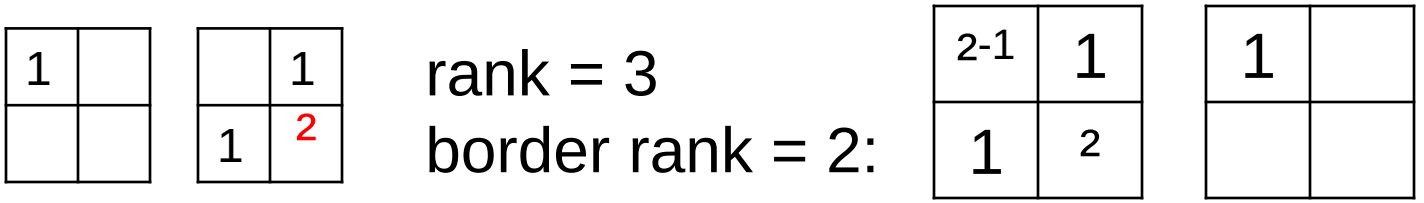
  - even computer search infeasible…

Strassen's example



Upper bounds on rank

- Border rank = rank of sequence of tensors approaching target tensor entrywise



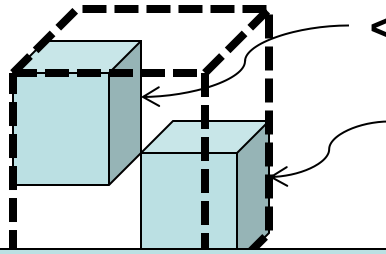
- Strategy II: bound *border rank* for small n
  - Lemma: $\underline{R}(\langle n,n,n\rangle) < r$) ! < $\log_n r$

    Idea: t-th tensor power is degree O(t) polynomial in ^{2}; interpolate to recover coefficient on ^{20}
  - $\underline{R}(\langle 2,2,3\rangle) \cdot 10$! < 2.79

Upper bounds on rank

- Direct sum of tensors
  <n,n,n> © <m,m,m>

  (multiple matrix multi



"Asymptotic Sum Inequality" and example (Schönhage 1981)

- Strategy III: bound (border) rank of *direct sums* of small matrix multiplication tensors

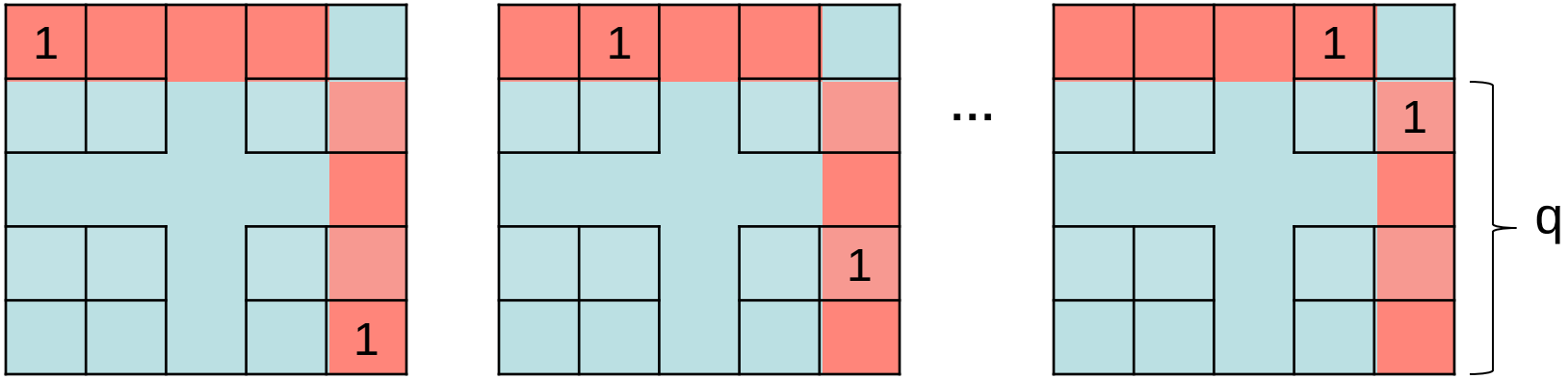
$$\underline{R}(\langle n_1,n_1,n_1\rangle \boxtimes \ldots \boxtimes \langle n_k,n_k,n_k\rangle) < r \;) \sum_i n_i^! < r$$

  – $\underline{R}(\langle 4,1,3\rangle \boxtimes \langle 1,6,1\rangle) \cdot 13 \quad ! < 2.55$

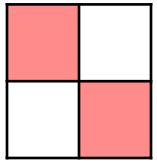
Upper bounds on rank

- Strategy IV: Strassen "laser method"
  - tensor with "course structure" of MM and "fine structure" components isomorphic to MM

    (many independent MMs in high tensor powers)



coarse structure
<1,2,1>

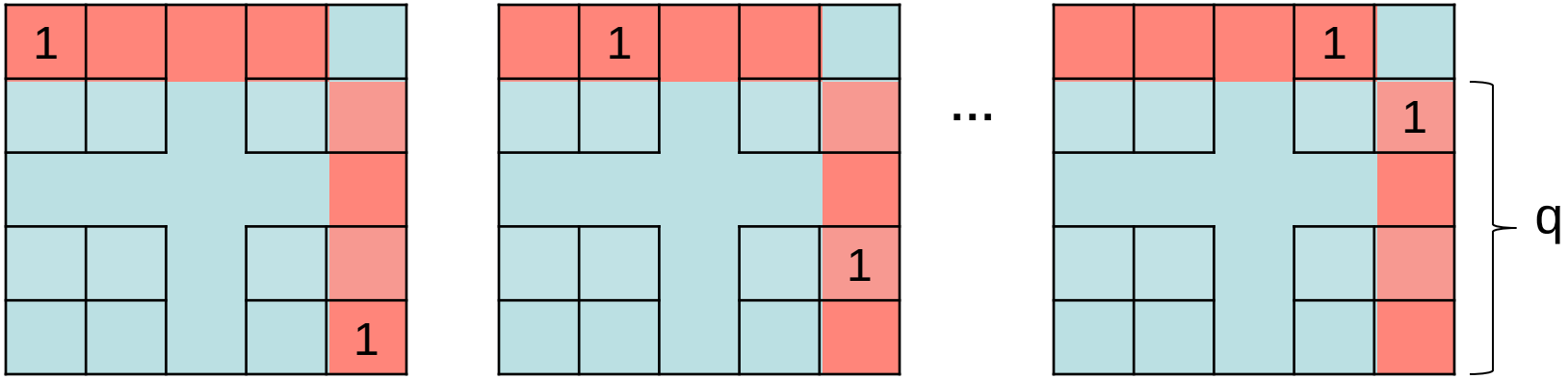


fine = scalar **x** row vector
col vector **x** scalar

Upper bounds on rank

- Strategy IV: Strassen “laser method”
  - tensor with “course structure” of MM and “fine structure” components isomorphic to MM

    (many independent MMs in high tensor powers)



border rank = q + 1; q = 5 yields ! < 2.48

Upper bounds on rank

- Strategy V : (C-W) border rank of this tensor is q+2:

$$\sum_{i=1\ldots q} X_0Y_iZ_i + X_iY_0Z_i + X_iY_iZ_0$$

  - zero-out variables leaving many independent MMs in high tensor power (sophisticated proof)
  - q = 6 yields ! < 2.41

Upper bounds on rank

- Strategy VI: (C-W) border rank of this tensor is q+2:

$$\sum_{i=1\ldots q} X_0Y_iZ_i + X_iY_0Z_i + X_iY_iZ_0 + X_0Y_0Z_{q+1} + X_0Y_{q+1}Z_0 + X_{q+1}Y_0Z_0$$

  - 6 “pieces”: target proportions in high tensor power affect # and size of independent MMs
  - optimize by hand
  - q = 6 yields ! < 2.388

Upper bounds on rank

- Strategy VII: border rank of *tensor powers* of T:

$$T = \sum_{i=1\ldots q} X_0 Y_i Z_i + X_i Y_0 Z_i + X_i Y_i Z_0 + X_0 Y_0 Z_{q+1} + X_0 Y_{q+1} Z_0 + X_{q+1} Y_0 Z_0$$

Tensor power	# pieces	bound	reference
2	36	2.375	C-W
4	1296	2.3737	Stothers
8	1679616	2.3729	Williams
16	2.82 x 10^12	2.3728640	Le Gall
32	7.95 x 10^24	2.3728639	Le Gall

Upper bounds on rank

- Strategy VII: border rank of *tensor powers*

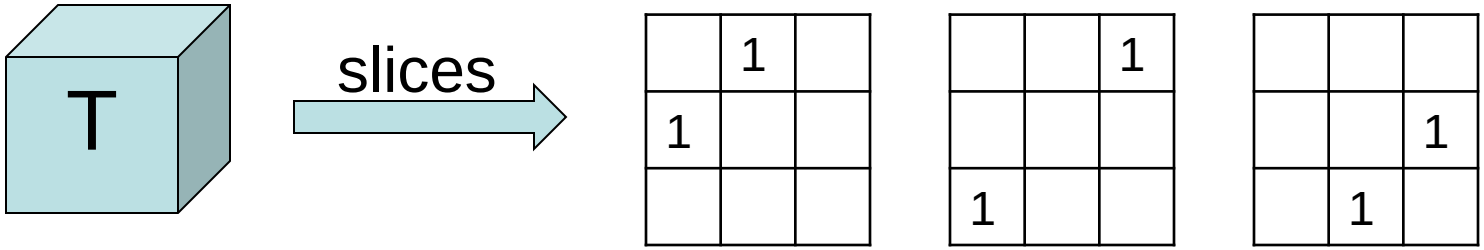
Tensor power	# pieces	bound	reference
2	36	2.375	C-W
4	1296	2.3737	Stothers
8	1679616	2.3729	Williams
16	2.82 x 10^12	2.3728640	Le Gall
32	7.95 x 10^24	2.3728639	Le Gall

- Ambainis-Filmus 2014: N-th tensor power cannot beat bound of 2.3078

Conjectures implying ! = 2

“Asymptotic Rank” conjecture [CW90]

$$T = \sum_{i=1,2} X_0 Y_i Z_i + X_i Y_0 Z_i + X_i Y_i Z_0$$



– border rank = 4

– *asymptotic* rank of T = $\lim_{n \to \infty} R(T^n)^{1/n}$

conjecture: asymptotic rank of T = 3

“no 3 disjoint equivoluminous subsets” conjecture [CW90]

one way to achieve asymptotic rank 3:

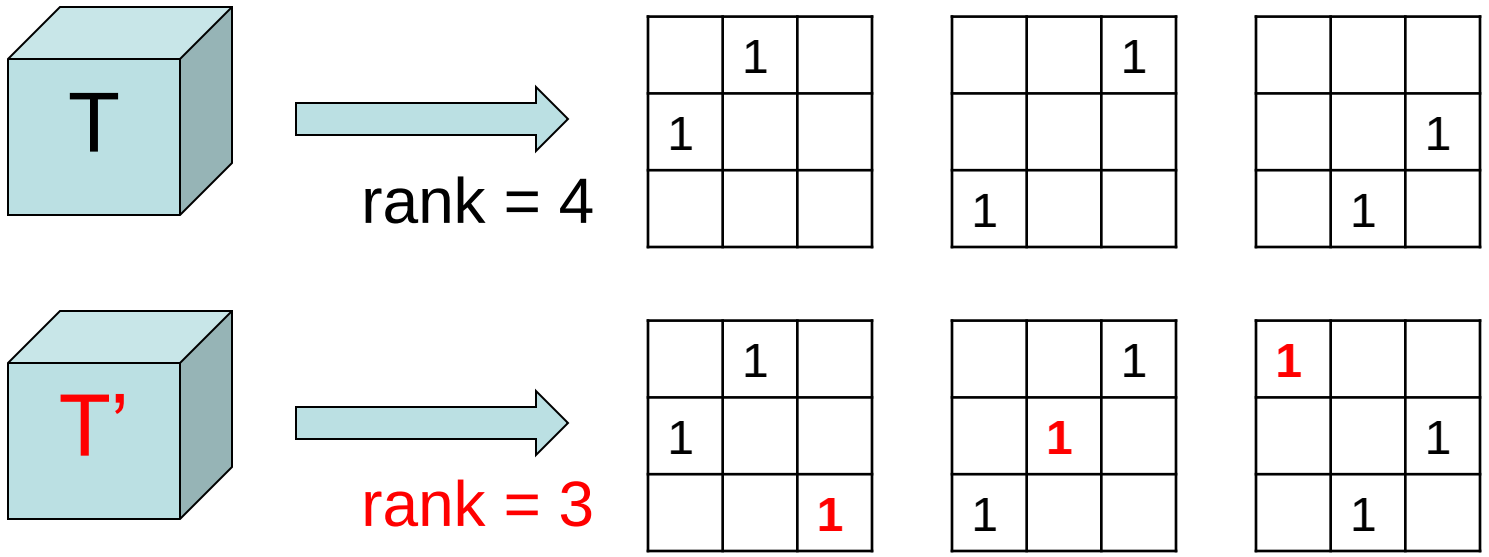
- $m_1, m_2, \ldots, m_n \in H$ (abelian group)
- for all disjoint S, T, U ⊆ [n] the sums

$$\sum_{i \in S} m_i \qquad \sum_{i \in U} m_i \qquad \sum_{i \in T} m_i$$

are not all equal

conjecture: can take $|H| \cdot 2^{o(n)}$

Strong Uniquely Solvable Puzzle Conjecture [CKSU05]



Strong Uniquely Solvable Puzzle:

gives a way to zero-out variables leaving many independent MMs in high tensor power of T' (instead of T)

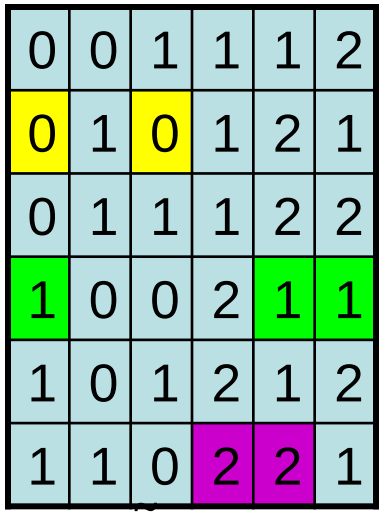
Strong Uniquely Solvable Puzzle Conjecture [CKSU05]

Uniquely Solvable Puzzle:

every unintended way of assembling pieces has overlap of 2 or 3 in some cell

Strong Uniquely Solvable Puzzle:

every unintended way of assembling pieces has overlap of exactly 2 in some cell



conjecture: Strong USPs exist with

$N = (w \text{ choose } w/3)^{1 - o(1)}$ rows

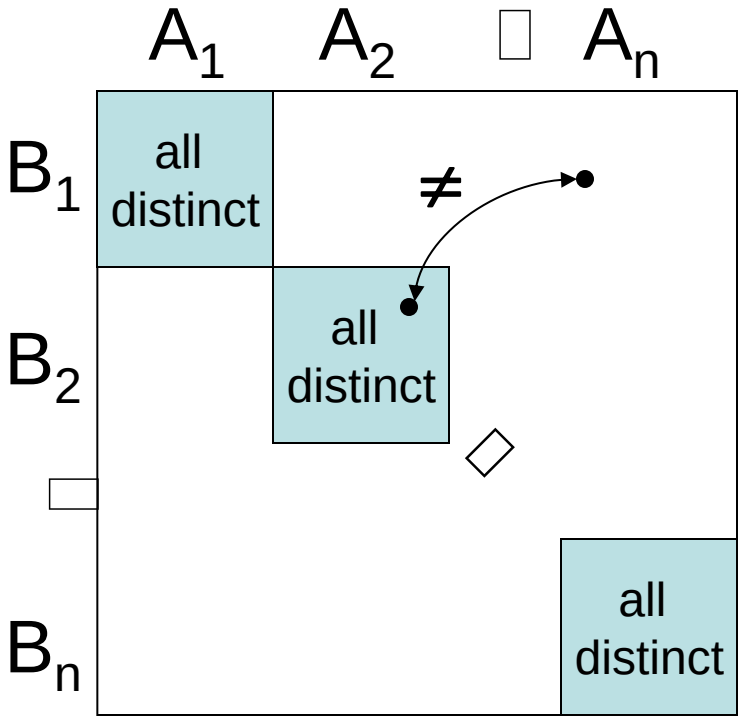
“Two Families” conjecture [CKSU05]

- subsets $A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_n$ of Abelian group H

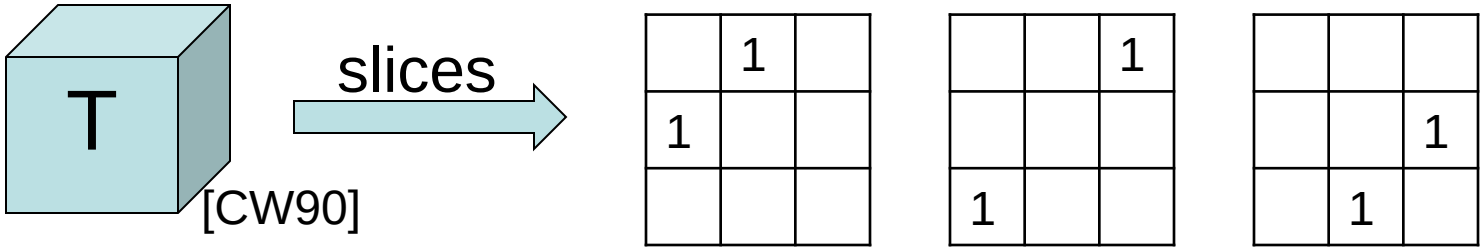
1. $|A_i + B_i| = |A_i| ¢ |B_i|$
2. $(A_i + B_i) Å (A_j + B_k) = ;$ if $j \neq k$

conjecture: can achieve

- $n = |H|^{1/2 - o(1)}$
- $|A_i| = |B_i| = |H|^{1/2 - o(1)}$



Status of conj's implying ! = 2



- asymptotic rank(T) = $R(T^n)^{1/n}$! 3?

- no 3 disjoint equivoluminous subsets [CW90]
- strong uniquely solvable puzzle [CKSU05]
- two families [CKSU05]

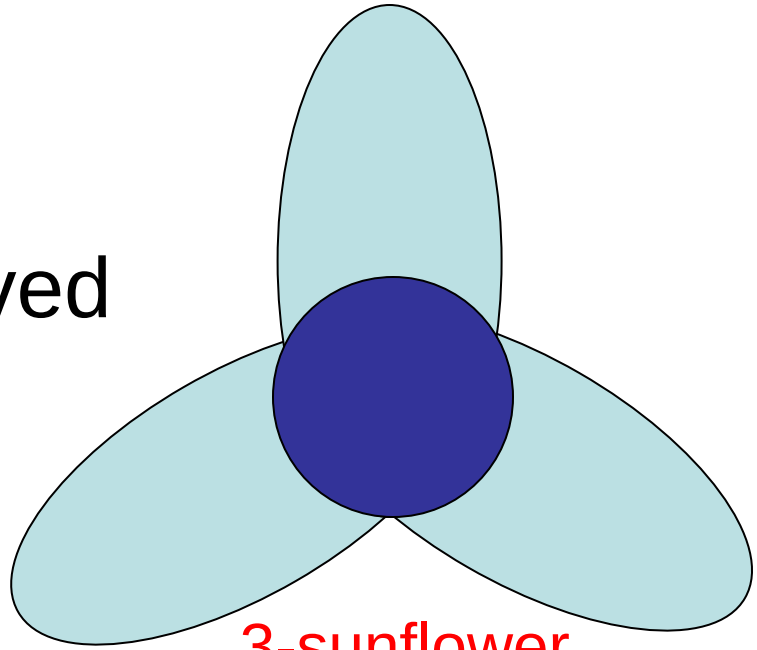
sunflower conj. #1 false

sunflower conj. #2 false

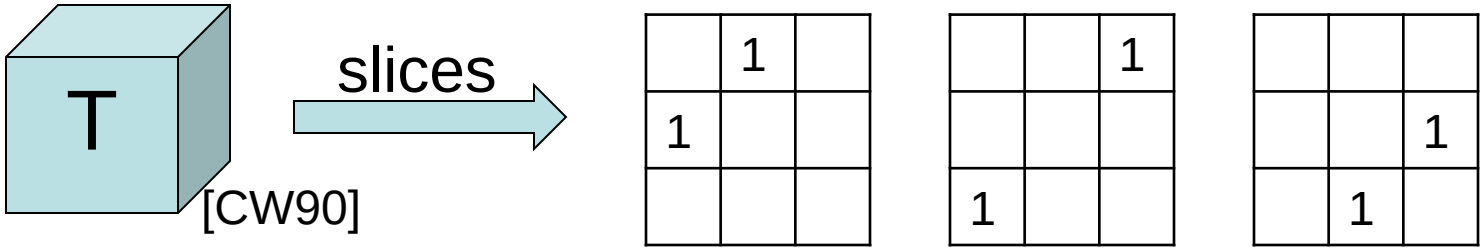
Sunflower conjecture #1

Erdos-Rado conjecture: every collection of const^s s-subsets contains a 3-sunflower

- $s!2^s$ known
- conjecture widely believed



Status of conj's implying ! = 2



- asymptotic rank(T) = $R(T^n)^{1/n}$! 3?
- no 3 disjoint equivoluminous subse[illegible]90]
- strong uniquely solvable puzzle [CKSU05]
- two families [CKSU05]

unlikely

sunflower conj. #1 false

sunflower conj. #2 false

Sunflower conjecture #2

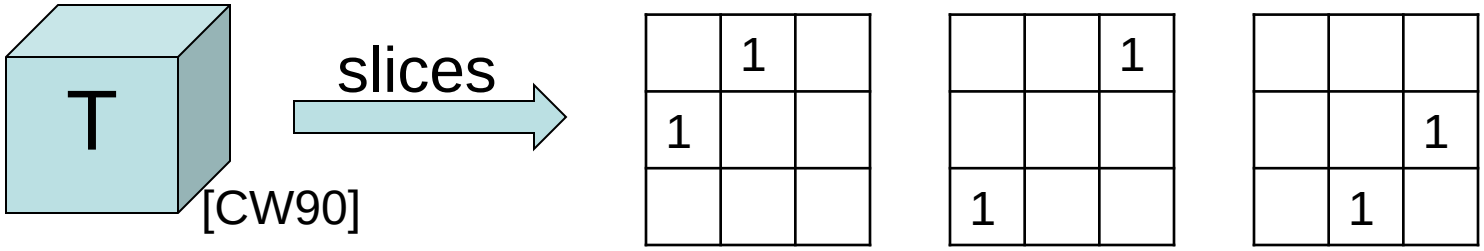
Every collection of $3^{n(1- \text{const})}$ vectors in Z_3^n contains u,v,w such that u + v + w = 0

- 50% chance of being false?

Every 3-colored collection* of $3^{n(1- \text{const})}$ vectors in Z_3^n contains u,v,w s.t. u+v+w=0

- 60% chance of being false?

Status of conj's implying ! = 2



- asymptotic rank(T) = $R(T^n)^{1/n}$! 3?
- no 3 disjoint equivolumi… bse… 90]
- strong uniquely solva… zle [CKSU05]
- two families [CKSU05]

Maybe unlikely

unlikely

sunflower conj. #1 false

sunflower conj. #2 false

A different approach

- So far...
  - bound border rank of small tensor (by hand)
  - asymptotic bound from high tensor powers
- Disadvantages
  - limited universe of “starting” tensors
  - high tensor powers hard to analyze

- Next: matrix multiplication via groups