III. Matrix multiplication using coherent configurations

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

Outline

Lecture I:

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

Lecture II:

- group-theoretic approach

This talk: extending to coherent configurations

But first…

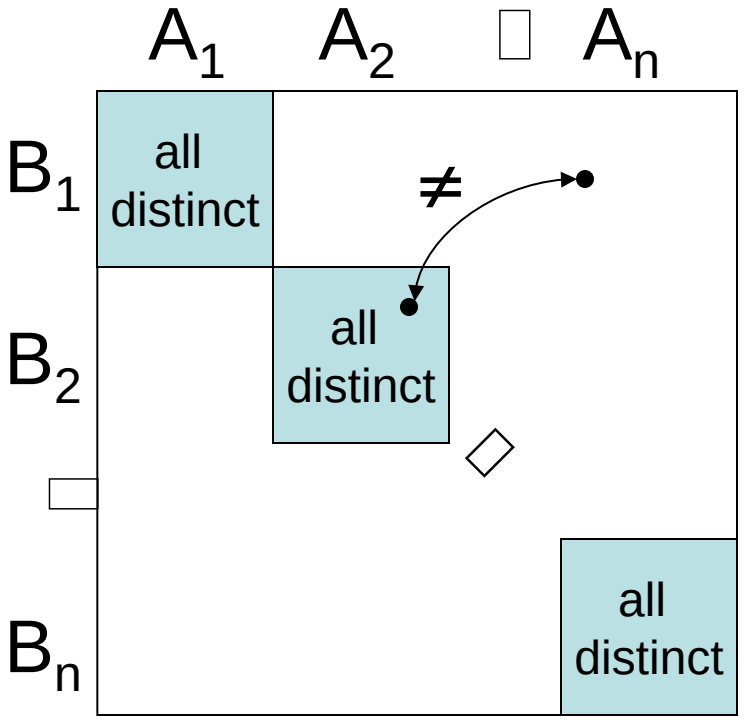
“near-solution” (?)
to the
two families conjecture

Two Families conjecture

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

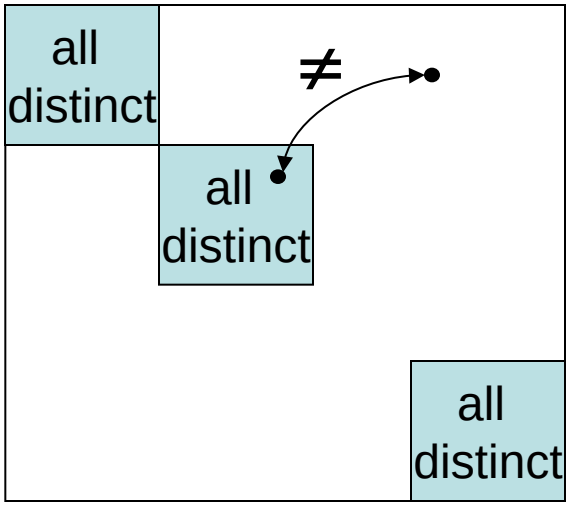
1. $|A_i + B_i| = |A_i| ¢ |B_i|$

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



Parameters

- **Optimal**:
  - $n = |H|^{1/2 - o(1)}$
  - $|A_i| = |B_i| = |H|^{1/2 - o(1)}$
  - yields $\omega = 2$



- Best construction we know:
  - $n = |H|^{0.3868\ldots}$
  - $|A_i| = |B_i| = |H|^{0.4491\ldots}$
  - yields $\omega \cdot 2.48\ldots$

Conjecture: exists optimal construction

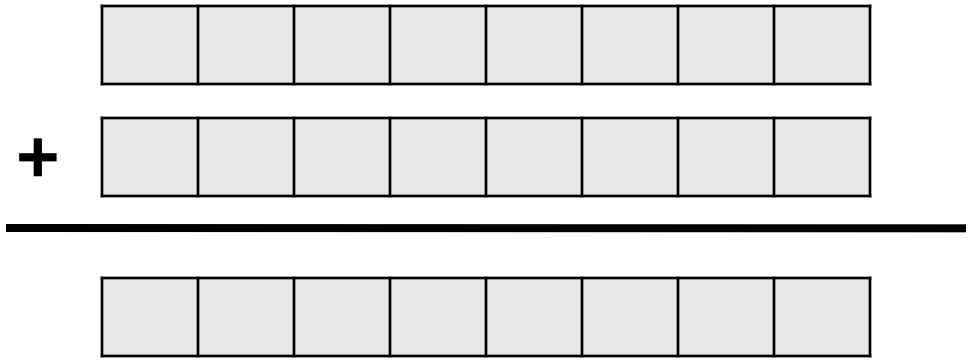
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



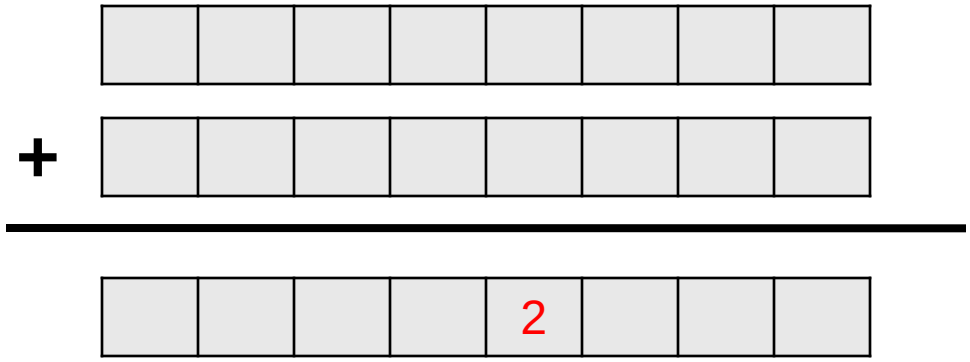
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



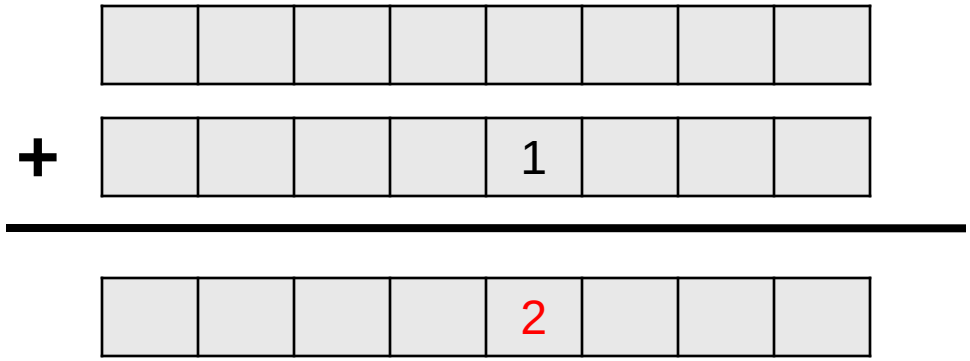
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



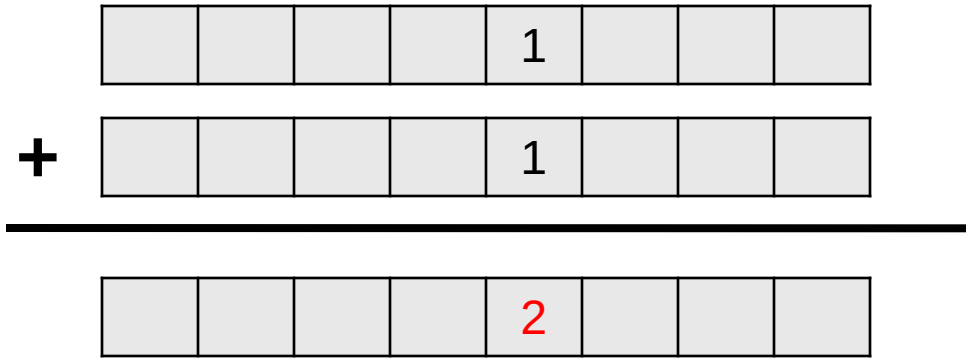
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



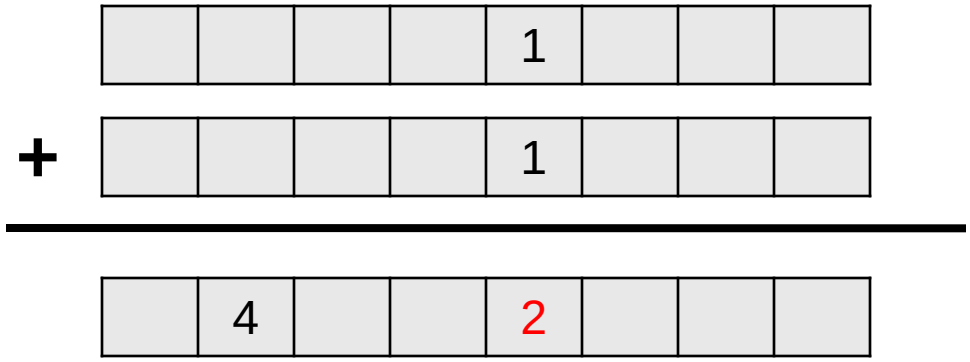
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



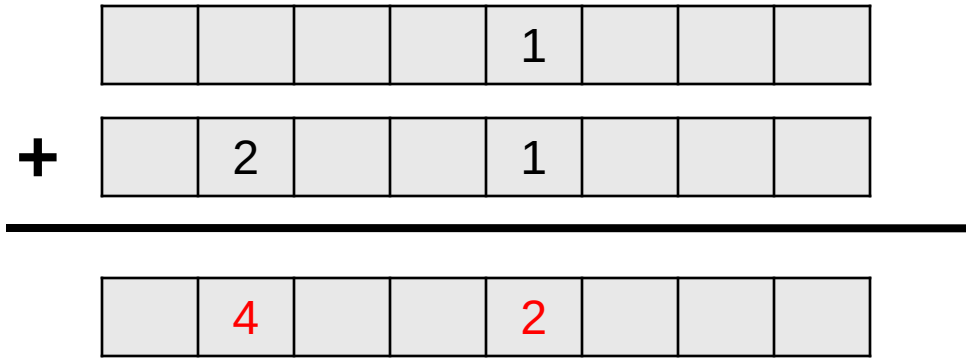
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



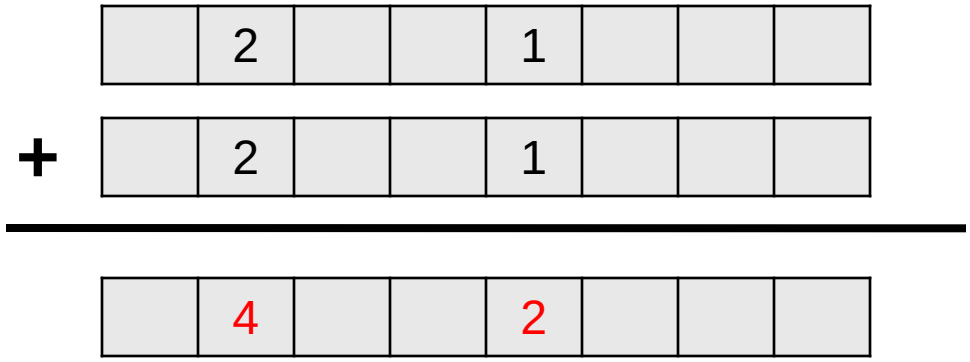
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



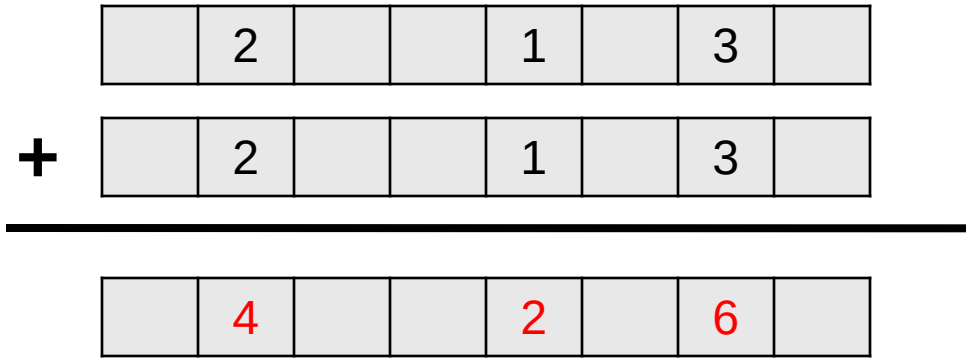
A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z



A “near-solution” (?)

- Warm up:

  subset S of abelian group $(Z_{10n})^n$:

  S = {all permutations of (1,2,3,…, n)}

Claim: x,y,z 2 S with x+y=2z) x = y = z

	4	2	5	8	1	7	3	6
+	4	2	5	8	1	7	3	6
	8	4	10	16	2	14	6	12

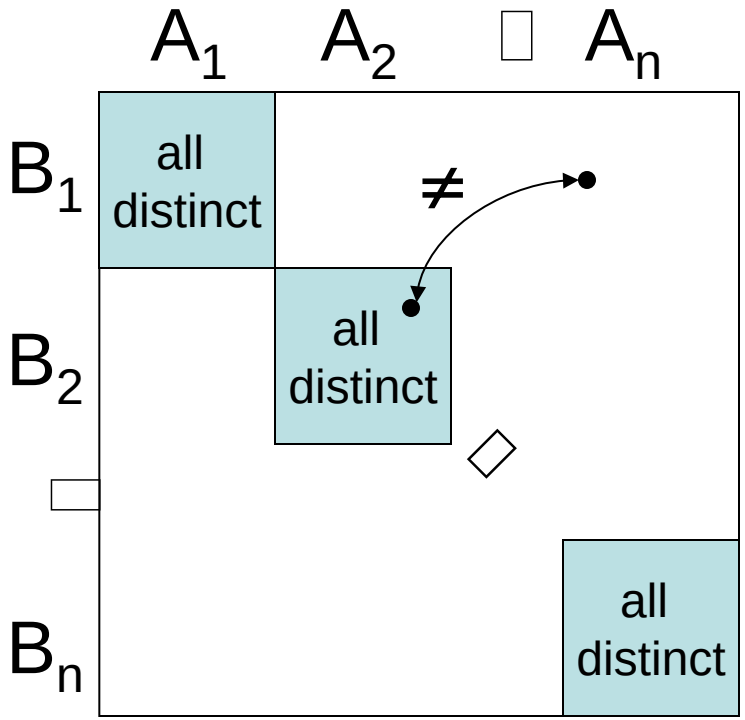
A “near-solution” (?)

$S=\{\text{permutations of } (1,2,3,\ldots, n)\} \mu H=(Z_{10n})^n$

for each ¼ 2 S_n:

$A_{¼} = \{(x, \ ¼x): x 2 S\}$

$B_{¼} = \{(y, -¼y): y 2 S\}$



Check sizes:

$|A_{¼}| = |B_{¼}| = \#\text{ pairs} = n!$

$|H^2| = ((10n)^n)^2 = (n!)^{2+o(1)}$

A “near-solution” (?)

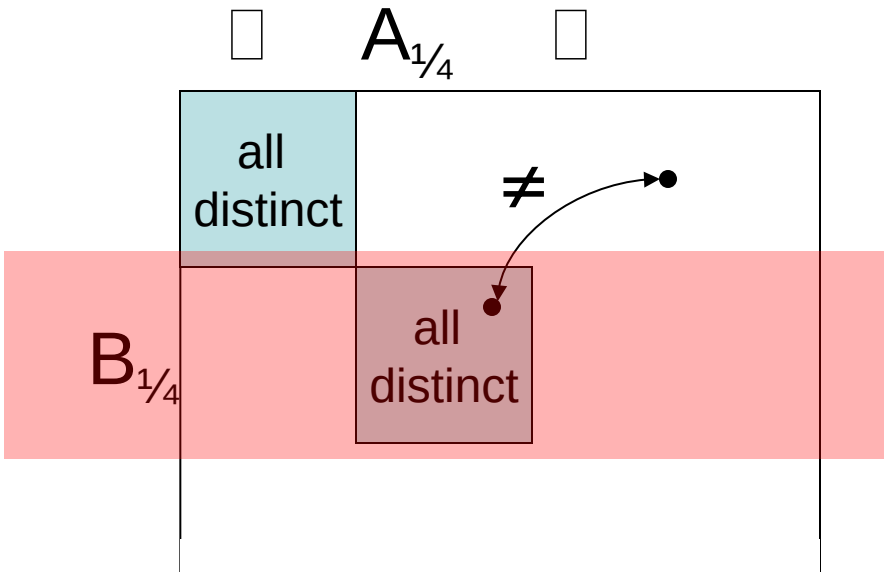
S={permutations of (1,2,3,…, n)} μ $H=(Z_{10n})^n$

for each ¼ 2 S_n:

$A_{¼}$ = {(x, ¼x): x 2 S}

$B_{¼}$ = {(y, -¼y): y 2 S}

	x	¼x
+	y	-¼y
	x+y	¼(x-y)



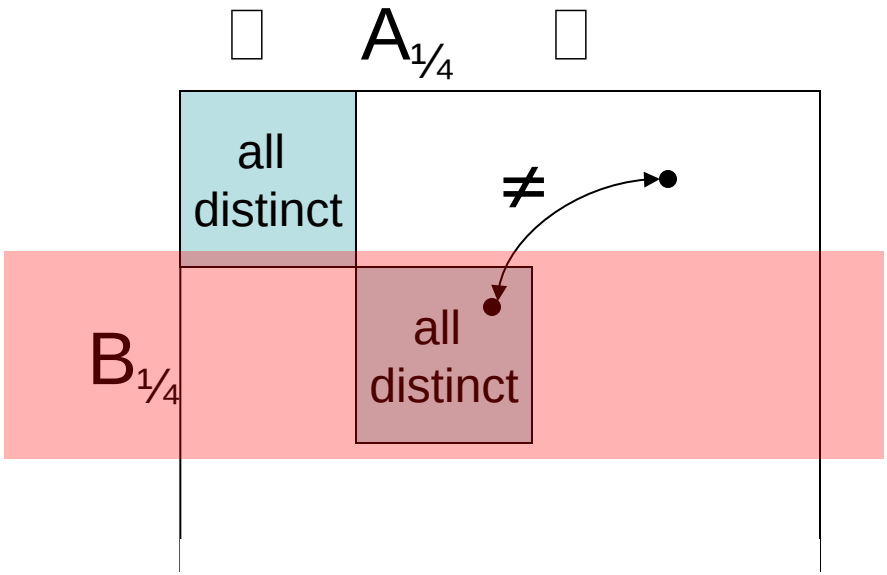
A “near-solution” (?)

$S=\{\text{permutations of } (1,2,3,\ldots,n)\} \mu H=(Z_{10n})^n$

for each ¼ 2 S_n:

$A_{¼} = \{(x, \ ¼x): x \in S\}$

$B_{¼} = \{(y, -¼y): y \in S\}$



	x	¼x
+	y	-¼y
	x+y	¼(x-y)

multiply 2nd coord by ¼-1

x+y	x-y

and add coords

2x

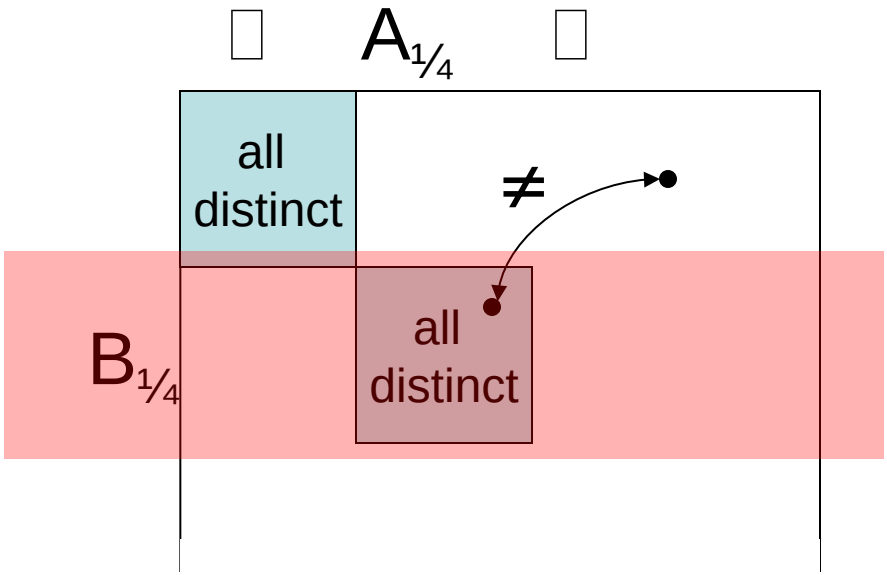
A “near-solution” (?)

$S=\{\text{permutations of } (1,2,3,\ldots,n)\}$ μ $H=(Z_{10n})^n$

for each ¼ 2 S_n:

$A_{¼} = \{(x,\ ¼x) : x \in S\}$

$B_{¼} = \{(y,\ -¼y) : y \in S\}$



	x	½x
+	y	-¼y
	x+y	½x - ¼y

multiply 2nd coord by ¼-1

x+y	¼-1½x - y

and add coords

x+ ¼-1½x

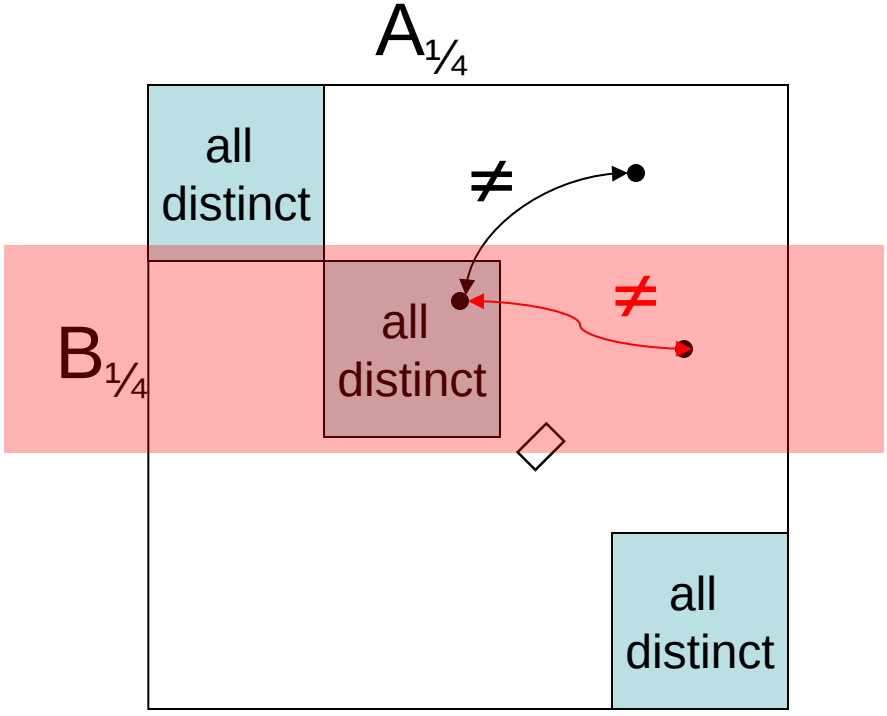
A “near-solution” (?)

S={permutations of (1,2,3,…, n)} μ $H=(Z_{10n})^n$

for each ¼ 2 S_n:

$A_{¼} = \{(x, \ ¼x): x \in S\}$

$B_{¼} = \{(y, -¼y): y \in S\}$



Conclude:

satisfy properties

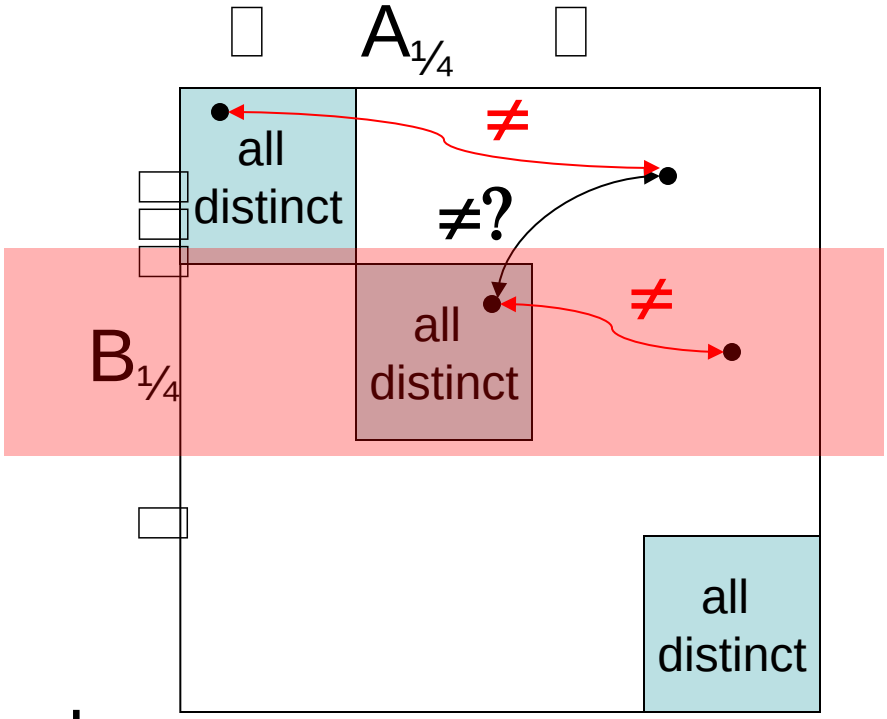
within each row…

A “near-solution” (?)

Conclude:

satisfy properties

within each row…



But…

most diagonal elements

appear in *most* diagonal blocks

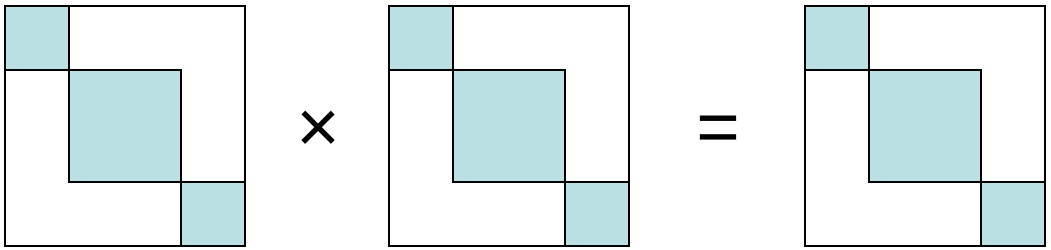
) *most* elements satisfy properties within *most* rows

(“most” = 1 - ² fraction)

matrix multiplication via coherent configurations

The general approach

- Cohn-Umans 2003, 2012:
  - *embed* n x n matrix multiplication into semi-simple algebra multiplication
  - semi-simple: isomorphic to block-diagonal MM



commutative, diagonal

  - key hope: “nice basis” w/ combinatorial structure
  - reduce n x n MM to smaller MMs; recurse

Example semi-simple algebra

- finite group G
- the group algebra C[G] has elements

$$\Sigma_g a_g g$$

with multiplication

$$(\Sigma_g a_g g)(\Sigma_h b_h h) = \Sigma_f (\Sigma_{gh = f} a_g b_h) f$$

- group elements are “nice basis”

The embedding:

$Q(S) = \{s^{-1}t: s, t \in S\}$

Subsets X, Y, Z of G satisfy the

triple product property

if for all $x \in Q(X)$, $y \in Q(Y)$, $z \in Q(Z)$:

$xyz = 1$ iff $x = y = z = 1$.

$\underline{\mathbf{A}} = \Sigma a_{x,y} (x\ y^{-1})$ $\underline{\mathbf{B}} = \Sigma b_{y,z} (y\ z^{-1})$

Claim: $(AB)_{x,z}$ = coeff. on $(x\ z^{-1})$ in **A***B.

Example semi-simple algebra

$$C[G] \,' (C^{d_1 \times d_1}) \times (C^{d_2 \times d_2}) \times \ldots \times (C^{d_k \times d_k})$$

- $d_1, d_2, \ldots, d_k$ are character degrees of G

Theorem: in group G realizing n£n matrix multiplication, with character degrees $d_1, d_2, d_3, \ldots$, we obtain:

$$R(<n,n,n>) \cdot \sum_i d_i^{\omega}$$

Example semi-simple algebra

$$C[G] \text{ ' } (C^{d_1 \times d_1}) \times (C^{d_2 \times d_2}) \times \ldots \times (C^{d_k \times d_k})$$

- $d_1, d_2, \ldots, d_k$ are character degrees of G

Theorem: in group G realizing n£n matrix multiplication, with character degrees $d_1, d_2, d_3, \ldots$, we obtain:

$$R(<n,n,n>) \cdot \sum_i d_i^{\omega} \cdot d_{max}^{\omega-2} ¢ |G|$$

goals: $|G| ¼ n^2$ *and* small d_{max}

Group algebra approach

- [CKSU 2005] wreath product groups yield :
  - $\omega < 2.48$, $\omega < 2.41$
  - key part of construction is combinatorial
  - two conjectures implying $\omega = 2$

- Main disadvantage:
  - non-trivial results *require* non-abelian groups
  - most ideas foiled by too-large char. degrees

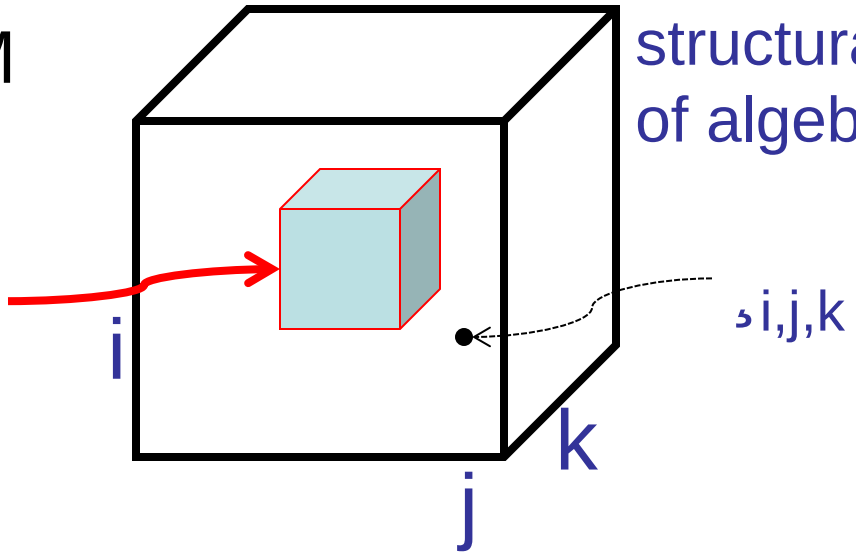
General semi-simple algebras

- (finite dimensional, complex) algebra specified by
  - "nice basis" $e_1, e_2, \ldots, e_r$
  - structure constants $c_{i,j,k}$ satisfying

$$e_i\, e_j = \sum_k c_{i,j,k}\, e_k$$

"realizes" MM if contains*:

MM tensor <n,n,n>



Weighted vs. unweighted MM

- Technical problem:
  - MM tensor <n,n,n> given by $\sum_{i,j,k} X_{i,j} Y_{j,k} Z_{k,i}$
  - embedding into algebra yields tensor given by

$$\sum_{i,j,k} \delta_{i,j,k} X_{i,j} Y_{j,k} Z_{k,l}$$
$$(\text{with } \delta_{i,j,k} \neq 0)$$

  - group algebra: $\delta_{i,j,k}$ always 0 or 1

Weighted vs. unweighted MM

s-rank of tensor T: minimum rank of tensor with same ***s****upport* as T

Does upper bound on s-rank of MM tensor imply upper bound on ordinary rank?

Example:

a_{11}	a_{12}
a_{21}	a_{22}

x

b_{11}	b_{12}
b_{21}	b_{22}

=

$a_{11}b_{11} + a_{12}b_{21}$	$a_{11}b_{11} + a_{12}b_{21}$
$a_{11}b_{11} + a_{12}b_{21}$	$a_{11}b_{11} + a_{12}b_{21}$

Weighted vs. unweighted MM

s-rank of tensor T: minimum rank of tensor with same ***s****upport* as T

Does upper bound on s-rank of MM tensor imply upper bound on ordinary rank?

Example:

a_{11}	a_{12}
a_{21}	a_{22}

x

b_{11}	b_{12}
b_{21}	b_{22}

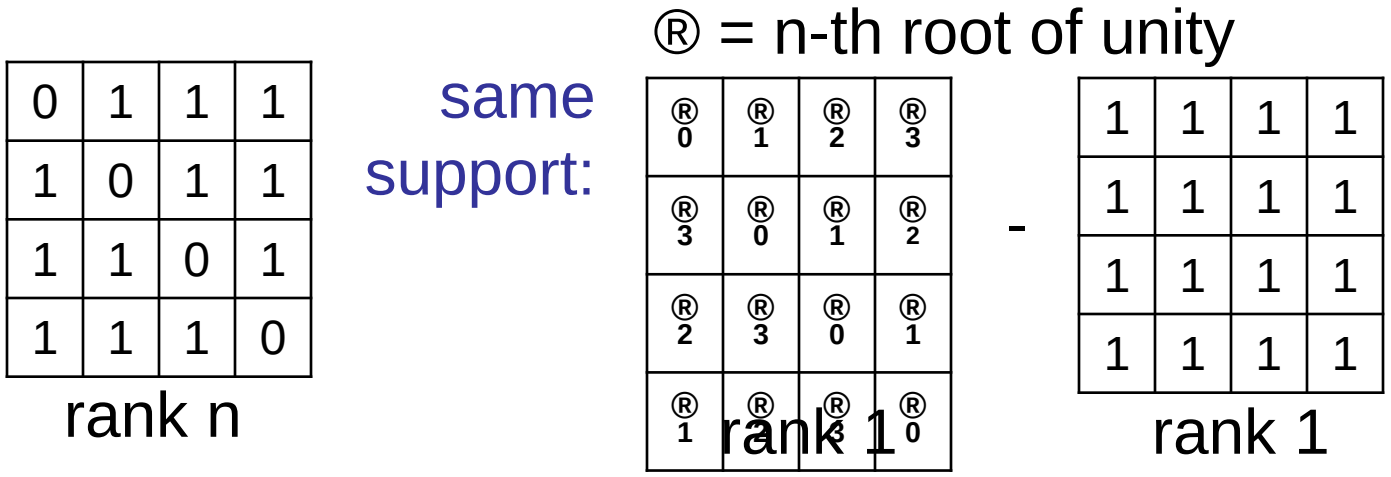
!

$a_{11}b_{11} + a_{12}b_{21}$	$a_{11}b_{11} + a_{12}b_{21}$
$a_{11}b_{11} + a_{12}b_{21}$	$a_{11}b_{11} +$ **2¢**$a_{12}b_{21}$

does it help if can compute this in 6 multiplications?

Weighted vs. unweighted MM

- s-rank can be much smaller than rank:



maybe it's easy to show s-rank of $n \times n$ matrix multiplication is n^2 (!!)

Weighted vs. unweighted MM

$$\omega = \inf \{\alpha : \text{rank}(\langle n,n,n \rangle) \cdot O(n^{\alpha})\}$$

$$\omega_s = \inf\{\alpha : \text{s-rank}(\langle n,n,n \rangle) \cdot O(n^{\alpha})\}$$

Theorem: $\omega \cdot (3\omega_s - 2)/2$

in particular, $\omega_s \cdot 2 + ^{2}) \omega \cdot 2 + (3/2)^2$

- Proof idea:
  - find ¼ n^2 copies of <n,n,n> in 3rd tensor power
  - when broken up this way, can rescale

A promising family of semisimple algebras

Coherent configurations

"group theory without groups"

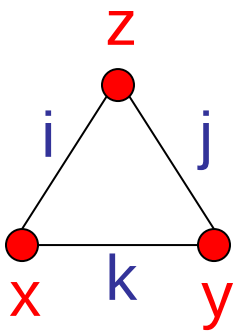
- points X, partition $R_1, R_2, \ldots, R_r$ of X^2
  - diagonal $\{(x,x) : x \in X\}$ is union of some classes
  - for each i, there is i^*

    $R_{i^*} = \{(y,x) : (x,y) \in R_i\}$
  - exist integers $p_{i,j}^{\ k}$ such that for all $(x,y) \in R_k$:

    $\#\{z: (x,z) \in R_i \text{ and } (z,y) \in R_j\} = p_{i,j}^{\ k}$

if *one* class: "association scheme"

$p_{i,j}^{\ k} = p_{j,i}^{\ k}$: commutative



Coherent configs: examples

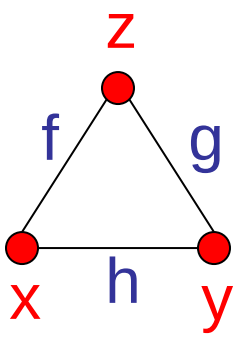
- Hamming scheme:
  - points 0/1 vectors
  - classes determined by hamming distance

- distance-regular graph:
  - points = vertices
  - classes determined by distance in graph metric

Coherent configs: examples

- scheme based on finite group G
  - set X = finite group G
  - classes $R_g = \{(x,xg) : x \in X\}$

$p_{f,g}^{h} = 1$ if $fg=h$, 0 otherwise



- "Schurian":
  - group G acts on set X
  - classes = orbits of (diagonal) G-action on X^2

Coherent configs: examples

- "Schurian":
  - group G acts on set X
  - classes = orbits of (diagonal) G-action on X^2
- one Schurian scheme: "group scheme"
  - group G x G acts on G via $(g,h)¢x = gxh^{-1}$
  - orbits all of the form $\{(x,y): xy^{-1} 2 C_i\}$ for conjugacy class C_i
  - always commutative!

Adjacency algebra

CC: points X, partition $R_1, R_2, \ldots, R_r$ of X^2

- for each class R_i, matrix A_i with

$$A_i[x,y] = 1 \text{ iff } (x,y) \; 2 \; R_i$$

- 3 CC axioms)

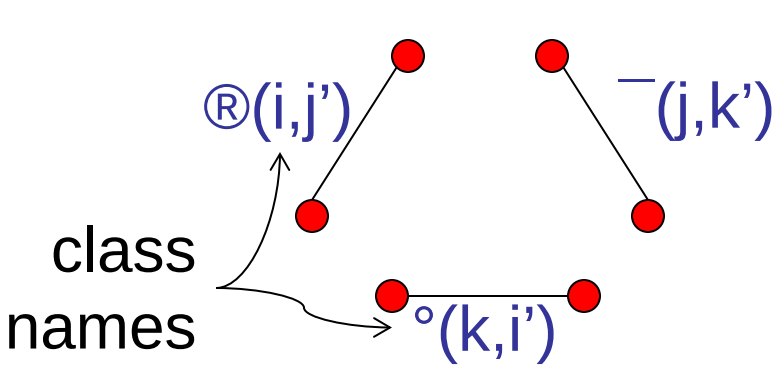
$\{A_i\}$ generate a semisimple algebra

  - e.g., 3rd axiom implies $A_i A_j = \sum_k p_{ij}^k A_k$
  - if the CC based on group G, algebra is C[G]

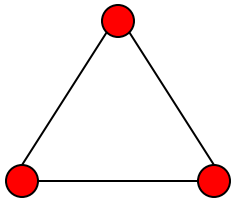
Nice basis conditions

- group algebra C[G]: “nice basis” yields triple product property

- adjacency algebras of CCs: “nice basis” yields triangle condition:



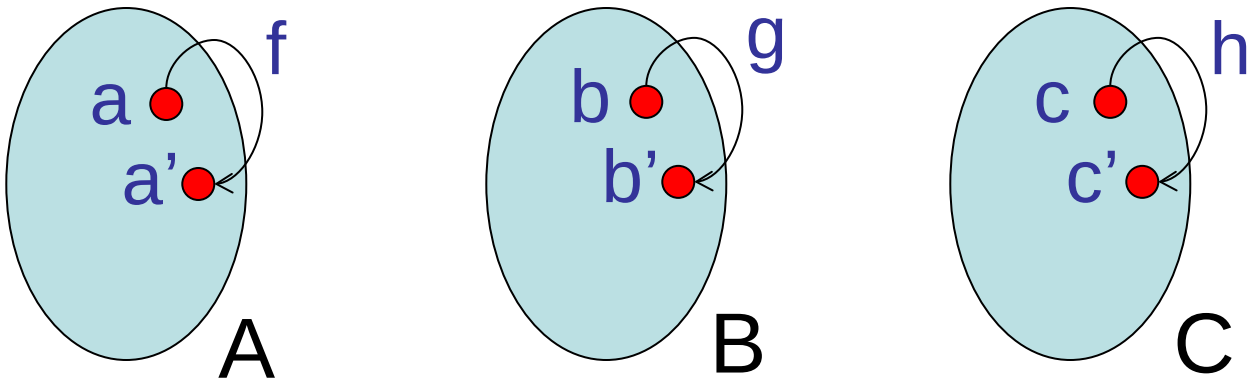
can look like



iff i = i’, j = j’, k = k’

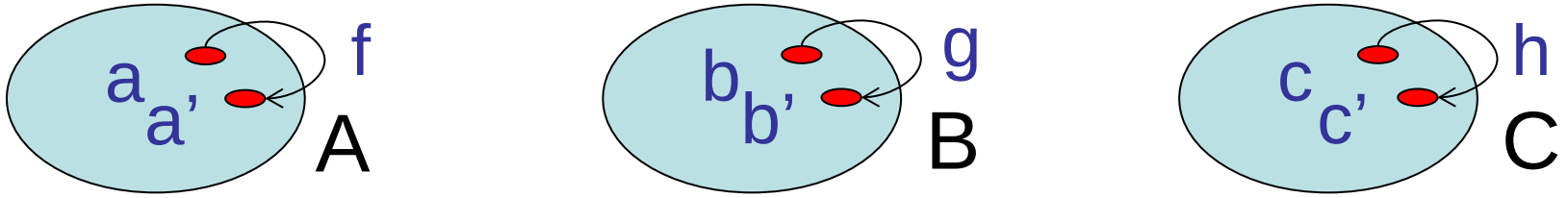
Nice basis conditions

- Schurian CCs: “nice basis” yields
  - group G acts on set X
  - subsets A,B,C of X realize <|A|, |B|, |C|> if:



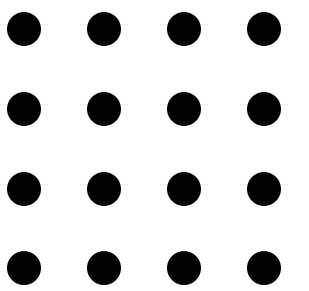
fgh = 1 implies a = a’, b = b’, c = c’

example in a Schurian CC

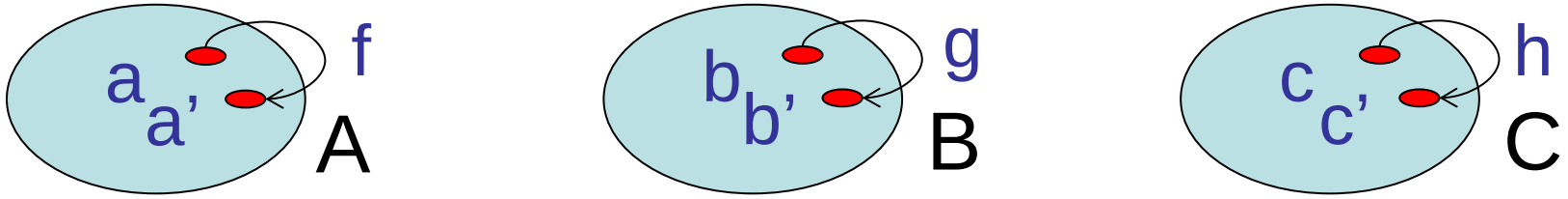


fgh = 1 implies a = a', b = b', c = c'

- S_n acts on $X = (Cyc_m)^n$
  - partition Cyc_m into $n^{1/2}$ parts
  - A = {v : part i used on row i}
  - B = {v : part i used on column i}
  - C = {v : part i used on diagonal i}

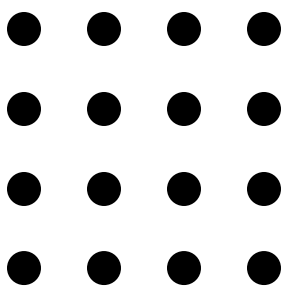


example in a Schurian CC



fgh = 1 implies a = a', b = b', c = c'

- S_n acts on $X = (Cyc_m)^n$
  - partition Cyc_m into $n^{1/2}$ parts
  - A = {v : part i used on row i}
  - B = {v : part i used on column i}
  - C = {v : part i used on diagonal i}



n points

orbits ¼ $(m^n)^2/n!$ ¼ $(m^2/n)^n$ $|A| = (m/n^{1/2})^n$

Coherent configs vs. groups

Generalization for generalization's sake?

- recall group framework:
  - non-commutative necessary

Theorem: in group G realizing n£n matrix multiplication, with character degrees $d_1, d_2, d_3, \ldots$, we obtain:

$$R(\langle n,n,n\rangle) \cdot \sum_i d_i^{\omega} \cdot d_{max}^{\omega-2} \text{¢} |G|$$

goals: $|G|$ ¼ n^2 *and* small d_{max}

Coherent configs vs. groups

Generalization for generalization's sake?

- coherent configuration framework:
  - commutative suffices!
  - combinatorial constructions from old setting yield
    $!_s < 2.48$, $!_s < 2.41$
  - conjectures from old setting (if true) would imply $!_s = 2$

in commutative Schurian CC's

even group schemes

even symmetric

Proof idea

we prove a general transformation:

if can realize several independent matrix multiplications in CC…

- can do this in abelian groups
- conjectures: can "pack optimally"

… then high symmetric power of CC realizes *single* matrix multiplication

– reproves Schönhage's Asymptotic Sum Inequality

preserves commutativity

Commutative CCs suffice

Main point

embedding n x n matrix multiplication into a commutative coherent configuration of rank ¼ n^2 is a viable route to ! = 2

(no representation theory needed)

Open problems

- find a construction in new framework that
  - proves non-trivial bound on $!_s$
  - is not based on constructions from old setting

- is the (border) s-rank of <2,2,2> = 6?

- embed n £ n MM into commutative coherent configuration of rank ¼ n^2