EFFICIENTLY PARALLELIZABLE STRASSEN-BASED MULTIPLICATION OF A MATRIX BY ITS TRANSPOSE

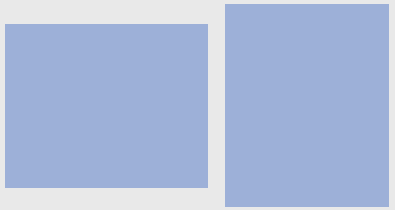
*Viviana Arrigoni**

Filippo Maggioli*

Annalisa Massini

Emanuele Rodolà

$A^T A$



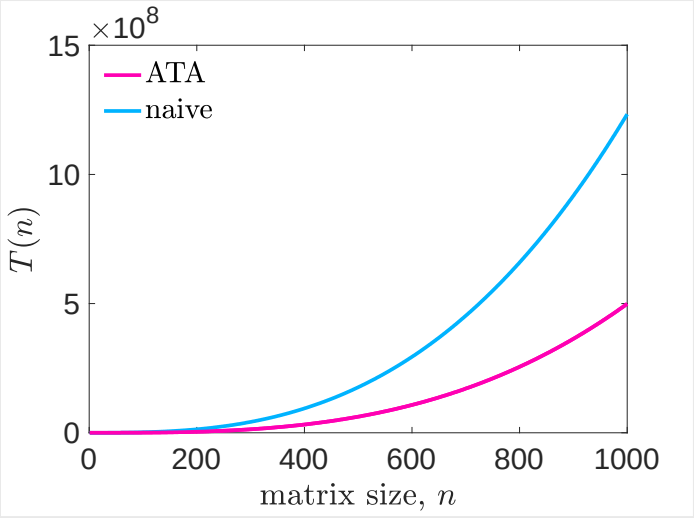
- Check for matrix orthogonality
- Project vectors onto the column space of A
- Vector orthogonalization (Gram-Schmidt)
- Normal equations for linear systems $Ax = b$
- SVD by eigenproblems of $A^T A$ or AA^T
- Discrete exterior Calculus and discrete differential Geometry

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

$$A^T A = \begin{bmatrix} A_{1,1}^T & A_{2,1}^T \\ A_{1,2}^T & A_{2,2}^T \end{bmatrix} \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} = \begin{bmatrix} A_{1,1}^T A_{1,1} + A_{2,1}^T A_{2,1} & A_{1,1}^T A_{1,2} + A_{2,1}^T A_{2,2} \\ A_{1,2}^T A_{1,1} + A_{2,2}^T A_{2,1} & A_{1,2}^T A_{1,2} + A_{2,2}^T A_{2,2} \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{2,1}^T \\ C_{2,1} & C_{2,2} \end{bmatrix}$$

computational complexity

$$T(n) = \frac{14}{3} n^{\log_2 7}$$



space complexity

$$S(n) = \frac{3}{2} n^2$$

cache complexity

$$C(n;\, M, b) = \Theta\left(1 + \frac{n^2}{b} + \frac{n^{\log_2(7)}}{b\sqrt{M}}\right)$$

$$C(1; M, b) = C_S(1; M, b)$$

$$C(n; M, b) \leq C_S(n; M, b)$$

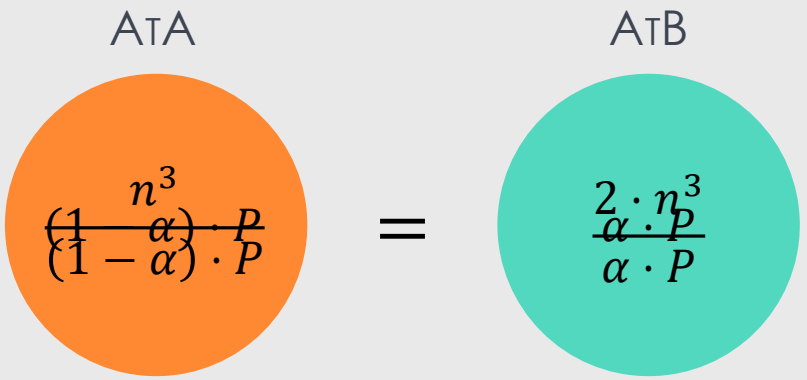
$$C(2n; M, b) \leq 4C_S(n; M, b) + 2C_S(2n; M, b)$$

Parallel implementation

SHARED
AᴛA-S

DISTRIBUTED
AᴛA-D

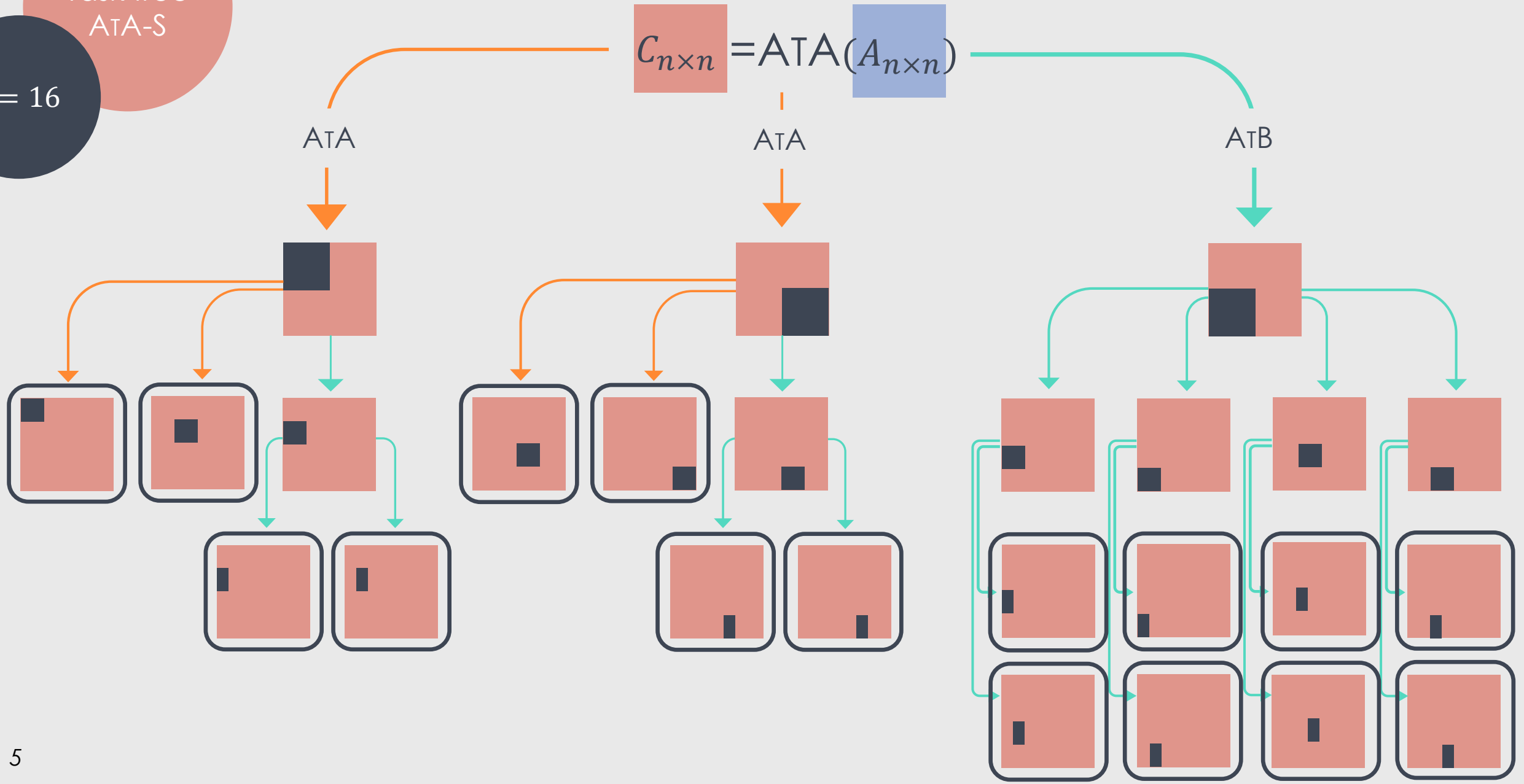
TASK TREE

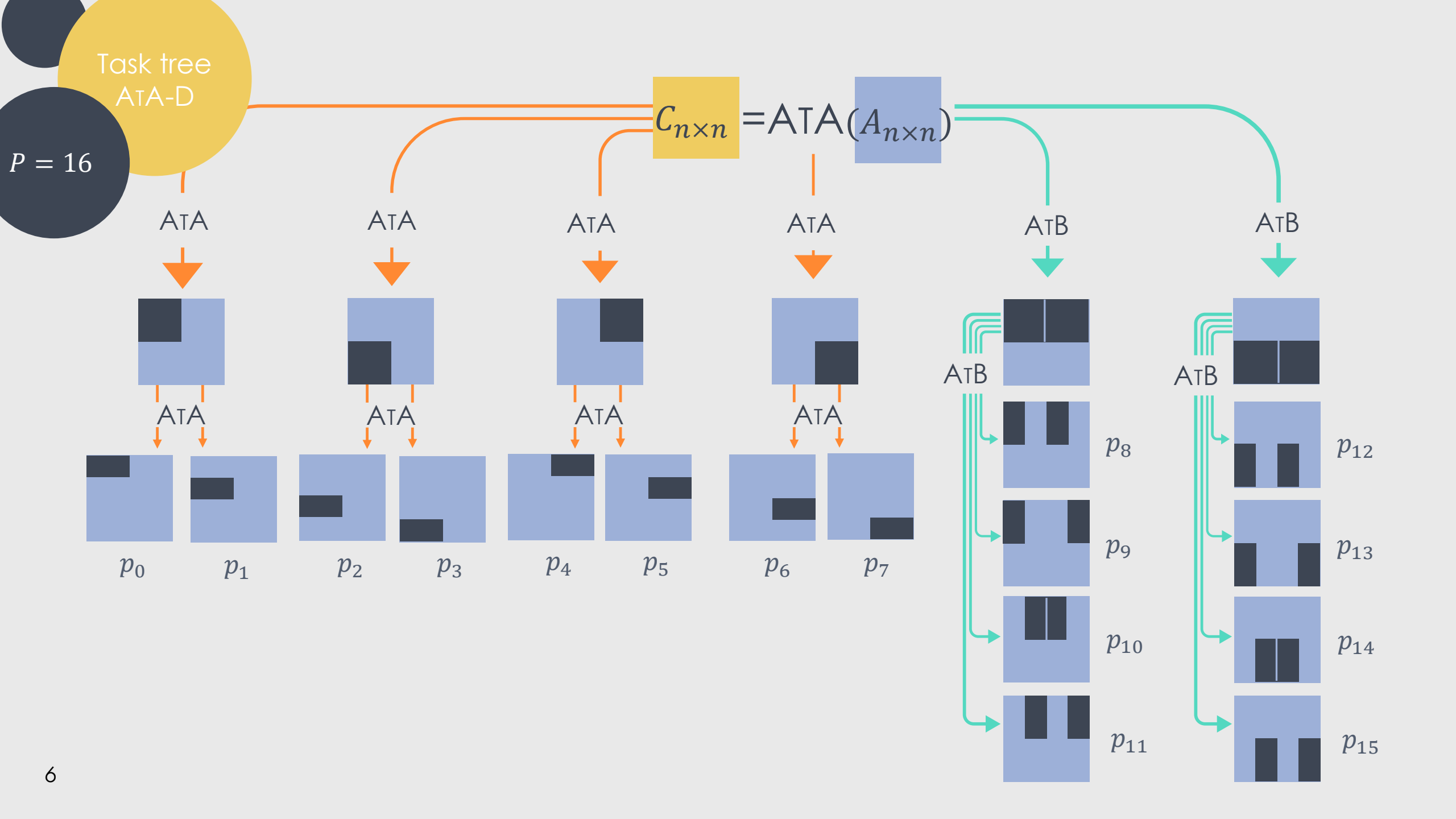


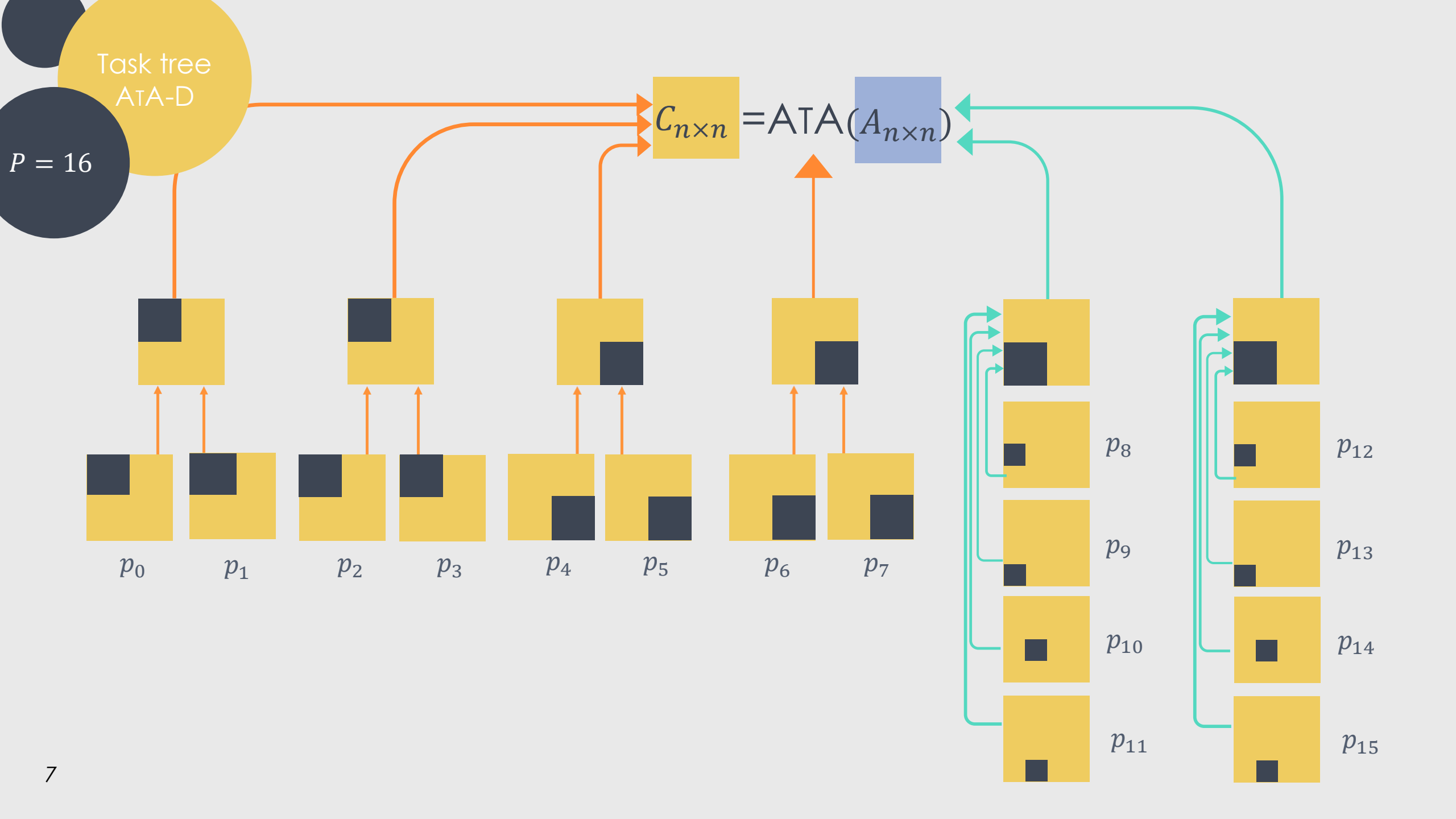
$$\alpha = \frac{1}{2}$$

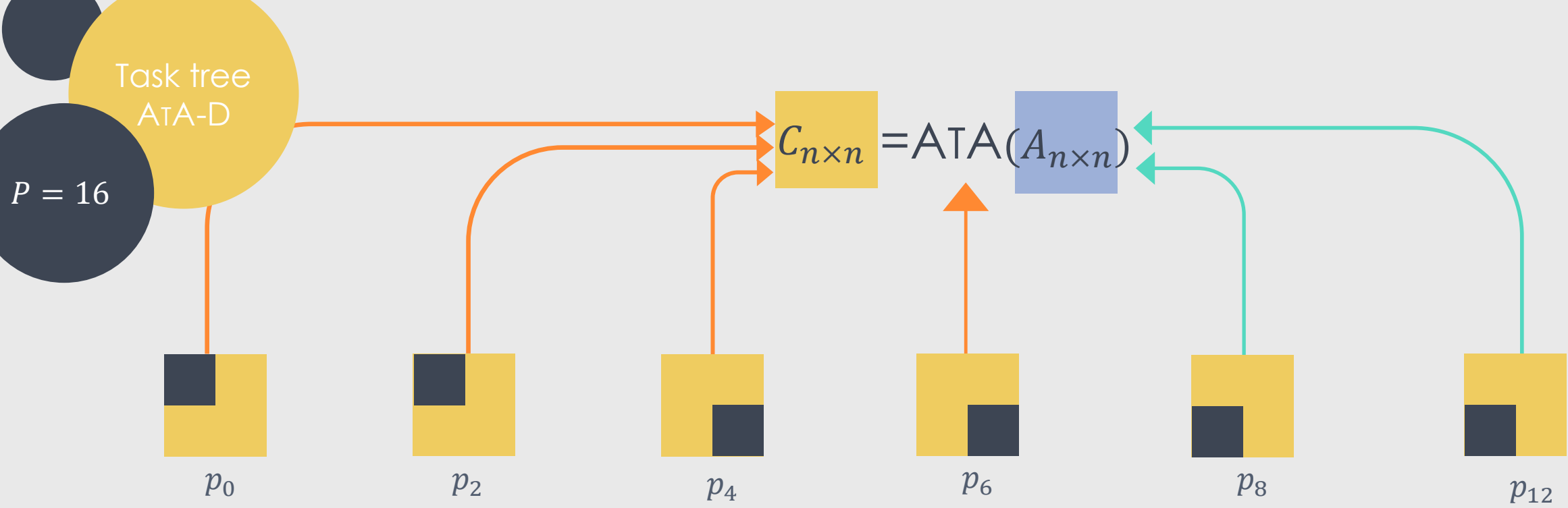
Task tree
AᵀA-S

$P = 16$




Task tree
AᵀA-D
$P = 16$
$C_{n\times n} = A^TA(A_{n\times n})$
AᵀA
AᵀA
AᵀA
AᵀA
AᵀB
AᵀB
AᵀA
AᵀA
AᵀA
AᵀA
AᵀB
AᵀB
p_0
p_1
p_2
p_3
p_4
p_5
p_6
p_7
p_8
p_9
p_{10}
p_{11}
p_{12}
p_{13}
p_{14}
p_{15}


Task tree
ATA-D
$P = 16$
$C_{n\times n}$ =ATA($A_{n\times n}$)
p_0
p_1
p_2
p_3
p_4
p_5
p_6
p_7
p_8
p_9
p_{10}
p_{11}
p_{12}
p_{13}
p_{14}
p_{15}


Task tree
ATA-D
$P = 16$
$C_{n\times n}$ =ATA($A_{n\times n}$)
p_0
p_2
p_4
p_6
p_8
p_{12}

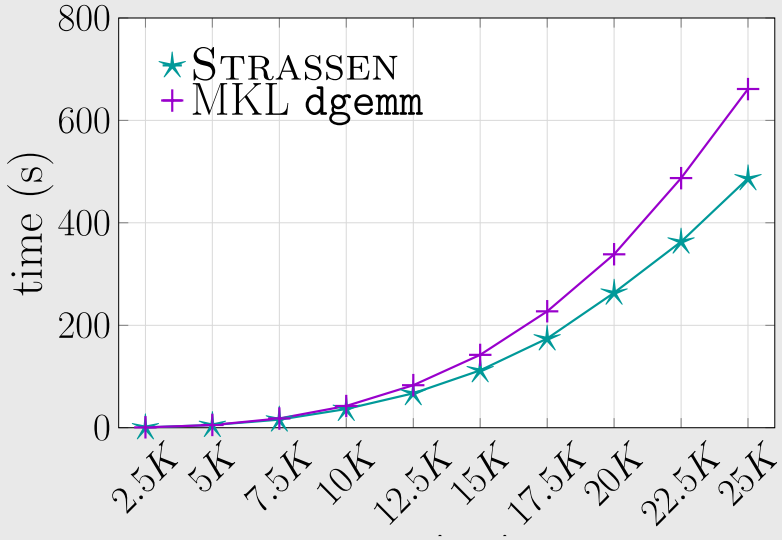
Task tree
AᵀA-D

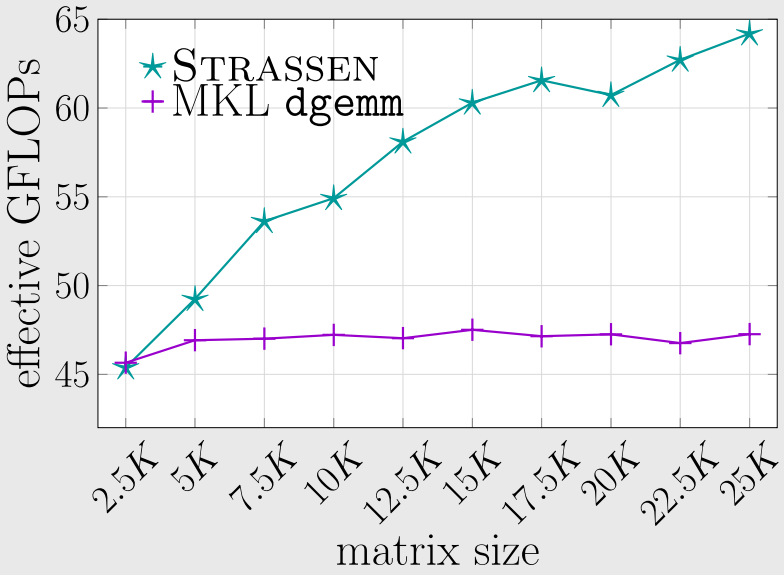
$P = 16$

$$C_{n\times n} = \mathrm{A^TA}(A_{n\times n})$$

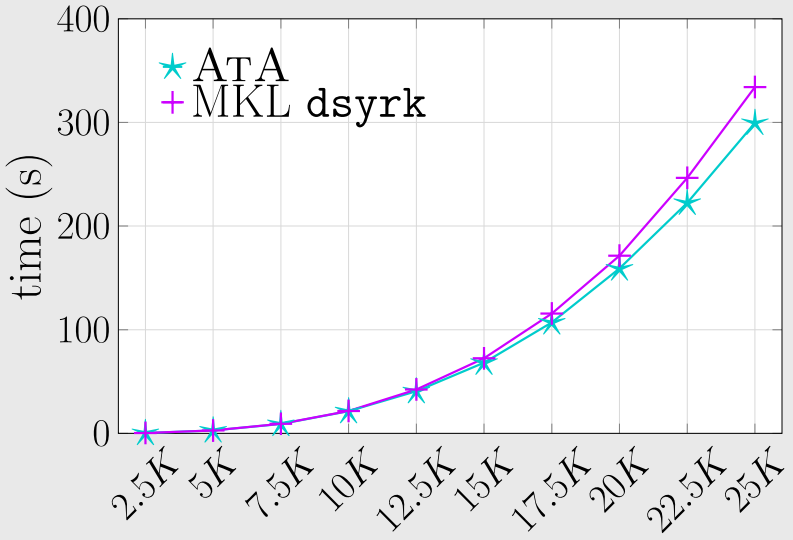
p_0

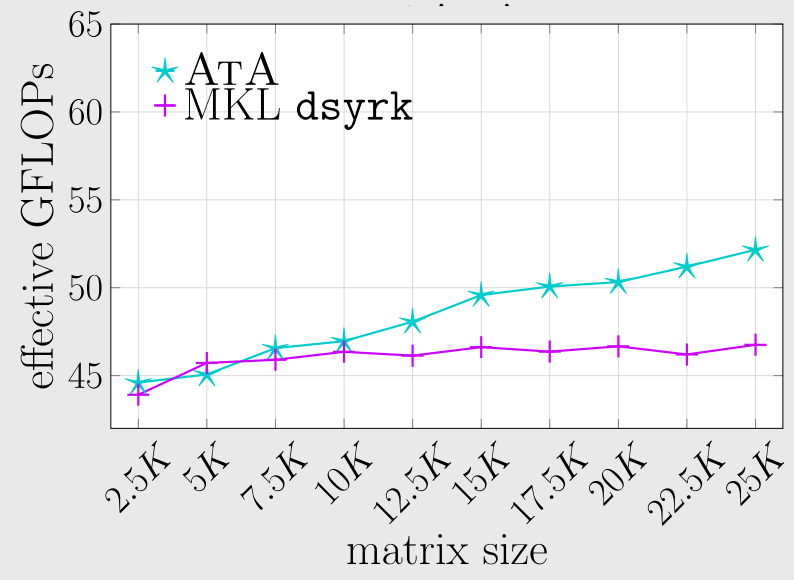
sequential implementation





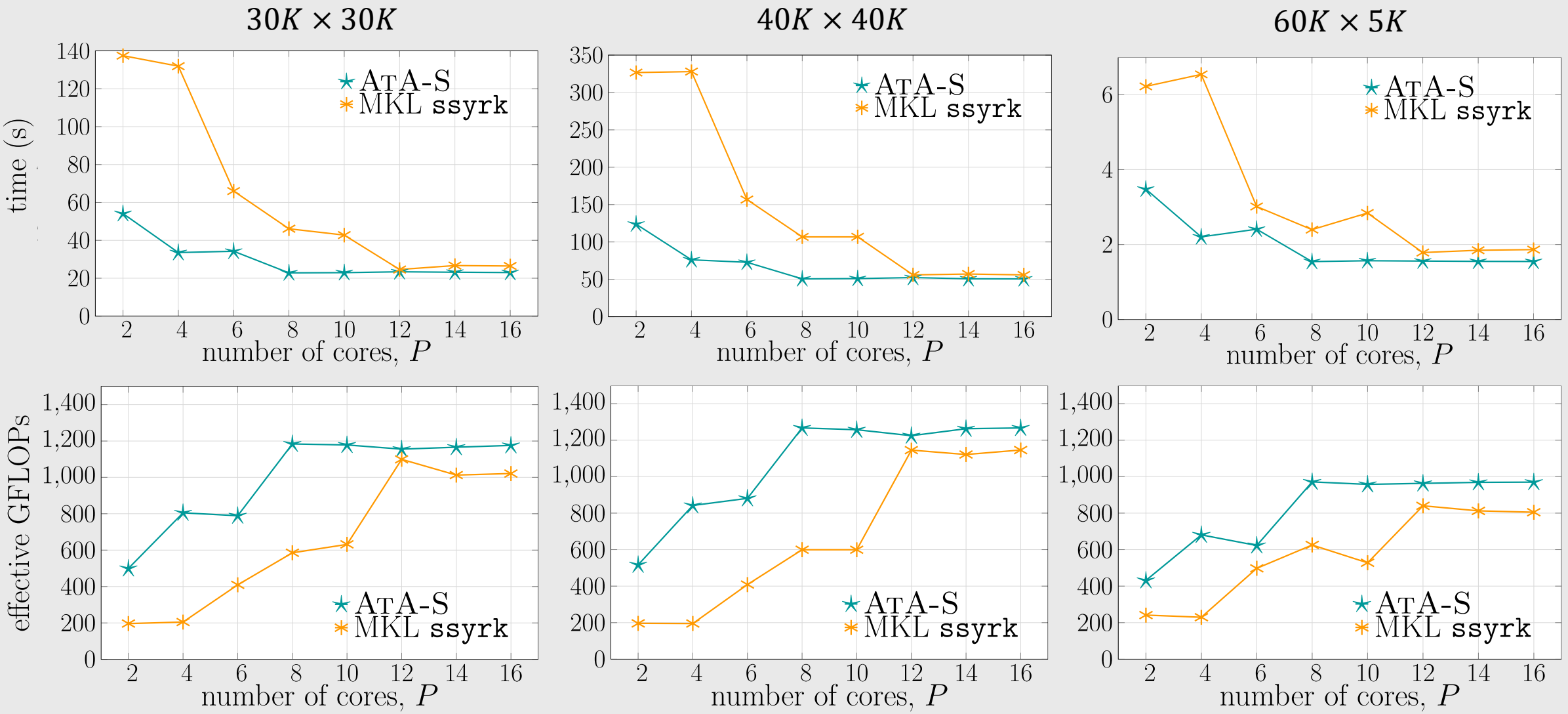
$$EG = \frac{r \cdot n^3}{10^9 \cdot t}$$





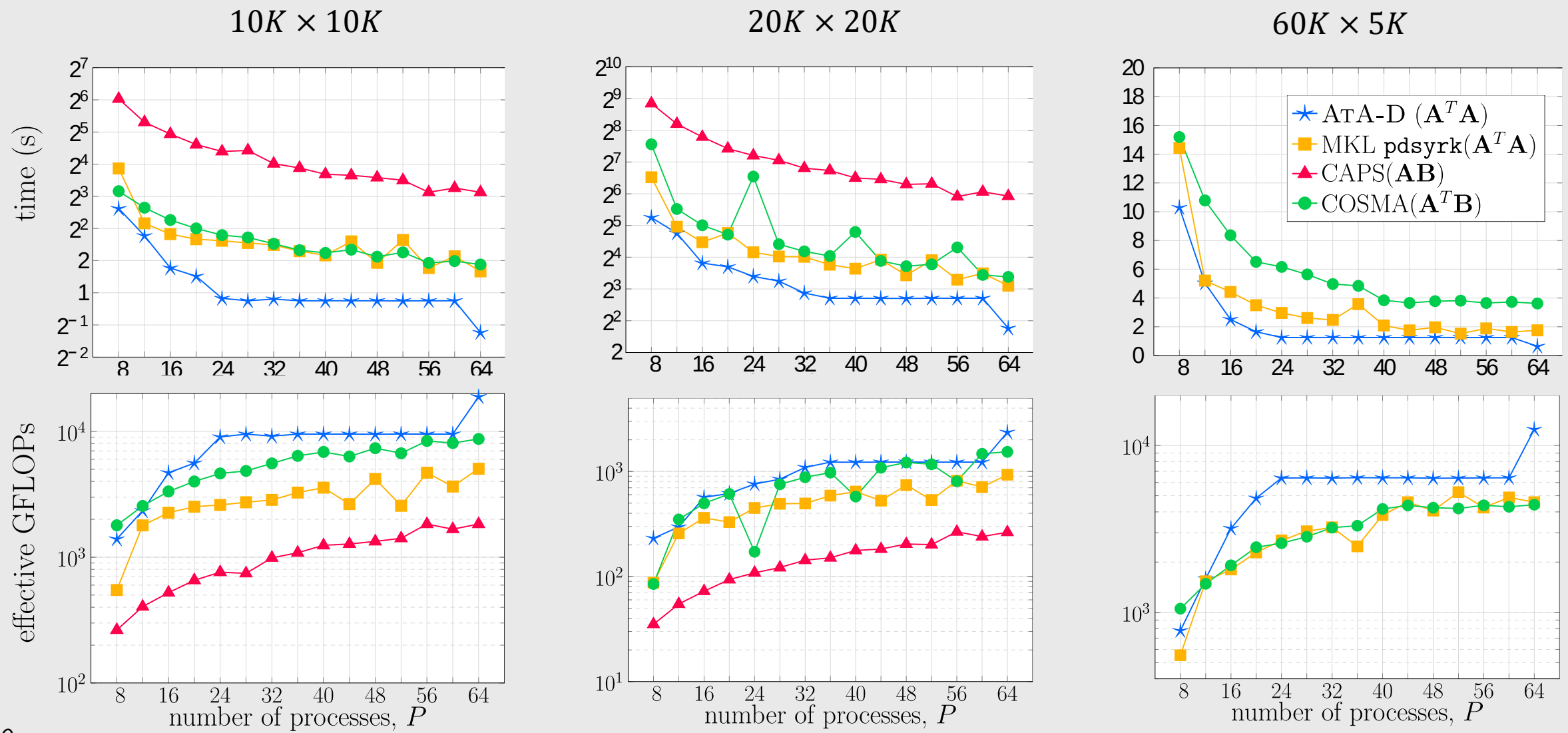
TERASTAT: 12 compute nodes, 2 sockets of Intel Xeon E5-2630v3 8 cores per node, 2.4 Ghz, 4 GB RAM per core

shared memory implementation



TERASTAT: 12 compute nodes, 2 sockets of Intel Xeon E5-2630v3 8 cores per node, 2.4 Ghz, 4 GB RAM per core

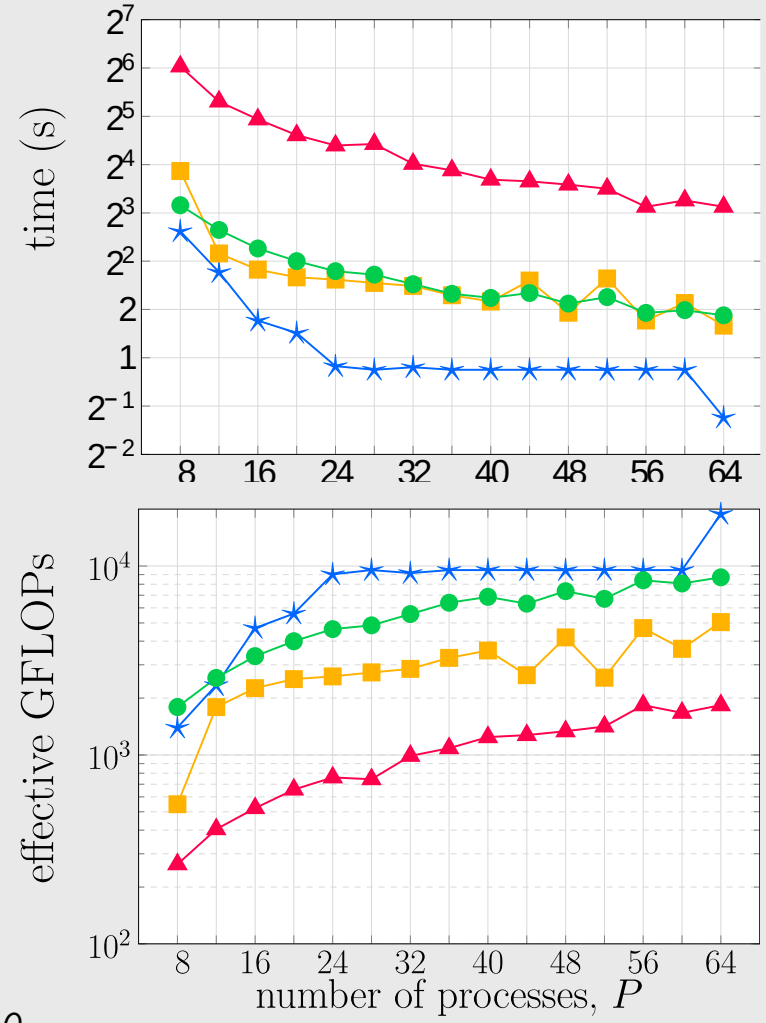
distributed memory implementation



TERASTAT: 12 compute nodes, 2 sockets of Intel Xeon E5-2630v3 8 cores per node, 2.4 Ghz, 4 GB RAM per core

distributed memory implementation

$10K \times 10K$



$$\begin{cases} \ell(1) = 0 \\ \ell(2 \le P \le 6) = 1 \\ \ell(P) = 1 + k + \text{sign}\left(\frac{P}{4} \bmod 8^{\max(k,1)}\right) \end{cases}$$

$$k = \max\left(k \in \mathrm{N} : \frac{P/4}{8^k} \ge 1\right)$$

TERASTAT: 12 compute nodes, 2 sockets of Intel Xeon E5-2630v3 8 cores per node, 2.4 Ghz, 4 GB RAM per core

*Viviana Arrigoni** arrigoni@di.uniroma1.it
Filippo Maggioli* maggioli@di.uniroma1.it
Annalisa Massini massini@di.uniroma1.it
Emanuele Rodolà rodola@di.uniroma1.it

https://github.com/filthynobleman/AtA

INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING