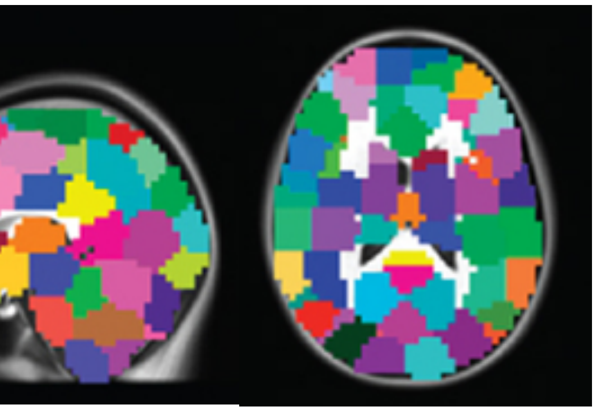
Shared Memory Parallelization of MTTKRP for Dense Tensors

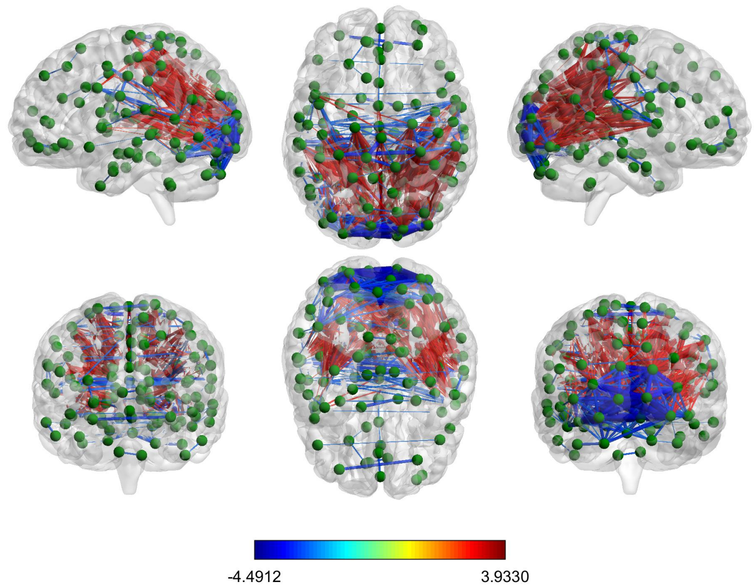
WAKE FOREST UNIVERSITY

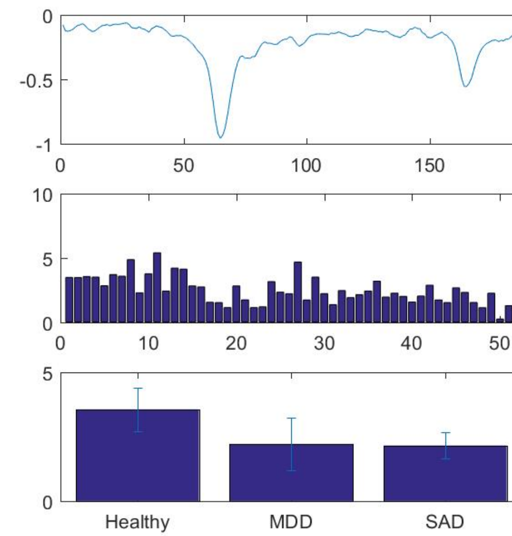
BLIS Retreat 2017,
September 18th

Koby Hayashi, Grey Ballard, Yujie Jiang, Michael Tobia
hayakb13@,ballard@,jiany14@,tobiamj@wfu.edu

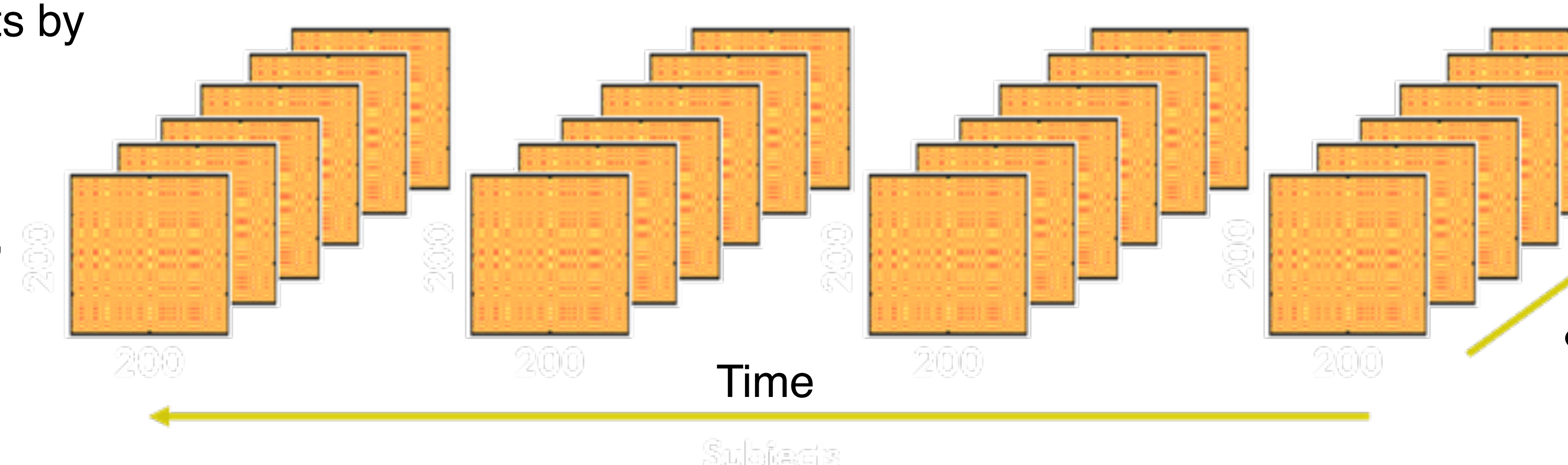
Neuroimaging Application





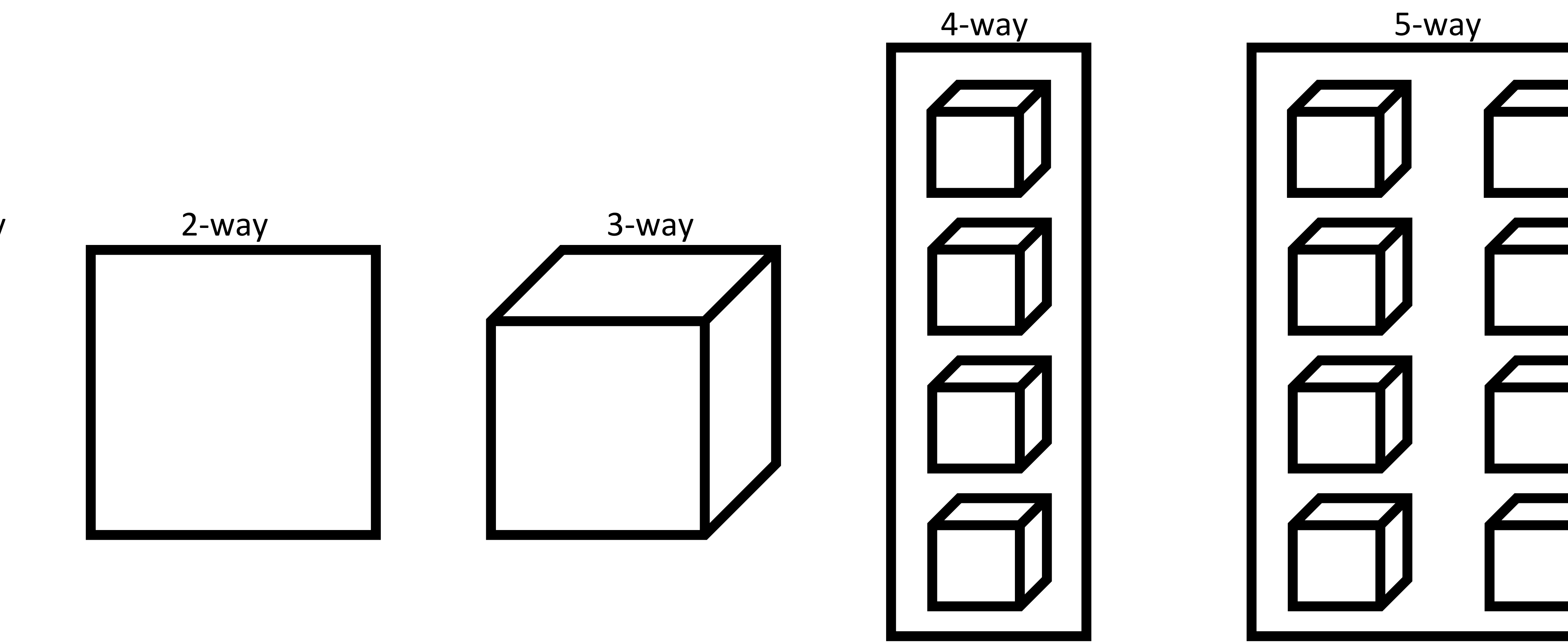


sor: Time by Subjects by
el Correlation Matrix
: Rest → Activity →
overy
jects: Control, MDD,
), COMO



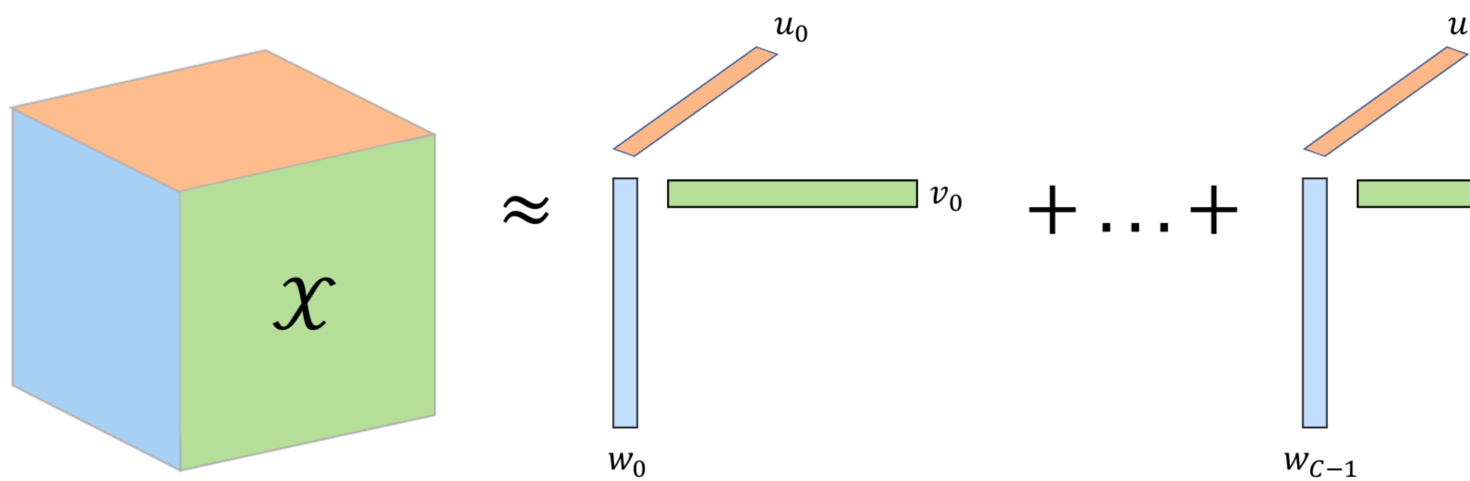
Quick Introduction to Tensors

idimensional arrays, an N-dimensional
ors is said to be N-way or order-N.



CP Decomposition

nical Polyadic
mposition (CP):
composes a tensor
o a sum of rank 1
sors



$$\mathcal{X} \approx \sum_{c=0}^{C-1} u_{ic} \circ v_{jc} \circ w_{kc}$$

$$\mathcal{X} \approx [\![U, V, W]\!]$$

CP via Alternating Least Squares

Algorithm 5 CP_ALS

Require: $\mathcal{X}$ is an N-way tensor with dimensions $I_0 \times I_1 \times \cdots \times I_{N-1}$, $n \in [N]$, $\mathbf{U}_{(n)}$ is the n^{th} factor matrix, and a rank C

function $\mathcal{Y} = \text{CP\_ALS}(\mathcal{X}, C)$
    while stopping conditions not met **do**
        for $n \in [N]$ **do**
            $\mathbf{H} = \mathbf{U}_{(0)}^{\mathsf{T}}\mathbf{U}_{(0)} * \cdots * \mathbf{U}_{(n-1)}^{\mathsf{T}}\mathbf{U}_{(n-1)} * \mathbf{U}_{(n+1)}^{\mathsf{T}}\mathbf{U}_{(n+1)} * \cdots * \mathbf{U}_{(N-1)}^{\mathsf{T}}\mathbf{U}_{(N-1)}$
            $\mathbf{M} = \mathbf{X}_{(n)}(\mathbf{U}_{(N-1)} \odot \cdots \odot \mathbf{U}_{(n+1)} \odot \mathbf{U}_{n-1} \odot \cdots \odot \mathbf{U}_{(0)})$
            solve $\mathbf{U}_n = \mathbf{M}\mathbf{H}^{\dagger}$
        end for
    end while
end function

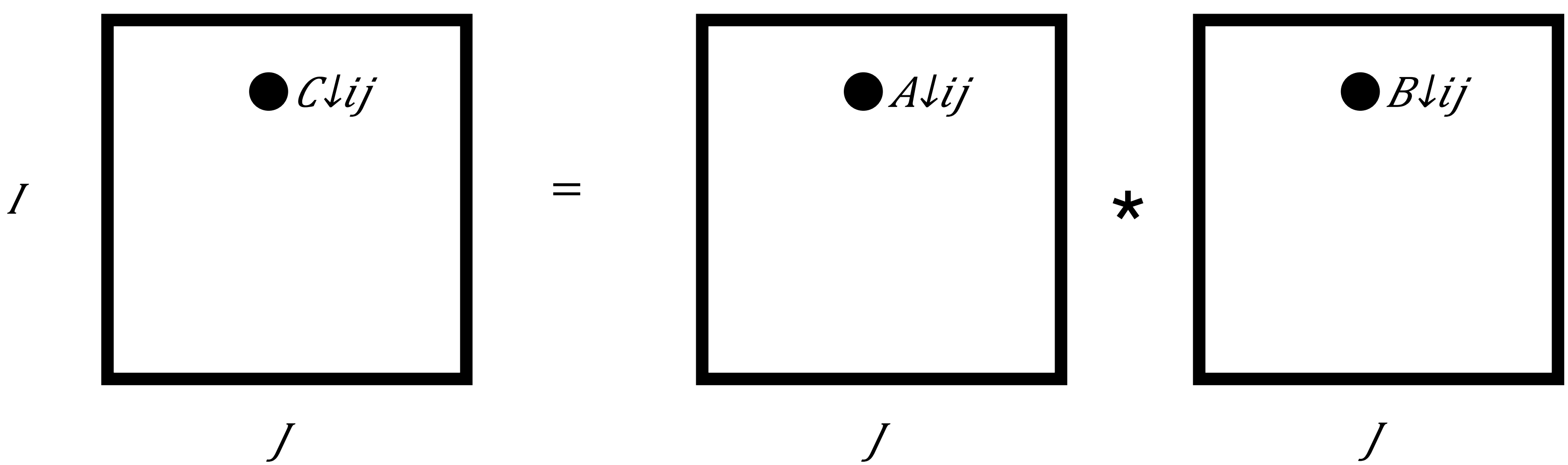
Ensure: $\mathcal{Y}$ is a rank C CP Model

Hadamard Product

$0\,)*\dots*(U_{n-1}^{T}\,U_{n-1}\,)*(U_{n+1}^{T}\,U_{n+1}\,)*\dots*(U_{N-1}^{T}\,U_{N-1}\,)$

*Element wise matrix product denoted **

$$C = A * B$$
$$C_{ij} = A_{ij} * B_{ij}$$



Khatri Rao Product

$(\boldsymbol{n})\ (\boldsymbol{U}_{N-1} \odot \ldots \odot \boldsymbol{U}_{n+1} \odot \boldsymbol{U}_{n-1} \odot \ldots \odot \boldsymbol{U}_{0})$

tri Rao Product (KRP):

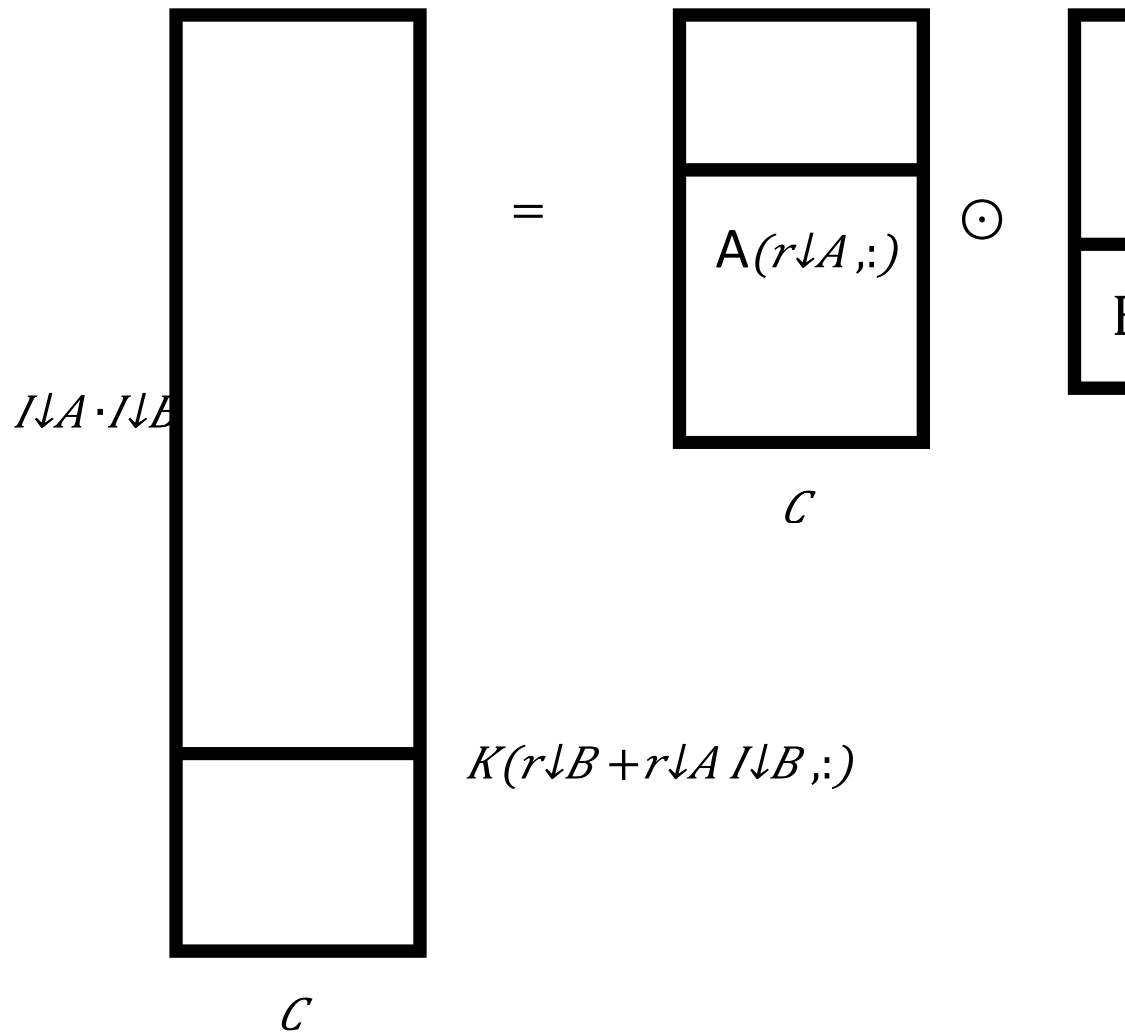
$\odot B$

umn-wise Kronecker Product

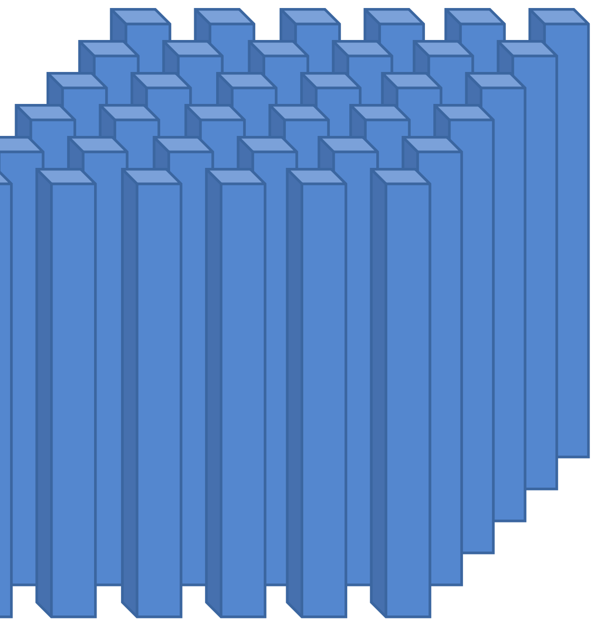
$= A(:,i) \odot B(:,i)$

Hadamard Product of Rows

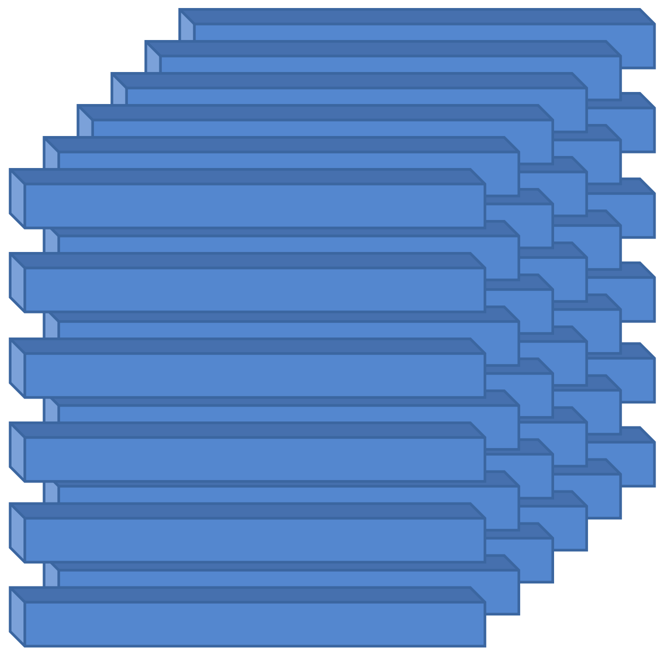
$B + r_A I_B ,:) = A(r_A ,:) * B(r_B ,:)$



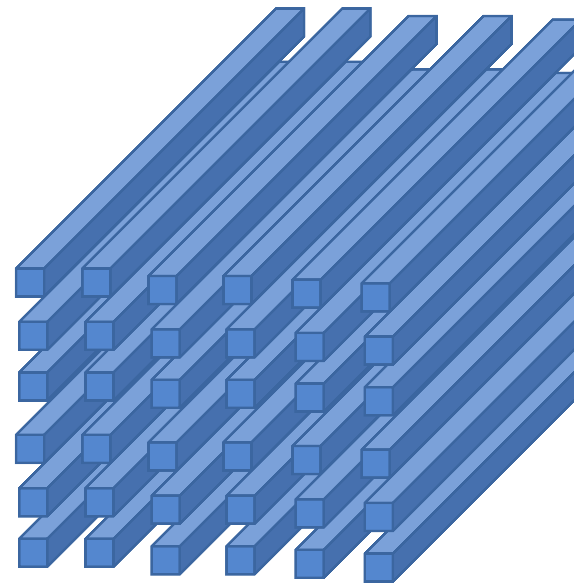
Tensor Fibers



$n=0,\ \mathcal{X}\downarrow(:jk)$



$n=1,\ \mathcal{X}\downarrow(i:k)$



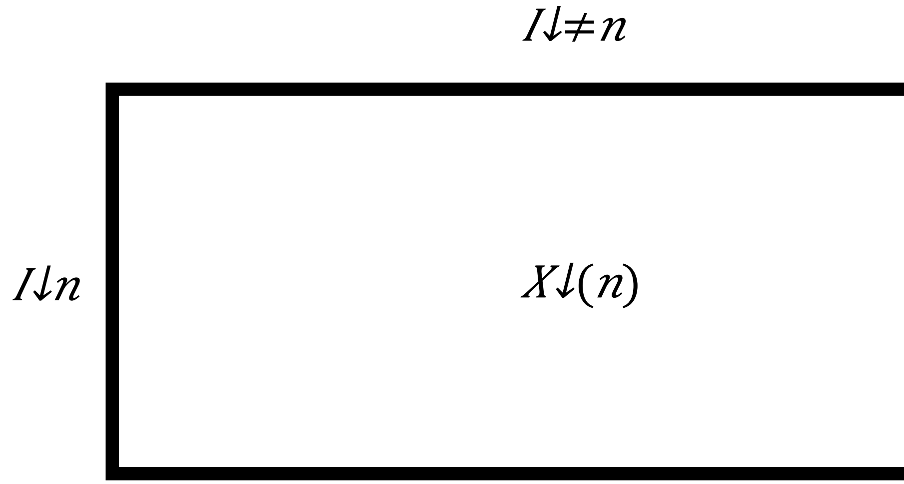
$n=2,\ \mathcal{X}\downarrow(ij:)$

Unfolding Tensors

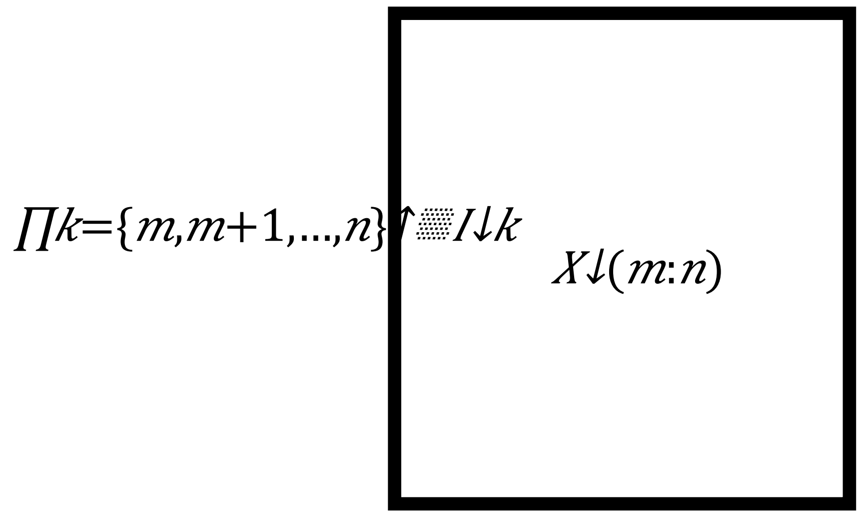
$\boldsymbol{{}_{(n)}\ (U_{N-1} \odot \dots \odot U_{n+1} \odot U_{n-1} \odot \dots \odot U_{0})}$

The n*th* mode matricization of a N-way tensor $\mathcal{X}$ that is $I_0 \times I_1 \times \dots \times I_{N-1}$ is denoted $X_{(n)}$ and is $I_n \times I_{\neq n}$

- $I_{\neq n} = \prod_{n \neq k \in [N]} I_k$



$X_{(m:n)}$ denotes a matricization where $\{m, m+1, \dots, n\}$ are the row modes



Matricized Tensor Times Khatri Rao Product

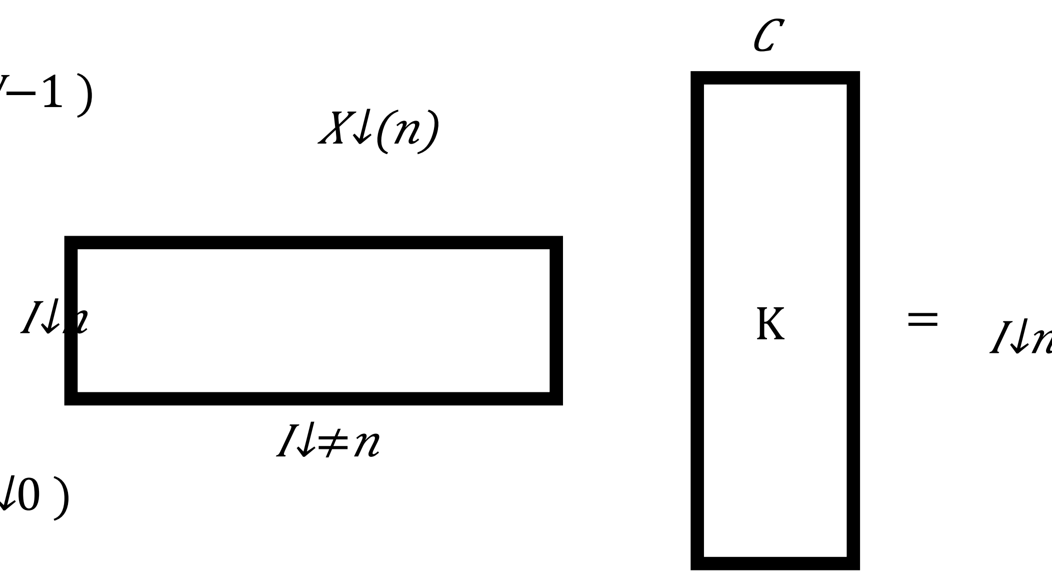
$_{(n)} (U_0 \odot \ldots \odot U_{n-1} \odot U_{n+1} \odot \ldots \odot U_{N-1})$

aïve algorithm

1. Permute $\mathcal{X}$ to $X_{(n)}$
2. Form $\mathrm{K} = (U_0 \odot \ldots \odot U_{n-1} \odot U_{n+1} \odot \ldots \odot U_{N-1})$
3. Call DGEMM

-Step and 2-Step MTTKRP

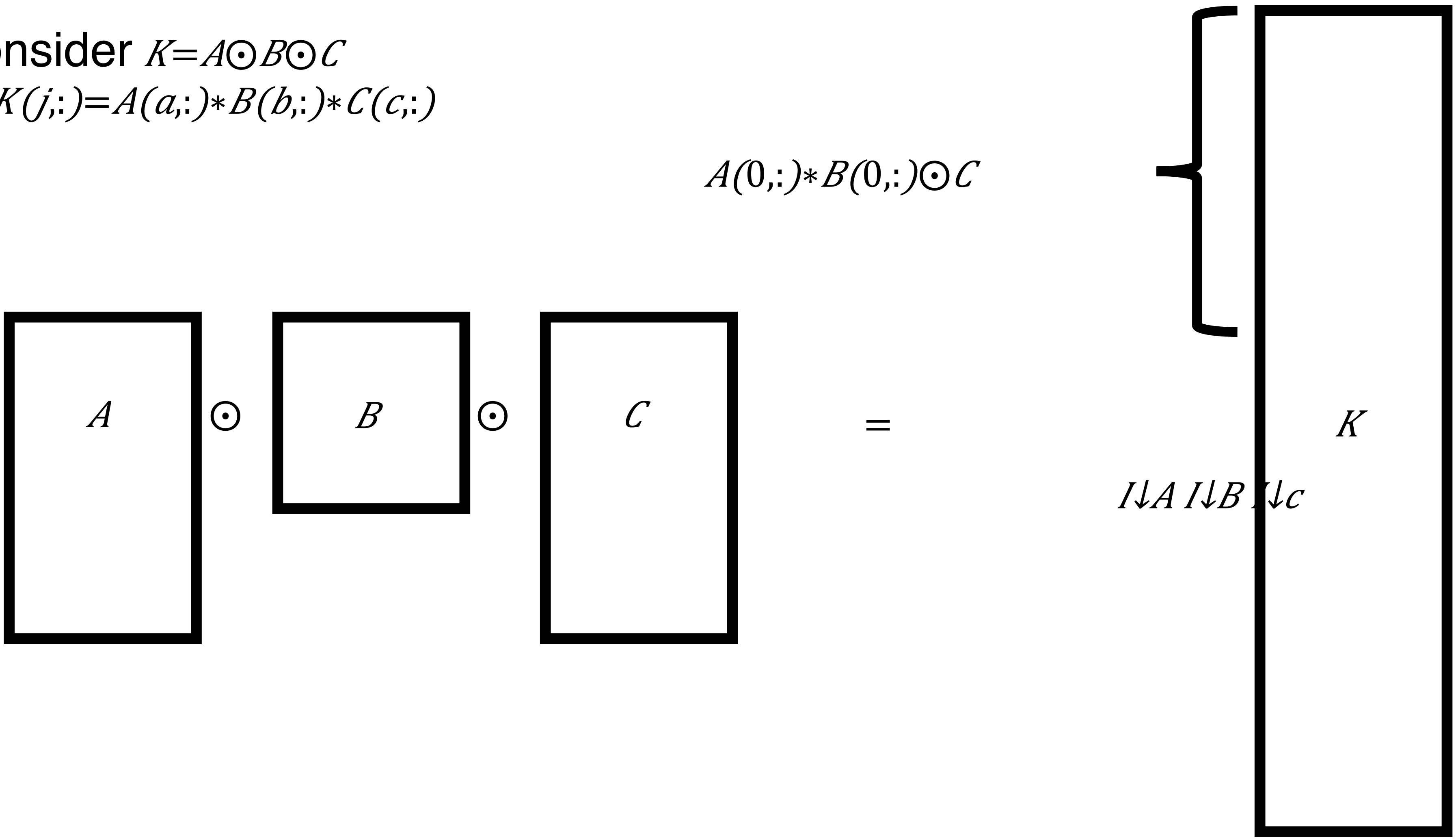
1. Avoid permuting $\mathcal{X}$
2. Efficiently form the KRP
   - 1Step
     - $(U_{N-1} \odot \ldots \odot U_{n+1} \odot U_{n-1} \odot \ldots \odot U_0)$
   - 2Step
     - $K_L = (U_0 \odot \ldots \odot U_{n-1})$
     - $K_R = (U_{n+1} \odot \ldots \odot U_N)$
3. Utilize BLAS

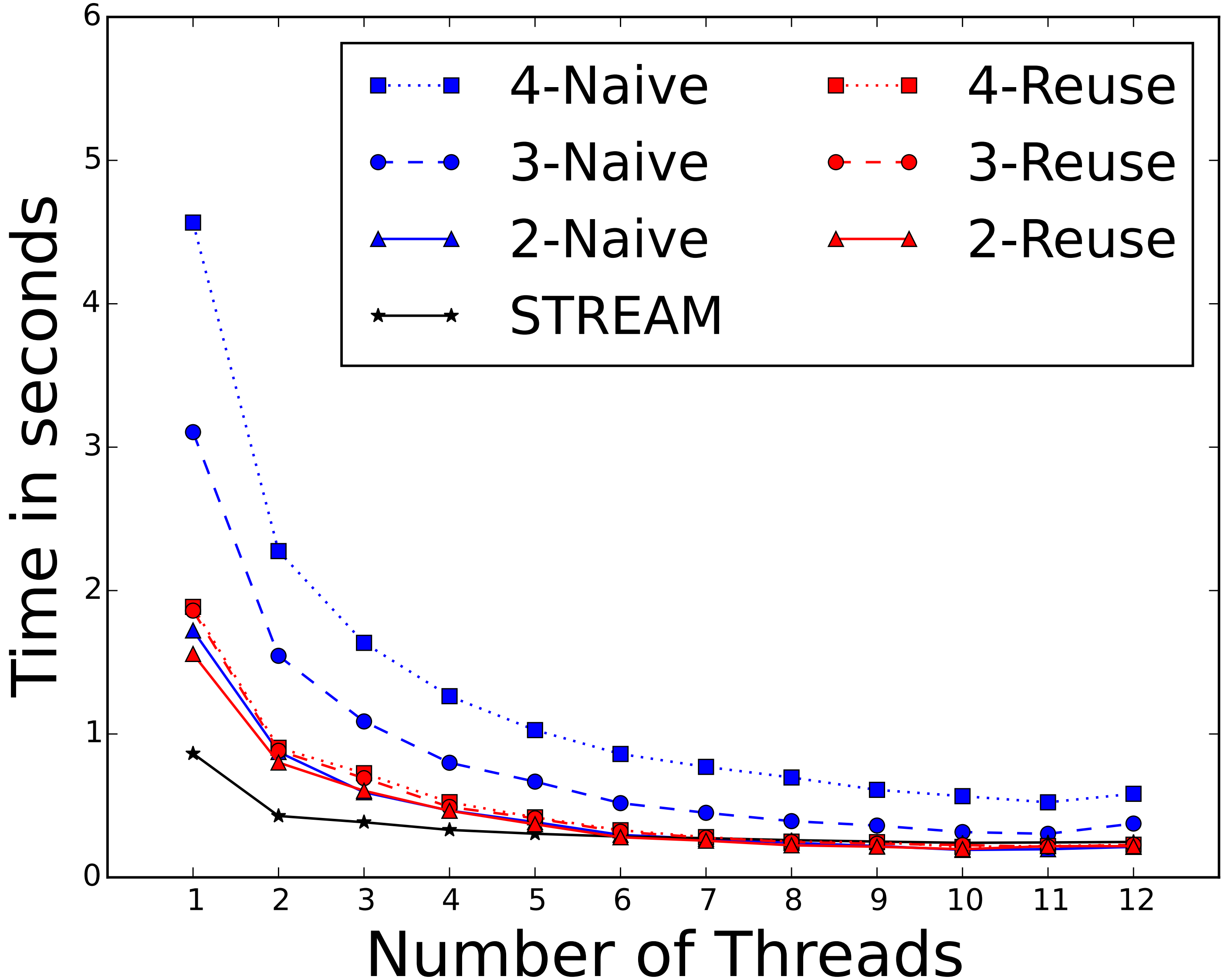


Computing the KRP

Consider $K = A \odot B \odot C$

$K(j,:) = A(a,:) * B(b,:) * C(c,:)$



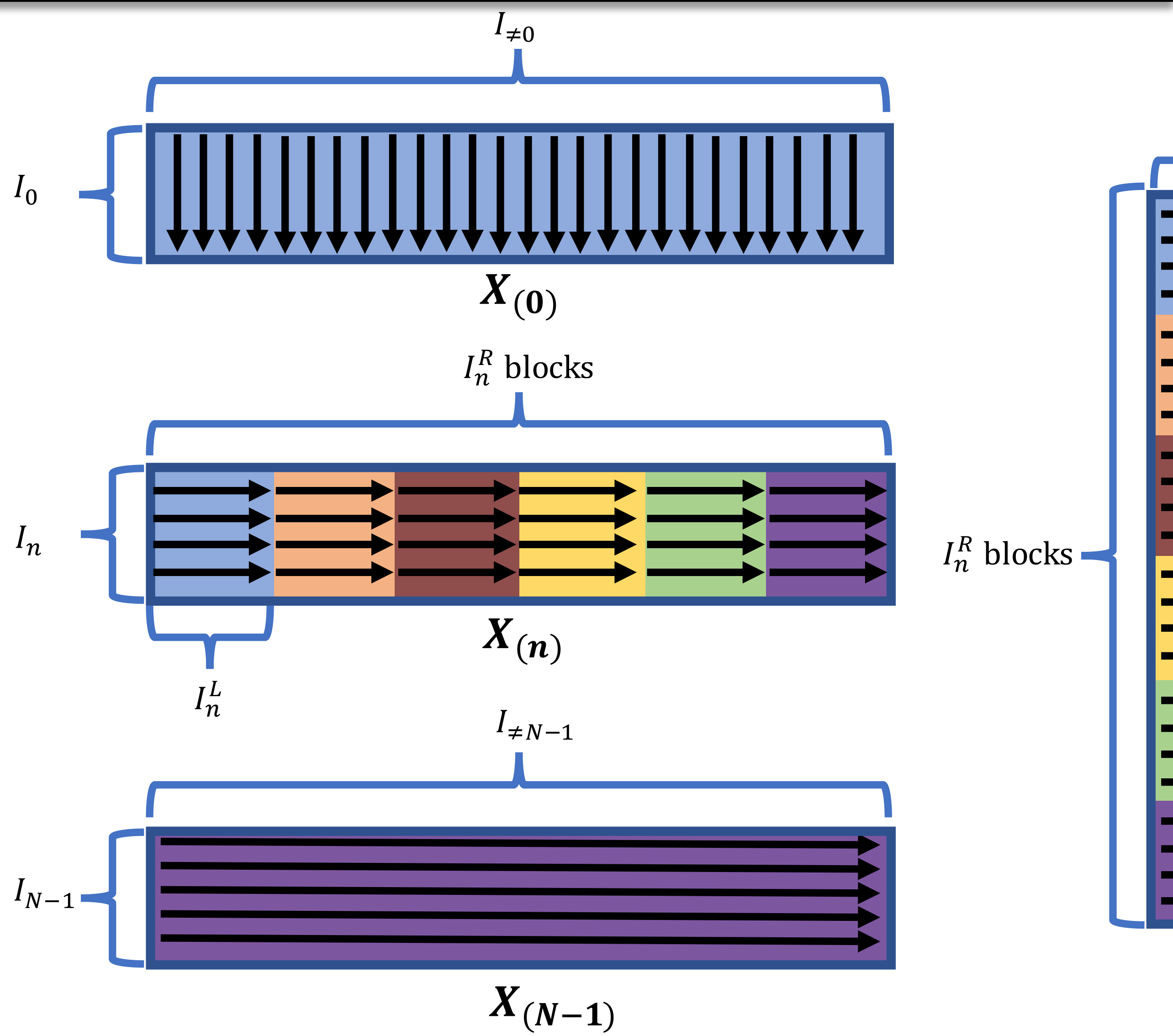


Timings for KRPs of naïve and reuse algorithms.

1-Step MTTKRP

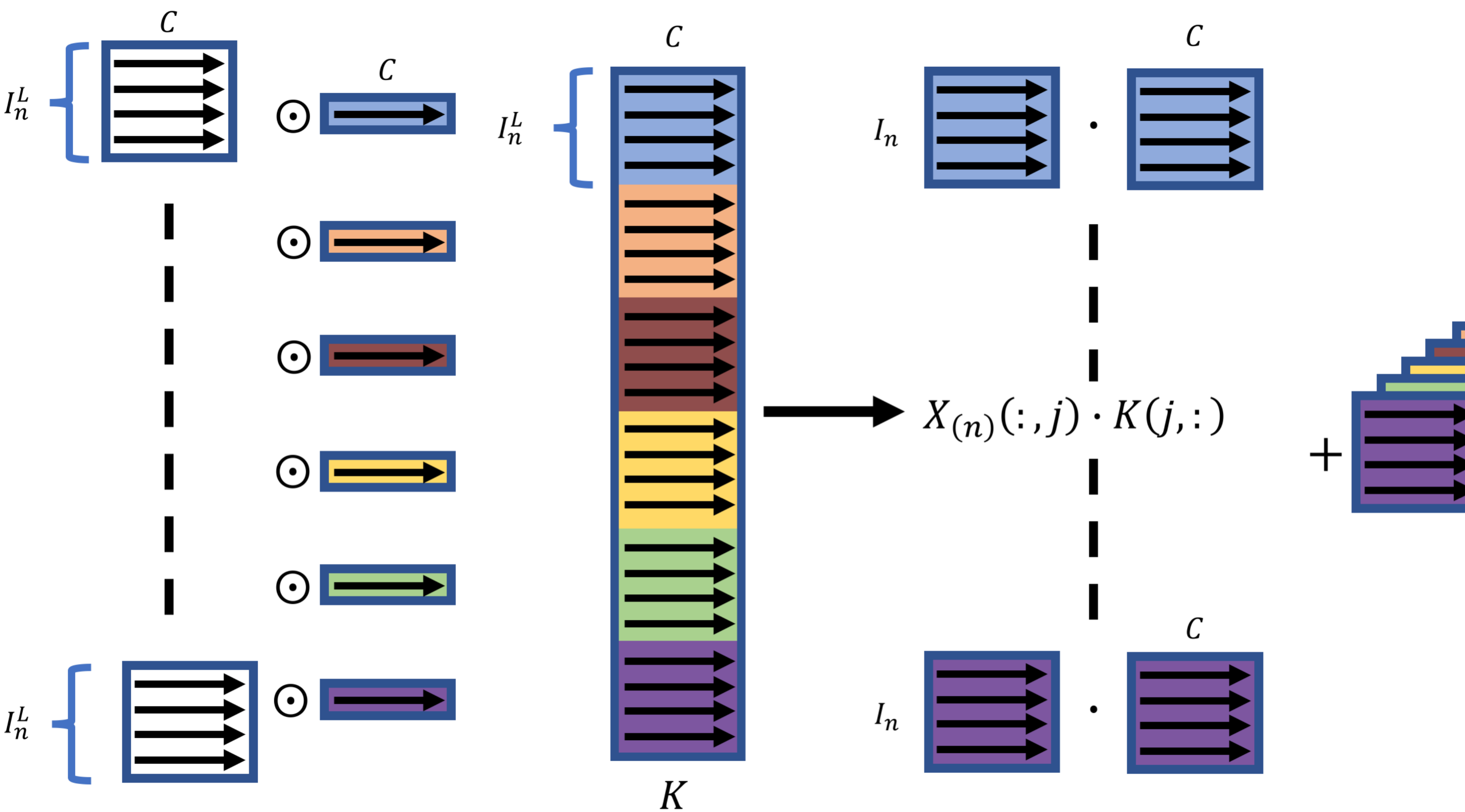
oid permuting tensor entries
st computation as matmul

bservation: the n*th* mode
cization of a tensor can be
ned by chunking the tensor
ontiguous submatrices of
size.



Parallel 1-Step MTTKRP

rm $K{\downarrow}L$
rm $K{\downarrow}R\,(j,:)$
rm $K(j,:)$
tMul
duce



2-Step MTTKRP

rst Compute a *Partial MTTKRP*

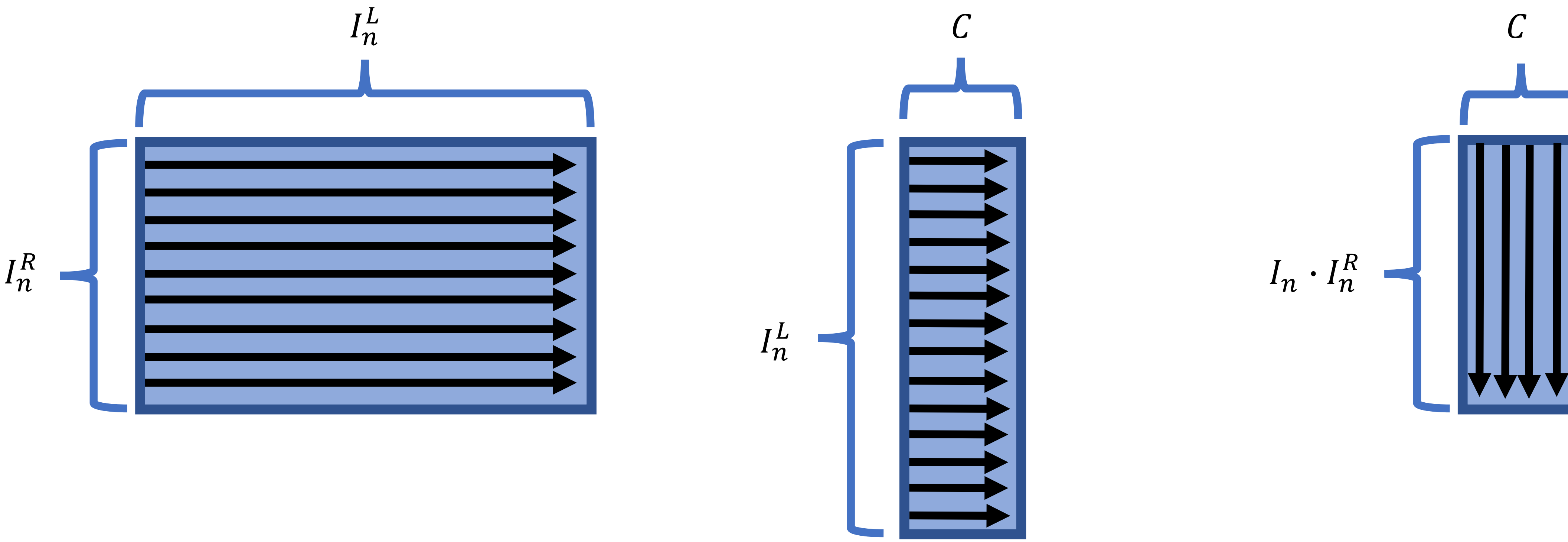
1. Compute K_L and K_R
2. $\mathcal{L} \leftarrow X_{(0:n-1)}^{T} \cdot K_L$
    - $\mathcal{L}$ is $I_n \times \ldots \times I_{N-1} \times C$

econd Compute a Series of ___?___ operations.

a. Tensor Times Vector (TTVs)
b. Tensor Times Matrix (TTMs)
c. Quasi-Tensor Times Matrix (q-TTMs)

2-Step MTTKRP: $\mathcal{L}$

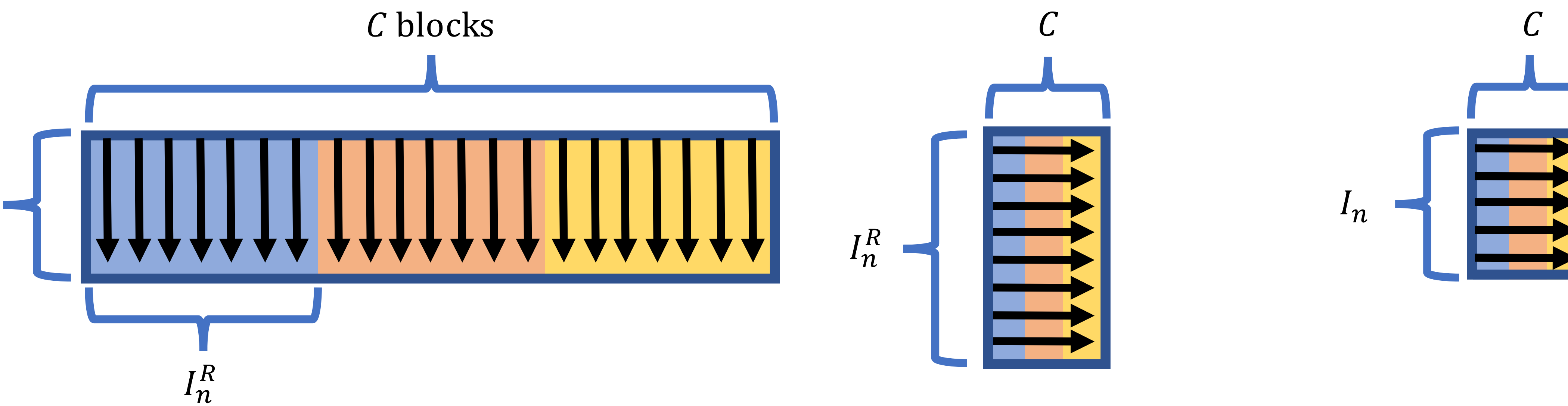
rst Compute a *Partial MTTKRP*



$$\boldsymbol{X}^{\boldsymbol{T}}_{(\boldsymbol{0}:\boldsymbol{n}-\boldsymbol{1})} \cdot \boldsymbol{K}_{\boldsymbol{L}} = \boldsymbol{L}_{(\boldsymbol{0}:\boldsymbol{N}-\boldsymbol{n}}$$

2-Step MTTKRP: $\mathcal{L}$

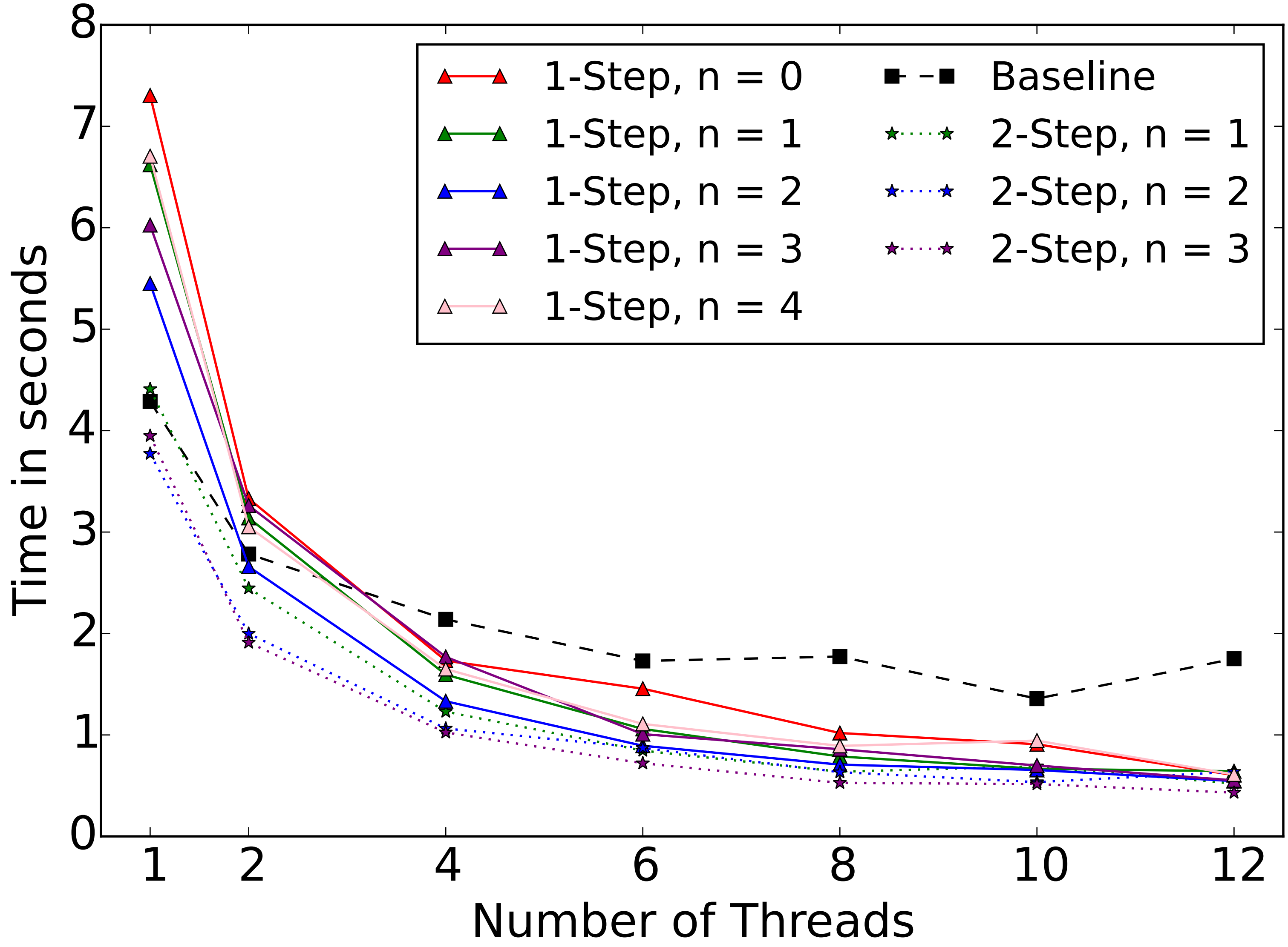
econd Compute a series of TTVs



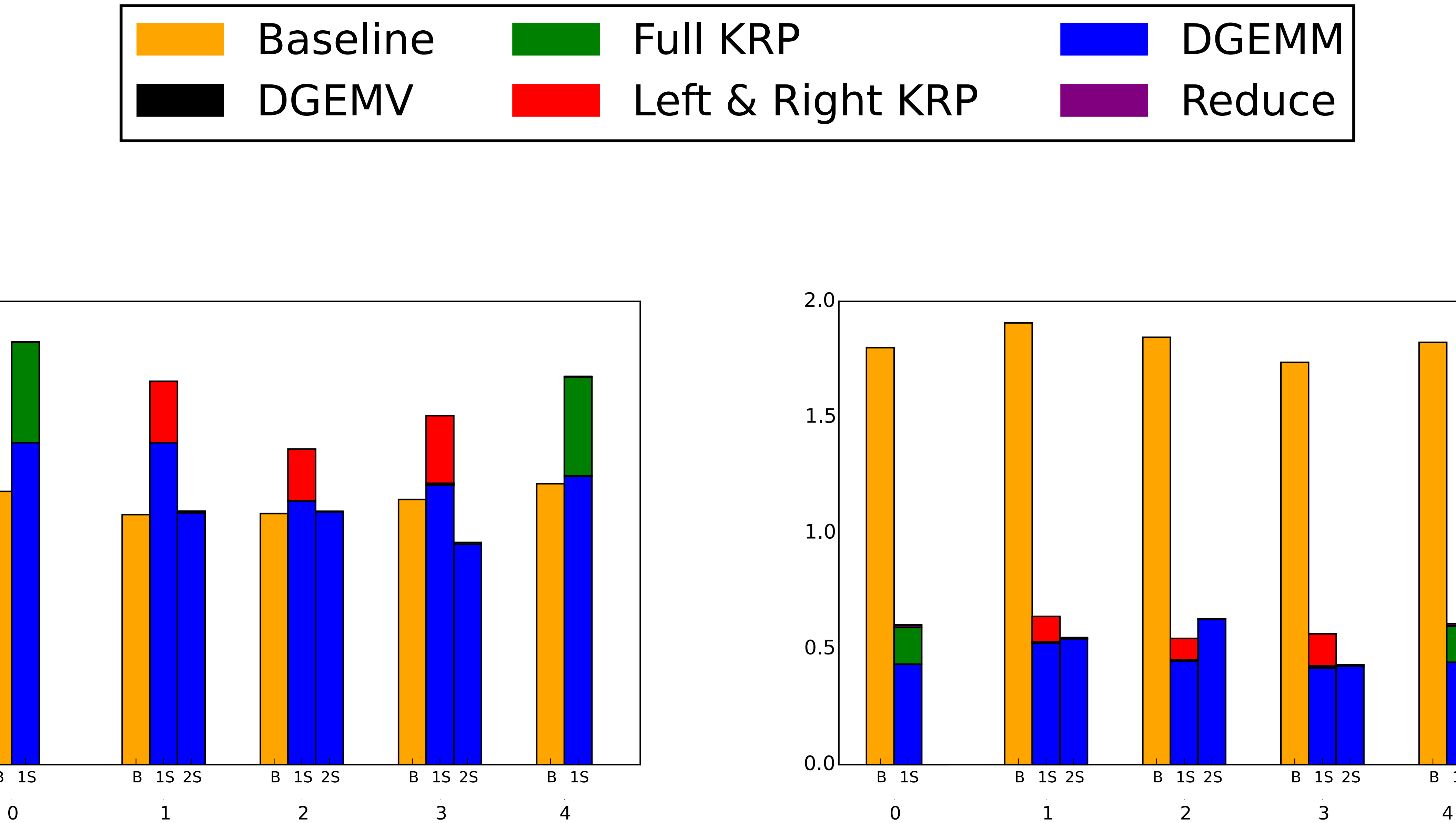
$$\boldsymbol{L}_{(0)}[j] \quad \cdot \quad \boldsymbol{K}_{\boldsymbol{R}}(:,\boldsymbol{j}) \quad = \quad \boldsymbol{M}(:,j$$

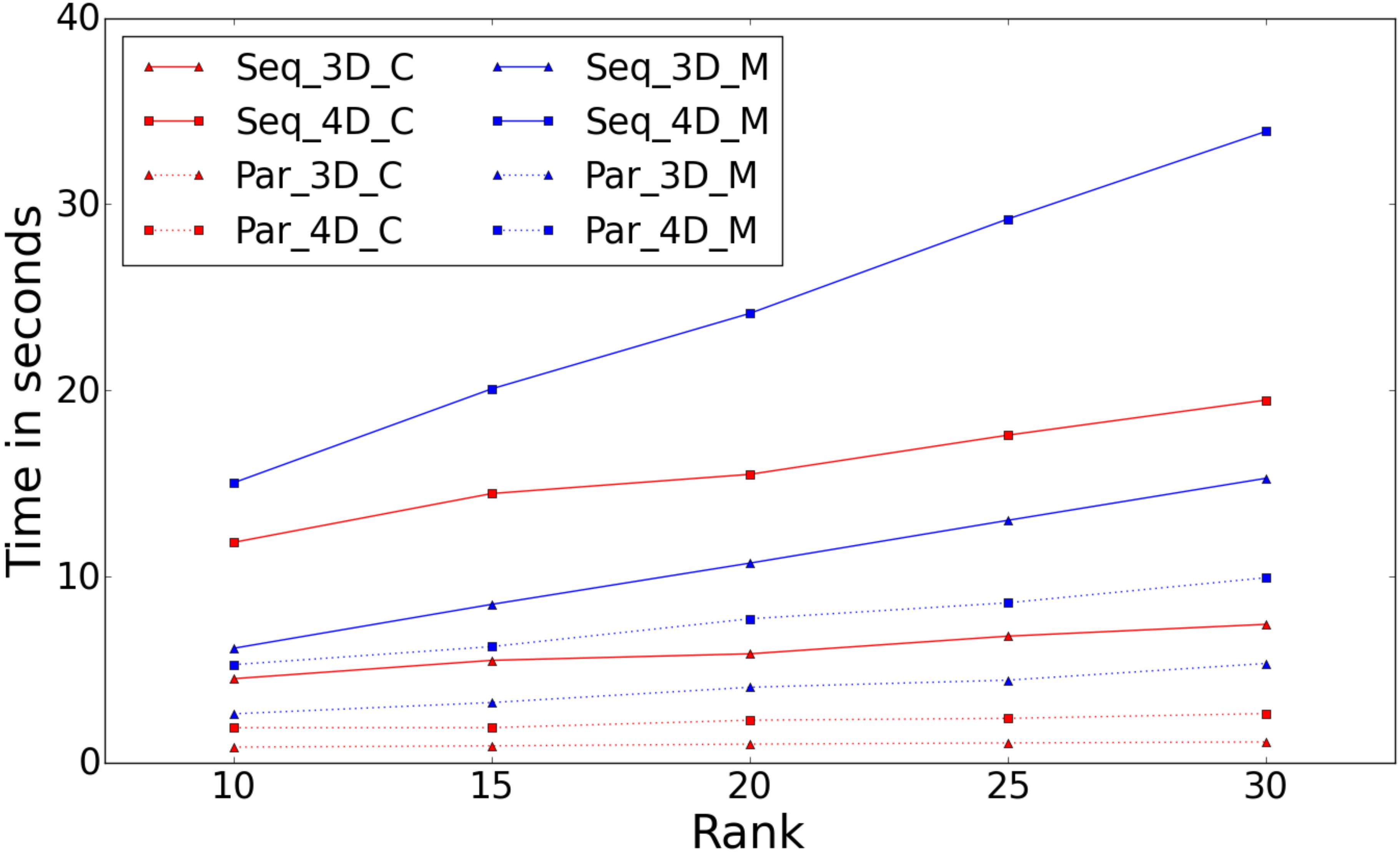
Parallel 2-Step MTTKRP

Call Parallel BLAS
WOW!!!



$60\times60\times60\times60\times60$

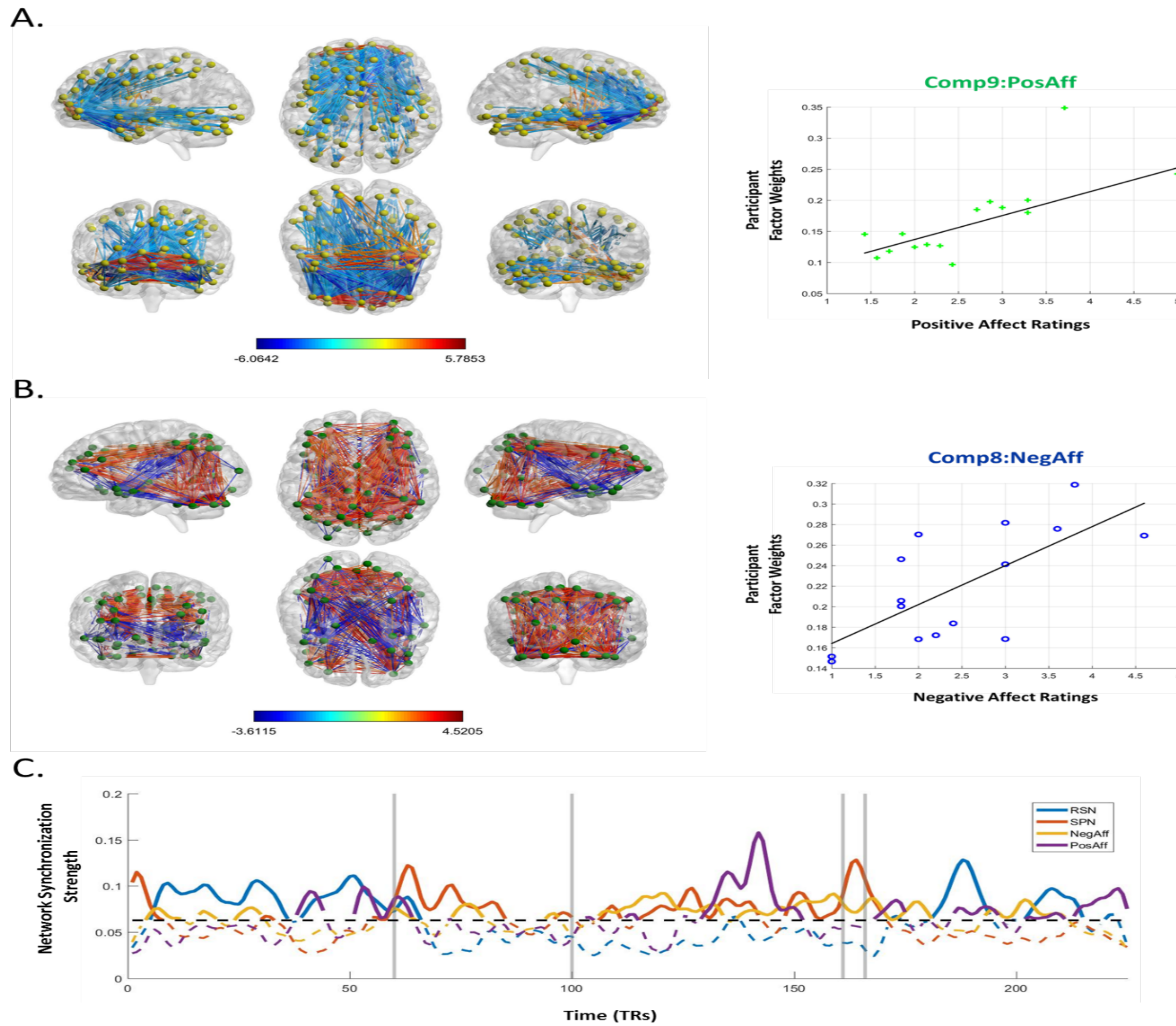

Baseline
Full KRP
DGEMM
DGEMV
Left & Right KRP
Reduce
2.0
1.5
1.0
0.5
0.0
B 1S 2S
0 1 2 3 4



Per iteration time of a CP decomposition via ALS. Matlab used the Tensor Toolbox cp_als function, version 2.6. [1]

Findings

- o interesting networks
  - Positive affect
  - Negative affect
- pia M., **Hayashi K.,** Ballard G., tlib I. *Dynamic Functional nnectivity and Individual ferences in Emotions During cial Stress* - to appear in man Brain Mapping



References

amara G. Kolda and Brett W. Bader. 2009. Tensor Decompositions and Applications. SIAM Rev. 51, 3 (Septem
009), 455–500. https://doi.org/10.1137/ 07070111X

ajia Li, Jee Choi, Ioakeim Perros, Jimeng Sun, and Richard Vuduc. 2017. Model Driven Sparse CP Decompositi
r Higher-Order Tensors. In IEEE International Parallel and Distributed Processing Symposium (IPDPS). 1048–1
tps://doi.org/10.1109/IPDPS.2017.80

haden Smith, Niranjay Ravindran, Nicholas D. Sidiropoulos, and George Karypis. 2015. SPLATT: Efficient and
arallel Sparse Tensor-Matrix Multiplication. In Proceedings of the 2015 IEEE International Parallel and Distribu
rocessing Symposium (IPDPS '15). IEEE Computer Society, Washington, DC, USA, 61–70. https://doi.org/10.11
DPS.2015.27

.C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T.E.J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S.V
urtiss, S. Della Penna, D. Feinberg, M.F. Glasser, N. Harel, A.C. Heath, L. Larson-Prior, D. Marcus, G. Michalare
Moeller, R. Oostenveld, S.E. Petersen, F. Prior, B.L. Schlaggar, S.M. Smith, A.Z. Snyder, J. Xu, and E. Yacoub. 2
e Human Connectome Project: a data acquisition perspective. Neuroimage 62, 4 (2012), 2222–2231. https:/
i.org/10. 1016/j.neuroimage.2012.02.018

nh-Huy Phan, Petr Tichavsky, and Andrzej Cichocki. 2013. Fast Alternating LS Algorithms for High Order
ANDECOMP/PARAFAC Tensor Factorizations. IEEE Transactions on Signal Processing 61, 19 (Oct 2013), 4834–
846. https://doi. org/10.1109/TSP.2013.2269903

End

Thanks for listening