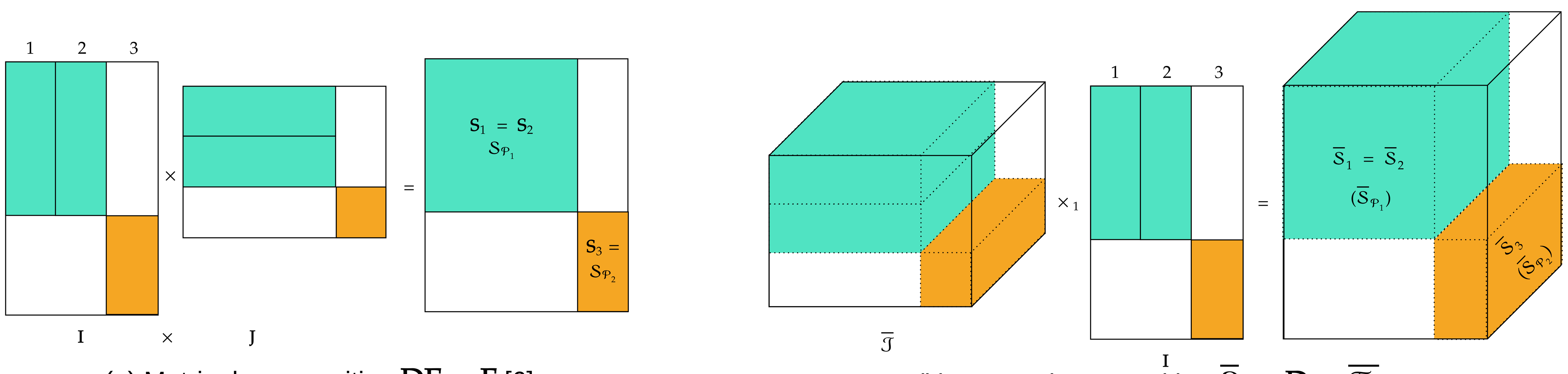
Tessellated Distributed Computing of Non-Linearly Separable Functions

Ali Khalesi*, Ahmad Tanha†, Derya Malak†, and Petros Elia†

* Institut Polytechnique des Sciences Avancées (IPSA), Paris, France.
† Communication Systems Department, EURECOM, Biot, Sophia Antipolis, France.
Emails: ali.khalesi@ipsa.fr, {tanha, malak, elia}@eurecom.fr

Motivation

- Accuracy in the analysis using real-valued floating-point arithmetic data representations.
- Computational overflows, quantization errors, and scalability barriers.



(a) Matrix decomposition $\mathbf{DE} = \mathbf{F}$ [6] (Two-dimensional tiles)

(b) Tensor decomposition $\bar{\mathcal{E}} \times_1 \mathbf{D} = \bar{\mathcal{F}}$ (High-dimensional tiles)

Capturing non-linearly separable functions by degrees

Related works

- Novel parallel computing frameworks to offload computations [1]
- Scalability [2, 3]
- Computation and communication efficient schemes [4, 5, 6]

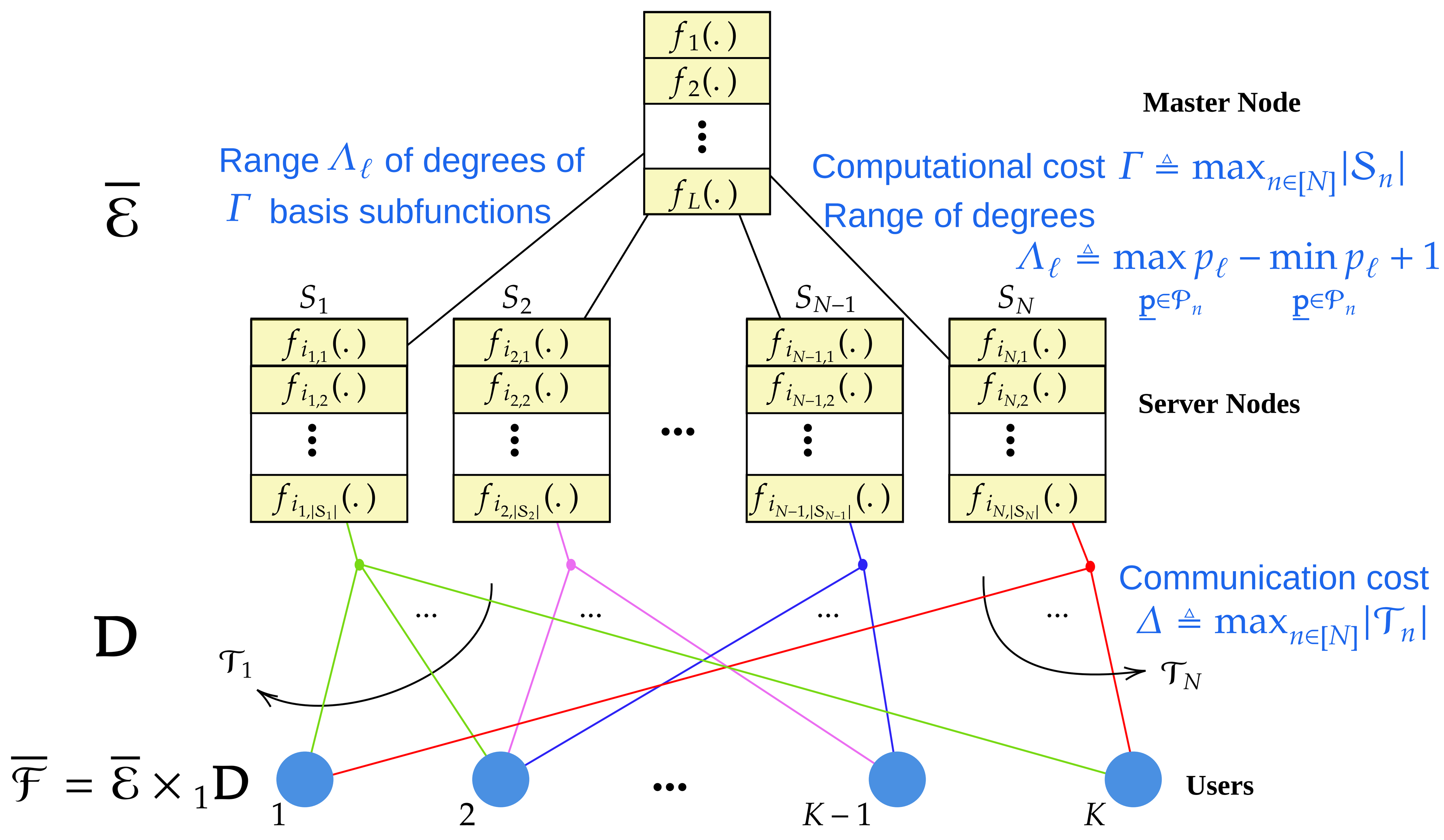
Our Contributions

- Fundamental limits of non-linear multi-user distributed-computation setting.
- Unearthing the connection between distributed computing and high-dimensional tiling, sparse tensor factorization, and large random matrices theory.

Future directions

- Optimal achievable rate
- Lossy setting with bounded error
- Converse bounds

System model and key parameters



$$\mathcal{P}_n \triangleq \{\underline{\mathbf{p}} \mid \textstyle\prod_{\ell\in\mathcal{S}_n} W_\ell^{p_\ell - 1} \text{ is computed in server } n\}, \quad \text{where} \quad \underline{\mathbf{p}} \triangleq (p_1, \ldots, p_L)$$

The system rate: $R \triangleq \frac{K}{N}$

The system capacity C: The supremum of all rates allowing error-free function reconstruction.

References

[1] Jeffrey Dean and Sanjay Ghemawat. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.

[2] Songze Li, Qian Yu, Mohammad Ali Maddah-Ali, and A Salman Avestimehr. A scalable framework for wireless distributed computing. *IEEE/ACM Trans. Netw.*, 25(5):2643–2654, 2017.

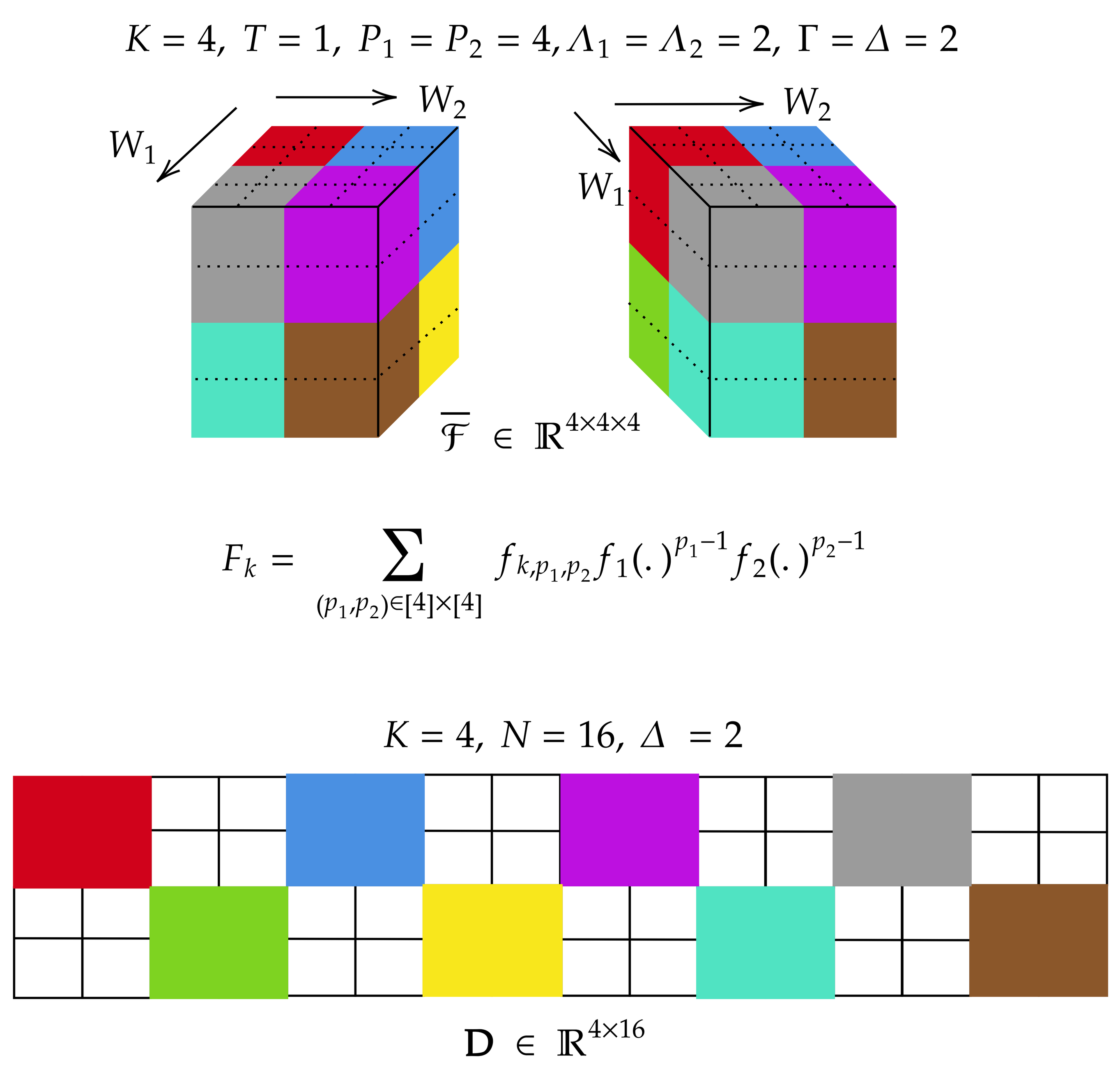
[3] Farzin Haddadpour, Mohammad Mahdi Kamani, Mehrdad Mahdavi, and Viveck Cadambe. Trading redundancy for communication: Speeding up distributed sgd for non-convex optimization. In *ICML*, pages 2545–2554. PMLR, 2019.

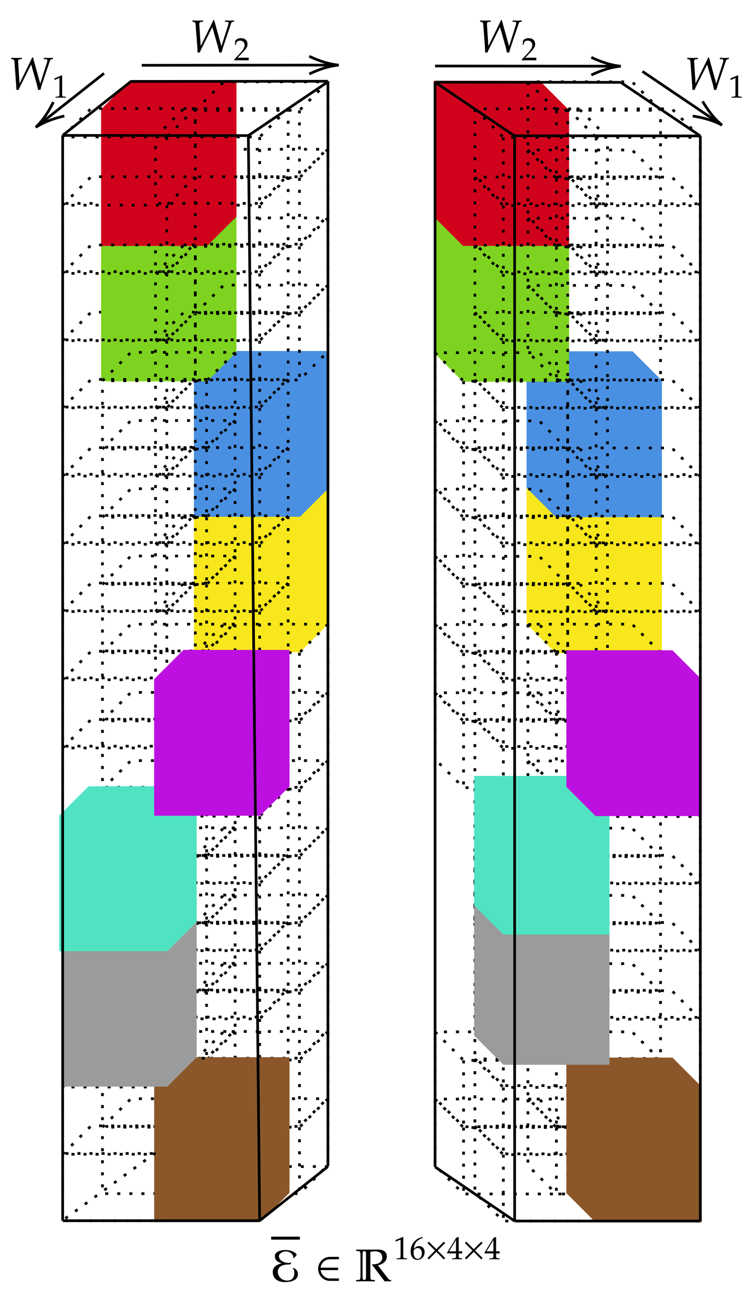
[4] Kai Wan, Hua Sun, Mingyue Ji, and Giuseppe Caire. Distributed linearly separable computation. *IEEE Trans. Inf. Theory*, 68(2):1259–1278, 2022.

[5] Qifa Yan, Sheng Yang, and Michèle Wigger. Storage-computation-communication tradeoff in distributed computing: Fundamental limits and complexity. *IEEE Trans. Inf. Theory*, 68(8):5496–5512, 2022.

[6] Ali Khalesi and Petros Elia. Tessellated distributed computing. *IEEE Trans. Inf. Theory*, 71(6):4754–4784, 2025.

Achievable lossless reconstruction: the tensor decomposition approach





- $\bar{\mathcal{F}}$ into 8 three-dimensional tiles of size $\Delta \times \Lambda_1 \times \Lambda_2$.
- The sparse tiling of $\mathbf{D}$ and $\bar{\mathcal{E}}$ with tiles $\mathbf{L}_j$ and $\bar{\mathcal{R}}_j$, respectively.
- $\bar{\mathcal{F}}$ is covered by the eight $\bar{\mathcal{S}}_j = \bar{\mathcal{R}}_j \times_1 \mathbf{L}_j$, $j \in [8]$, guaranteeing sparsity $\delta = \lambda_1 = \lambda_2 = \frac{1}{2}$, $\gamma = 1$ for $\mathbf{D}$ and $\bar{\mathcal{E}}$, respectively. This yields $N = 8 \times 2 = 16$.
- $L_M = \prod_{\ell\in[L]} P_\ell - 1 = 15$ basis subfunctions by the matrix factorization approach [6]. The number of required servers with the same computation and communication cost:

$$\frac{K}{\Delta}\left[\left\lfloor\frac{L_M}{\Gamma}\right\rfloor \times \frac{\min(\Delta,\Gamma)}{T} + \frac{\min(\Delta, \operatorname{mod}(L_M,\Gamma))}{T}\right] = 30\,.$$

- The achievable gain in the number of required servers: $\approx 47\%$.