NTK-DFL: Enhancing Decentralized Federated Learning in Heterogeneous Settings via Neural Tangent Kernel

*Gabriel Thompson**, Kai Yue, Chau-Wai Wong, Huaiyu Dai

Presented at ICML 2025

July 17, 2025

Overview

Introduction

NTK-DFL Overview

Results

Takeaway + Future Directions

Introduction

- **Federated Learning (FL)** jointly trains a machine learning model across multiple clients, without sharing training data
- **Decentralized Federated Learning (DFL)** does this without a central server

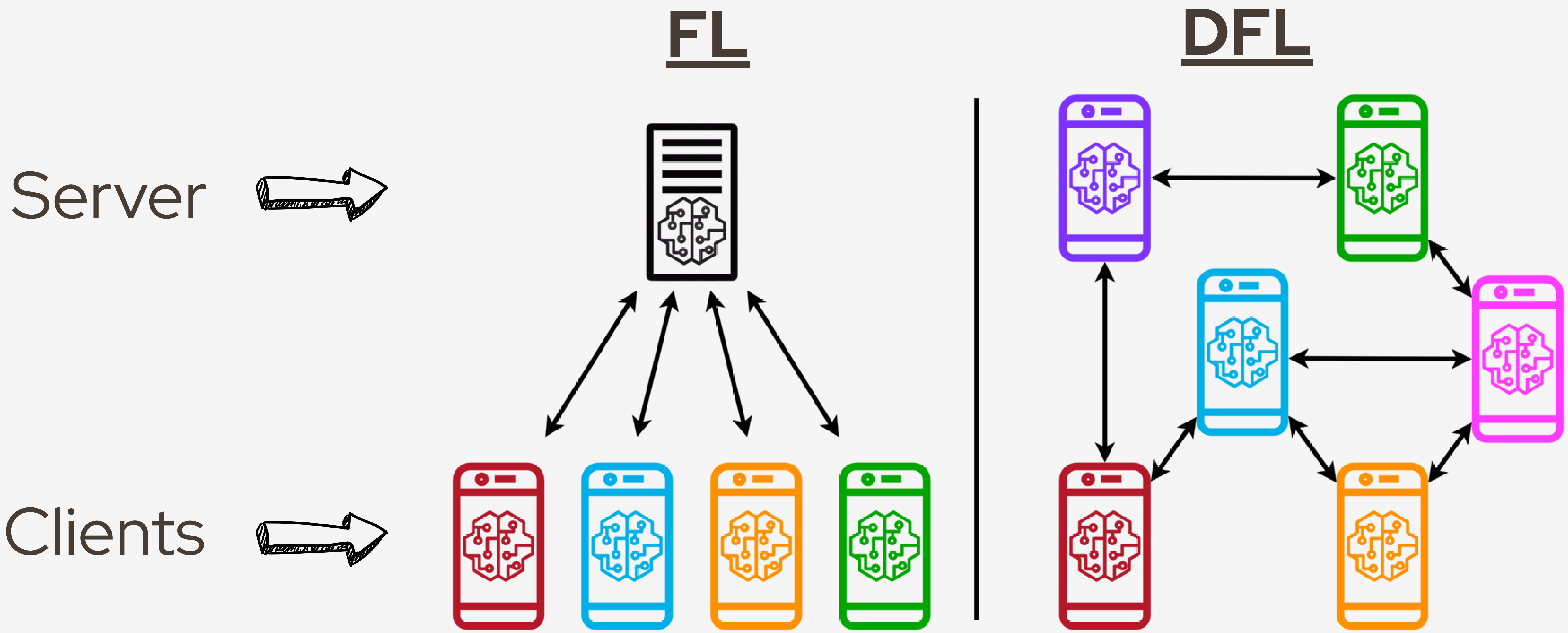


Image credit: https://marc.damie.eu

Data Heterogeneity

- Clients posses various, non-IID data distributions

- This can disrupt training and hurt convergence

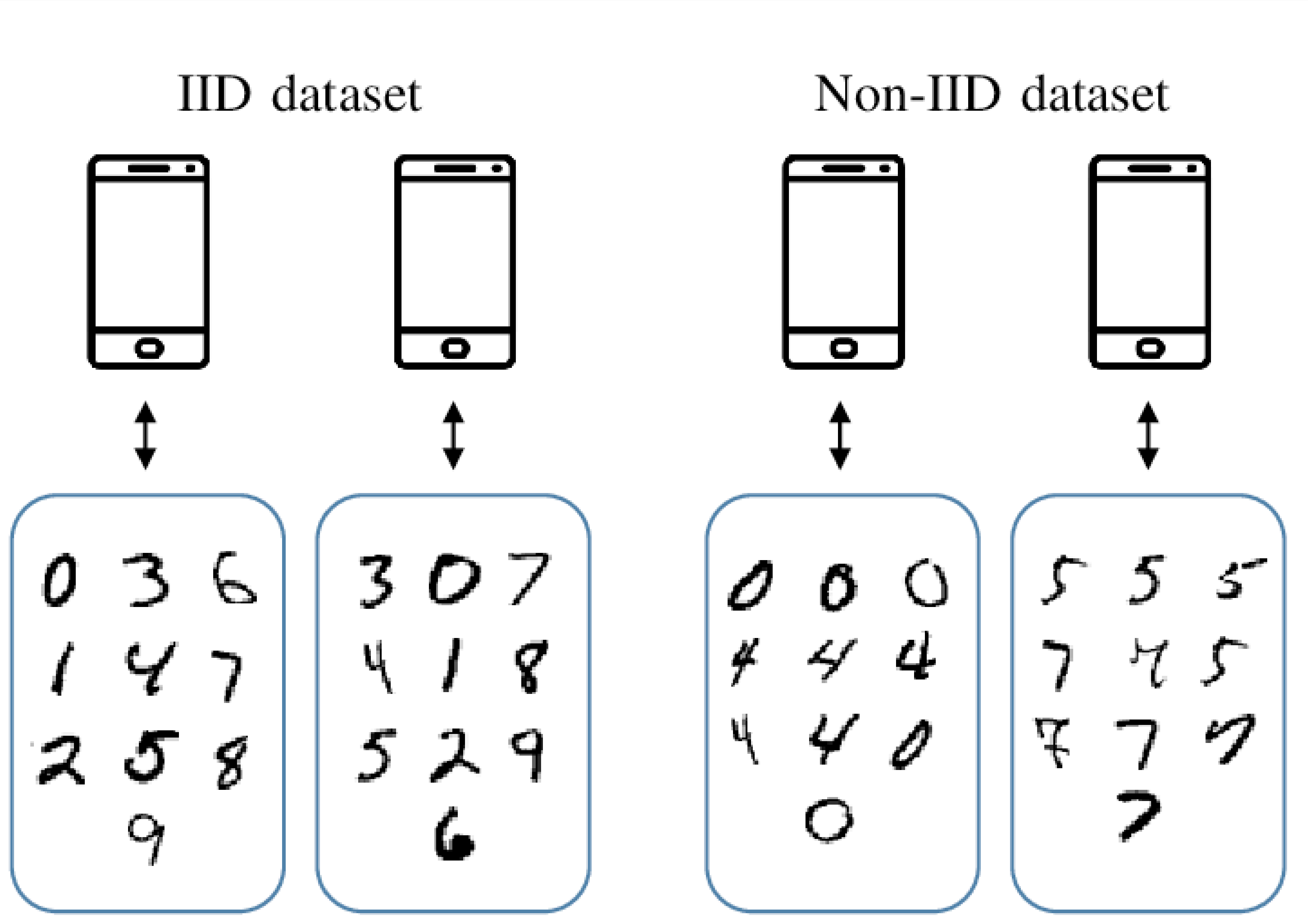


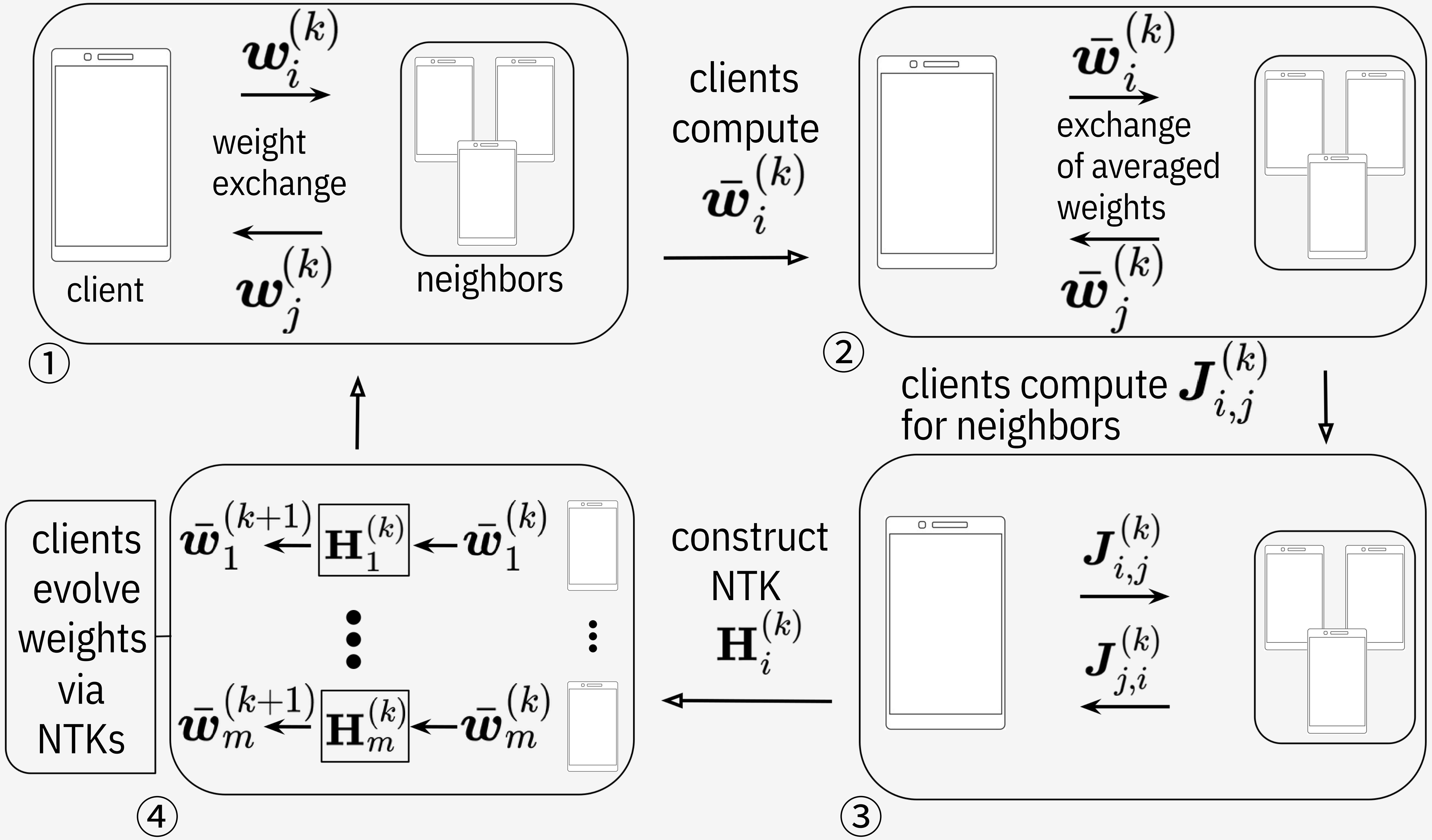
Image credit: Hellström, Henrik & Barros da Silva Júnior, José Mairton & Fodor, Viktoria & Fischione, Carlo. (2020). Wireless for Machine Learning.

NC STATE UNIVERSITY

NTK-DFL Overview

ICML International Conference On Machine Learning

- Yue et al. (2022) showed success of **neural tangent kernel** methods in centralized federated learning
- We replace gradient-based optimization from Sun et al. (2021) with neural tangent kernel optimization



- Yue et al. Neural tangent kernel empowered federated learning. 39th International Conference on Machine Learning, Jul 2022.
- Sun, T., Li, D., and Wang, B. Decentralized federated averaging. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021

NTK-DFL Overview

- NTK-DFL allows clients to communicate **more expressive updates**
- **Decreases** the number of communication rounds by **increasing** the communication volume per round

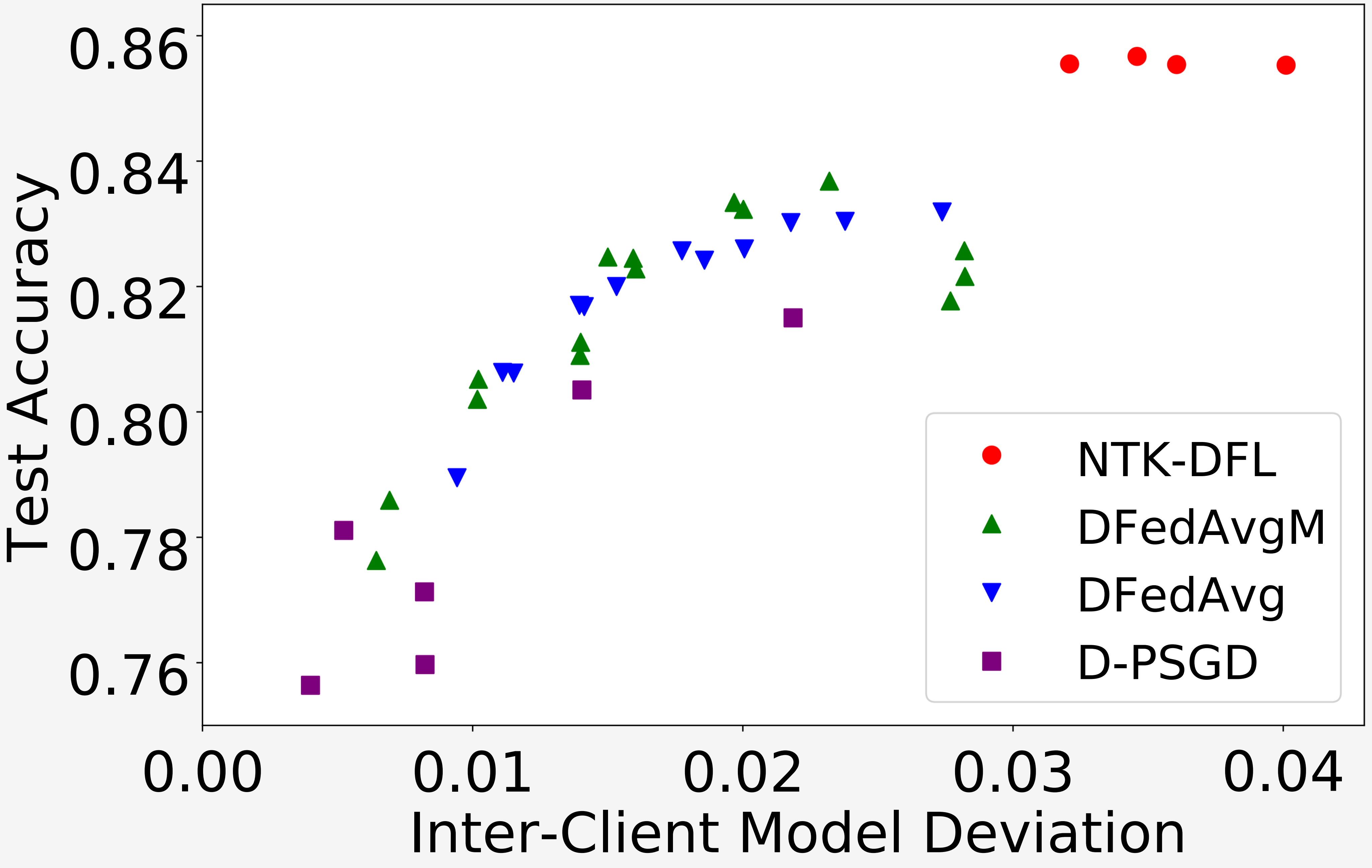
Useful in scenarios with:

- Large per-round communication latency
- Limited device availability
- Synchronization delays

NTK-DFL Overview

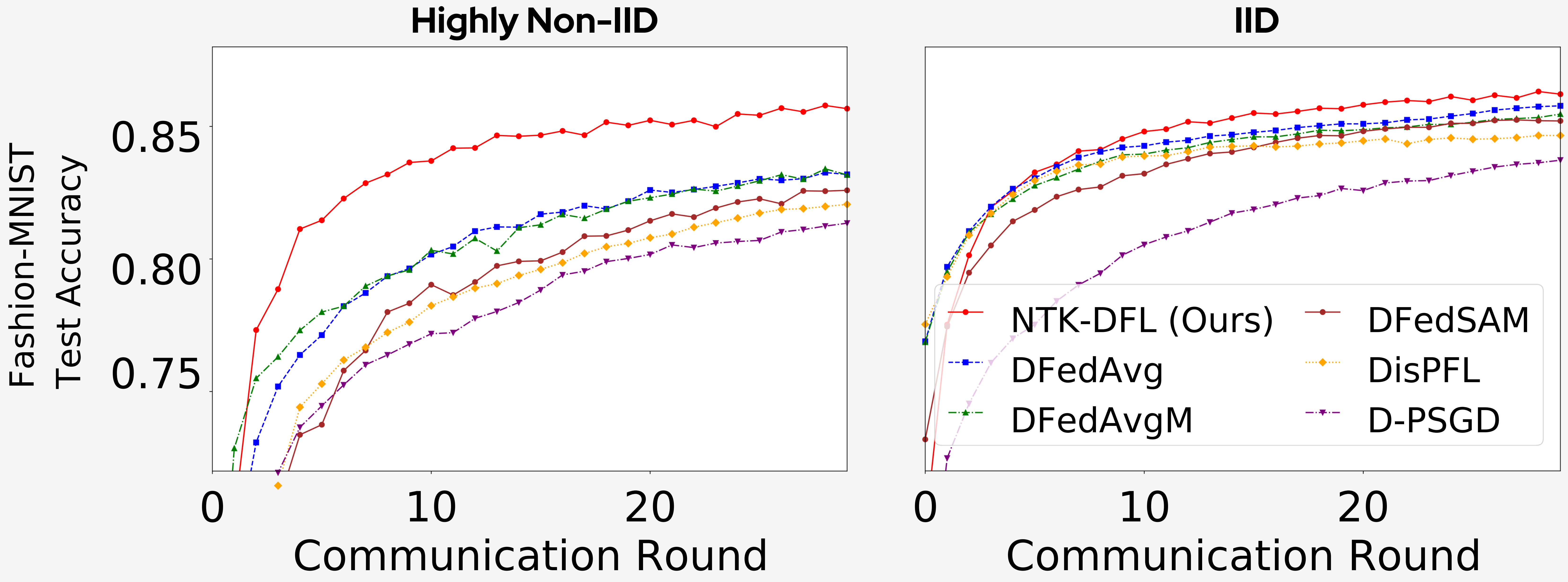
*NTK weight evolution is **synergetic** with model averaging in DFL*

- NTK updates encourage a useful **inter-client model deviation**, improving generalization in the decentralized setting



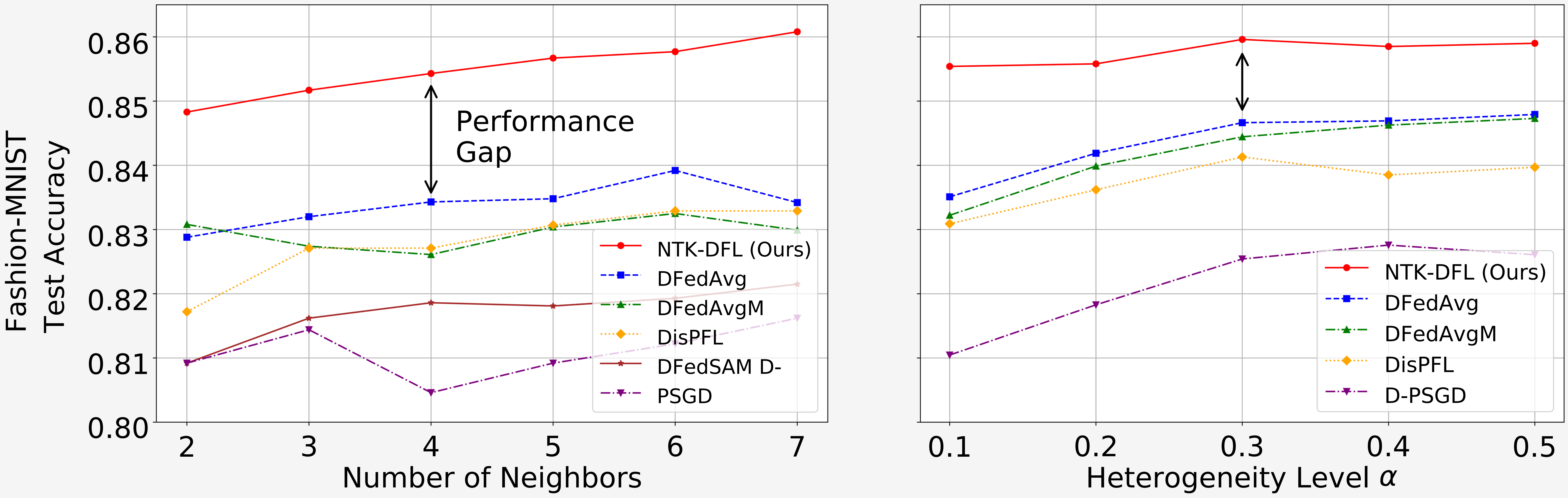
Results

- NTK-DFL **reduces communication rounds** needed for convergence
- Particularly effective with **non-IID** data



Results

- NTK-DFL performs consistently better across various **heterogeneity levels** and **topological sparsity** compared to SOTA methods



NC STATE UNIVERSITY

ICML International Conference On Machine Learning

Results

Theoretical analysis highlights possible source of improved convergence compared to DFedAvg

NTK-DFL (Proposed) ⇒

$$\min_{1\le k\le K}\left\|\nabla\mathcal{L}\left(\bar{w}^{(k)}\right)\right\|^2 \le O\left(\frac{\mathcal{L}(\bar{w}^{(1)})-\mathcal{L}^*}{\sqrt{KT}}+\dots\right)$$

vs.

DFedAvg (Sun et al., 2021) ⇒

$$\min_{1\le k\le K}\left\|\nabla\mathcal{L}\left(\bar{w}^{(k)}\right)\right\|^2 \le O\left(\frac{\mathcal{L}(\bar{w}^{(1)})-\mathcal{L}^*}{\sqrt{K}}+\dots\right)$$

Takeaway + Future Directions

Takeaway

- NTK evolution enables more expressive model updates
    - Trades *communication volume* for *communication rounds*
- NTK-DFL is synergetic with DFL model averaging

Future Directions

- Adopt to more modern ML architectures
- Apply NTK-DFL to suitable real-world applications (cross-silo DFL, healthcare)

THANK YOU

Please refer to our paper for more experimental results, ablation studies, the theoretical analysis, and analysis of communication overhead.

Gabriel Thompson, Kai Yue, Chau-Wai Wong, and Huaiyu Dai, "NTK-DFL: Enhancing decentralized federated learning in heterogeneous settings via neural tangent kernel," ICML 2025.