

Revisiting Ensembling in One-Shot Federated Learning

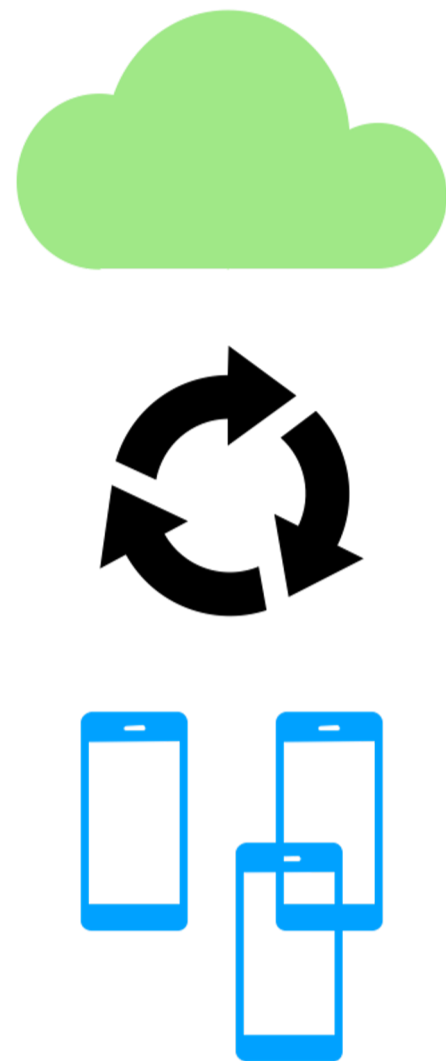
Youssef Allouah¹, Akash Dhasade¹, Rachid Guerraoui¹, Nirupam Gupta²,
 Anne-Marie Kermarrec¹, Rafael Pinot³, Rafael Pires¹, Rishi Sharma¹
¹EPFL, Switzerland ²University of Copenhagen
³Sorbonne Université and Université Paris Cité, CNRS, LPSM



Motivation

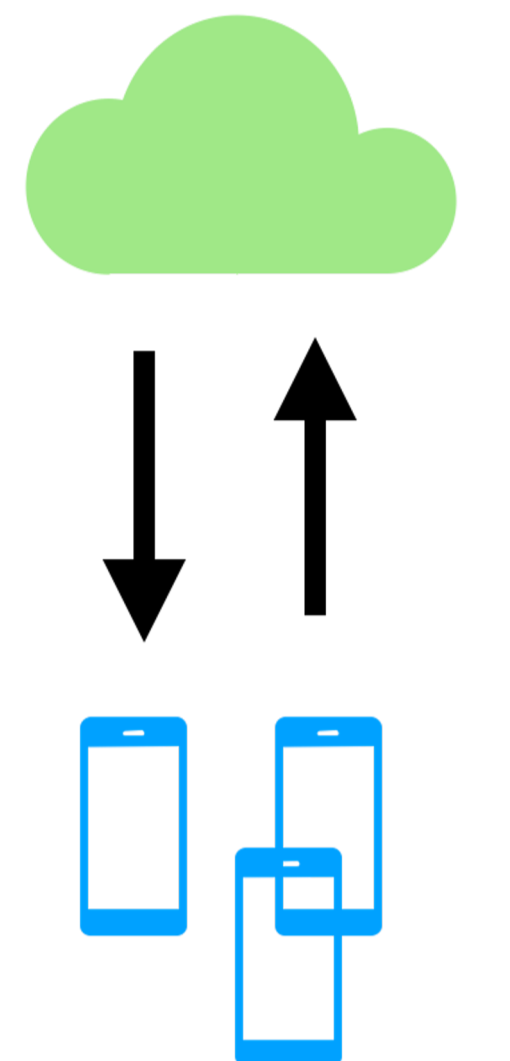
Federated Learning (FL)

1. **Iterative** training of models
2. Generally, achieves **good accuracy**
3. **Significant communication costs**



One-Shot Federated Learning (OFL)

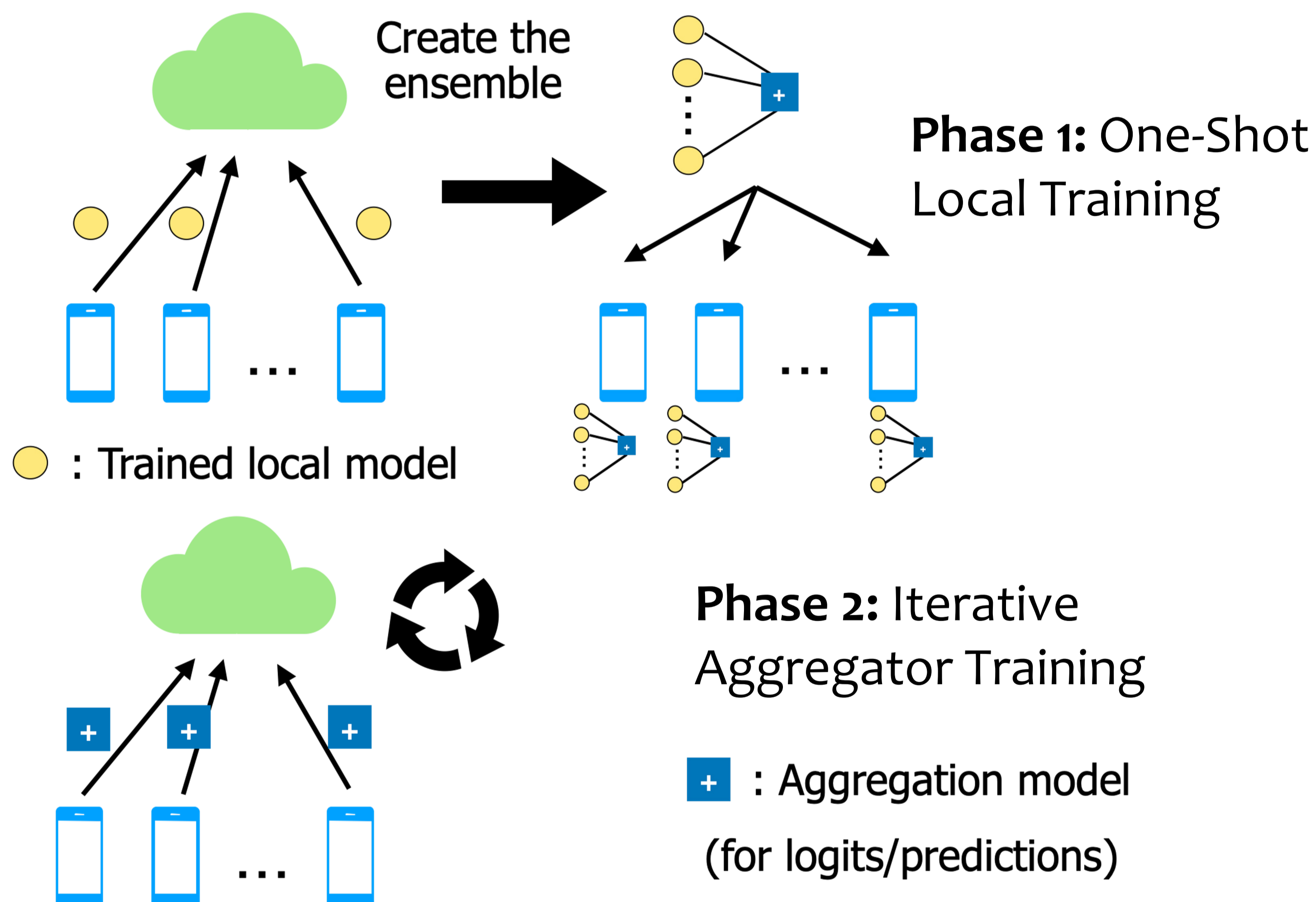
1. **One round** of transfers
2. **Low communication costs**
3. **Accuracy drop** compared to FL



How do we neatly combine the two to achieve best of the both?

Fens: Hybrid of FL and OFL

Training in Two Phases



Fens: Key Properties

Fens **stacks an aggregator model** (a neural network) atop client local models in the ensemble

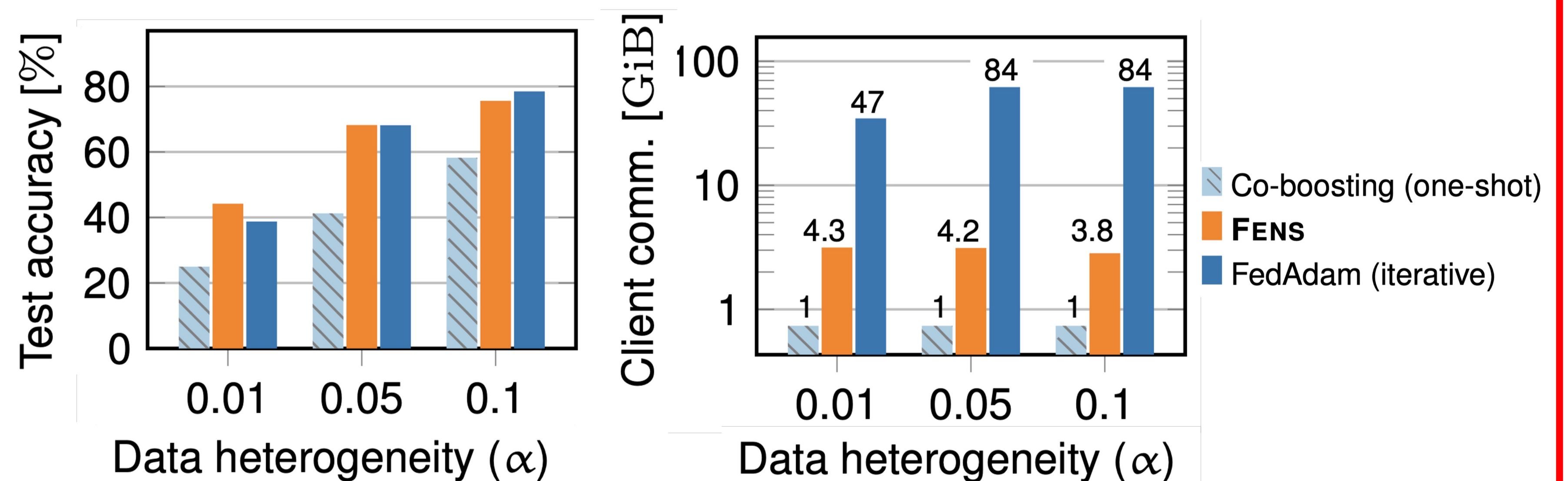
1. **MLP suffices** as the aggregator
 2. Size of aggregator \ll Size of client model
 3. Phase 2 induces very low additional communication costs
1. Stacked model is trained on two separate partitions of local data
 2. Aggregator training effectively resolves biases of client local models^[1]
 3. **Stacking significantly improves generalization**

Low comm. costs

Improved generalization

Evaluation

1. **20 clients, ResNet-8**
2. **CIFAR-10 Non-IID Partitioning**
 - Dirichlet Distribution ($\alpha = 0.01, 0.05, 0.1$)
3. **Co-Boosting^[2] (OFL) and FedAdam^[3] (FL)**
 - State-of-the-art algorithms in each family
4. **Fens (almost) achieves FedAdam's accuracy**
 - Communication cost $\sim 4x$ that of OFL, FL is 47-84x that of OFL



[1] Wolpert, D.H., 1992. Stacked generalization. *Neural networks*, 5(2), pp.241-259.

[2] Dai, Rong, et al. "Enhancing One-Shot Federated Learning Through Data and Ensemble Co-Boosting." In ICLR 2024.

[3] Reddi, Sashank, et al. "Adaptive federated optimization." In ICLR 2021.

