Epidemic Learning: Boosting Decentralized Learning with Randomized Communication

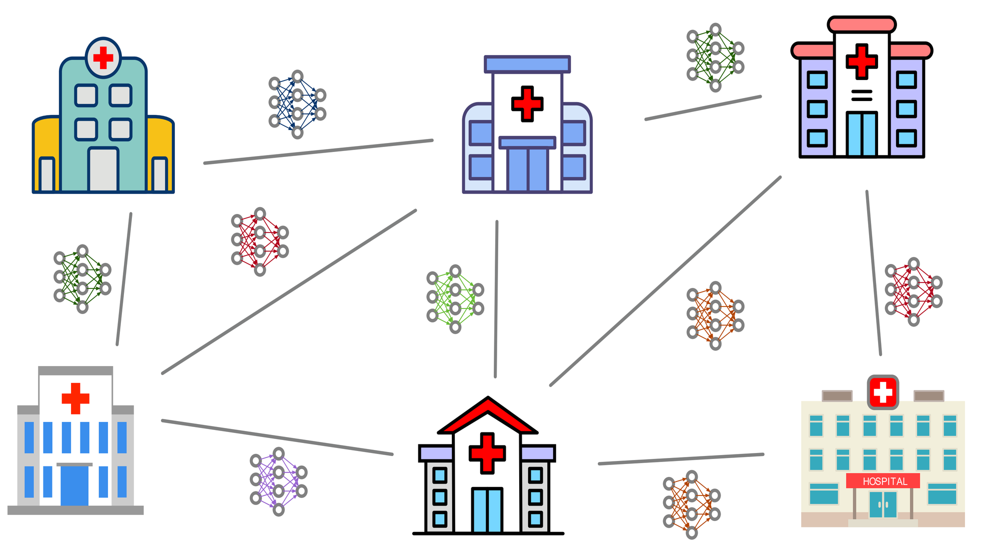
Martijn de Vos, Sadegh Farhadkhani, Rachid Guerraoui, Anne-Marie Kermarrec, Rafael Pires, Rishi Sharma

EPFL, Switzerland

Motivation

Decentralized Learning

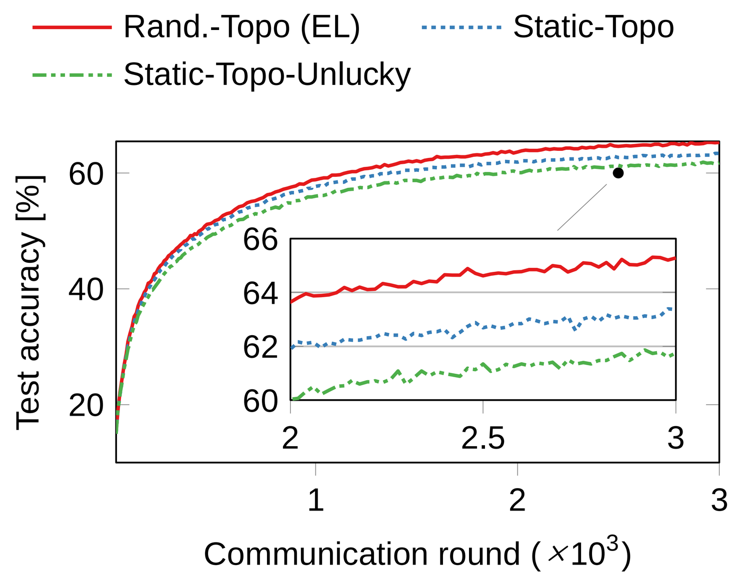
1. **Peer-to-peer** network of n nodes
2. **Data** stays where it is produced
3. **Neighbors** iteratively train and exchange models



DL at node i: Train → Share → Aggregate

Learning Topology

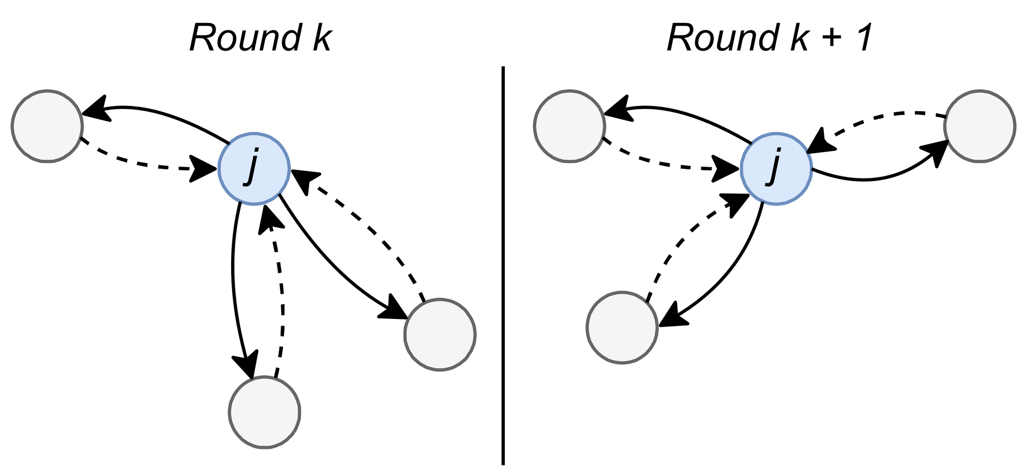
1. Topology affects the **convergence speed**
2. Convergence can be boosted through **randomization**
3. Randomization through **peer-samplers** [1]



DL with 96 nodes on CIFAR-10

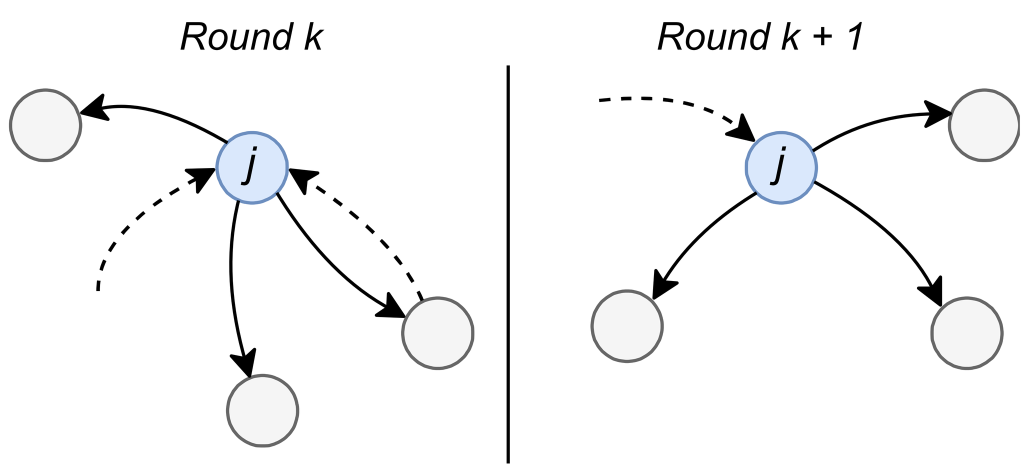
Epidemic Learning (EL)

Nodes randomly sample neighbors in each round



EL-Oracle (forming a s-regular graph)

- Balanced
- Global coordination



EL-Local (forming a s-out graph)

- Local decision
- (Slightly) unbalanced

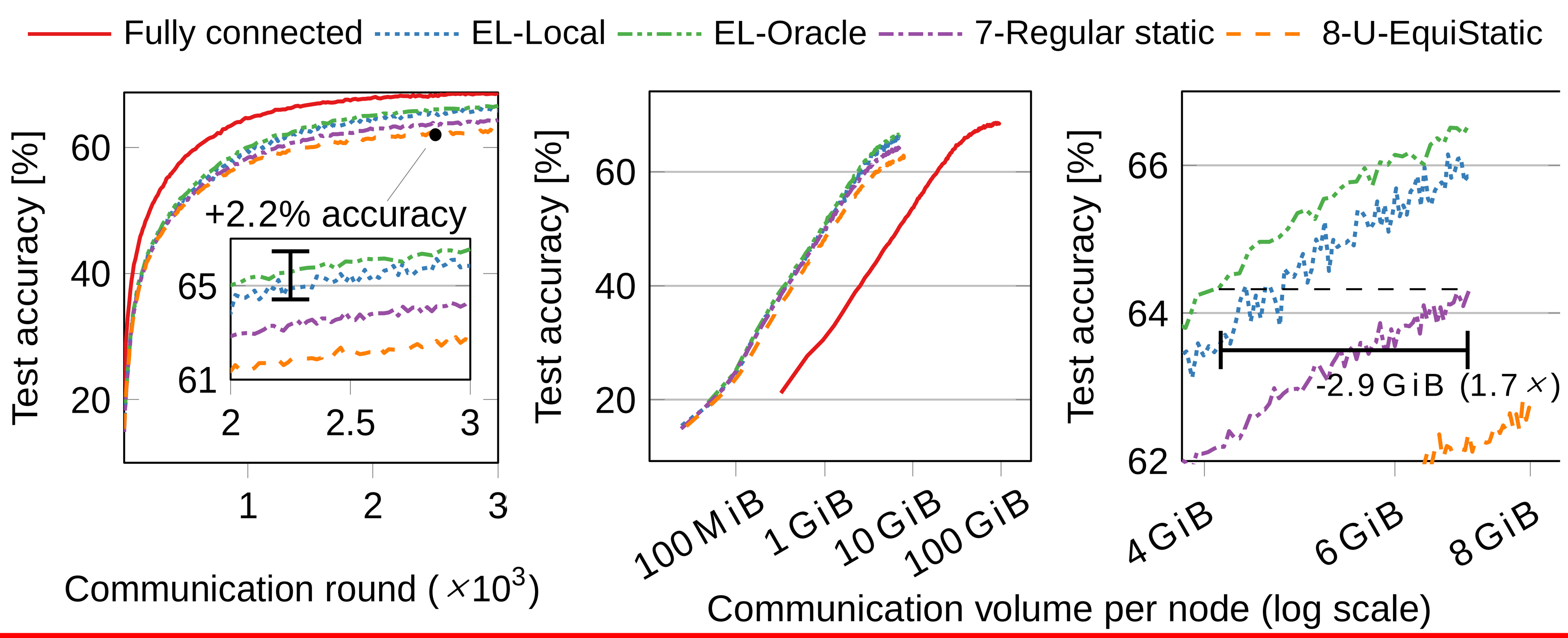
Convergence Guarantee

$$O\left(\frac{1}{\sqrt{nT}} + \frac{1}{\sqrt[3]{sT^2}} + \frac{1}{T}\right)$$

1. **Linear speedup**
   - First term: preserved from D-PSGD[2]
2. **Transient iterations**
   - Superior second term: $O\left({n^3}/{s^2}\right)$
   - Number of rounds for the first term to dominate
3. **Assumptions**
   - Smooth non-convex loss with bounded stochastic noise and data heterogeneity

Evaluation

1. **96 node Decentralized Learning**
   - Fully connected is the upper bound: high comm.
2. **CIFAR-10 Non-IID Partitioning**
   - Dirichlet Distribution ($\alpha = 0.1$)
3. **GN-LeNet with SGD**
4. **EL outperforms baselines**
   - Higher accuracy at a lower cost



[1] Jelasity, Márk, et al. "Gossip-based peer sampling" ACM *transactions on computer systems*, 2007.
[2] Lian, Xiangru, et al. "Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent." *Advances in neural information processing systems*. NeurIPS, 2017.

EPFL

NEURAL INFORMATION PROCESSING SYSTEMS