Boosting Asynchronous Decentralized Learning with Model Fragmentation

Sayan Biswas[1], Anne-Marie Kermarrec[1], Alexis Marouani[2], Rafael Pires[1], Rishi Sharma[1], Martijn de Vos[1]

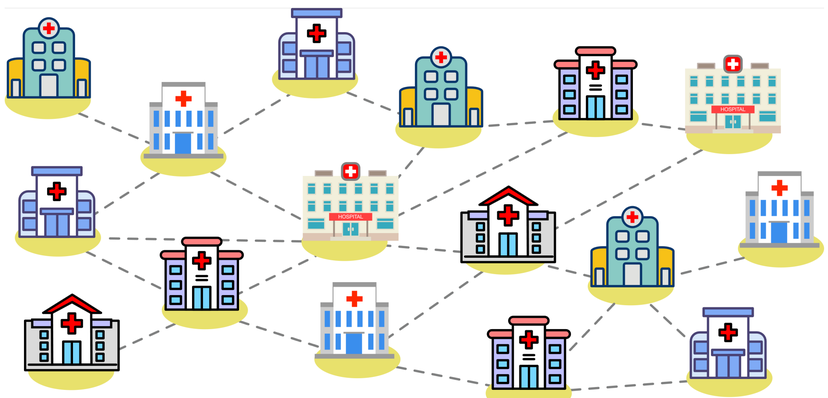
[1] EPFL, [2] École Polytechnique

Artifacts Available

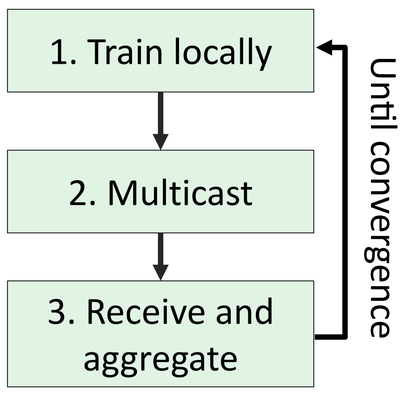
Full paper

Decentralized Learning (DL)

Collaborative learning without a central server [1]



▲ DL Topology

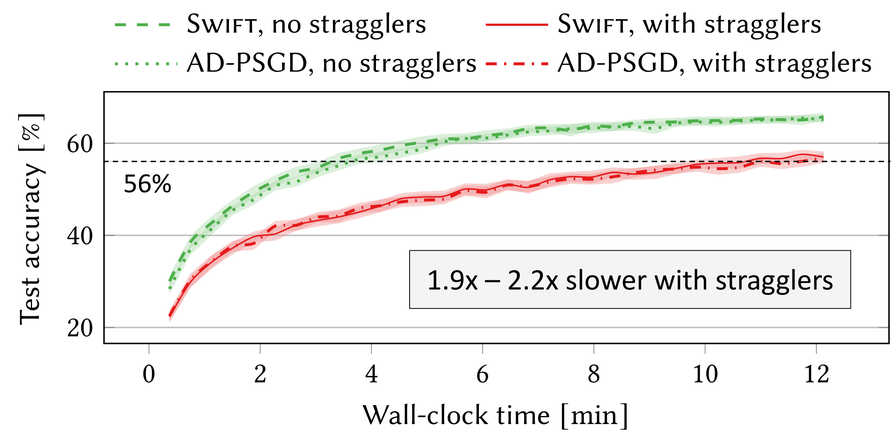


▲ DL Workflow

Asynchronous DL: nodes proceed independently without waiting for others

Problem: slow nodes (stragglers)

! Stragglers adversely affect convergence

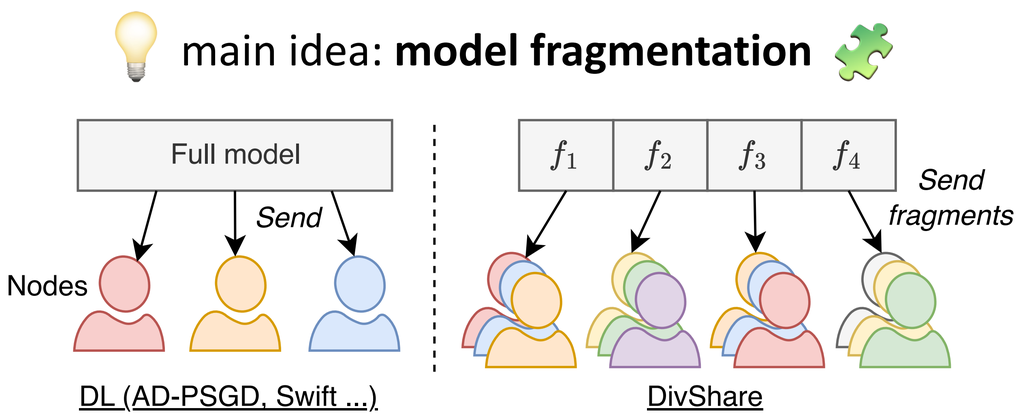


▲ Convergence of SOTA asynchronous DL with stragglers

? **How can we design a DL algorithm that can deal with communication stragglers?**

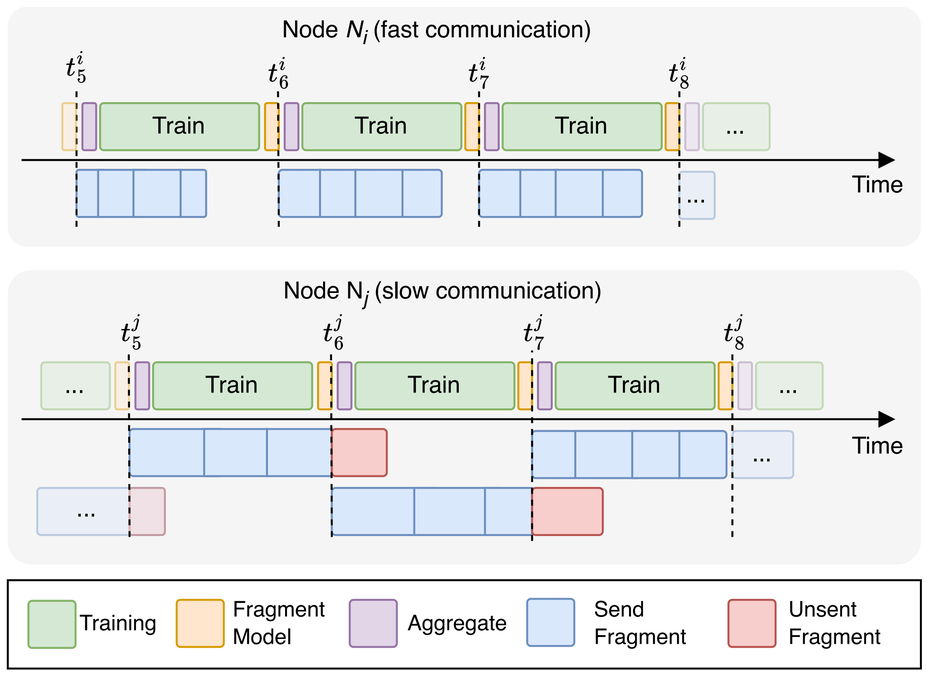
Our solution: DivShare

main idea: **model fragmentation**



- Allow slow nodes to contribute some of their model updates
- Sending *less* model updates to *more* nodes converges quicker
- Overlap communication and computation

Convergence analysis in the paper



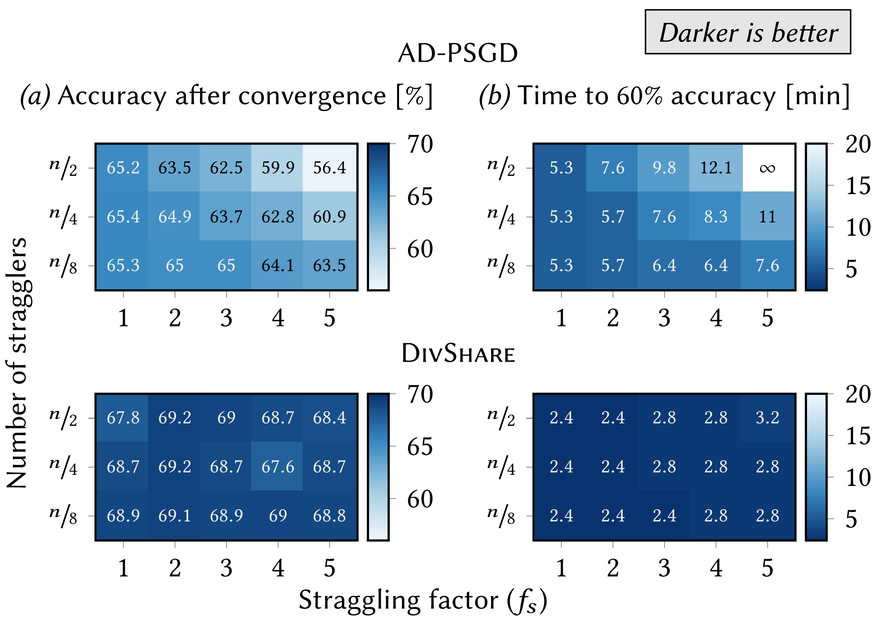
▲ DivShare timeline

Evaluation

Datasets	CIFAR-10 & MovieLens
Data distribution	Non-IID
Network size (n)	60 nodes
Topology degree (J)	6
Baselines	AD-PSGD [2] and Swift [3]

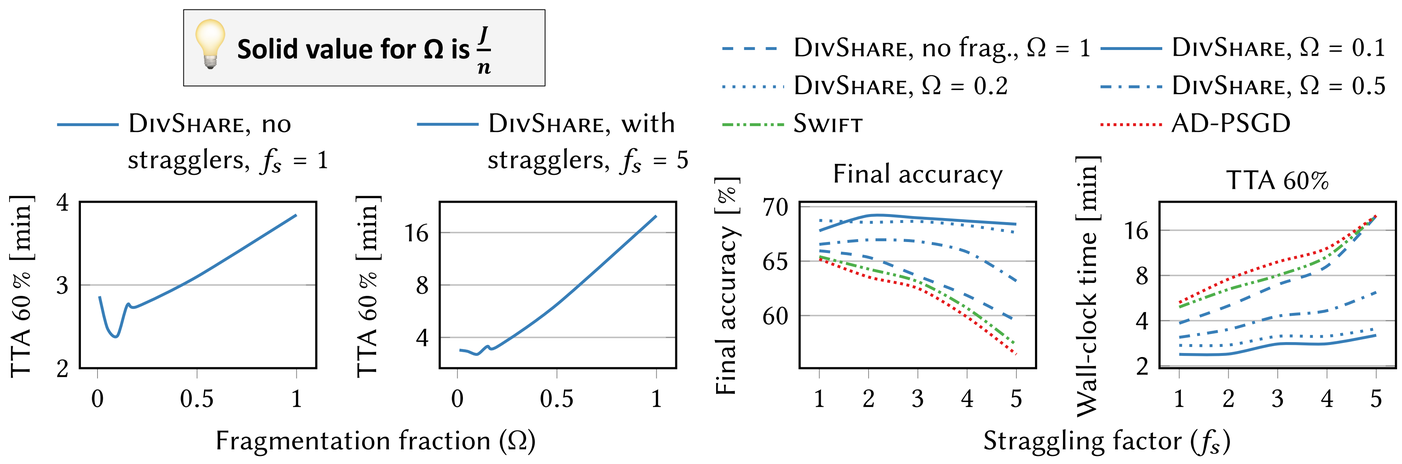
▲ Experiment setup

Darker is better



▲ Performance of DivShare against AD-PSGD

Solid value for Ω is $\frac{J}{n}$



▲ Varying Ω

▲ DivShare against baselines for varying values of Ω

Experiments show that DivShare lowers time-to-accuracy by up to **3.9x** and yields up to **19.4%** better test accuracy.

DivShare enables robust asynchronous DL, improving test accuracy and convergence time.

[1] Lian, Xiangru, et al. "Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent." *NeurIPS (2017)*.
[2] Lian, Xiangru, et al. "Asynchronous decentralized parallel stochastic gradient descent." *ICML (2018)*.
[3] Bornstein, Marco, et al. "SWIFT: Rapid Decentralized Federated Learning via Wait-Free Model Communication." *ICLR (2023)*.