Energy-Aware Decentralized Learning with Intermittent Model Training

Martijn de Vos[1], Akash Dhasade[1], Paolo Dini[2], **Elia Guerra[1,2]**, Anne-Marie Kermarrec[1], Marco Miozzo[2], Rafael Pires[1], and Rishi Sharma[1]

[1]**Scalable Computing Systems Laboratory, École polytechnique fédérale de Lausanne (EPFL)**

[2]**Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)**

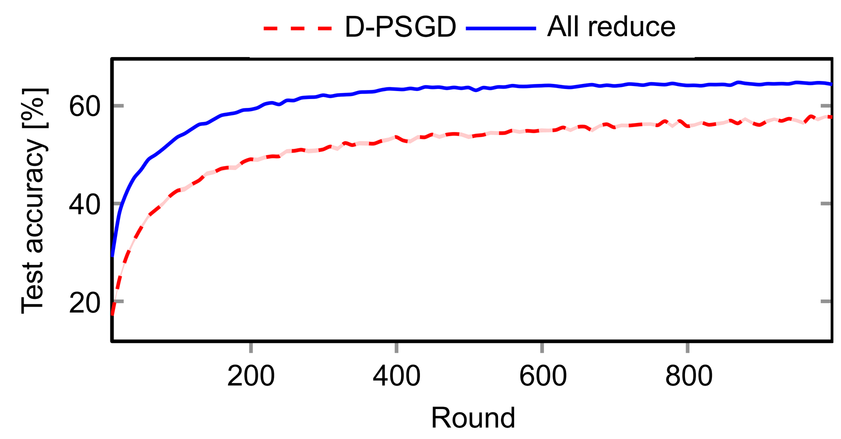
Introduction

Setting

- In decentralized learning (DL), nodes have intermittent network and constrained energy budgets
- Several efforts in the literature to improve communication efficiency
- Energy spent on the training process is often overlooked

Insights

- Training is more energy-demanding than sharing and aggregation
- DL convergence is faster with all-reduce aggregation
- Energy budgets distributed over the duration of the learning process



SkipTrain

Synchronization rounds

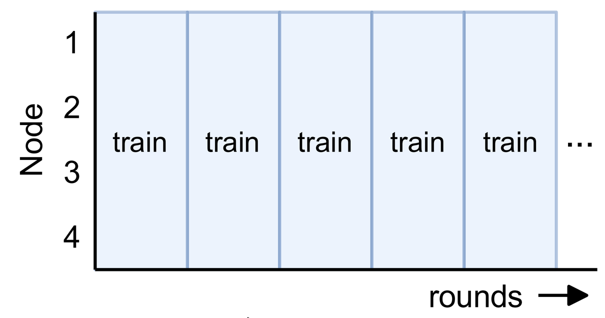
- In D-PSGD [1] each node iteratively executes training, sharing, and aggregation
- **Synchronization rounds:** sharing and aggregation
  - Each model converges towards the global average model [2]
- **Training rounds:** training, sharing, and aggregation
  - Train the model on the local dataset

Participation probabilities

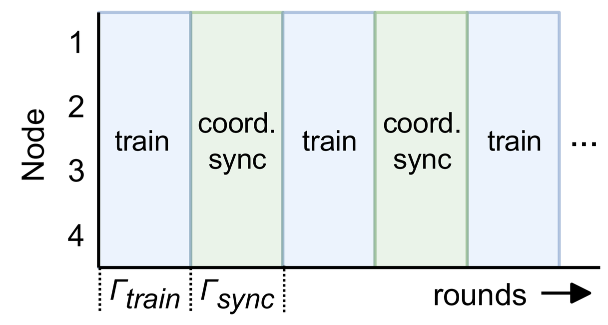
- Each client has an upper bound on the percentage of battery that can be used for training
- Participation probabilities distribute the energy budgets (τ_i) across total number of rounds (T)

$$T_{\text{train}} = \frac{\Gamma_{\text{train}}}{\Gamma_{\text{train}}+\Gamma_{\text{sync}}} T$$

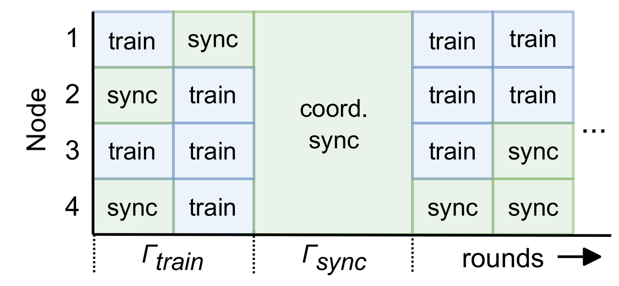
$$p_i = \min\left(\frac{\tau_i}{T_{\text{train}}}, 1\right)$$



1) D-PSGD



2) SkipTrain

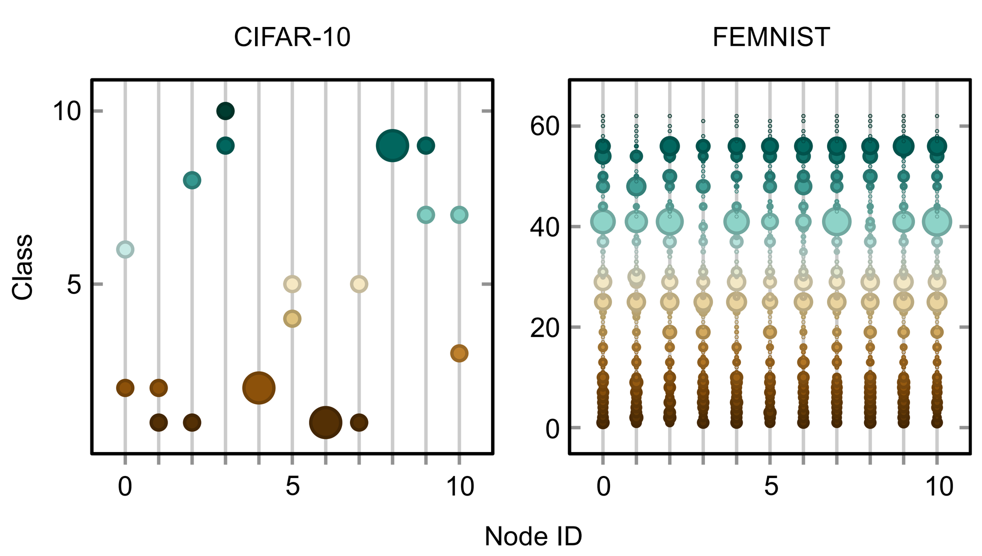


3) SkipTrain-Constrained

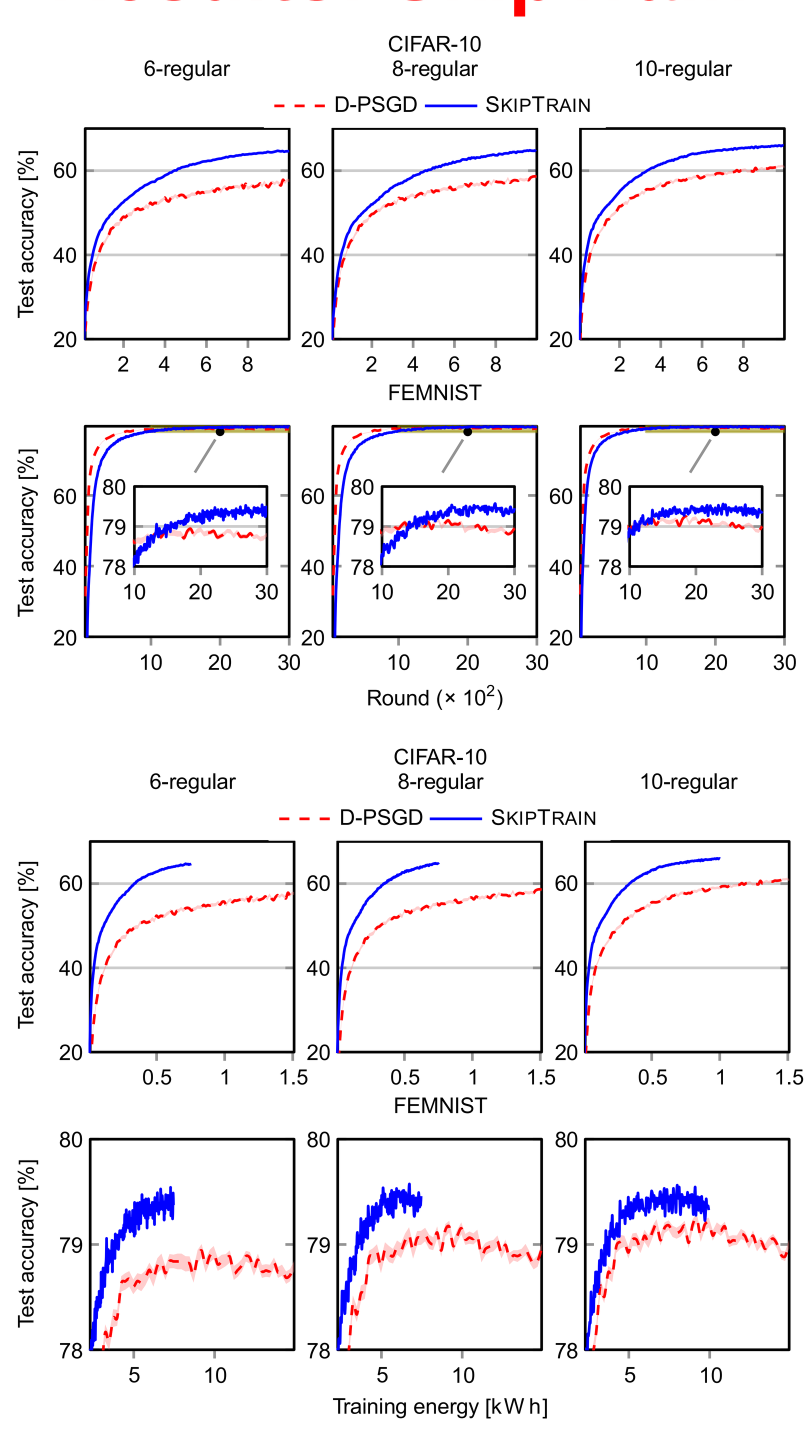
Evaluation

Experimental setting

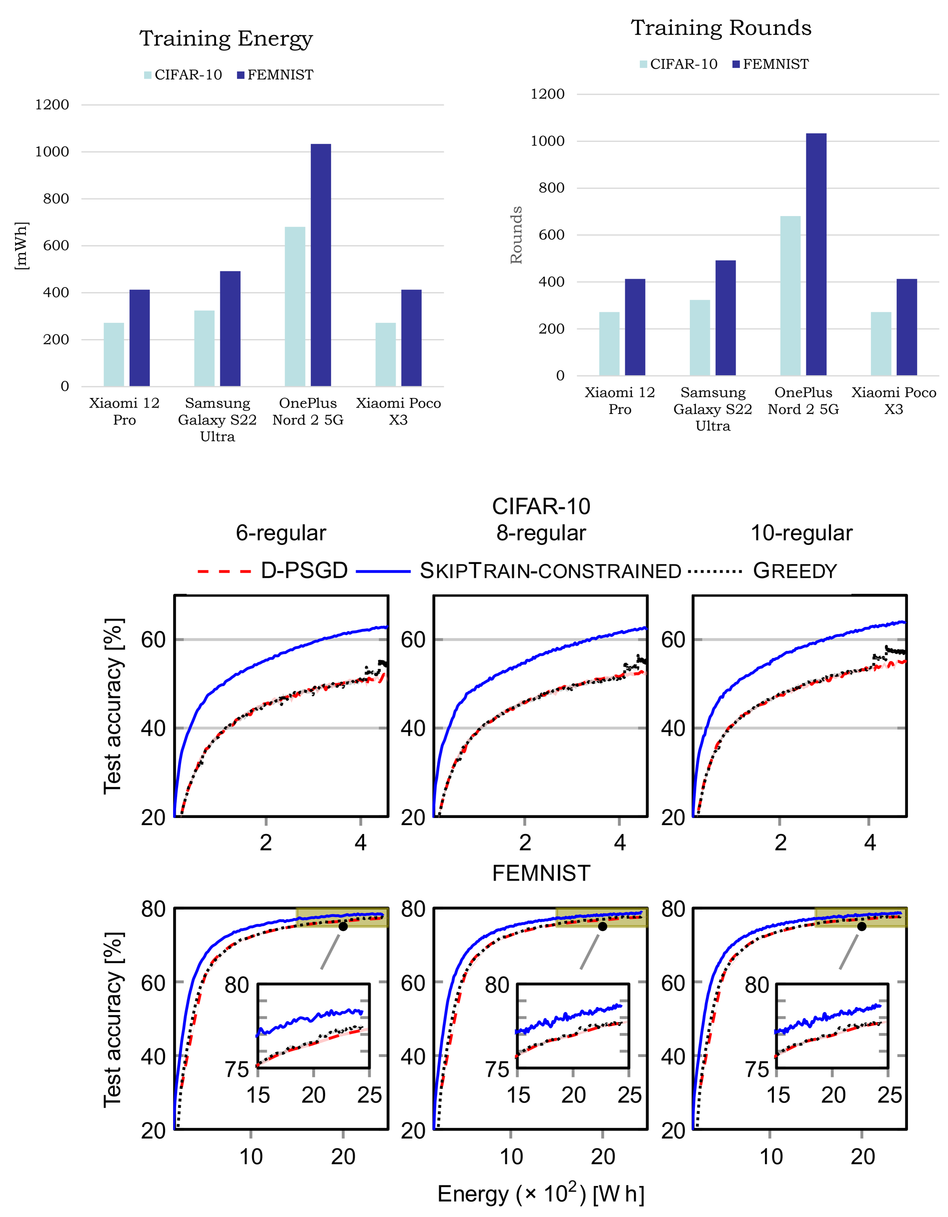
- Datasets: CIFAR-10 and FEMNIST
- Models: CNN
- **SkipTrain:** fixed number of rounds
- **SkipTrain-Constrained:** fixed energy budgets to compare accuracy



Results: SkipTrain

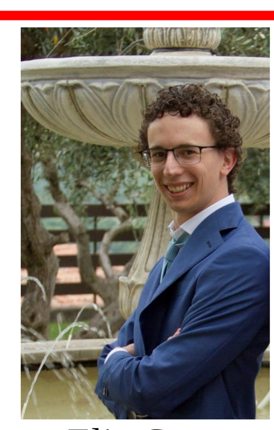


Results: SkipTrain Constrained



[1] X. Lian, et al. "Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent," Advances in neural information processing systems, vol. 30, 2017.

[2] L. Xiao, et al. "Fast linear iterations for distributed averaging", Systems & Control Letters, vol. 53, no. 1, pp. 65–78, 2004.



Elia Guerra
eliaguerra97@gmail.com

Advanced research for everyday life