DLion

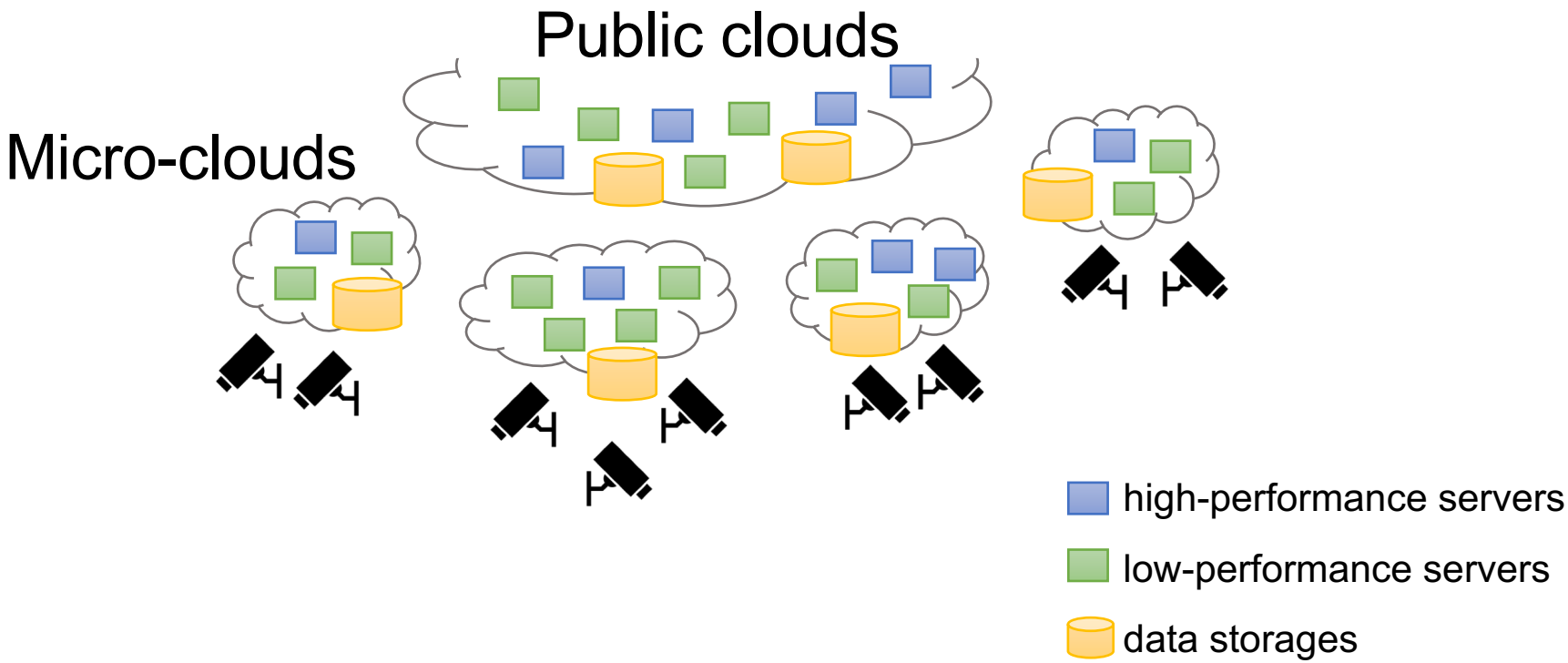
Decentralized Distributed Deep Learning in Micro-Clouds

Rankyung Hong *and Abhishek Chandra*

HotCloud, July 8, 2019

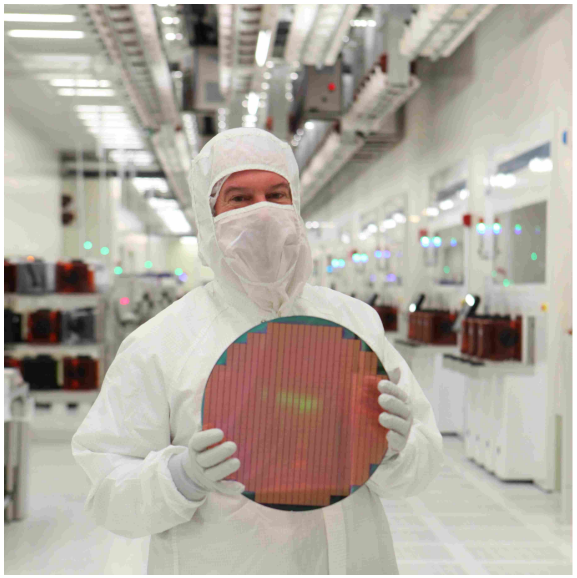
UNIVERSITY OF MINNESOTA

Micro-clouds for exponentially growing large amounts of data generated by lots of edge devices



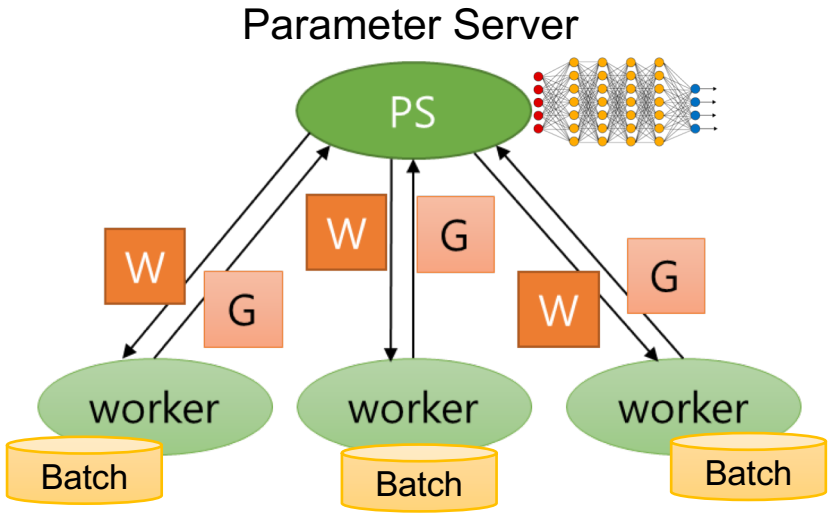
Using micro clouds for deep learning

- Deep learning tech. based QA for Wafer fabrication
  - Confidential data



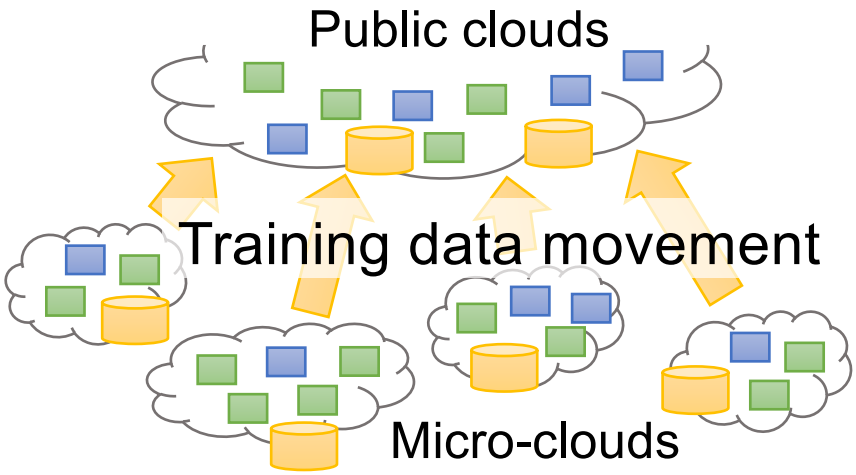
- Incremental deep learning over user data
  - Data privacy

Traditional Distributed Deep Learning (DDL)



Centralized DDL

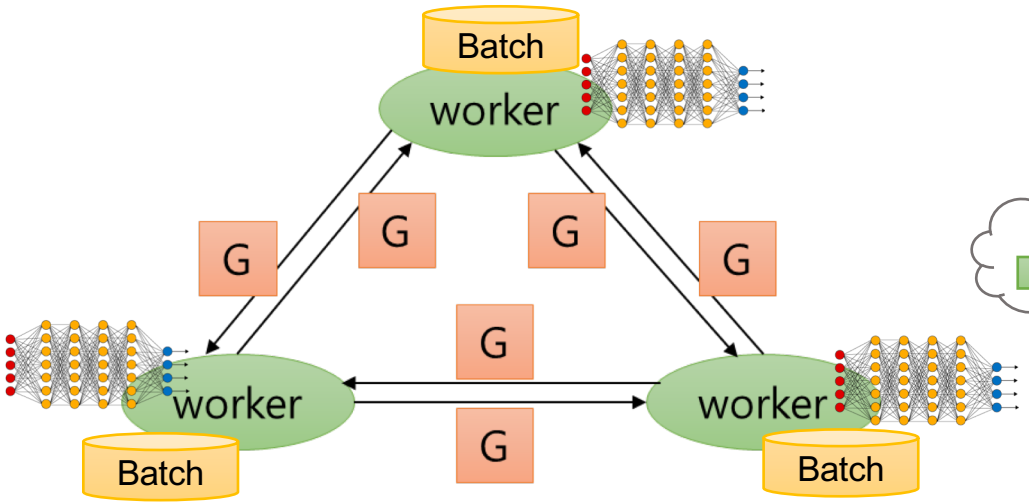
- One-time learning
- Fixed training data
- Single location



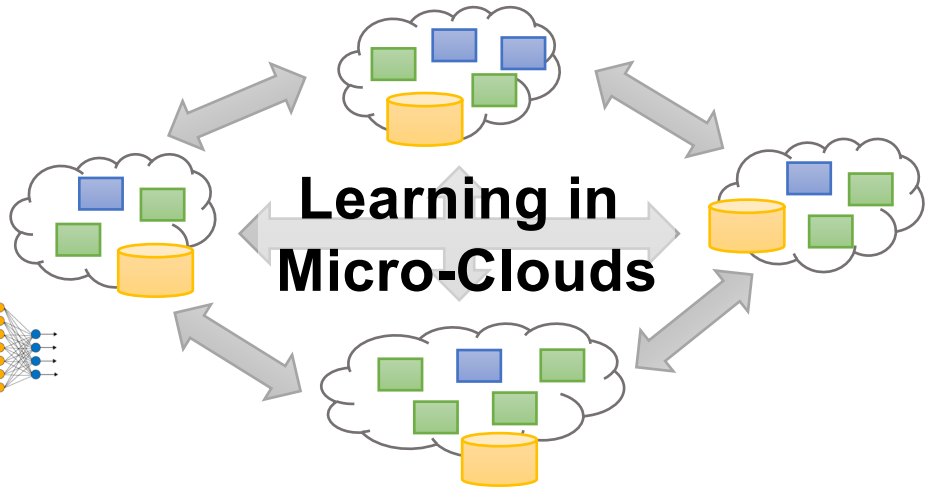
Decentralized DDL

+ New considerations

- Data movement restrictions
- Geo-distributed evolving data at many locations
- Online or incremental learning

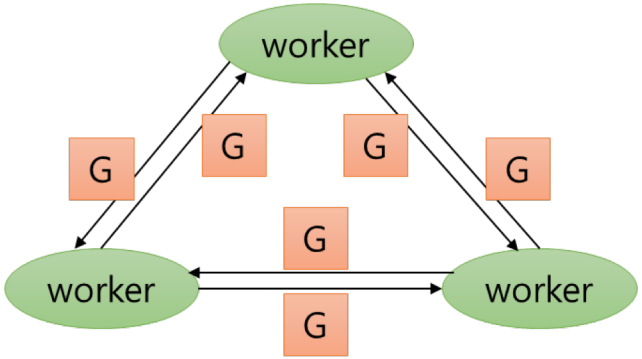


Decentralized DDL

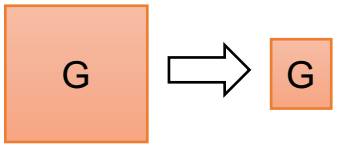


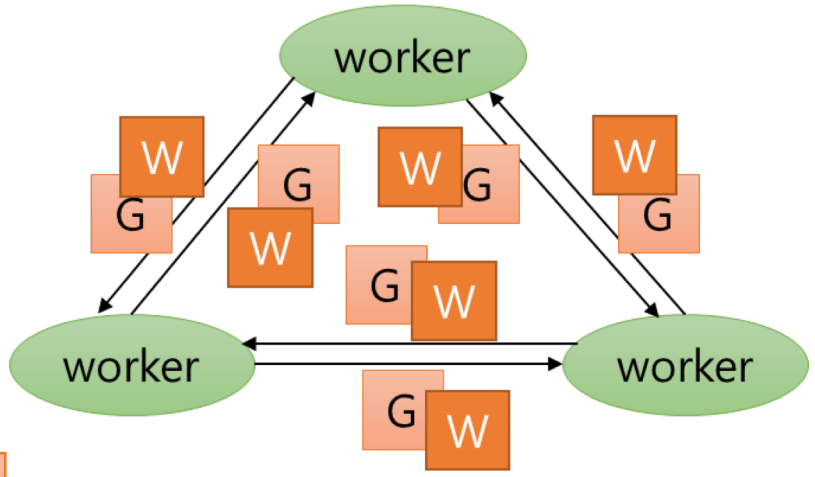
DLion : Decentralized DDL in micro clouds

- **Goals**
  - **[Time]** Faster training time
  - **[Accuracy]** More accurate model



Gaia & Ako

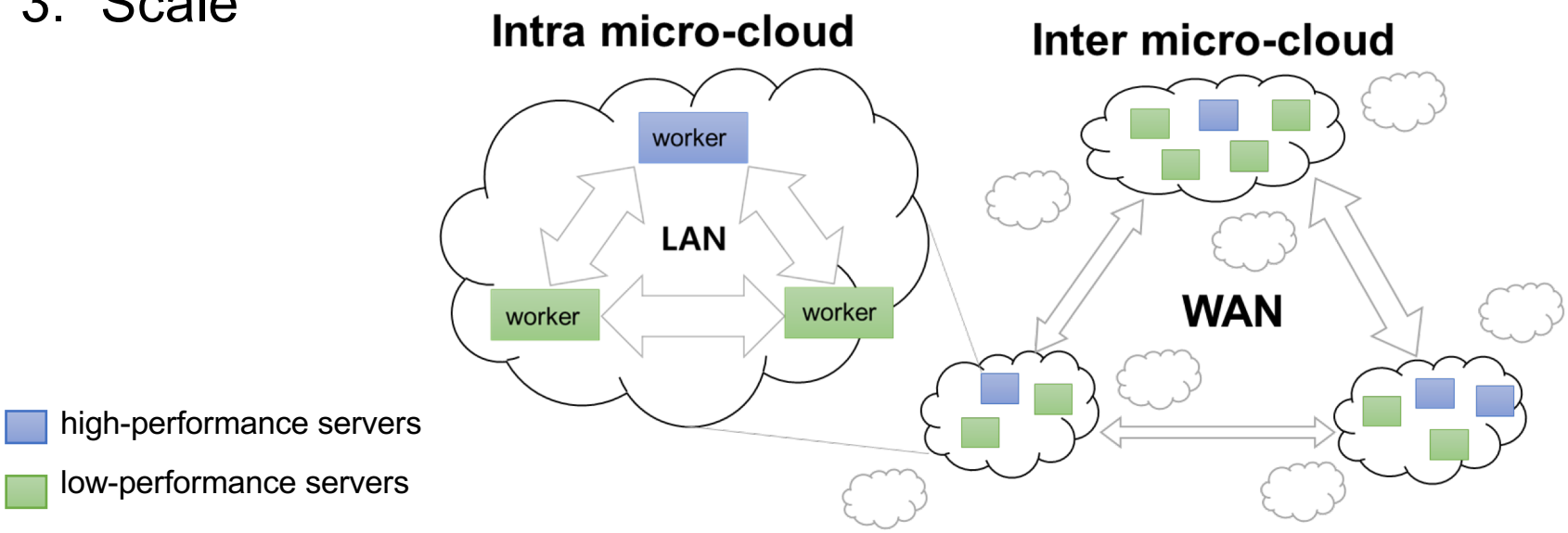




DLion

Challenges of learning in micro clouds

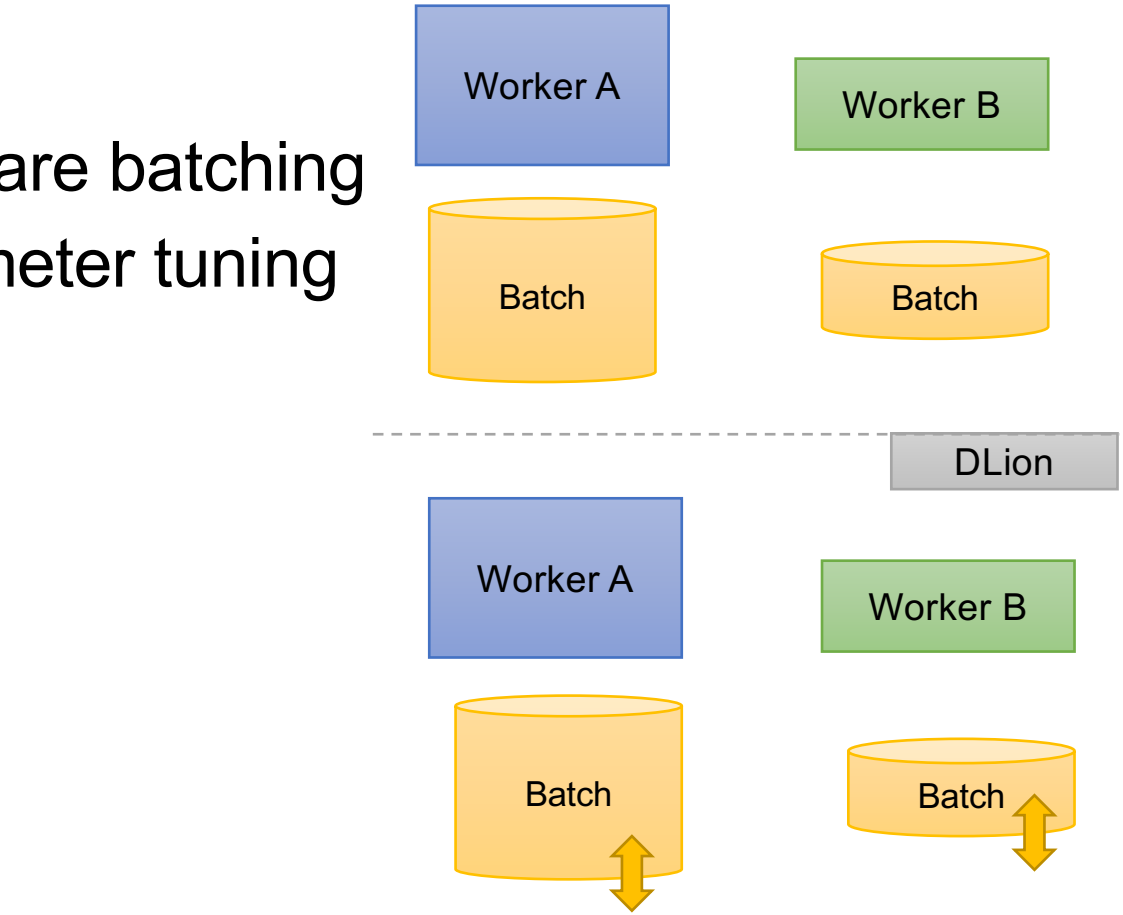
1. Compute resource heterogeneity
2. Network resource heterogeneity
3. Scale



Challenge 1: Compute resource heterogeneity

- Compute capacity-aware batching
- Adaptive model parameter tuning

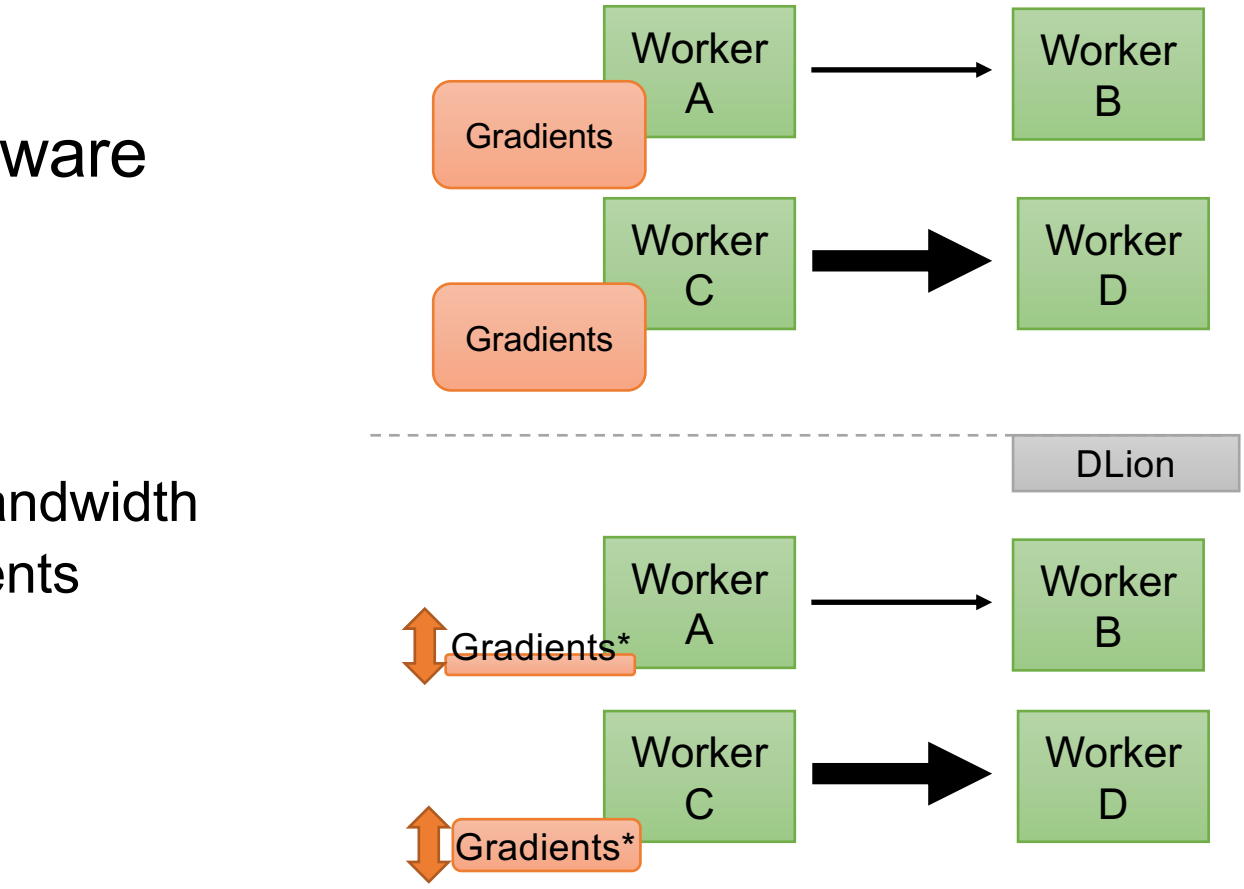
- Considerations
  - Computation Capacity
  - # of workers
  - Training progress



Challenge 2: Network resource heterogeneity

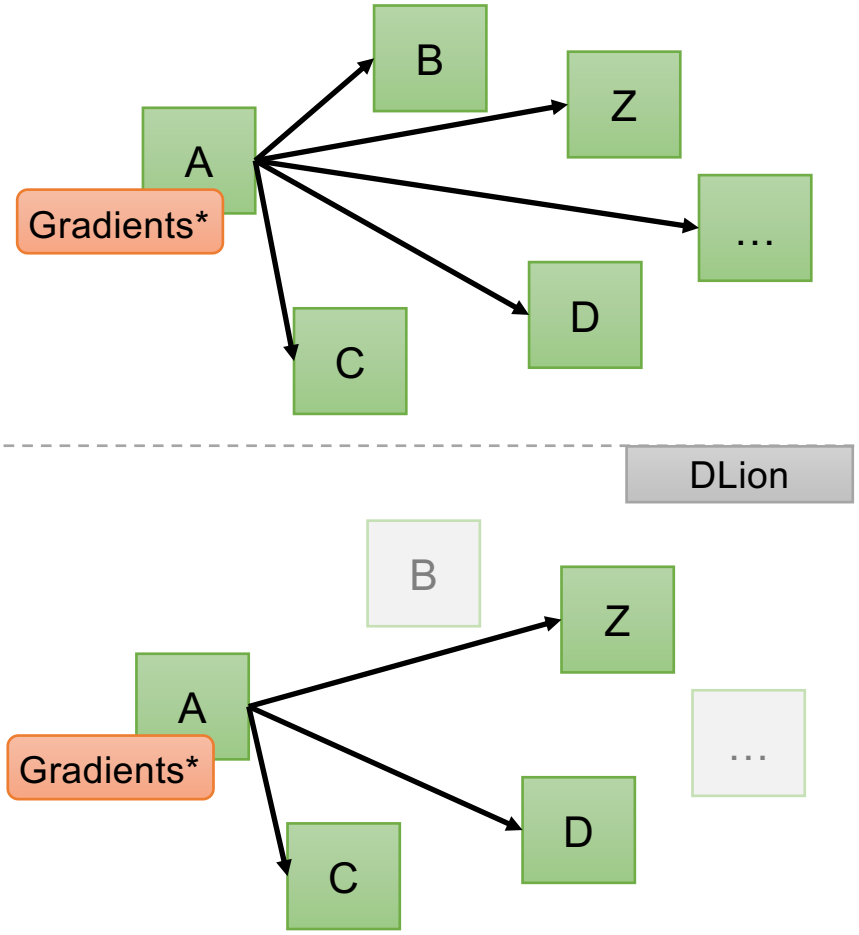
- Network capacity-aware data exchange

- Considerations
  - Available network bandwidth
  - Importance of gradients
  - # of workers

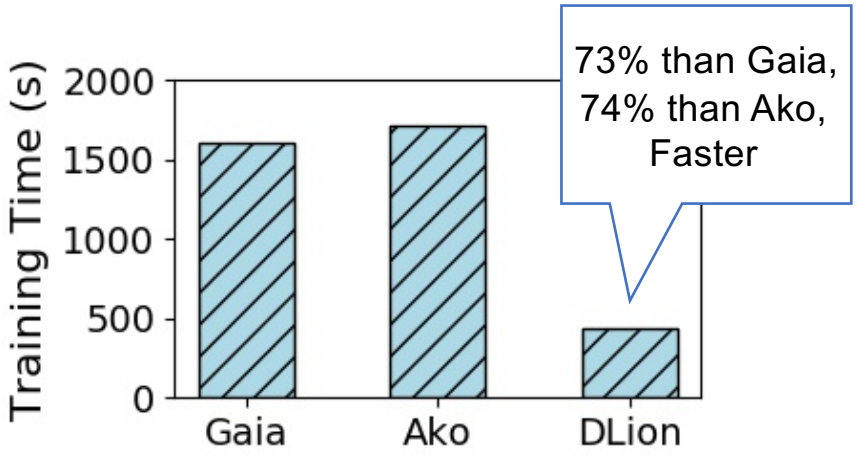


Challenge 3: Scale

- Selective data propagation
- Considerations
  - Amount of Information in gradients
  - Available network bandwidth



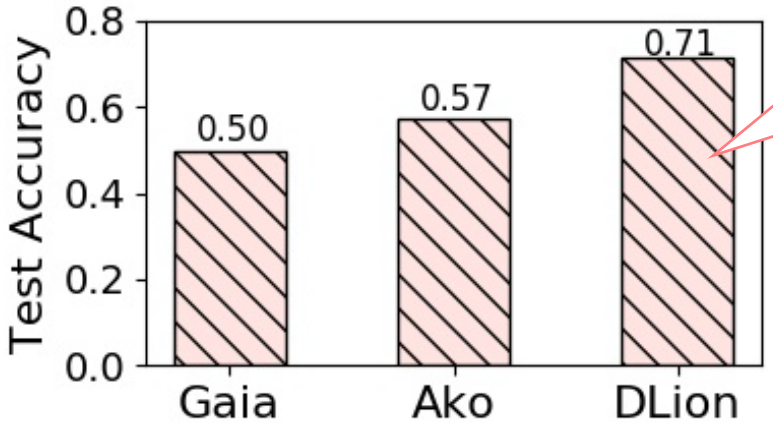
Evaluation: Handling compute capacity



Training Time Comparison
(60% accuracy)

3 low- and 1 high-performance workers, Homogeneous network bandwidth
2conv + 2fx model (17MB), CIFAR10 dataset

Evaluation: Handling network capacity



42% than Gaia,
25% than Ako,
Higher

Model Accuracy Comparison
(30 minutes)

2 micro-clouds (3 workers, 1 worker each), Homogeneous compute capacity
2conv + 2fx model (17MB), CIFAR10 dataset

Conclusion

- Challenges of learning in micro clouds
  - Heterogeneous computation capacity
  - Heterogeneous network capacity
  - Scale
- DLion: Decentralized Distributed Deep Learning System
  - To train a DL model in micro-clouds faster and get a higher accuracy
- Techniques
  - Weight exchange
  - + Computation capacity aware batching + adaptive model param. tuning
  - + Network capacity aware data exchange
  - Selective data propagation

Thanks!