Georgia Tech School of Electrical and Computer Engineering
College of Engineering

SYNERGY

http://synergy.ece.gatech.edu

STRA SIM

Exercise 4: Implementing a New Training Loop

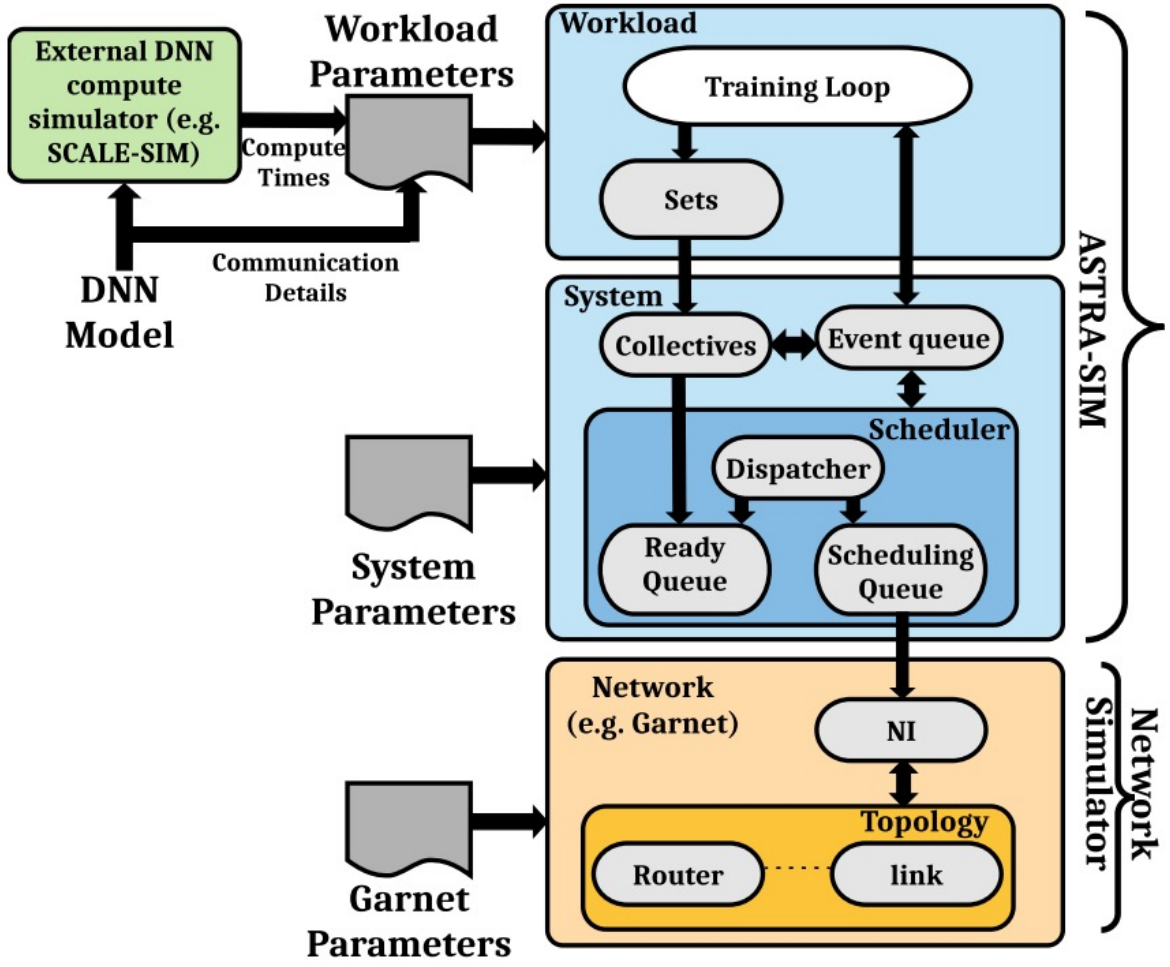
Taekyung Heo
Postdoctoral Fellow, School of ECE
Georgia Institute of Technology
tkheo@casys.kaist.ac.kr

Acknowledgments: William Won (GT), Srinivas Sridharan (Facebook), Sudarshan Srinivasan (Intel)

Objective

To demonstrate how you can implement a new training loop in ASTRA-sim

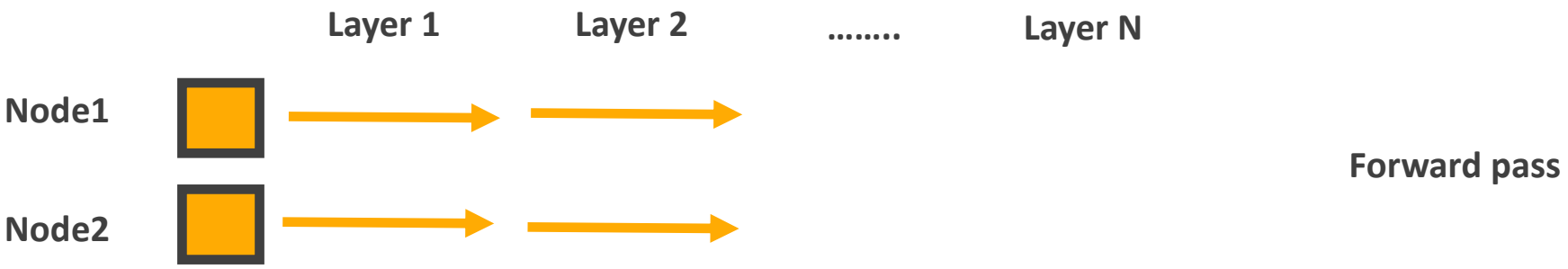
Training Loops



- Training loop determines the behavior of a workload
  - Parallelization Strategy
  - Computation Order
  - Communication Order
- Supported training loops
  - Data parallel
  - Model parallel
  - DLRM
  - Transformer
- You can implement a new training loop to support other models

Training Loop Analysis – Data Parallel

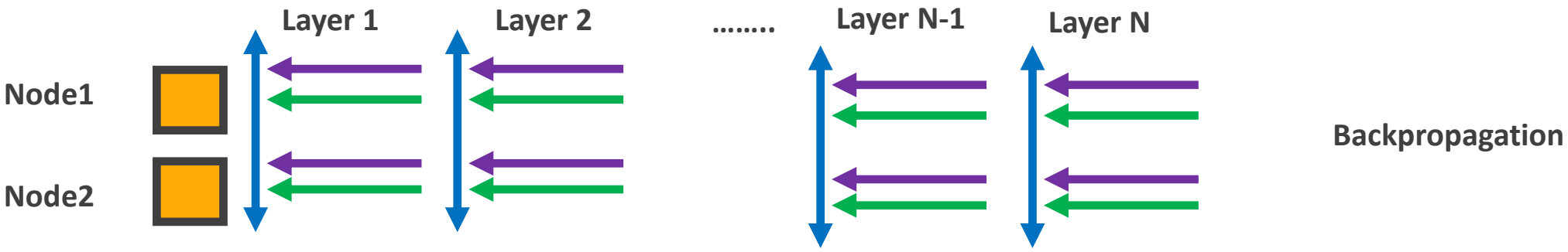
- Distribute Data across multiple nodes and replicate model (network) along all nodes.
- **No communication** during the forward pass.



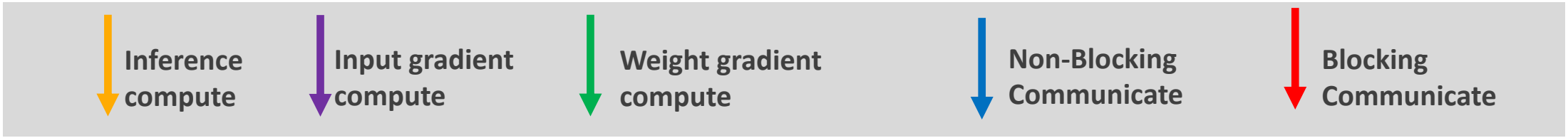
Flow-per-layer: 1.Compute output -> 2. go to the next layer

Training Loop Analysis – Data Parallel

- Distribute Data across multiple nodes and replicate model (network) along all nodes.
- **Communicate weight gradients** during the backpropagation pass.
  - Blocking wait during forward pass for collective of previous backpropagation for that layer.



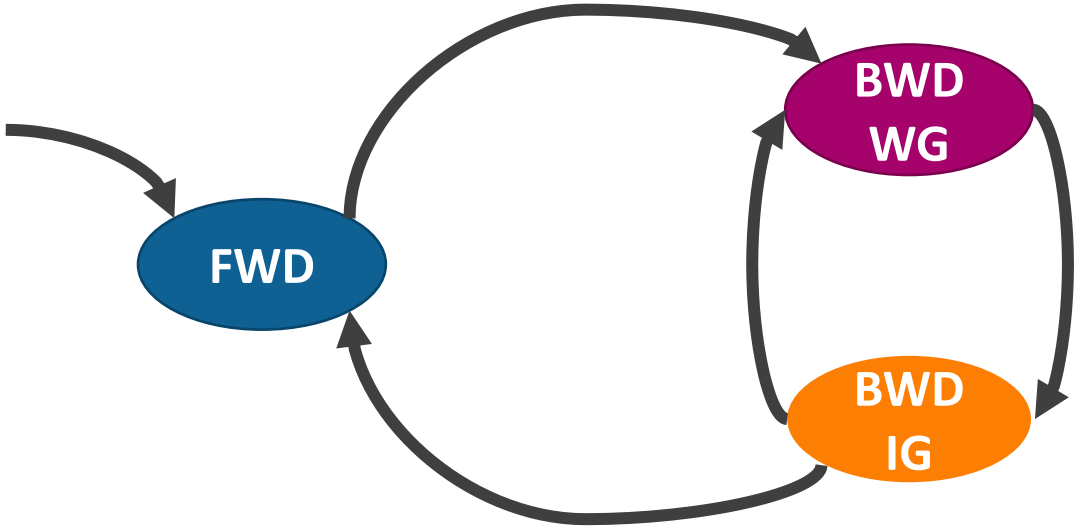
Flow-per-layer: 1.Compute weight gradient-> 2.issue weight gradient comm -> 3.compute input gradient -> 4. go to previous layer



Training Loop Analysis – Data Parallel

Vanilla Data-parallel Training Schedule

FSM Diagram



Exercise: Reorder Computation of Data-parallel Loop

Reorder Data-parallel Training Schedule

FSM Diagram