Deep Learning at 15PF

Supervised and Semi-Supervised Classification for Scientific Data

Matsuoka-Lab M1
Toshiki_Tsuchikawa

SC17
https://dl.acm.org/citation.cfm?id=3126916

ABSTRACT

- Present Deep Learning system for solving scientific pattern classification problems on two HPC architectures
  - supervised convolutional architectures for discriminating signals in high-energy physics data (HPE)
  - semi-supervised architectures for localizing and classifying extreme weather in climate data
- Use a hybrid strategy employing synchronous node-groups, while using asynchronous communication across groups
- Obtain peak performance of 11.73-15.07 PFLOP/s by using 9600 Xeon-Phi nodes

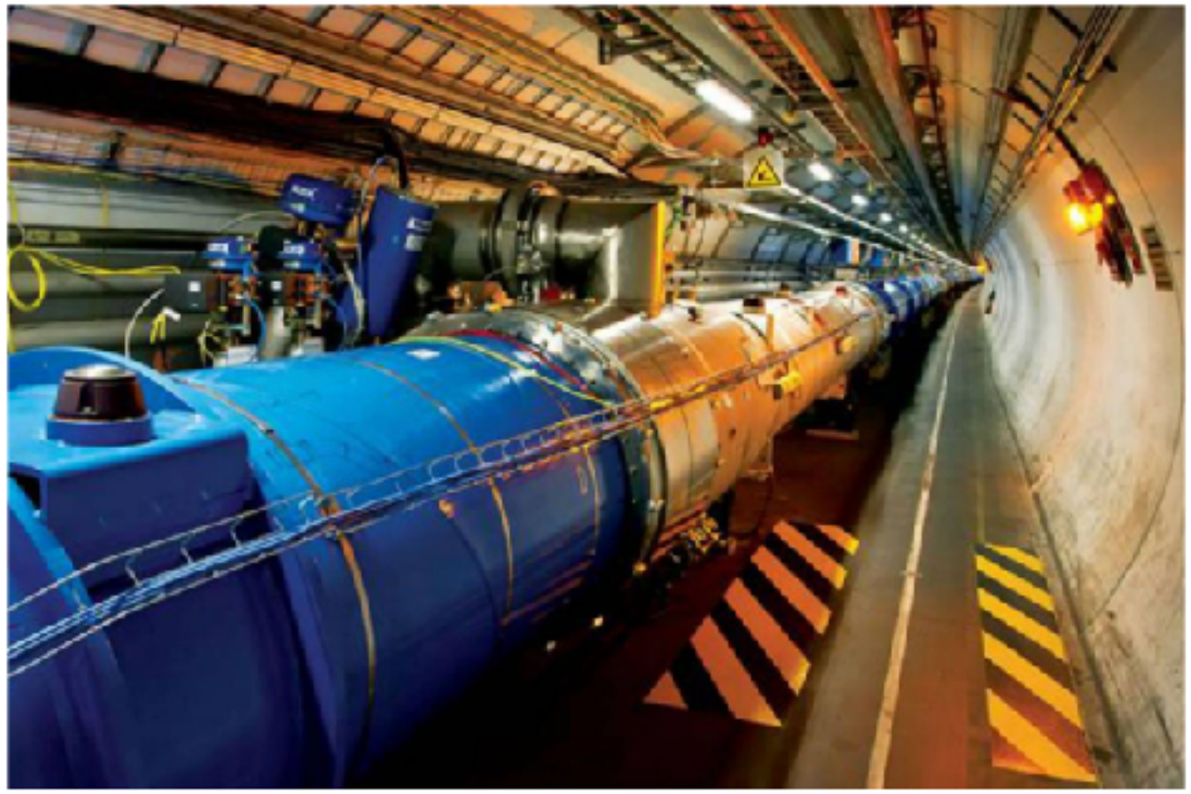
Supervised Learning for HEP

- find rare signals of new particles produced at accelerators such as the Large Hadron Collider (LHC) at CERN

- Data from the surface of the cylindrical detector can be represented as a sparse 2D image(228×228)

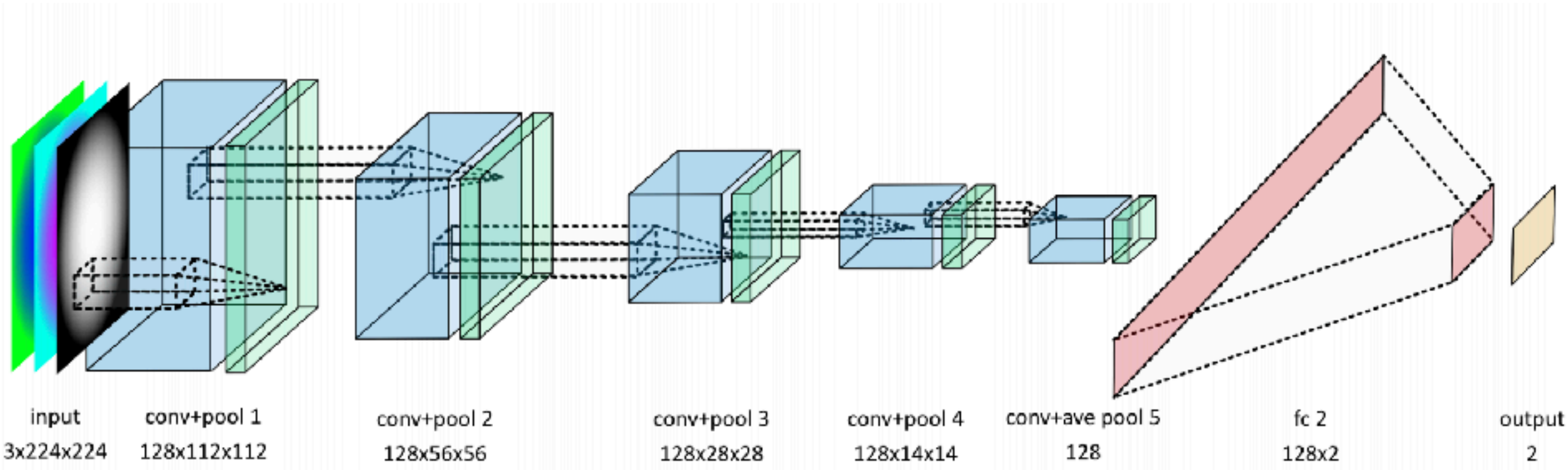
- Data size is 7.4TB

- #images is 10M



https://www.wired.co.uk/article/large-hadron-collider-explained

Supervised Learning for HEP



https://supercomputersfordl2017.github.io/Presentations/ThorstenLargeScaleDeepLearning.pdf

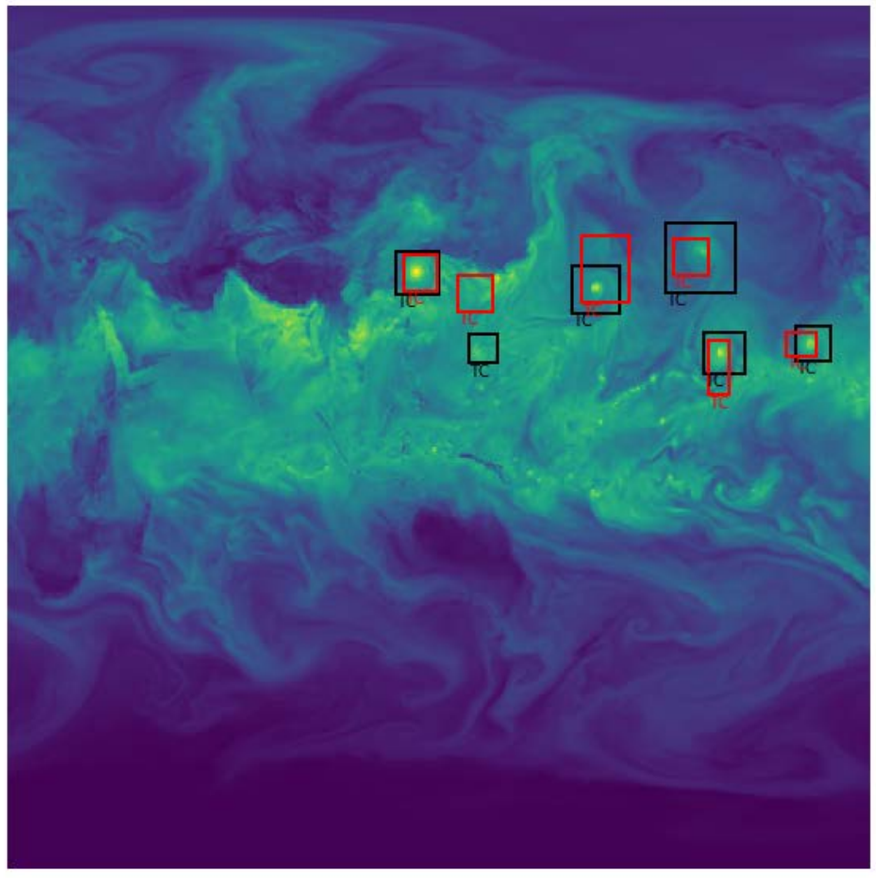
- Input image is made of 3-channels
- Use a Convolutional Neural Net comprised of 5 convolution+pooling units and 1 fully-connected layer with rectified linear unit (ReLU) activation functions
- Parameters size is 2.3MiB

Semi-Supervised Learning for Climate

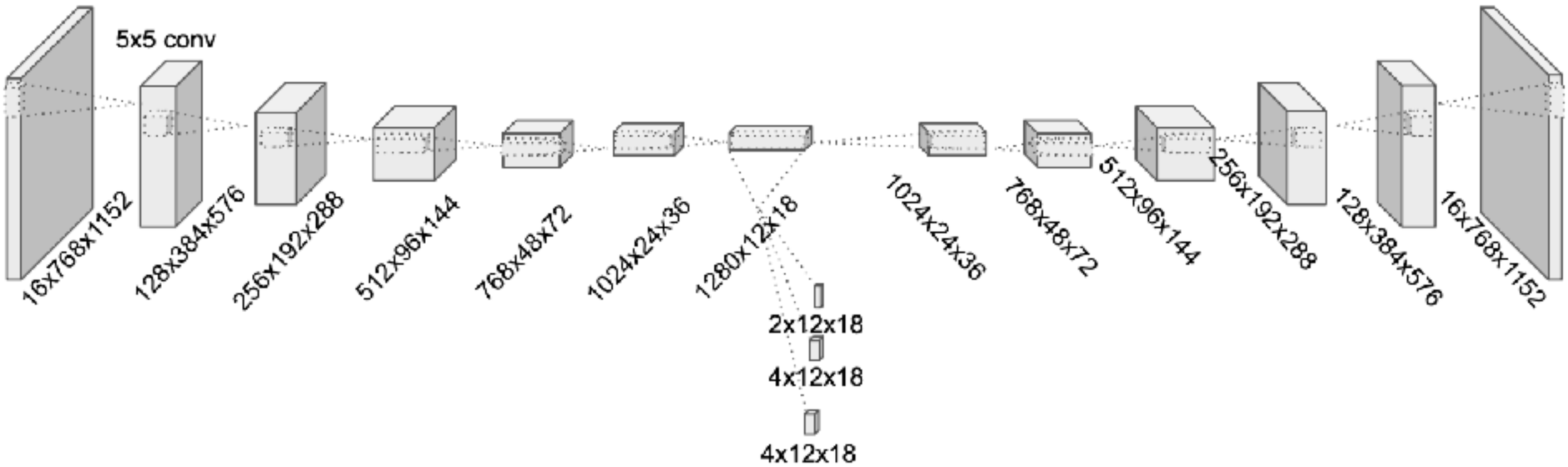
- interested in the task of finding extreme weather events in a 15 TB climate data

- The field of climate science typically relies on heuristics

- have a fully supervised convolutional network for bounding box regression and an unsupervised convolutional auto-encoder



Semi-Supervised Learning for Climate



https://supercomputersfordl2017.github.io/Presentations/ThorstenLargeScaleDeepLearning.pdf

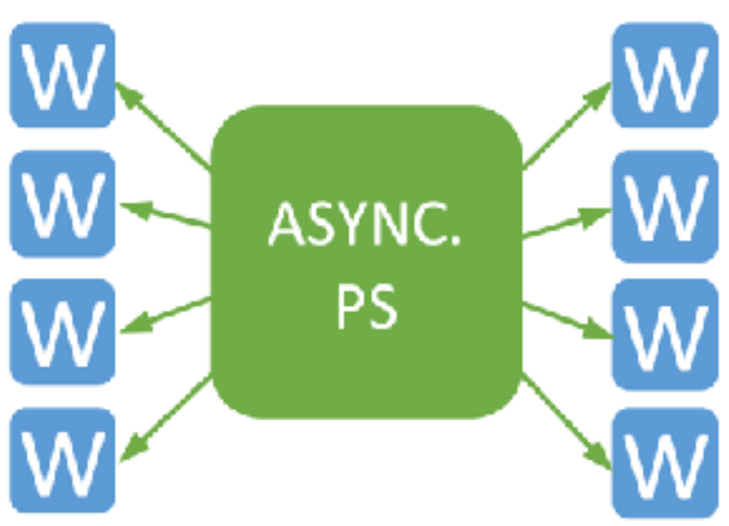
- Input image is made of 16-channels
- Use 9xConvolutions and 5xDeconvolutions
- Parameter size is 302.1 MiB

Deep Learning on multiple nodes

- Use data-parallel.
- SYNCHRONOUS
  - use synchronization barriers and force computational nodes to perform every update step

SYNCHRONOUS

- ASYNCHRONOUS
  - Each node works on its own iteration (mini-batch) and produces independent updates to the model
  - PS(parameter sever) applies the updates to the model in the order they are received, and sends back the updated model to the worker

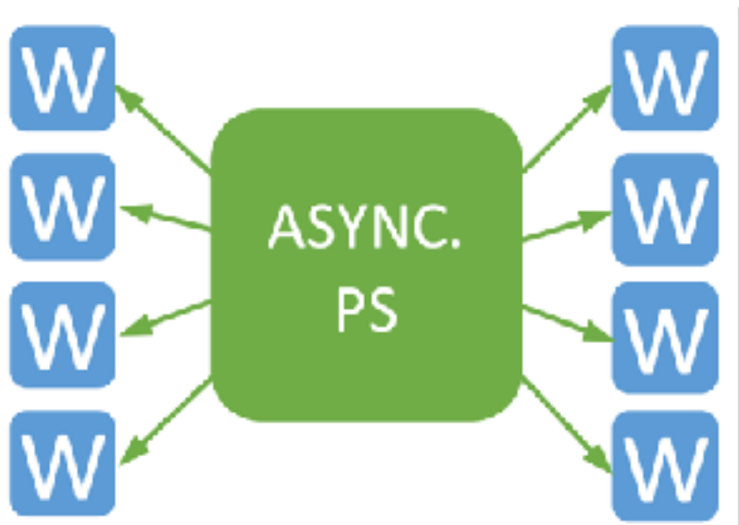


ASYNCHRONOUS

Deep Learning on multiple nodes

- SYNCHRONOUS
  - The batch size is a limit on the number of nodes in data-parallel synchronous systems
  - The duration of the iteration depends on the slowest node
- ASYNCHRONOUS
  - Not need to wait slowest node. So can have many iteration and not limit batch size
  - Use of out-of-date gradients

SYNCHRONOUS



ASYNCHRONOUS

Multi-node scaling with hybrid approach

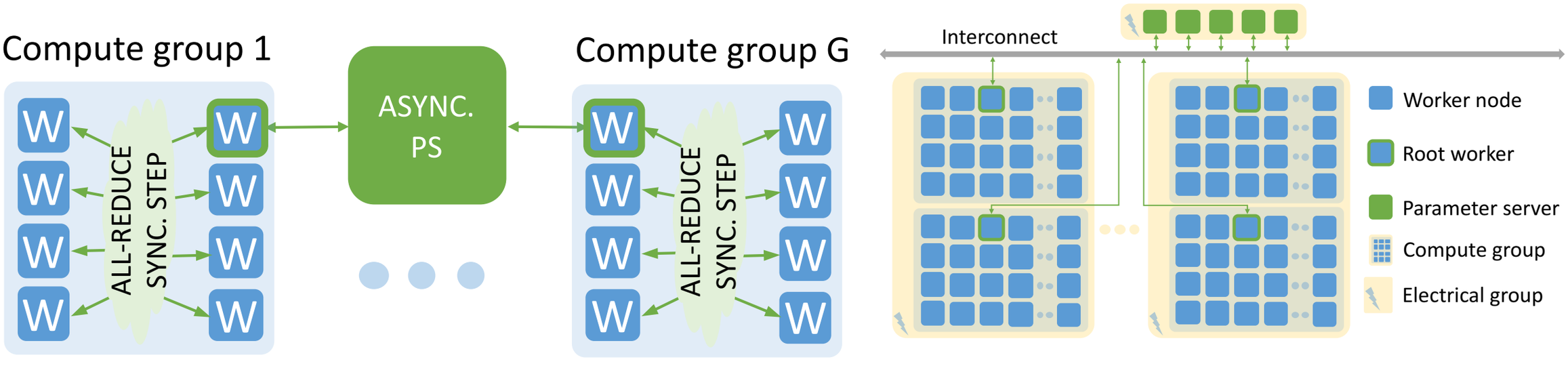
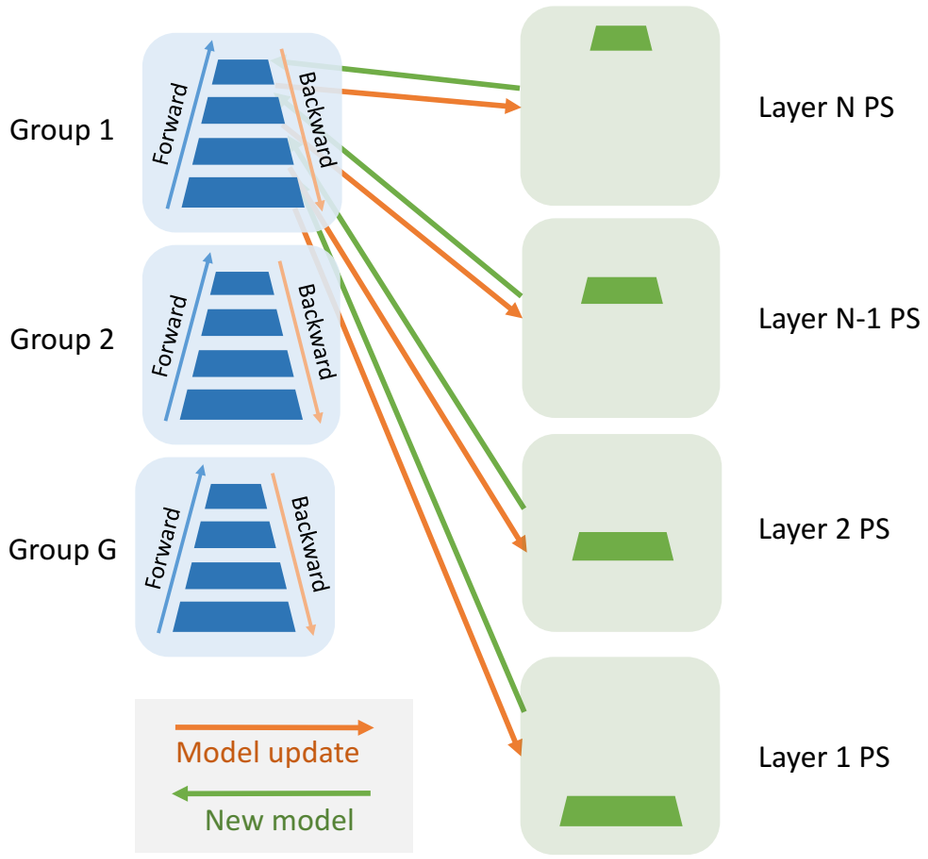


Figure 2: Hybrid architecture example.

- In this architecture, worker nodes coalesce into separate compute groups. Each compute group follows a synchro-nous architecture
- the number of compute groups (and their size) is a knob that controls the amount of asynchrony in the system
- assign a dedicated parameter server to each trainable layer of the network

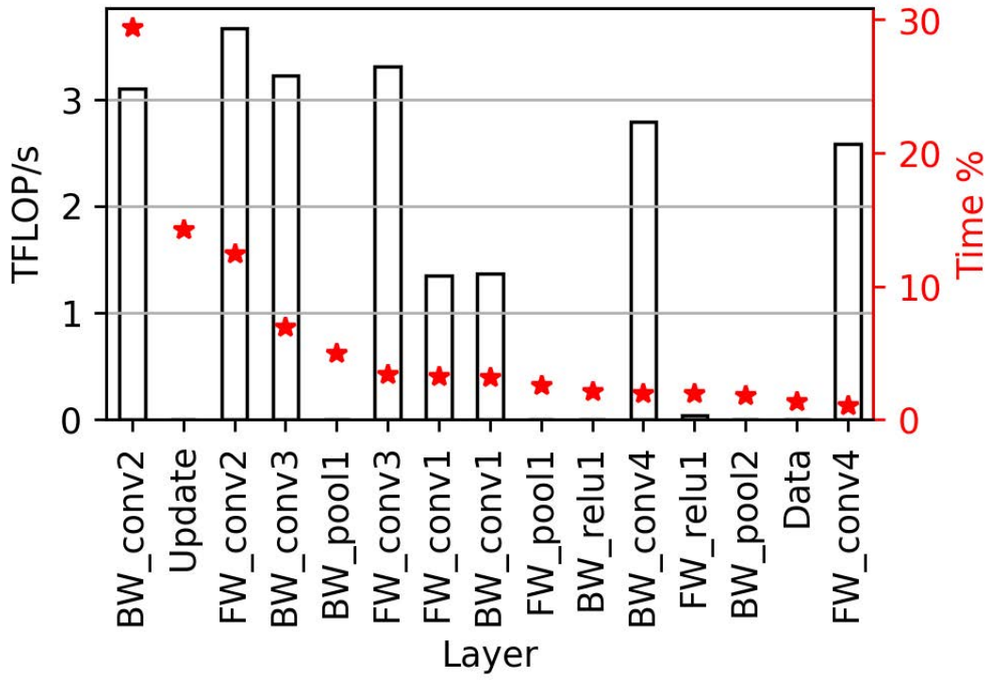


Software and Hardware

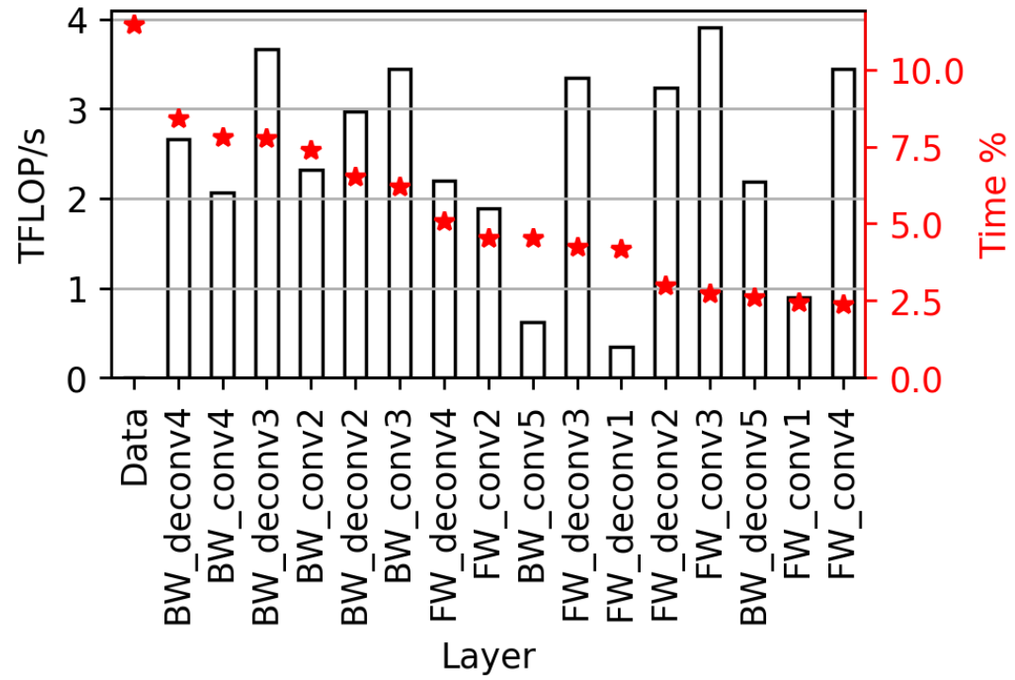
- Intel Distribution of Caffe
- Intel Machine Learning Scaling Library (MLSL)
- Cori-KNL HPC system
  - 9688 Intel® Xeon Phi™ 7250 processor nodes (Knight's Landing)
  - 68 cores per node with support for 4 hardware threads each (272 threads total)
  - The peak performance for single precision can be computed as: (9688 KNLs) x (68 Cores) x (1.4 GHz Clock Speed) x (64 FLOPs / Cycle) = 59 PetaFLOP/s

https://phys.org/news/2016-06-nersc-staff-users-readying-delivery.html

Single node performance



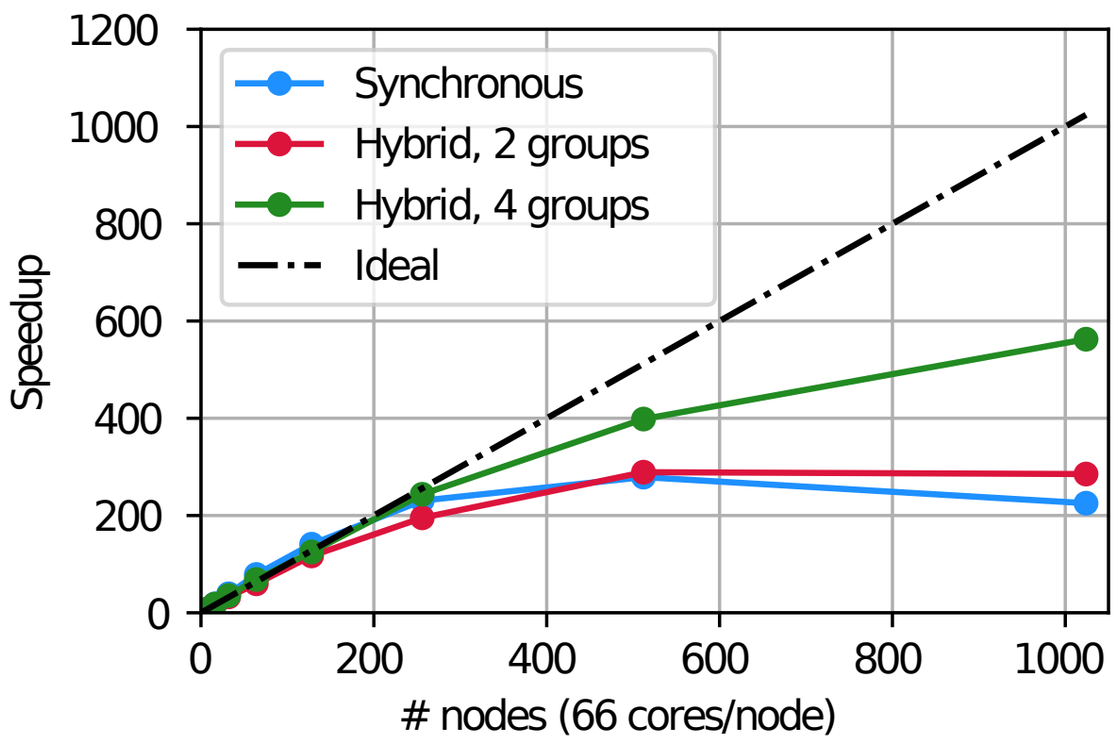
(a) HEP



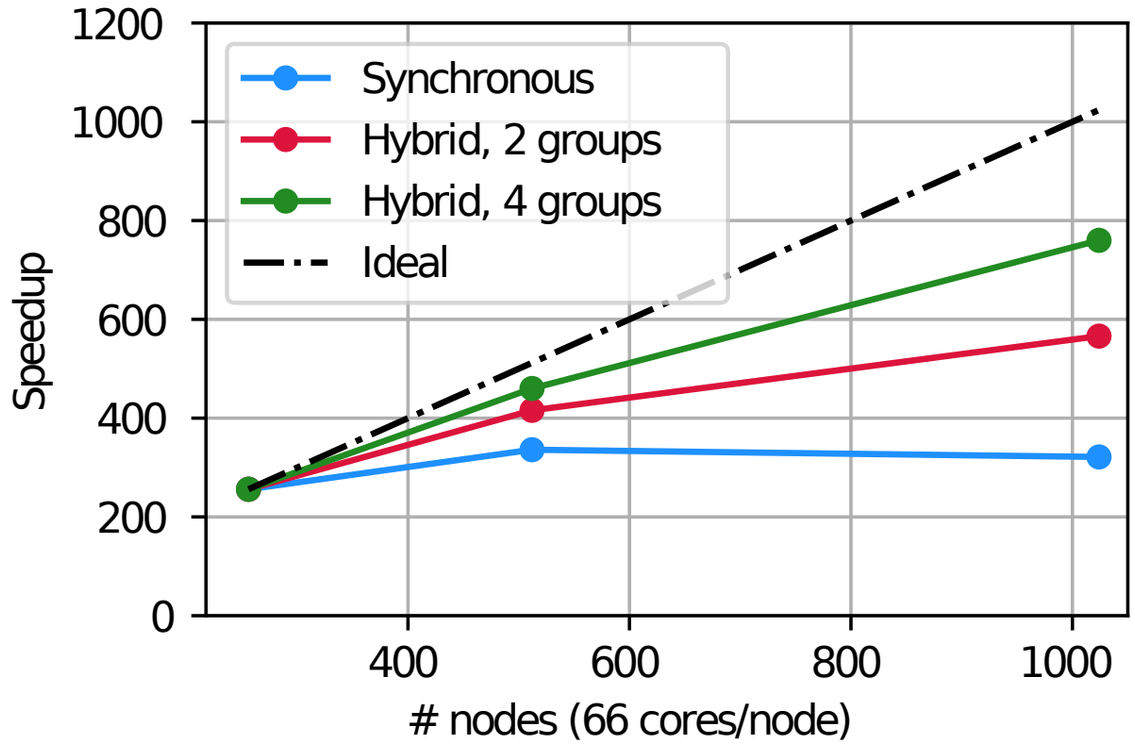
(b) Climate

- batch size of 8 images
- the overall flop rate of the HEP network stands at 1.90 TFLOP/s, while that of the Climate network stands at 2.09 TFLOP/s

Strong Scaling Results



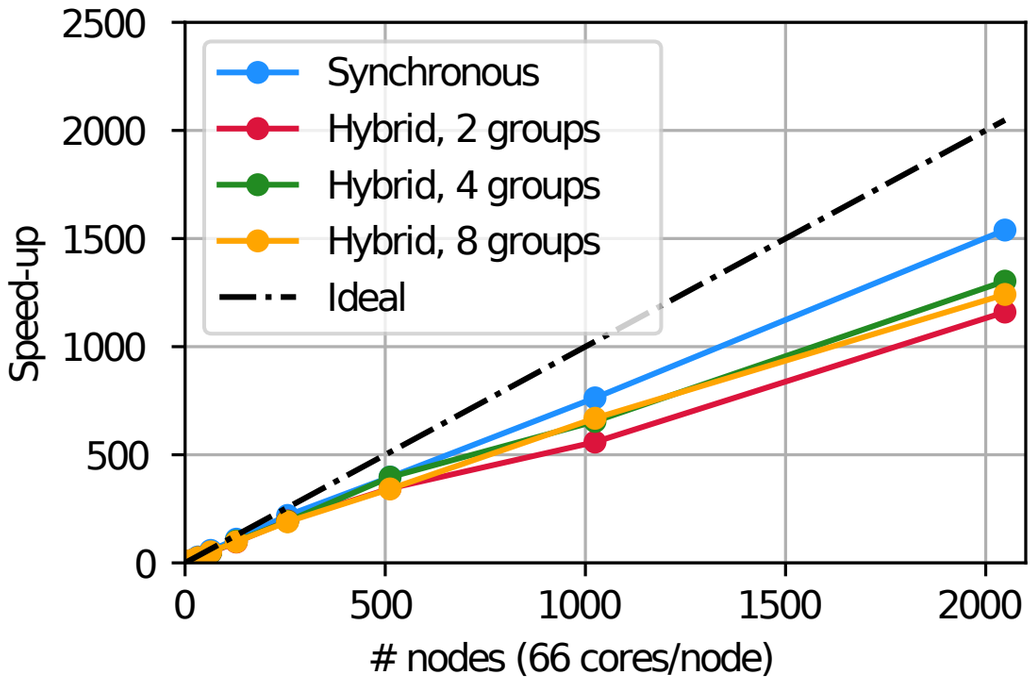
(a) HEP



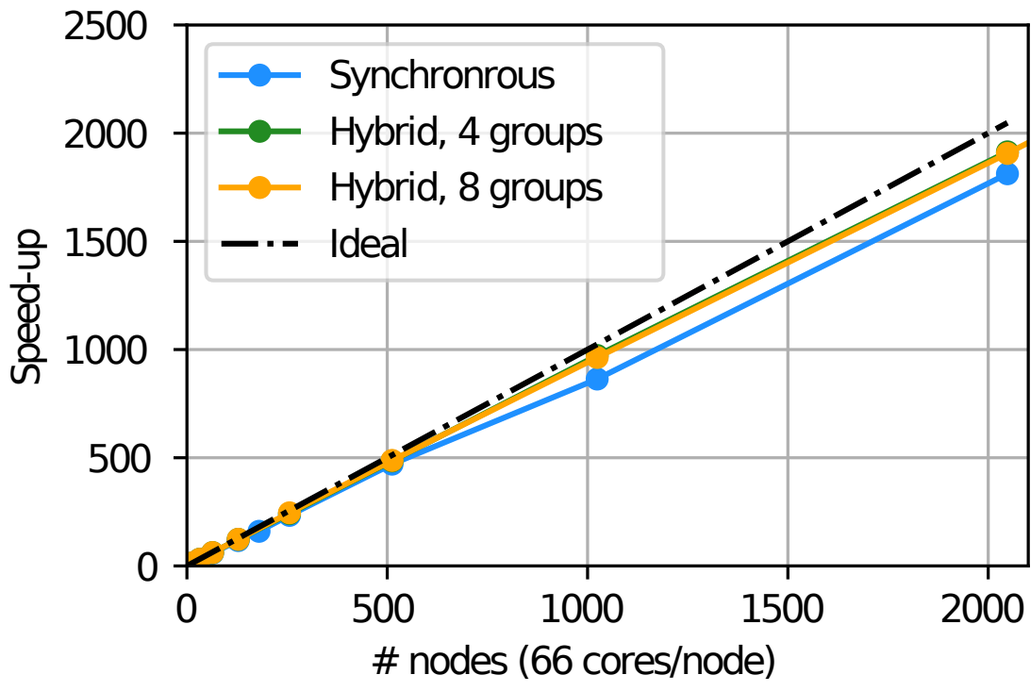
(b) Climate

- batch size = 2048 per synchronous group
- (a) shows that the synchronous algorithm does not scale past 256 nodes
- (b) shows the synchronous algorithm scales only to a maximum of 320x at 512 nodes and stops scaling beyond that point

Weak Scaling Results



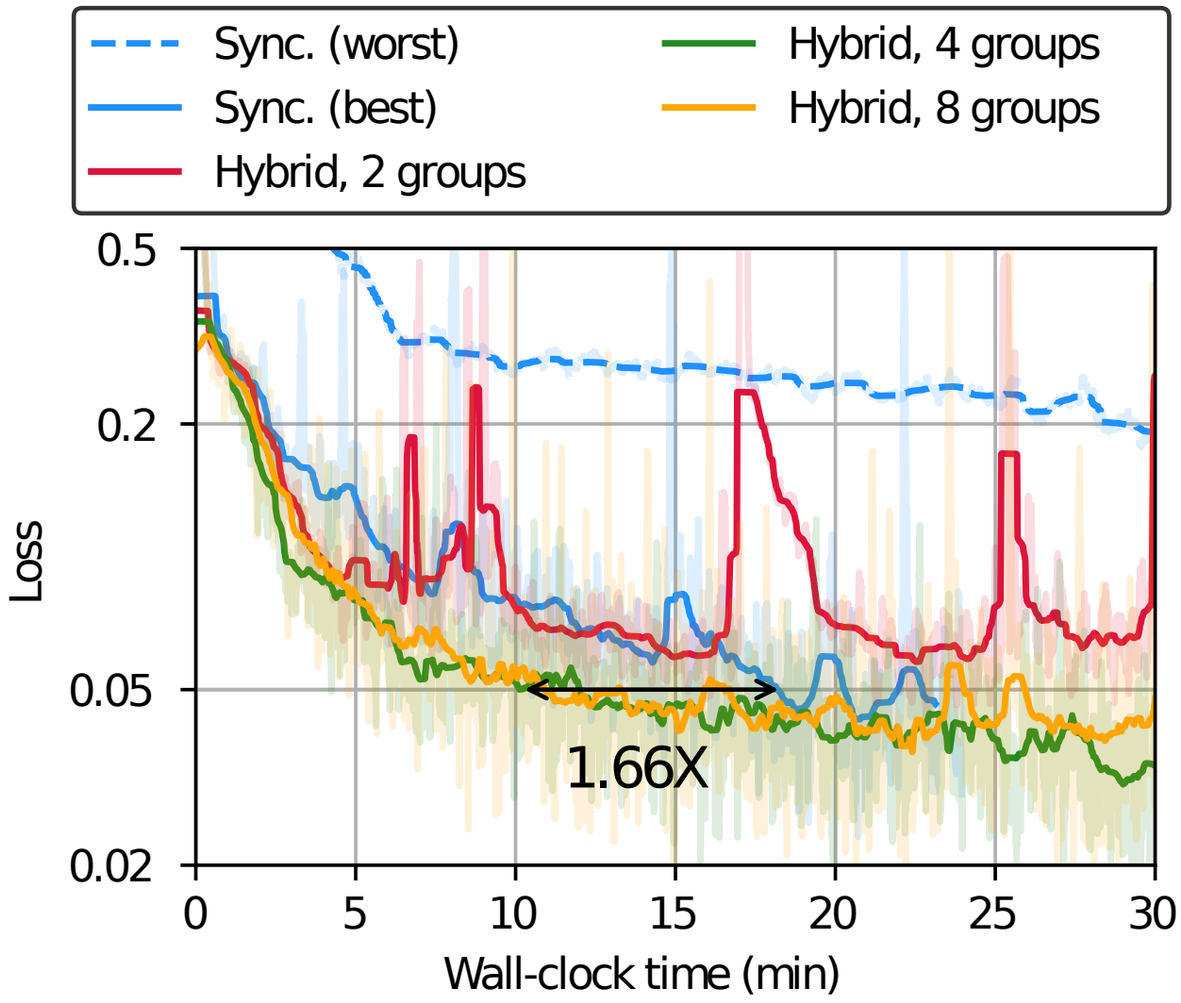
(a) HEP



(b) Climate

- batch size 8 per node
- climate network shows almost ideal scaling
- observe slightly better scaling for hybrid over synchronous configurations due to reduced straggler effects.

Training losses vs wall clock time for HEP



- the best hybrid configuration achieves the target loss(0.05) in about 10 minutes, which is about 1.66X faster than the best sync run

- some of the jumps are observed in the loss curves of the 2-group

Overall Performance

- HEP network
    - obtained a peak throughput of 11.73 PFLOP/s
    - 9594 compute nodes plus 6 parameter servers split into 9 groups
    - each group using a minibatch of 8528
    - corresponds to a speedup of 6173x over single node performance
- Climate network
    - obtained a peak throughput of 15.07 PFLOP/s
    - 9608 compute nodes plus 14 parameter servers split into 8 groups
    - each group using a minibatch of 9608
    - corresponds to a speedup of 7205X over single node performance