14th ANNUAL WORKSHOP 2018

USING FABRICS TO ACCELERATE DEEP LEARNING

Todd Rimmer, Senior Principal Engineer
Ira Weiny, Software Engineer
Intel Corporation

[April 10, 2018]

NOTICES AND DISCLAIMERS

- **Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.**
- **No computer system can be absolutely secure.**
- **Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.**
- **Intel, the Intel logo, Xeon and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.**
- **© 2018 Intel Corporation.**

NOTICES AND DISCLAIMER

- **Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.**

- **Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.**

- **Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.**

- **Copyright© 2018, Intel Corporation. All rights reserved. Intel, the Intel logo, Atom, Xeon, Xeon Phi, Core, VTune, and Cilk are trademarks of Intel Corporation in the U.S. and other countries.**

AGENDA

- **What is AI**
- **Where can fabrics help**
- **Some real world examples**

WHAT IS AI

AI IS TRANSFORMING INDUSTRIES







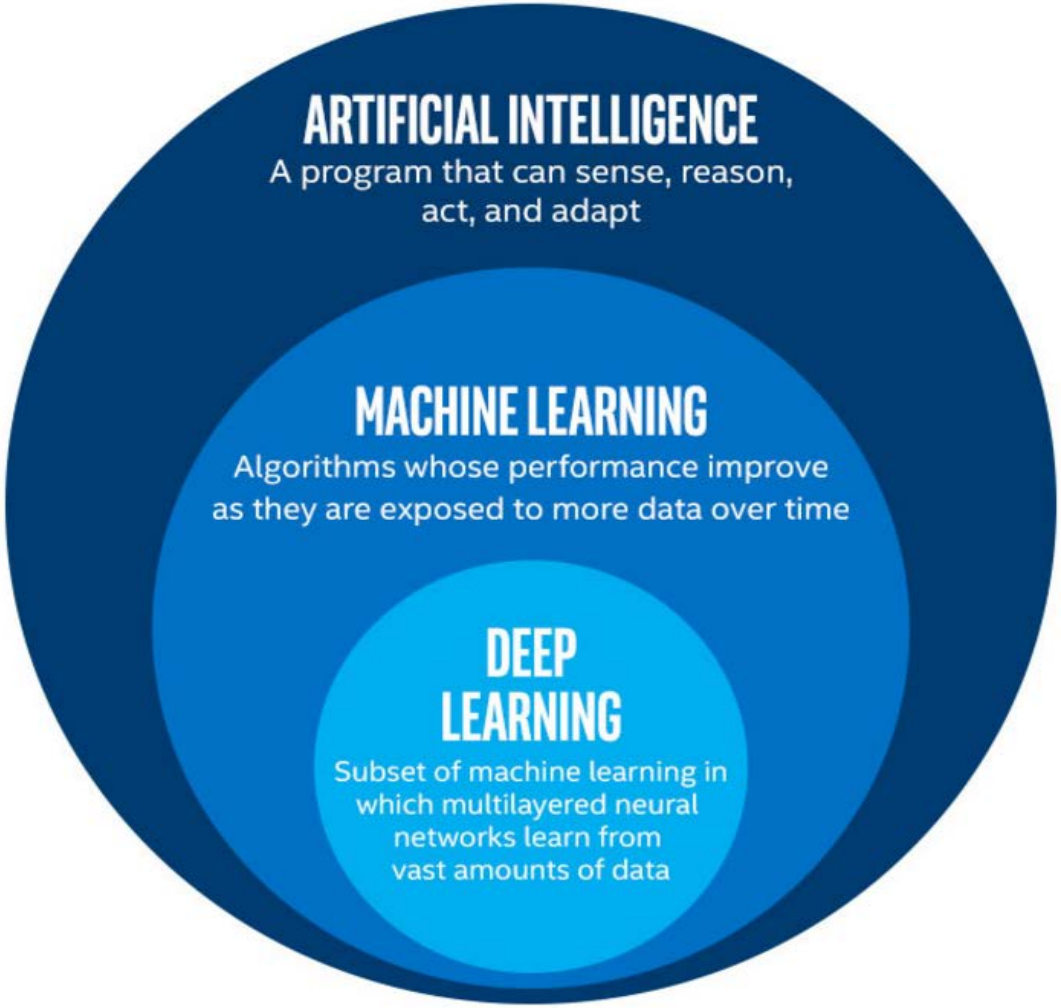

examples

Consumer	Health	Finance	Retail	Government	Energy	Transport	Industrial	Other
Smart Assistants	Enhanced Diagnostics	Algorithmic Trading	Support	Defense	Oil & Gas Exploration	Automated Cars	Factory Automation	Advertising
Chatbots	Drug Discovery	Fraud Detection	Experience	Data Insights	Smart Grid	Automated Trucking	Predictive Maintenance	Education
Search	Patient Care	Research	Marketing	Safety & Security	Operational Improvement	Aerospace	Precision Agriculture	Gaming
Personalization	Research	Personal Finance	Merchandising	Resident Engagement	Conservation	Shipping	Field Automation	Professional & IT Services
Augmented Reality	Sensory Aids	Risk Mitigation	Loyalty	Smarter Cities		Search & Rescue		Telco/Media
Robots			Supply Chain					Sports
			Security					

Early adoption

Source: Intel forecast

ARTIFICIAL INTELLIGENCE



MACHINE LEARNING

Machine Learning

How do you engineer the best features?

$N \times N$

$(f_1, f_2, \ldots, f_K)$

Roundness of face
Dist between eyes
Nose width
Eye socket depth
Cheek bone structure
Jaw line length
…etc.

CLASSIFIER ALGORITHM

SVM
Random Forest
Naïve Bayes
Decision Trees
Logistic Regression
Ensemble methods

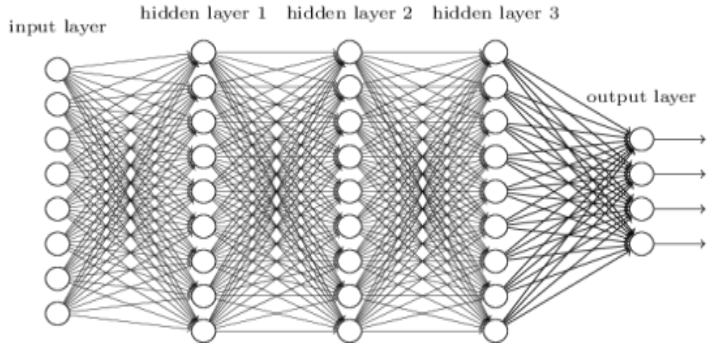
Arjun

Deep Learning

How do you guide the model to find the best

$N \times N$

NEURAL NETWORK



Arjun

WHERE CAN FABRICS HELP

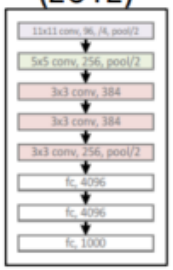
MORE SOPHISTICATED ALGORITHMS

Some Popular Topologies (*Not to be confused with Network topologies!)*

Topologies: Network Depth is of importance

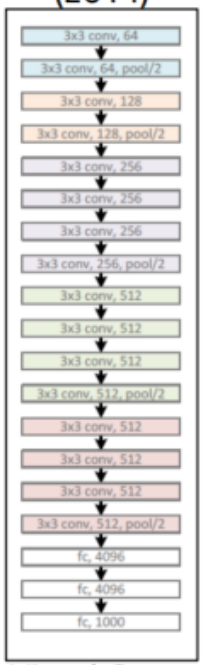
Faster & Less Accurate ←<<<<<<<<<<<<<<<<< >>>>>>>>>>>>>>>>>>>>→ More Accurate

AlexNet, 8 Layers (2012)



U Toronto

VGG, 19 Layers (2014)

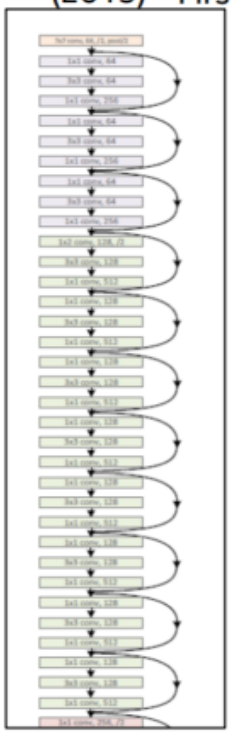


Visual Geom. Group (VGG)

GoogleNet, 22 Layers (2014)

Google

ResNet-50
ResNet-152 Layers (2015) First 35 layers of 152



Microsoft

WHAT IS A FRAMEWORK

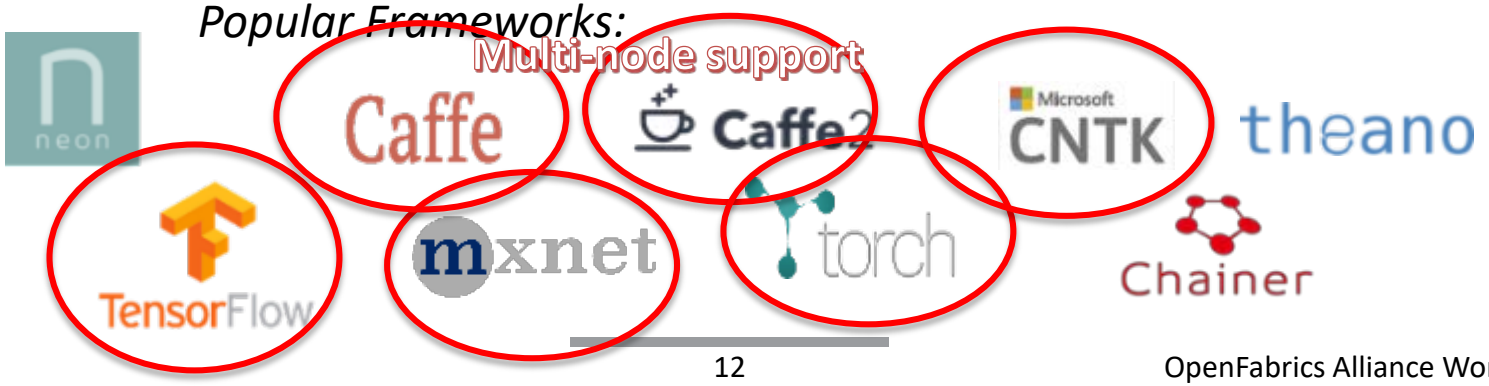
A high-level software tool, typically open-sourced, which is used by most deep learning practitioners to facilitate model development. Each framework includes essential deep learning building blocks such as model libraries, computational graphing, APIs, tools and more. Each framework also has unique advantages/disadvantages in stability, speed, scalability, data integration, flexibility, iteration speed, debuggability, etc.

Popular Frameworks:

WHAT IS A FRAMEWORK

A high-level software tool, typically open-sourced, which is used by most deep learning practitioners to facilitate model development. Each framework includes essential deep learning building blocks such as model libraries, computational graphing, APIs, tools and more. Each framework also has unique advantages/disadvantages in stability, speed, scalability, data integration, flexibility, iteration speed, debuggability, etc.

Popular Frameworks:



REDUCED TIME TO TRAIN DRIVE INNOVATION

Properties of Neural Networks[1]:

Results get better with

- **More Data**
- **Bigger models**
- **More computation**

Training is not a one time effort

- **Many operational neural networks as part of different applications**
- **Each neural network may be trained with domain specific training sets**
- **Evolving input data sets drives the need for re-training**

Google Brain's Jeff Dean quantifies benefits of reducing the time to train:

Minutes, hours:

- Interactive research
- Instant gratification of results

1-4 Days:

- Tolerable
- Interactivity replaced by running many experiments in parallel

1-4 weeks:

- High value experiments only
- Progress stalls

>1 Month:

- Don't even try.

Jeff Dean: Large-Scale Deep Learning for Intelligent Computer Systems (GoogleBrain)

Striving for interactive research drives need for more computational power and multi-node training options.

[1] https://static.googleusercontent.com/media/research.google.com/en//people/jeff/BayLearn2015.pdf

TOWARDS MULTI-NODE TRAINING

Let's solve this problem using DL

Start with a single server

Interesting results. Drive for better accuracy

- More data for better results
- Now training takes much longer
- How to scale and keep time-to-train manageable

Options

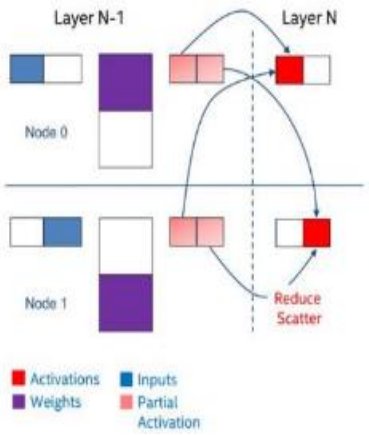
Single server takes too long to train now

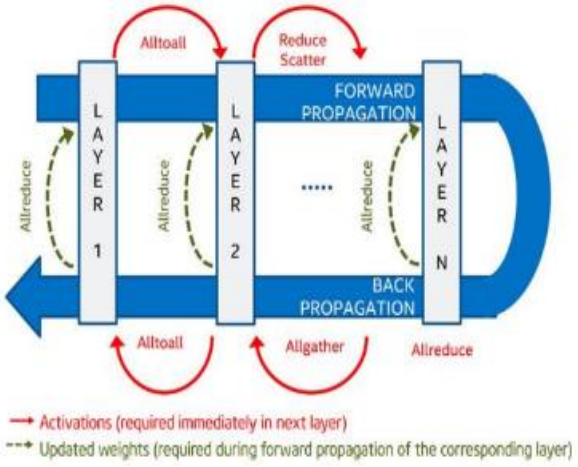
- Scale out training infrastructure
- Multi-Node
- Fabric interconnected training

Increased accuracy and reduced time to train

WHY MULTI-NODE TRAINING

- Data size and model complexity growing
- Need timely training or re-training of models
- AI shifting from fast single nodes to clusters
  - Data and model parallelism, and smart node grouping can keep scaling efficient
  - Proven near linear scalability; 97% scalability up to 256 Xeon Phi™ servers with Omni-Path and Resnet-50[1]



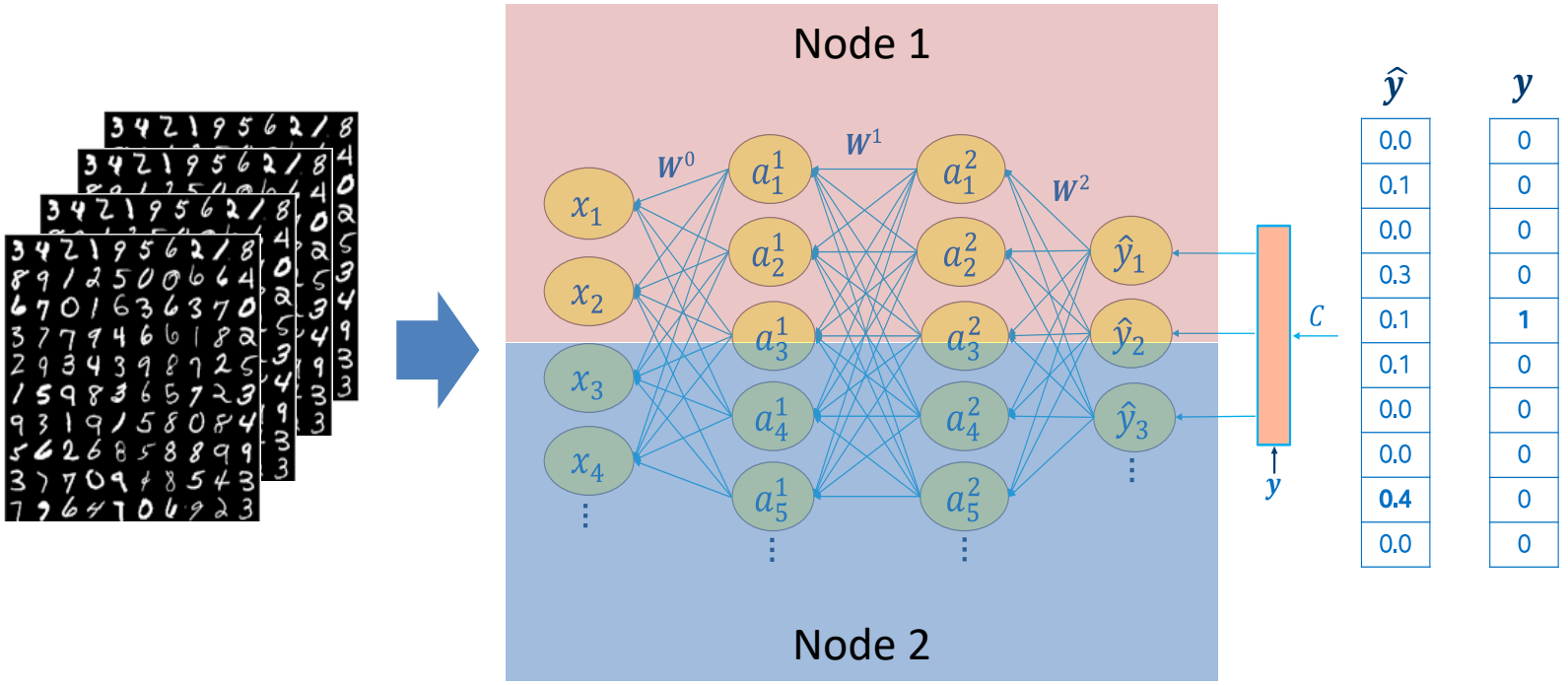


As scale node count, interconnect performance and scalability is critical

- Models are trained in a iterative manner requiring inter-node communication to proceed
- Injection bandwidth and scalable latency are key for iterative global weight updates
- Multi-MB Collective operations on the critical path for communications

1. https://blog.surf.nl/en/imagenet-1k-training-on-intel-xeon-phi-in-less-than-40-minutes/

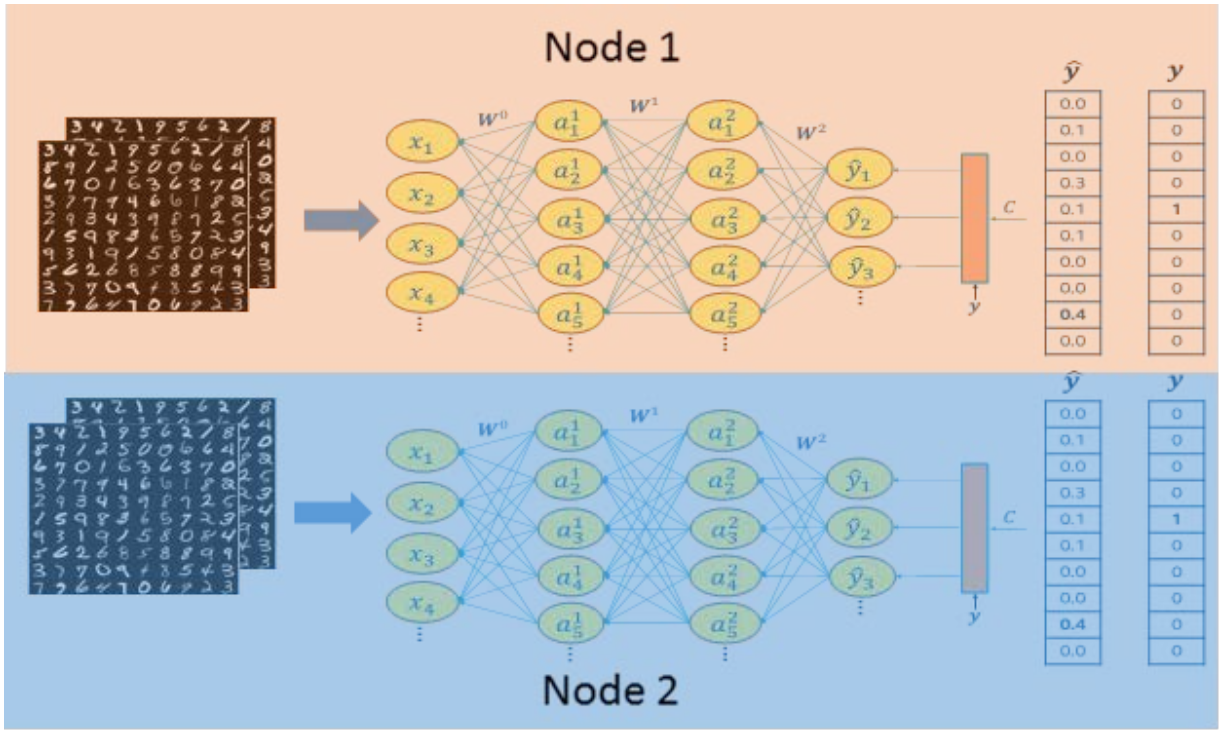
HOW TO PARALLELIZE (MODEL)



Model parallelism

- Share the Neural Network across many nodes
- Communication occurs for layers in each iteration; creates lots of communication
- Only as fast as slowest machine due to interactivity of code

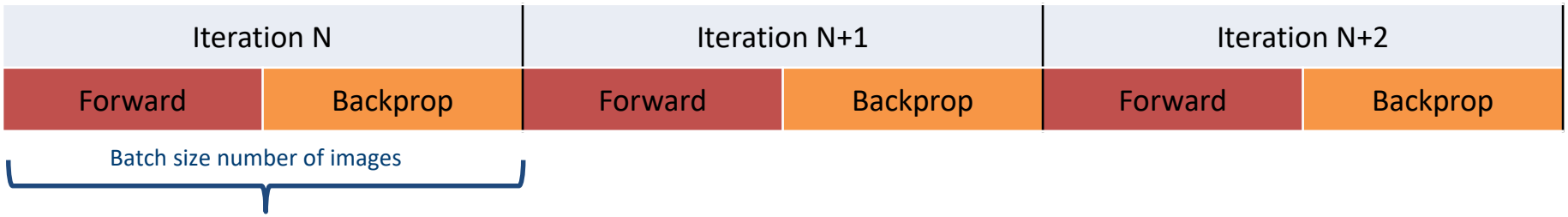
HOW TO PARALLELIZE (DATA)



Data parallelism

- Each system runs on its own dataset
- Communication occurs at each iteration, but less frequently then model parallelism
- Fast, non-blocking communication best to insure computation is not waiting on data.

DATA PARALLELISM



- **Forward prop is all compute, communication during back propagation***
- **Size of weights exchanged during comm phase depends only on topology[1], *not* on framework**
- **Implied barrier at iteration boundary***
- **Compute/comm overlap heavily dependent on framework implementation**

[1] Neural network topology! ;-)

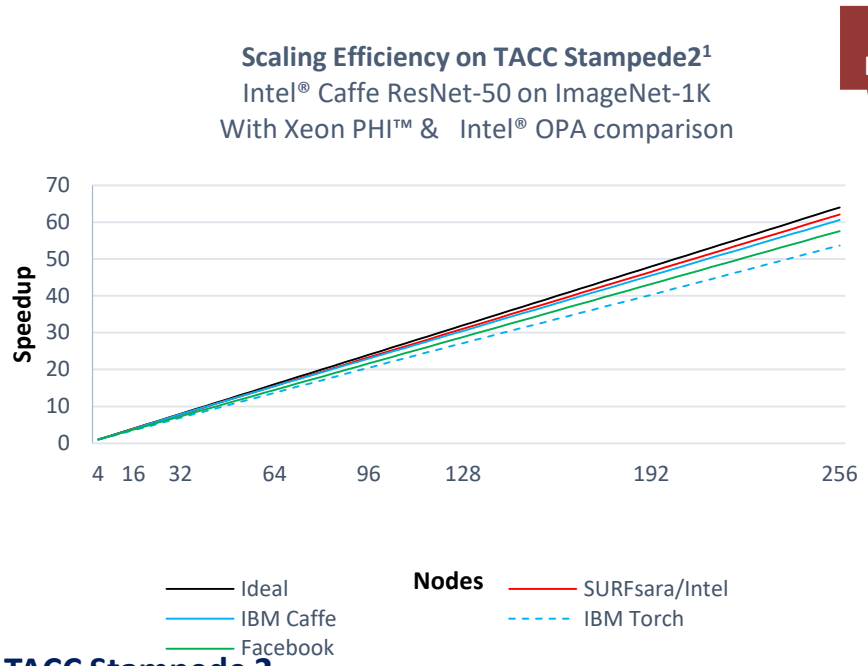
INTERESTING FABRIC FEATURES FOR AI

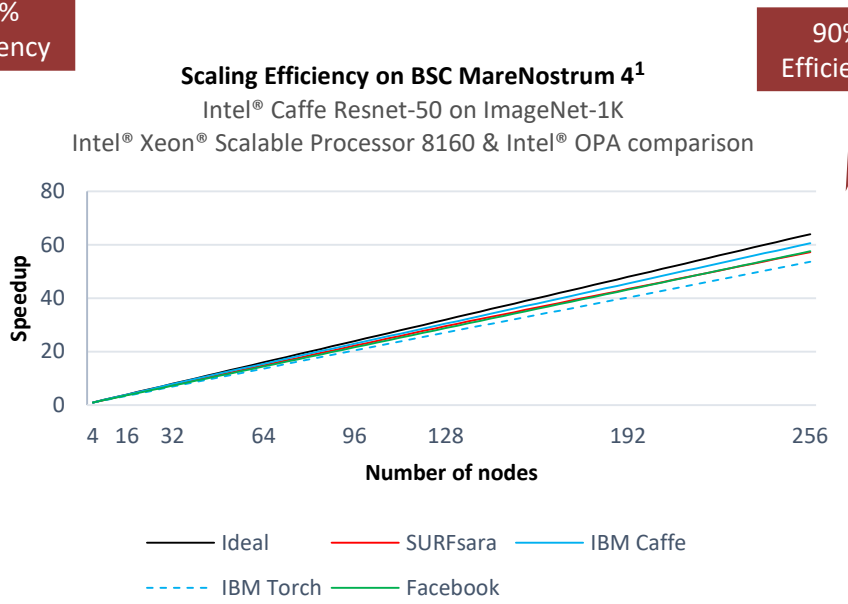
- **Higher Bandwidth**
  - data set ingest into training
  - large hyper-parameter set communications

- **Increased MPI message rate per core – helps model parallelism**
- **scalable latency**

AI ON HPC

SCALING EFFICIENCY





- **TACC Stampede 2**
  - **97% scaling efficiency** from 4 to 256 Intel® Xeon Phi™ 7250 nodes interconnected with Intel® OPA
  - Convergence with Top1/5 > **74%/92%**
  - 4 - 256 node runs: batch size of 16 per node, scaling efficiency of 97% in **63 minutes**

- **BSC MareNostrum 4**
  - Convergence with Top1/5 > **74%/92%**
  - 4 - 256 node runs: Batch size of 32 per node, 90% scaling efficiency, Total time to train: 70 Minutes

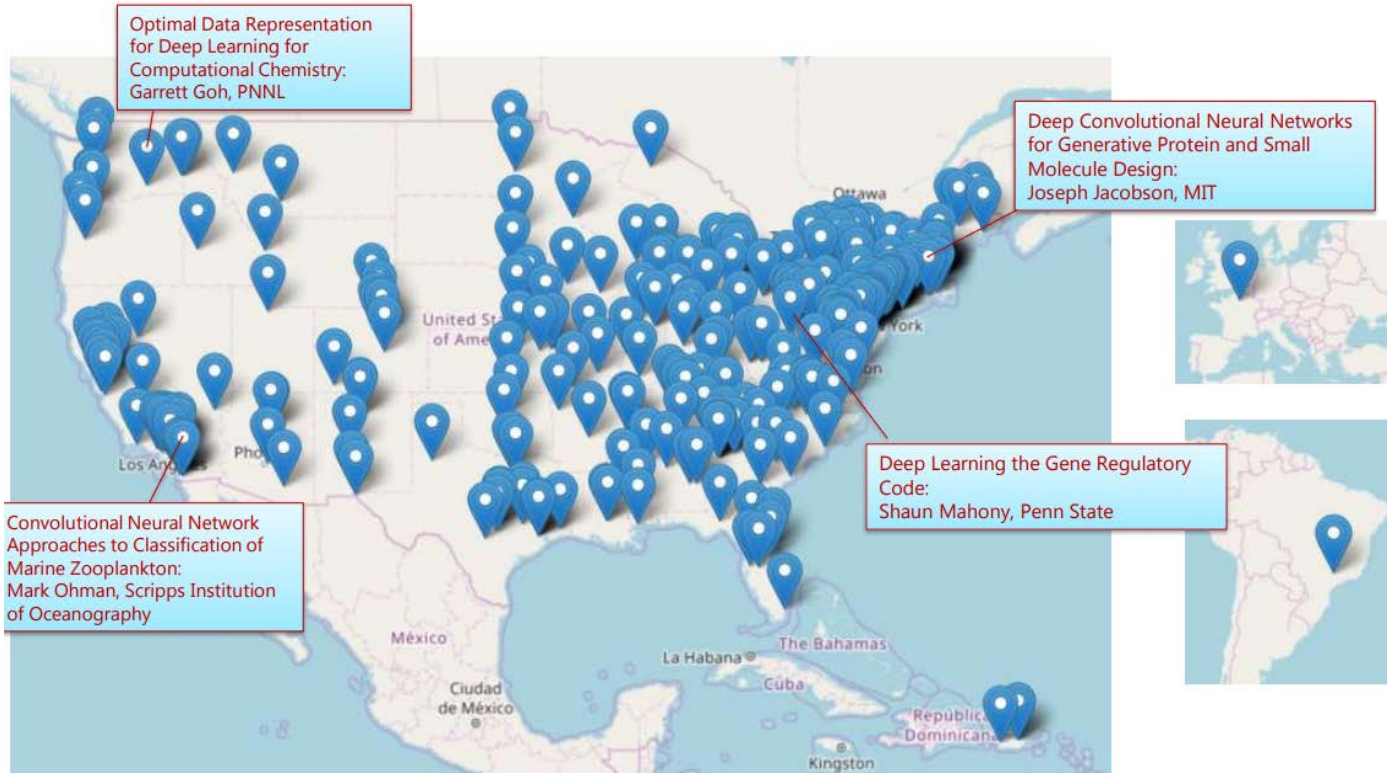
Strong multi-node training, with high accuracy with Intel® OPA

1. https://blog.surf.nl/en/imagenet-1k-training-on-intel-xeon-phi-in-less-than-40-minutes/

More Information
- https://www.bsc.es/user-support/mn4.php
- http://portal.tacc.utexas.edu/user-guides/stampede2
- Goyal, Priya, et al. "Accurate, Larg Minibatch SGD: Training ImagNetin 1 Hour." arXiv preprint arXiv:1706.02677 (2017)
- Cho, Minsik, et al. "powerAI DDL." arXiv preprint arXiv:1708.02188 (2017)
- IBM claims 95% scaling efficiency and Facebook claims 89%

AI ON HPC

- **MIT Lincoln Lab Supercomputing Center (LLSC)**
  - Address the learning phase of DL/AI especially for autonomous systems and device physics with enormous amounts of Big Data
- **Pittsburg Supercomputing Center: Bridges[1]**
  - 301 institutions
  - 1008 projects
  - 3682 users
  - 101 fields of science



[1] Nick Nystrom, PSC presentation at HPC User Forum, Milwaukee, Sept 2017

14th ANNUAL WORKSHOP 2018

THANK YOU

Todd Rimmer, Senior Principal Engineer
Ira Weiny, Software Engineer

Intel Corporation

[LOGO HERE]