Evolving Highly-Adapted AI with Supercomputers

Travis Johnston

Collaborators:
Robert Patton, Steven Young, Catherine Schuman,
Don March, Thomas Potok, Derek Rose,
Seung-Hwan Lim, Thomas Karnowski, Maxim Ziatdinov,
and Sergei Kalinin

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

Deep Learning is Pervasive in Commercial Applications

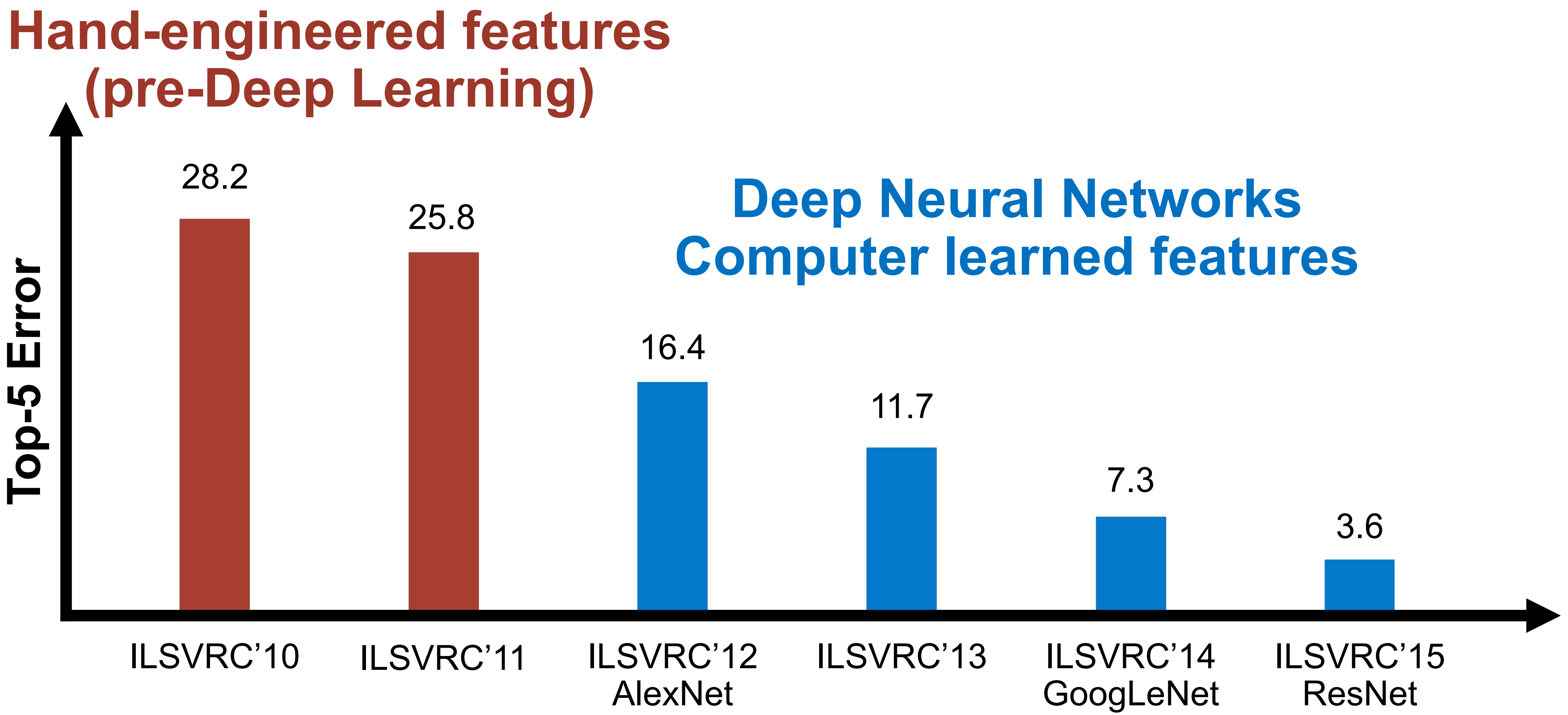
https://cs.stanford.edu/people/karpathy/cnnembed/

- **Computer Vision**
  Object Recognition
  Object Detection
  Semantic Segmentation
  Face Detection
  Facial Recognition

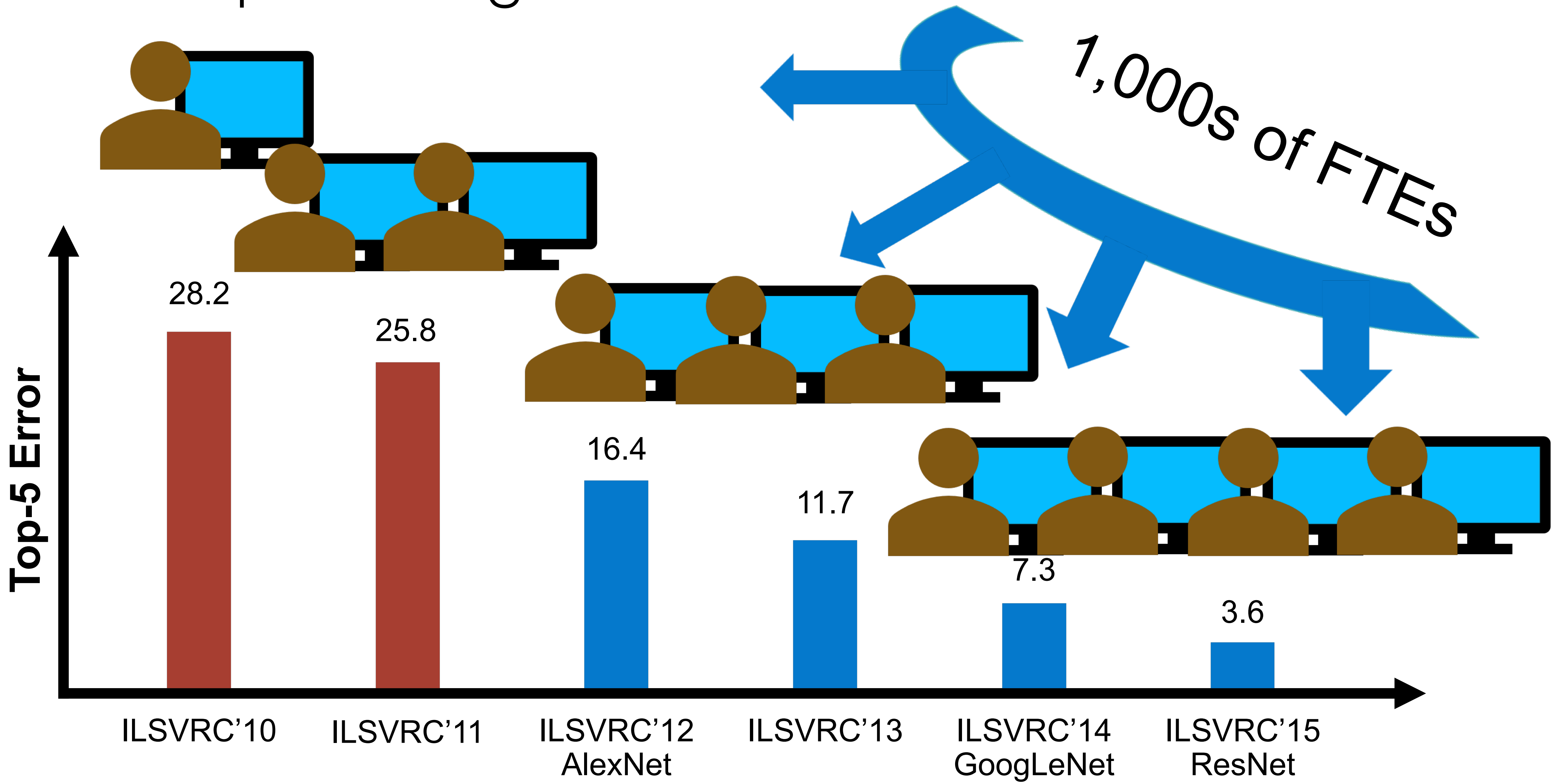
- **Natural Language Processing**
  Text translation
  Text generation (e.g. chat bots)
  Speech recognition

And many others…

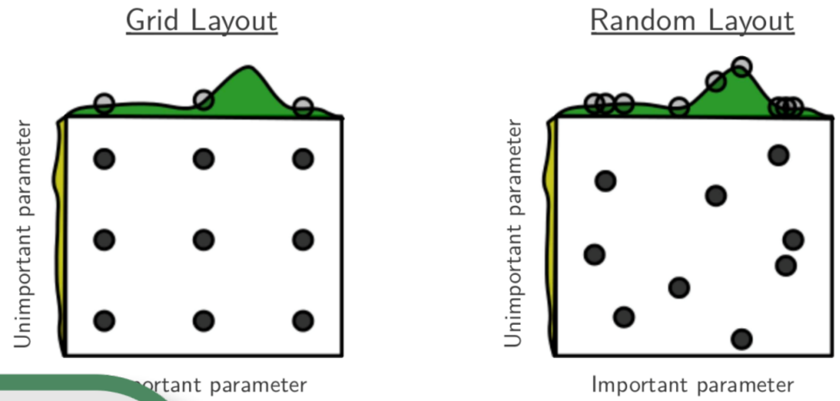
How Deep Learning Became State-of-the-Art



How Deep Learning Became State-of-the-Art



What Can Scientists Do?



…stra, J, and Bengio, Y. Random Search for Hyperparameter …ation, Journal of Machine Learning Research, Feb. 2012.

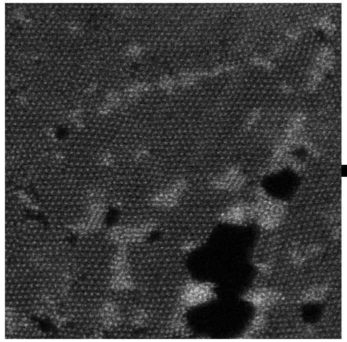
- *Tune* an out-of-the-box network
  - **Hyper-parameter sweeps**
    Assumes independence of hyper-parameters
  - **Grid search**
    Requires training
  - **Random search**
    Significant impro… …of information learned during training.

What if none of this works well enough?

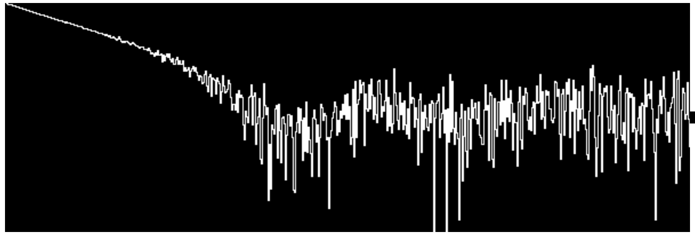
Why Do Scientific Applications of Deep Learning Trail Commercial Applications?

A cat sitting on a bench

A fabric surface



A picture containing an animal



A Christmas tree

Close-up of a logo

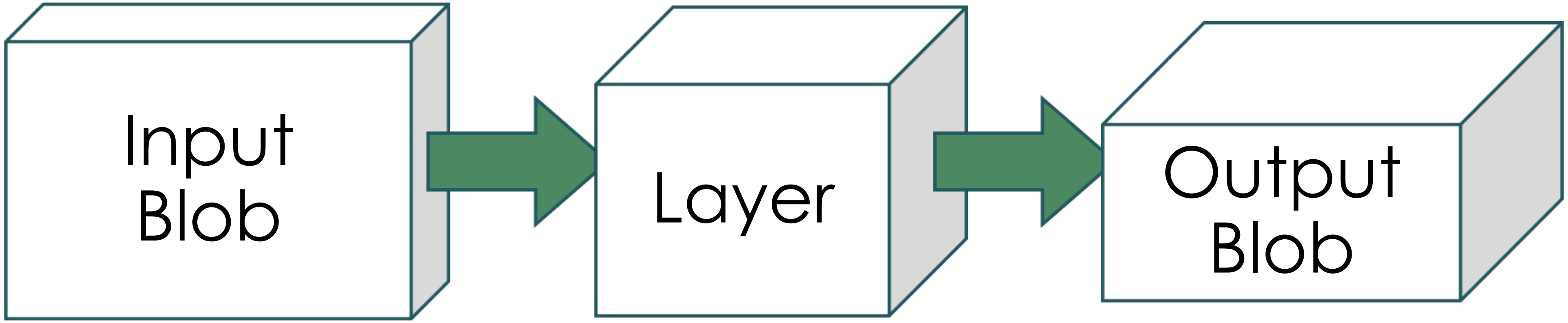
True Fact:
Scientific imagery or data

True Fact:
Fewer people can access

True Fact:
Fewer people are qualified

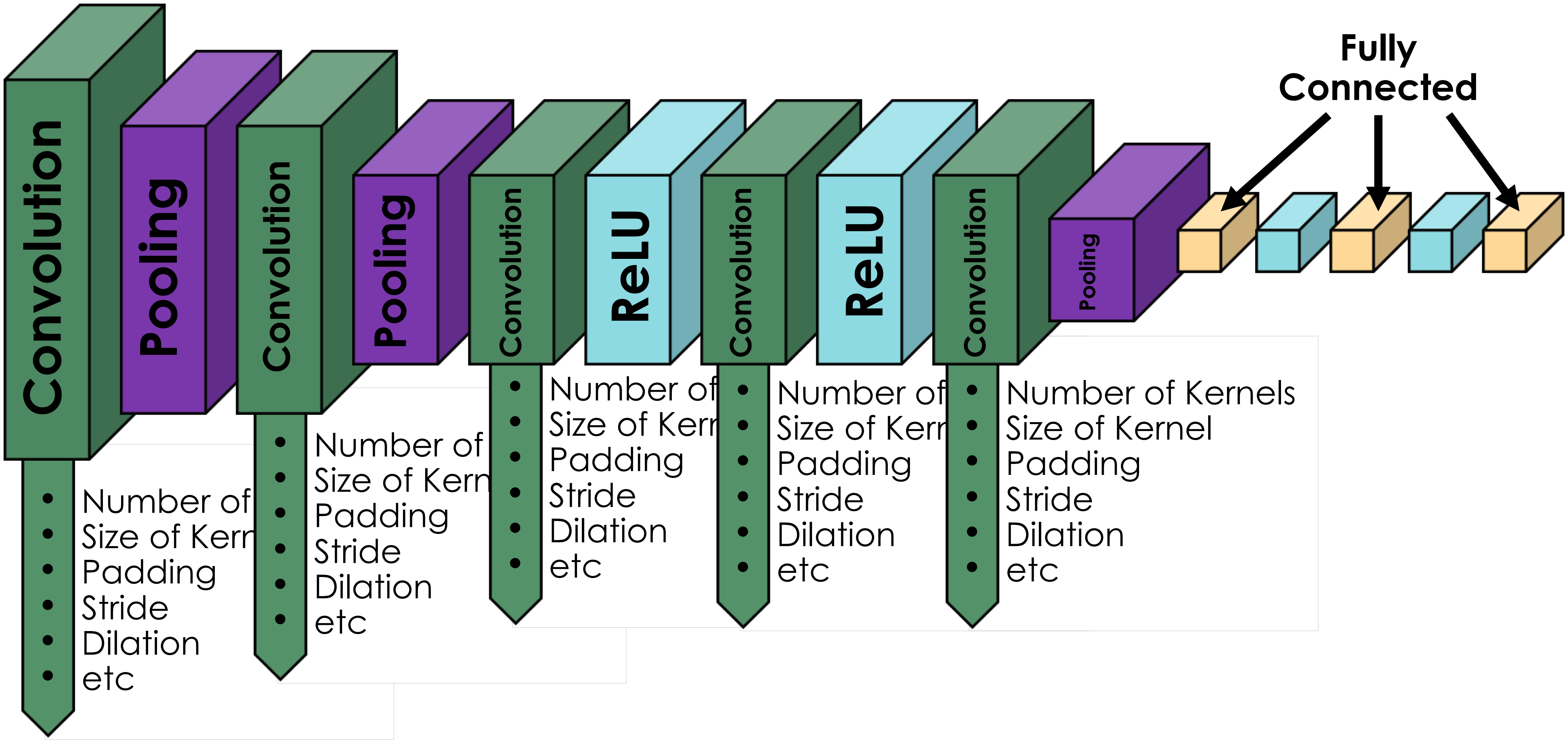
True Fact:
Commercial applications (typically) provide much greater financial incentive.

Designing a Neural Network (from scratch)



Convolution	Pooling	Inner Product, or Fully Connected
Number of Kernels		Number of Neurons
Size of Kernel	Size of Kernel	
Stride	Stride	
Pad	Pad	
Dilation	Type: MAX, AVG, etc	

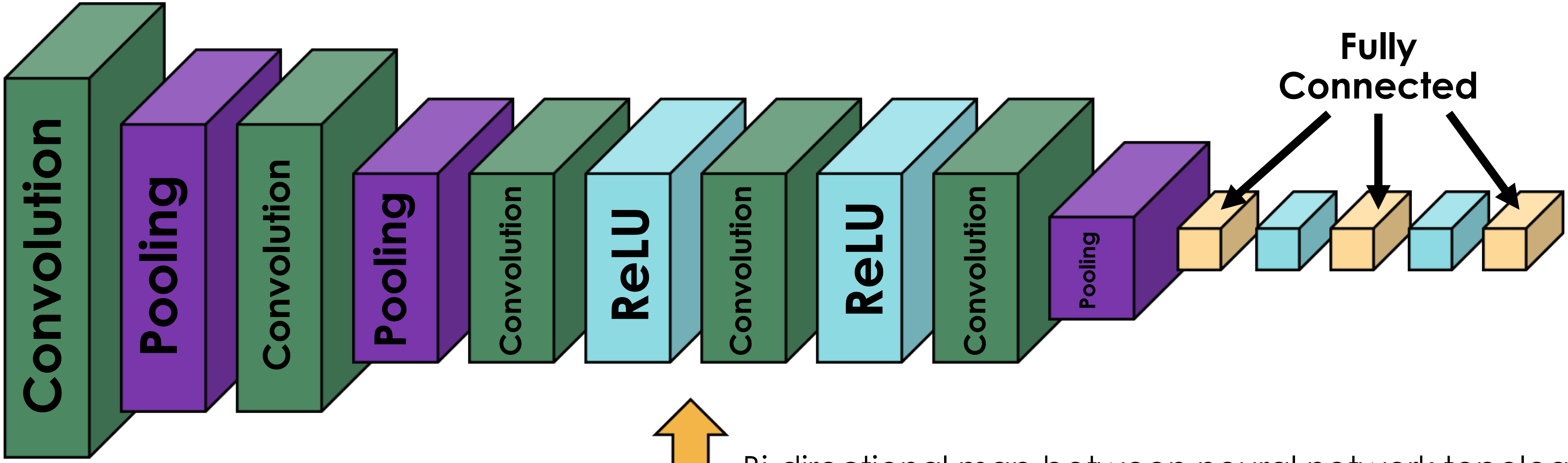
Designing a Neural Network (from scratch)



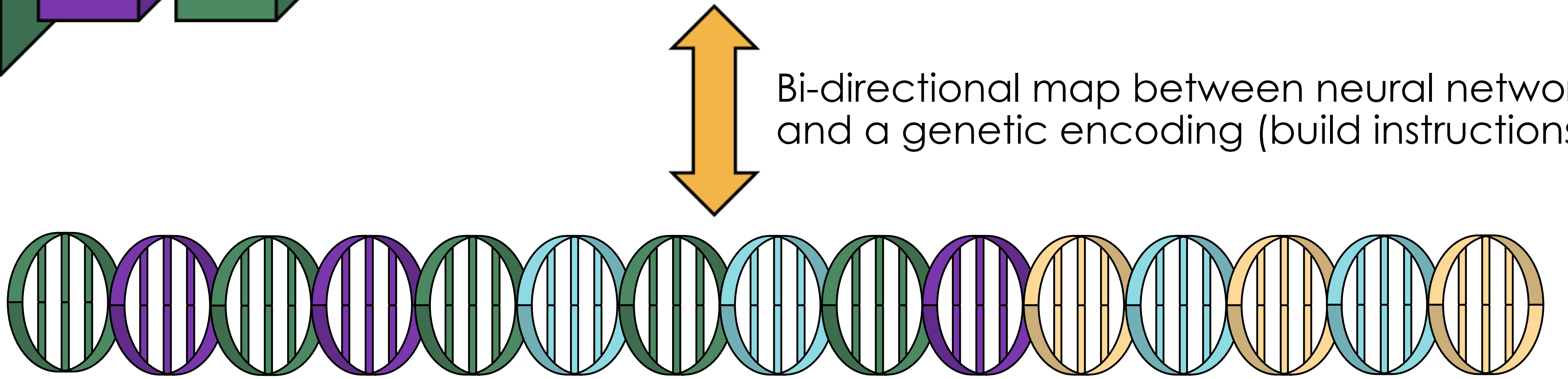
MENNDL: Multi-node Evolution of Neural Networks for Deep Learning

- Asynchronous, evolutionary algorithm used to explore and search hyper-parameter space for deep learning
  - Evolve *only* the network topology
  - Evaluate individual topologies through training process (e.g. SGD)
  - *Scalable* and *adaptability* for many data sets and compute platforms
- Leverage many GPUs
  - ~~Titan (18,688 K20 GPUs) evaluates about 5-900,000 networks per day~~
  - **Summit (about 27,600 Volta V100 GPUs) easily evaluating millions of networks per day**

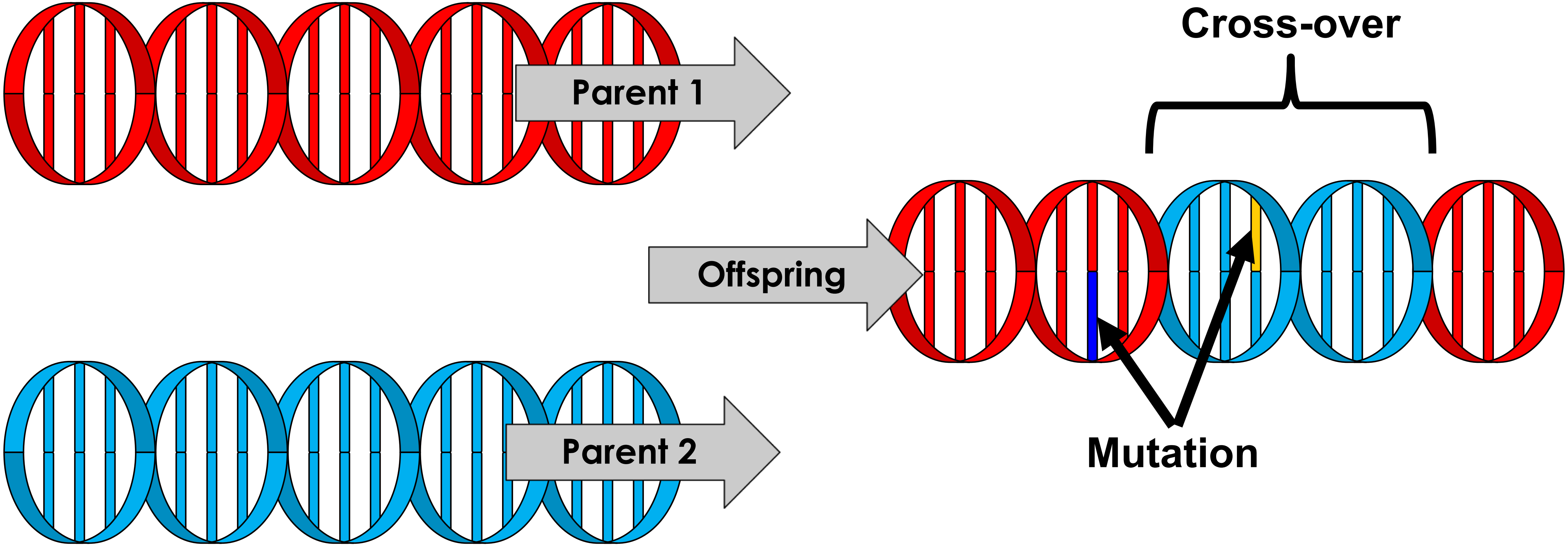
MENNDL:
Multi-node Evolution of Neural Networks for Deep Learning



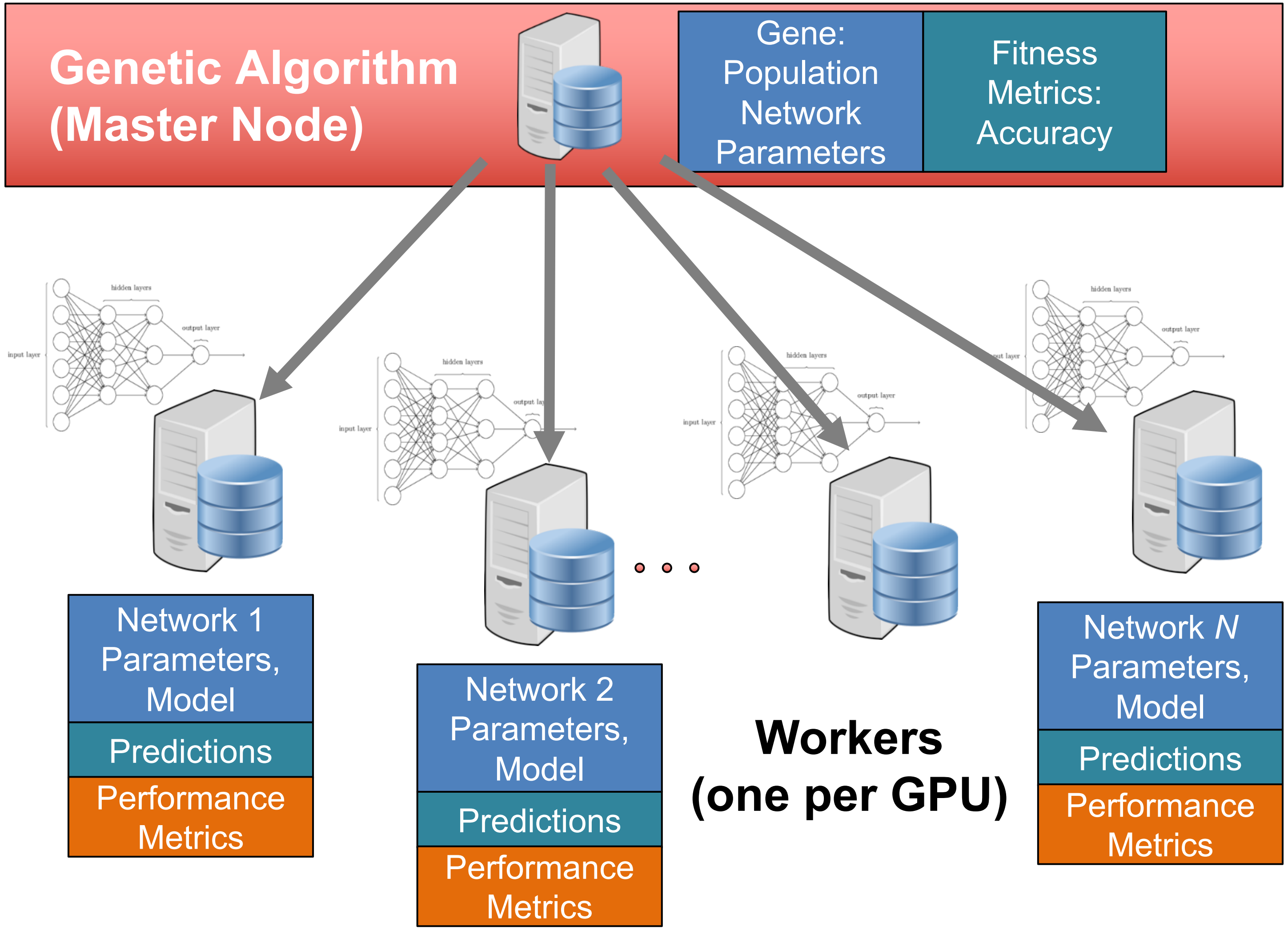
Bi-directional map between neural network topologies and a genetic encoding (build instructions)



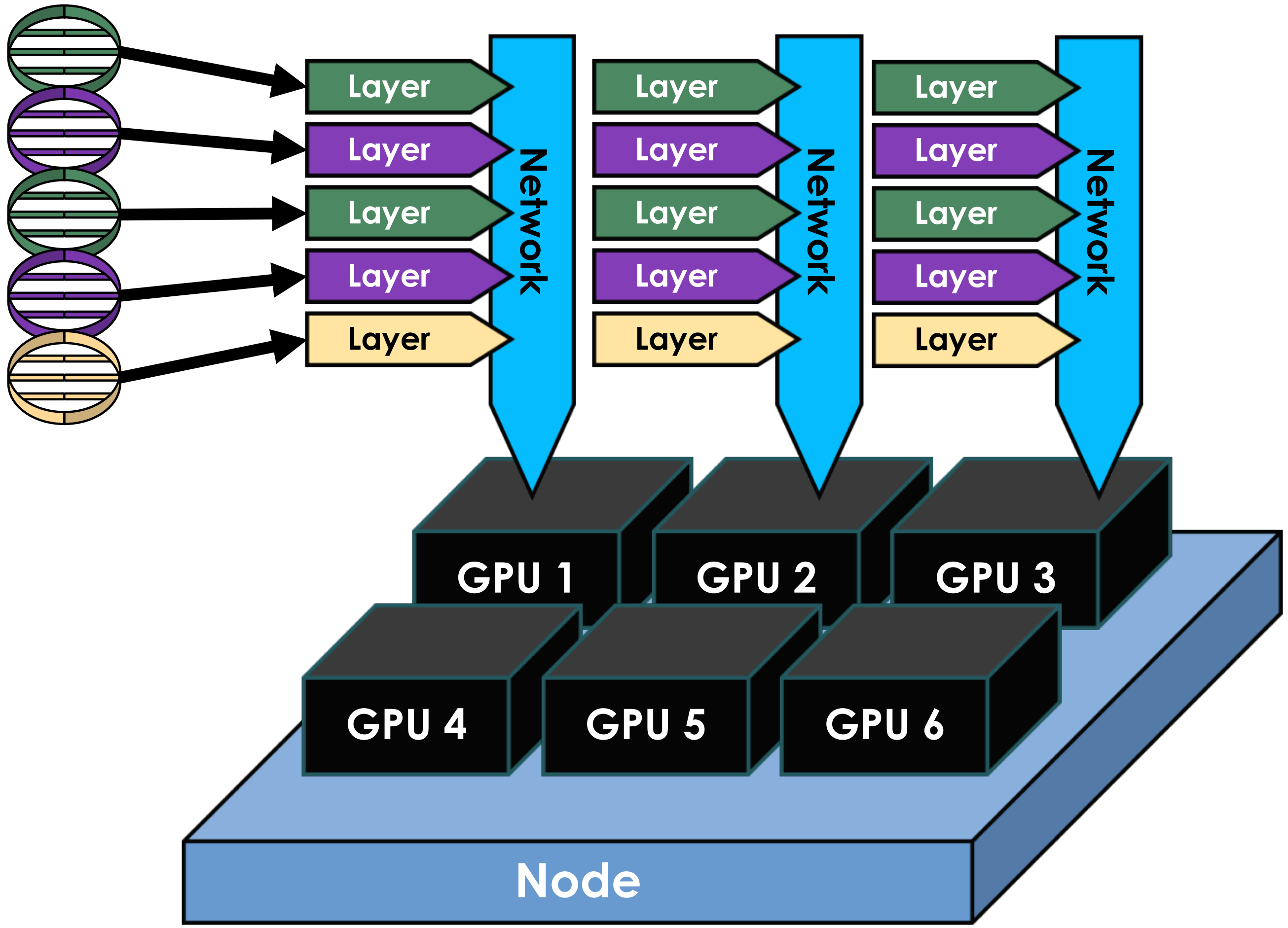
MENNDL: Evolution through Crossover and Mutation



MENNDL: Asynchronous Evaluation of Networks



Measuring the Performance of MENNDL



Measuring the Performance of MENNDL

Layer

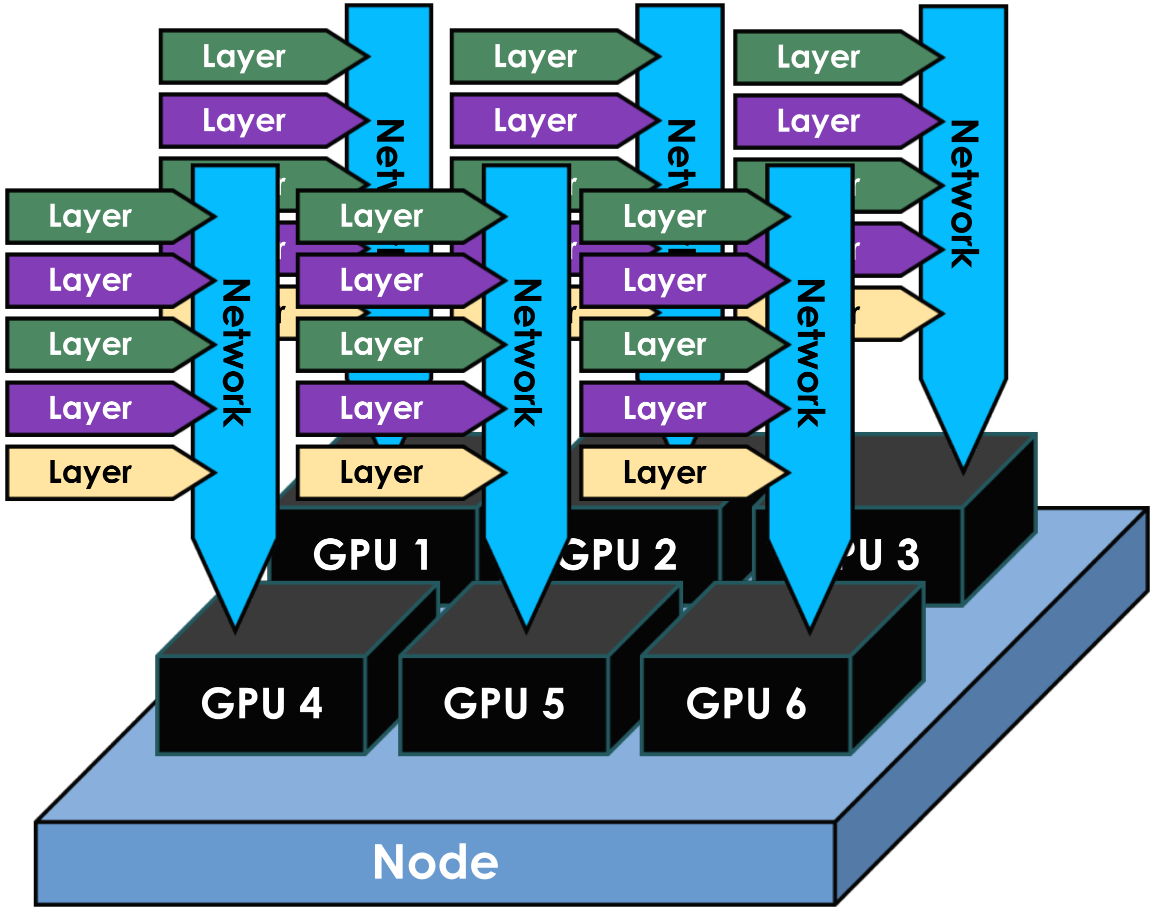
Combinations of layer hyper-parameters affect:
- Computational cost
  (i.e. number of operations to transform data)
- What features *may* be learned

Network

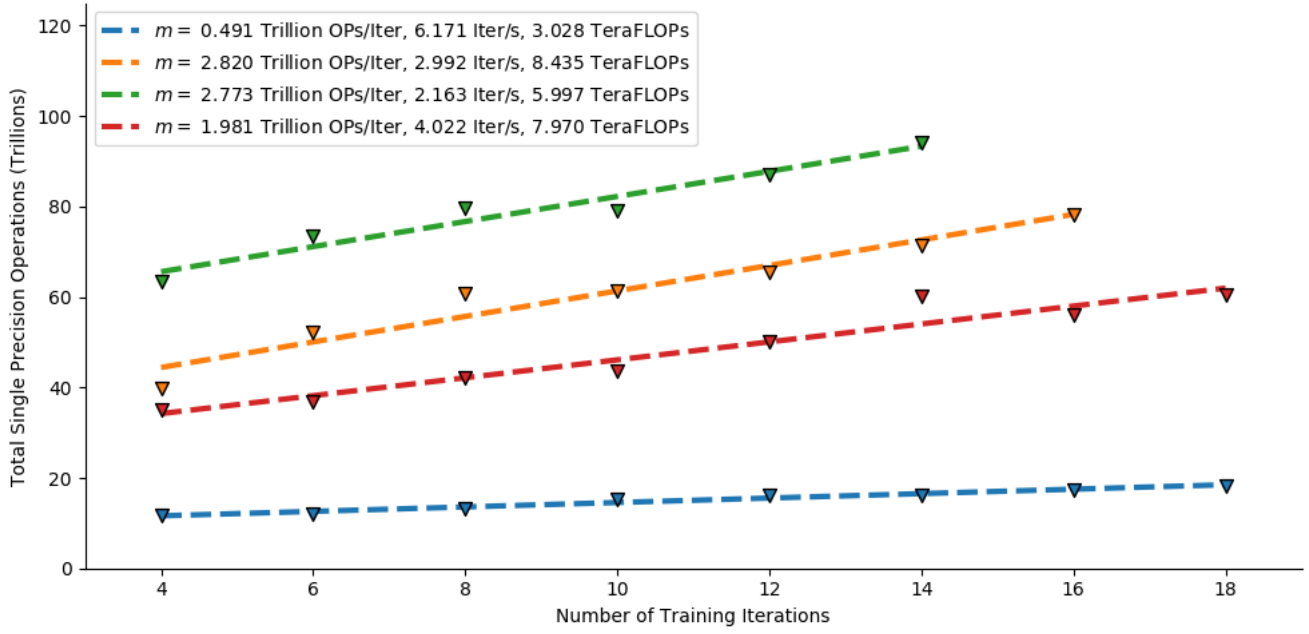
Combinations of layers affect:
- Single-GPU computational cost (**FLOPS**)
  (i.e. number of operations to transform data)
- How features are aggregated
- **Accuracy** of *single* network

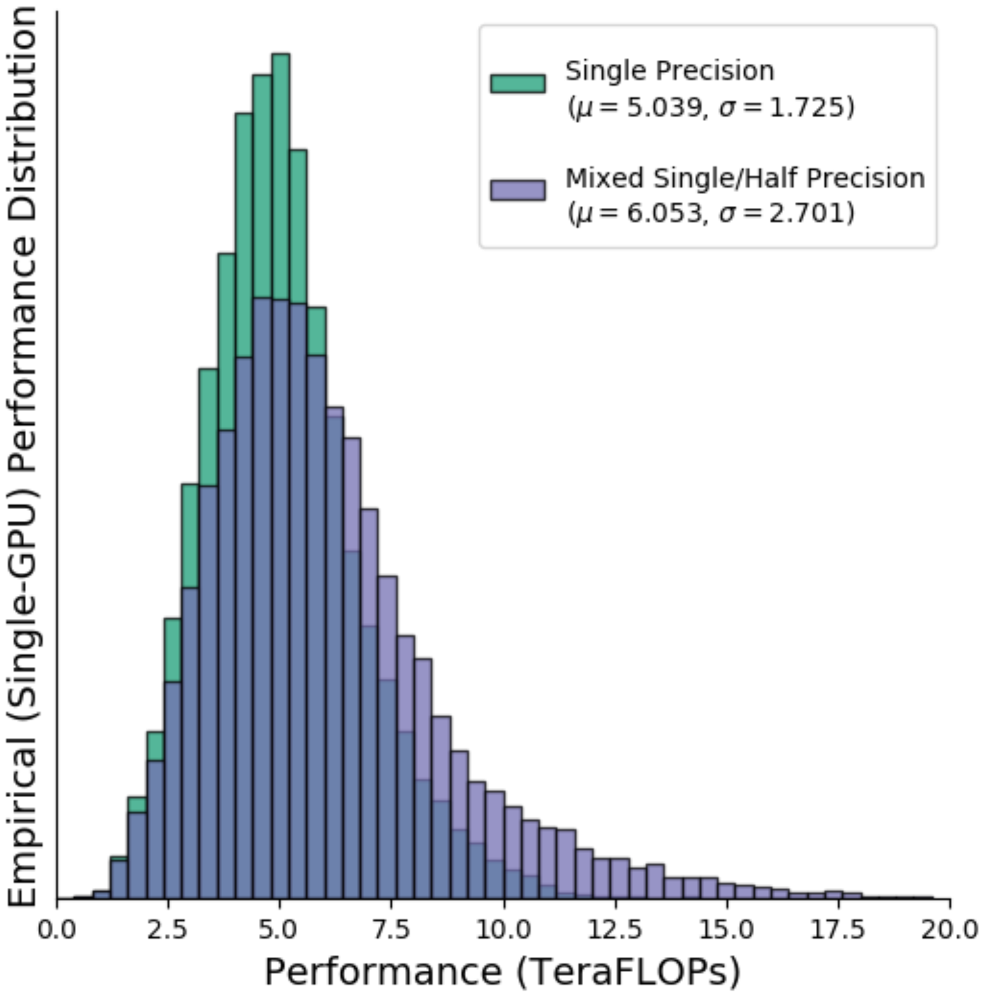
Whole System

Population of networks affect:
- Overall computational cost (**FLOPS**)
- **Overall Accuracy**
  (fittest individual, rate of convergence, etc)



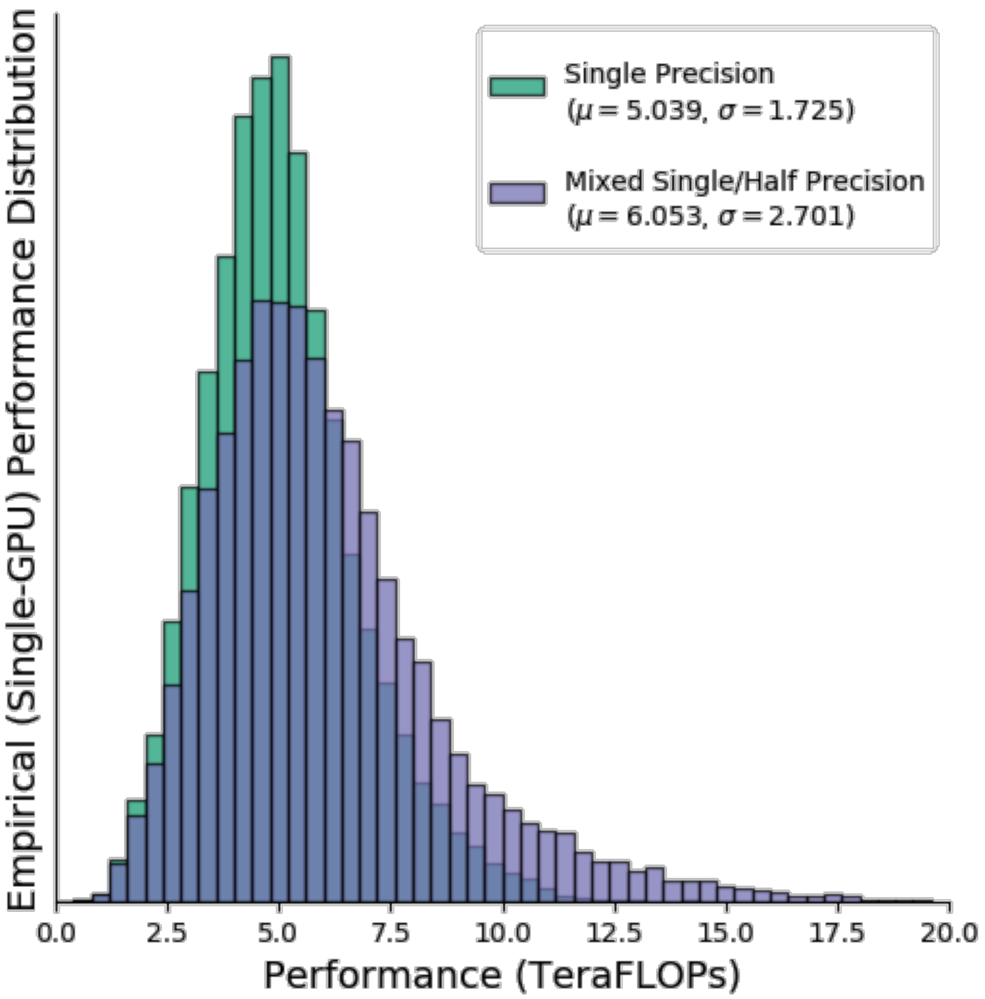
Node-Level Compute Performance (FLOPS)

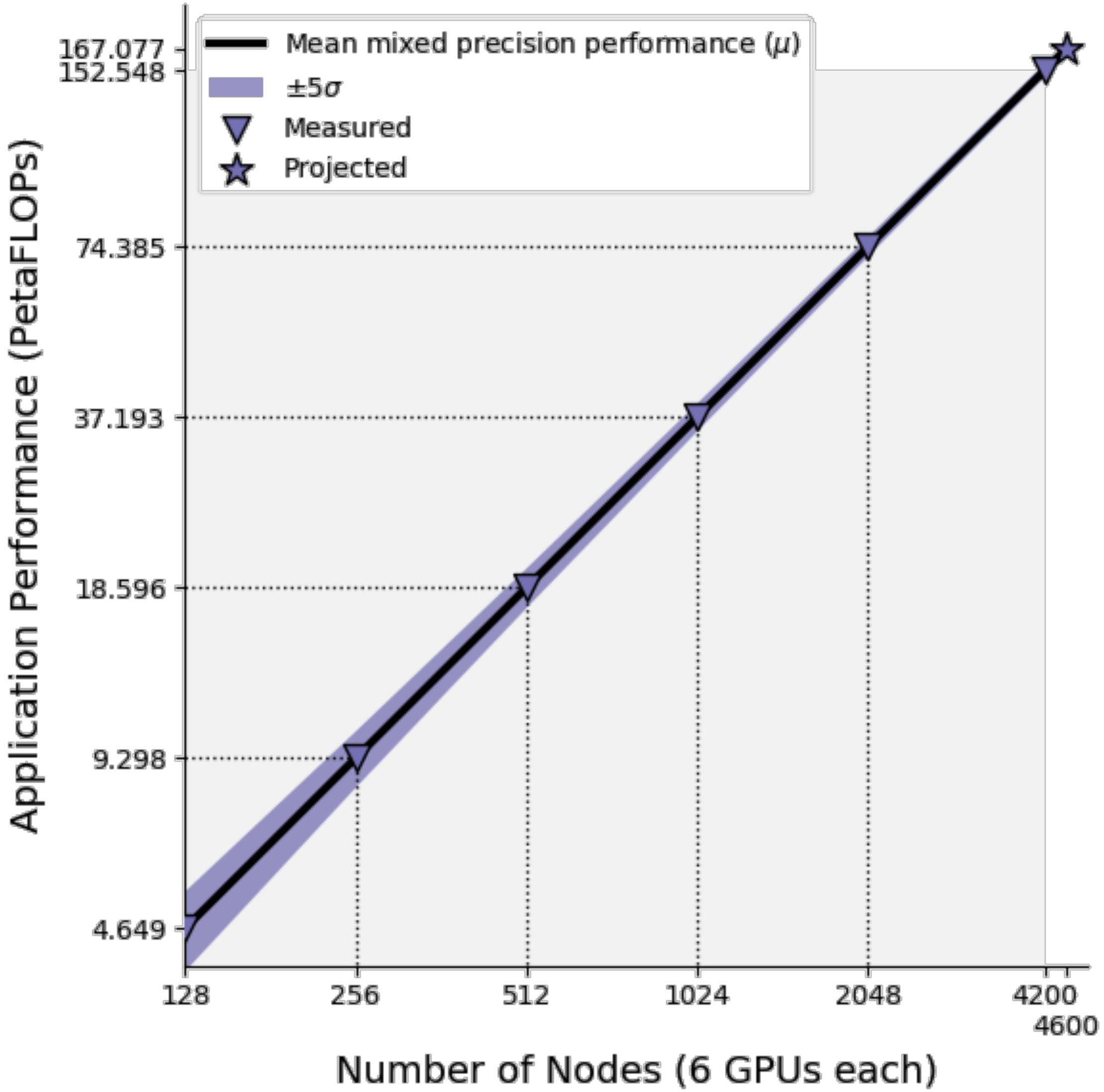




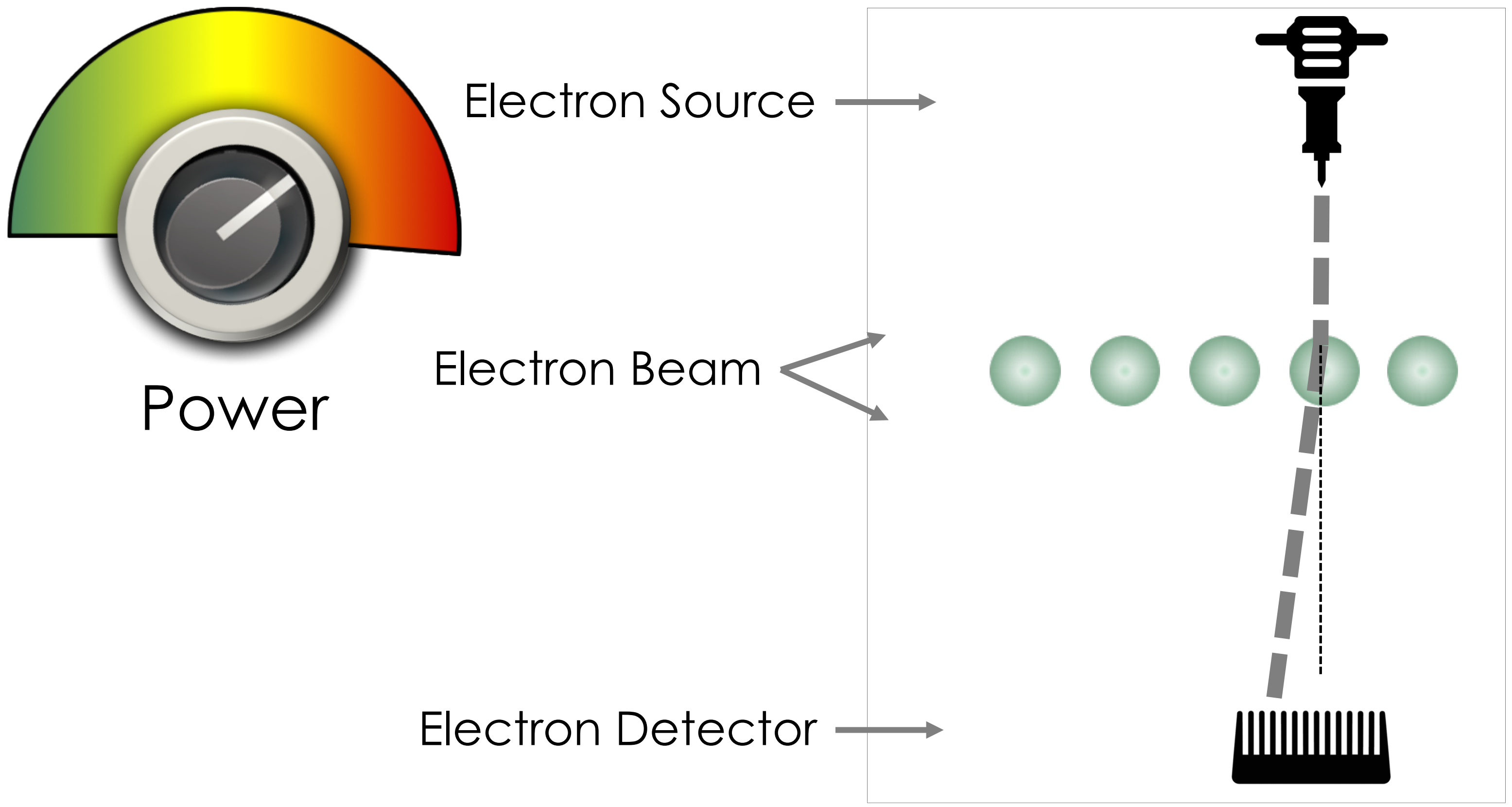
OAK RIDGE National Laboratory

System-Level Performance (FLOPS) and Scaling

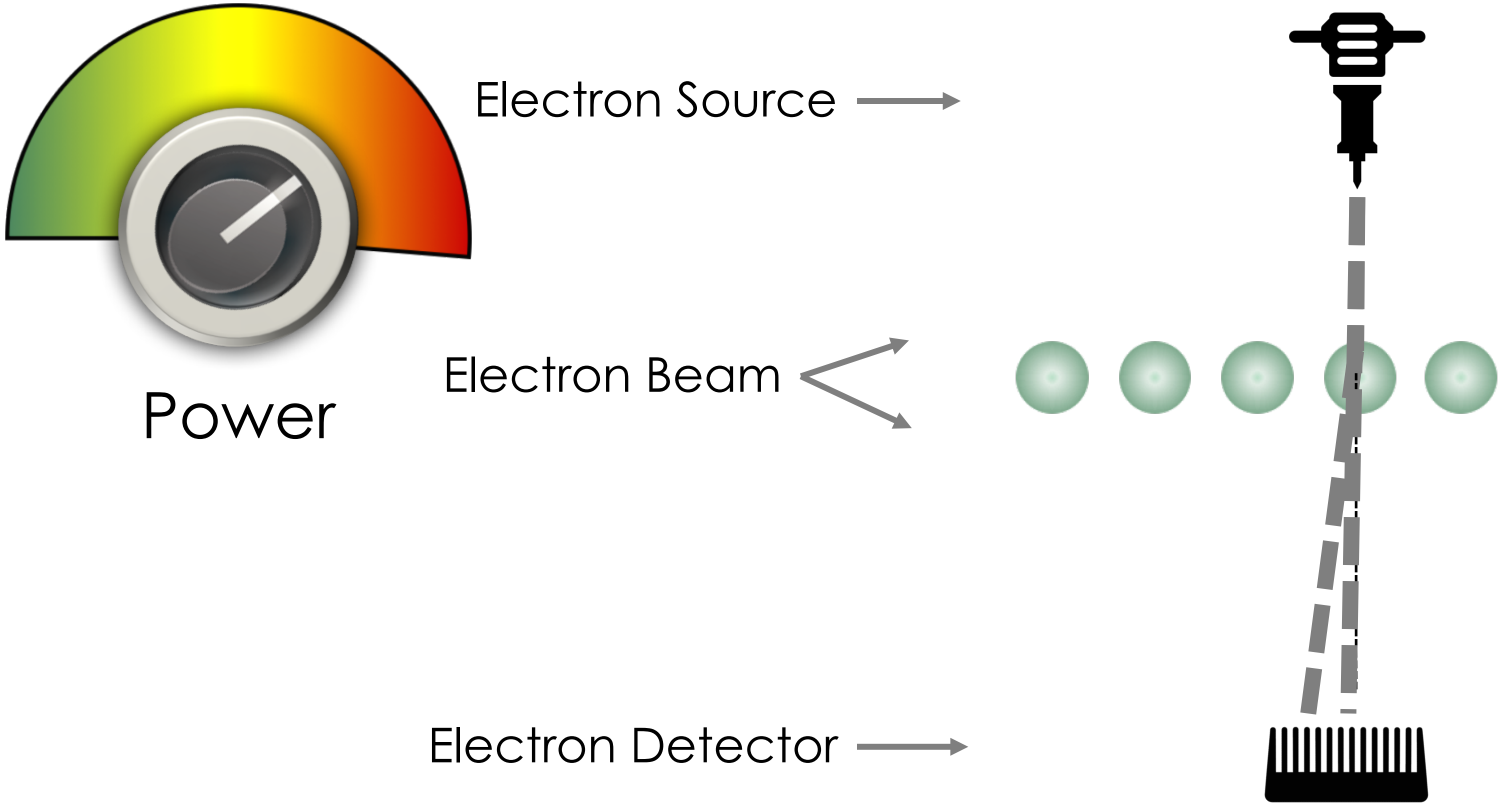




Scanning Transmission Electron Miscroscopy (STEM)

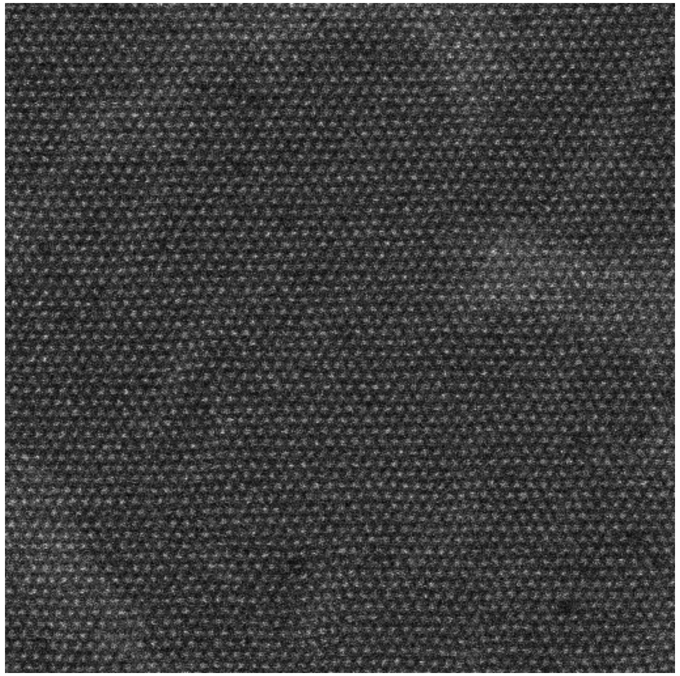


Scanning Transmission Electron Miscroscopy (STEM)

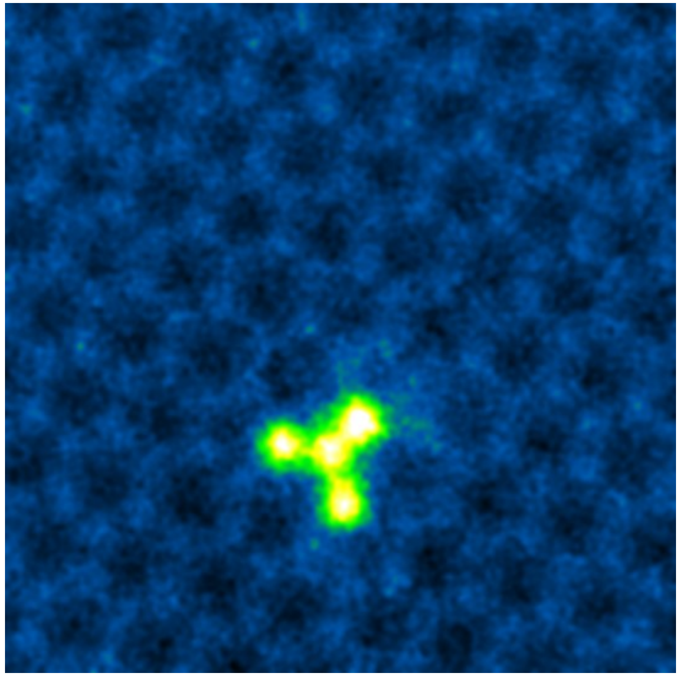


Vision: Use *AI* to Drive the Electron Microscope

Atomic Imaging of Materials



Manual Atomic Manipulation



https://www.ornl.gov/news/scientists-forge-ahead-electron-microscopy-build-quantum-materials-atom-atom
Ondrej Dyck, Sergei Kalinin, Stephen Jesse, Albina Borisevich

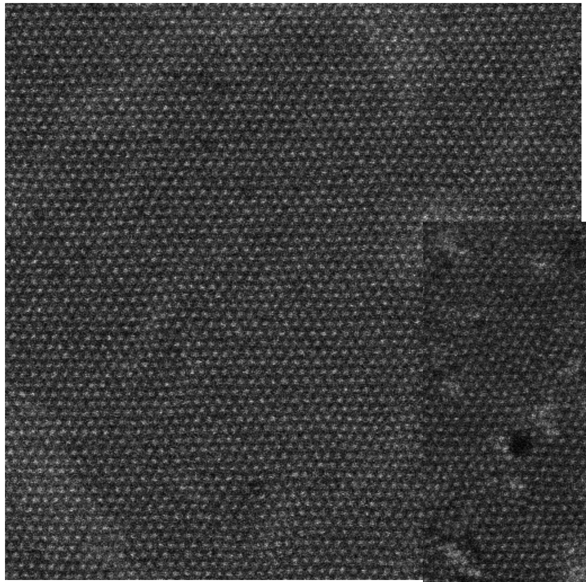
AI-Driven Atomic Manipulation

- Enable large scale production of materials customized at the atomic scale

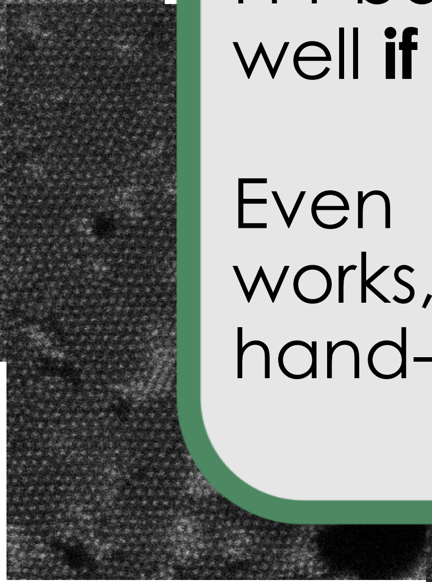
Application: Electron Microscopy
1st step towards building a smart, self-driving microscope

Atomic Level Data

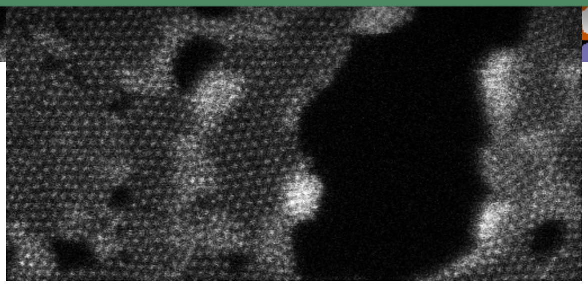
FFT-Generated Labels



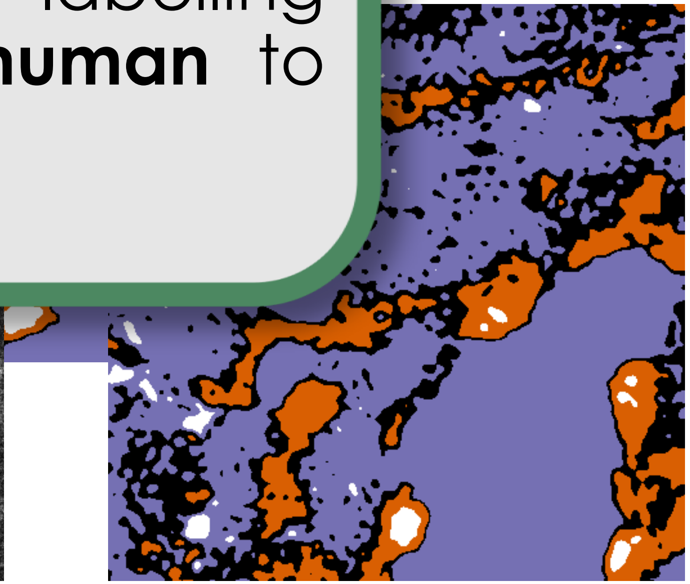
Frame 1



Frame 88



Frame 99



= Defect (positive)

= Not defect (negative)

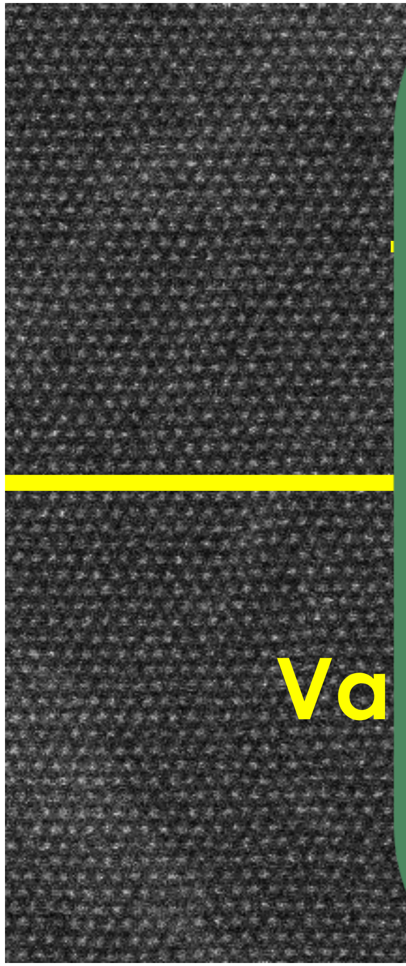
= False positive (sort of bad)

= False Negative (really bad)

FFT-based labelling **only** works well **if the lattice is regular.**

Even if FFT-based labelling works, **it requires a human** to hand-tune parameters

Application: Electron Microscopy
1st step towards building a smart, self-driving microscope



Provide data to MENNDL:

- ***Accuracy*** run
  1,000 Nodes
  6,000 GPUs
  6 hours
  Train ~200,000 Networks
- ***Benchmarking*** run
  4,200 Nodes
  25,200 GPUs
  2 hours
  Profile 25,200 Networks

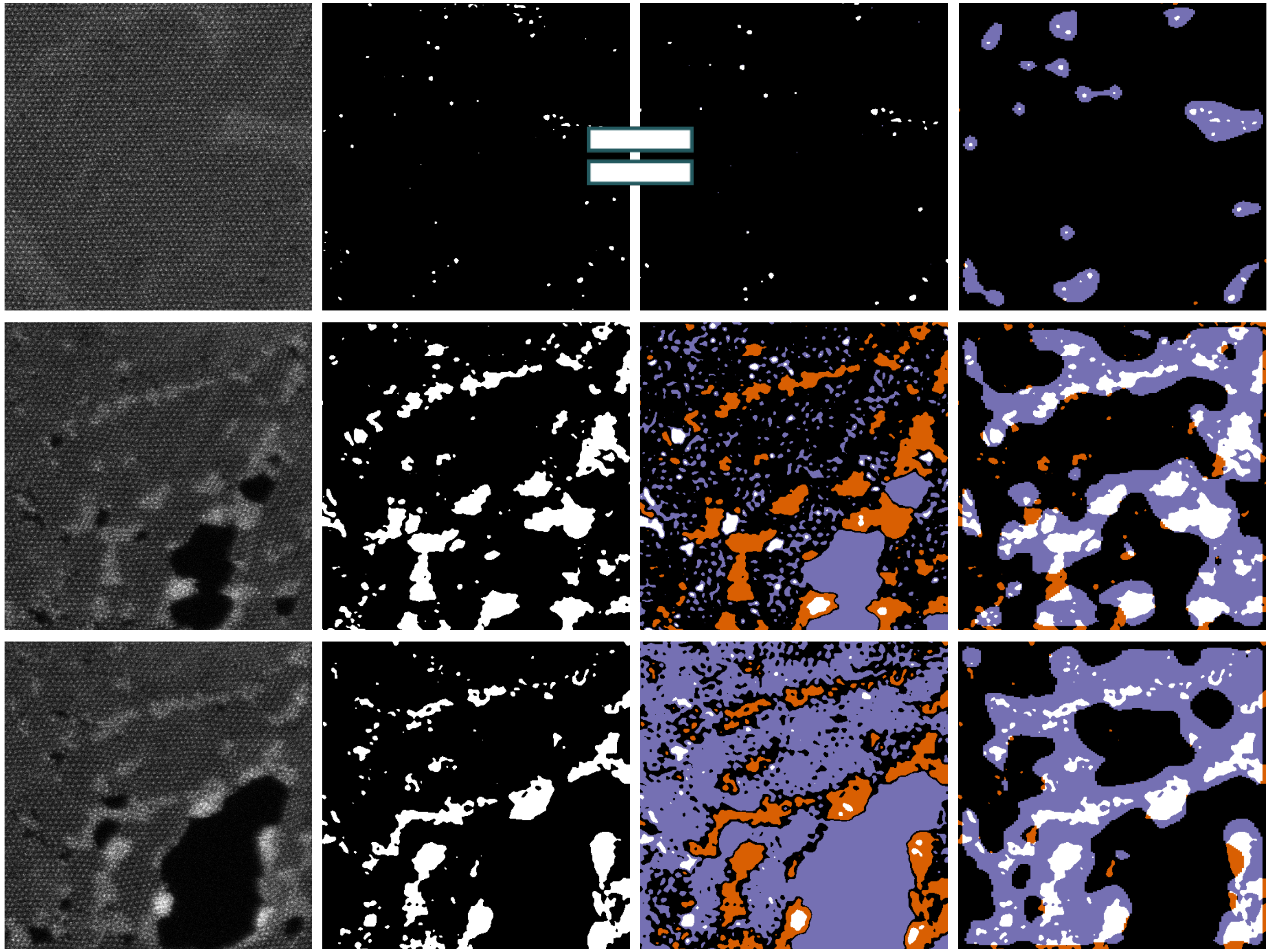
Results

Instrument Data | Human Labels | FFT-based Labels | CNN-based Labels

Frame 0

Frame 88

Frame 99



= True Positive

= True Negative

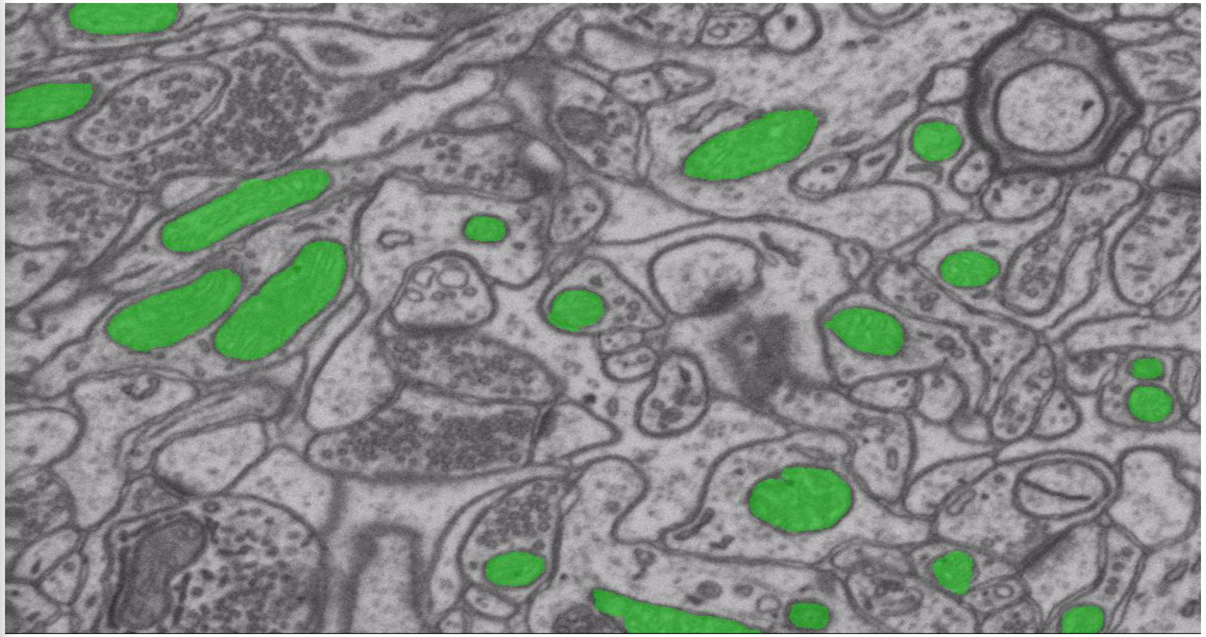
= False Positive

= False Negative

Application: 3D Electron Microscopy

MENNDL:
- 24 hrs on Titan, 18,000 Nodes
- Evaluated 900,000+ networks
- 93.8% Accuracy

  Reduction in error of more than 30% over *standard* networks and human hand-crafted networks

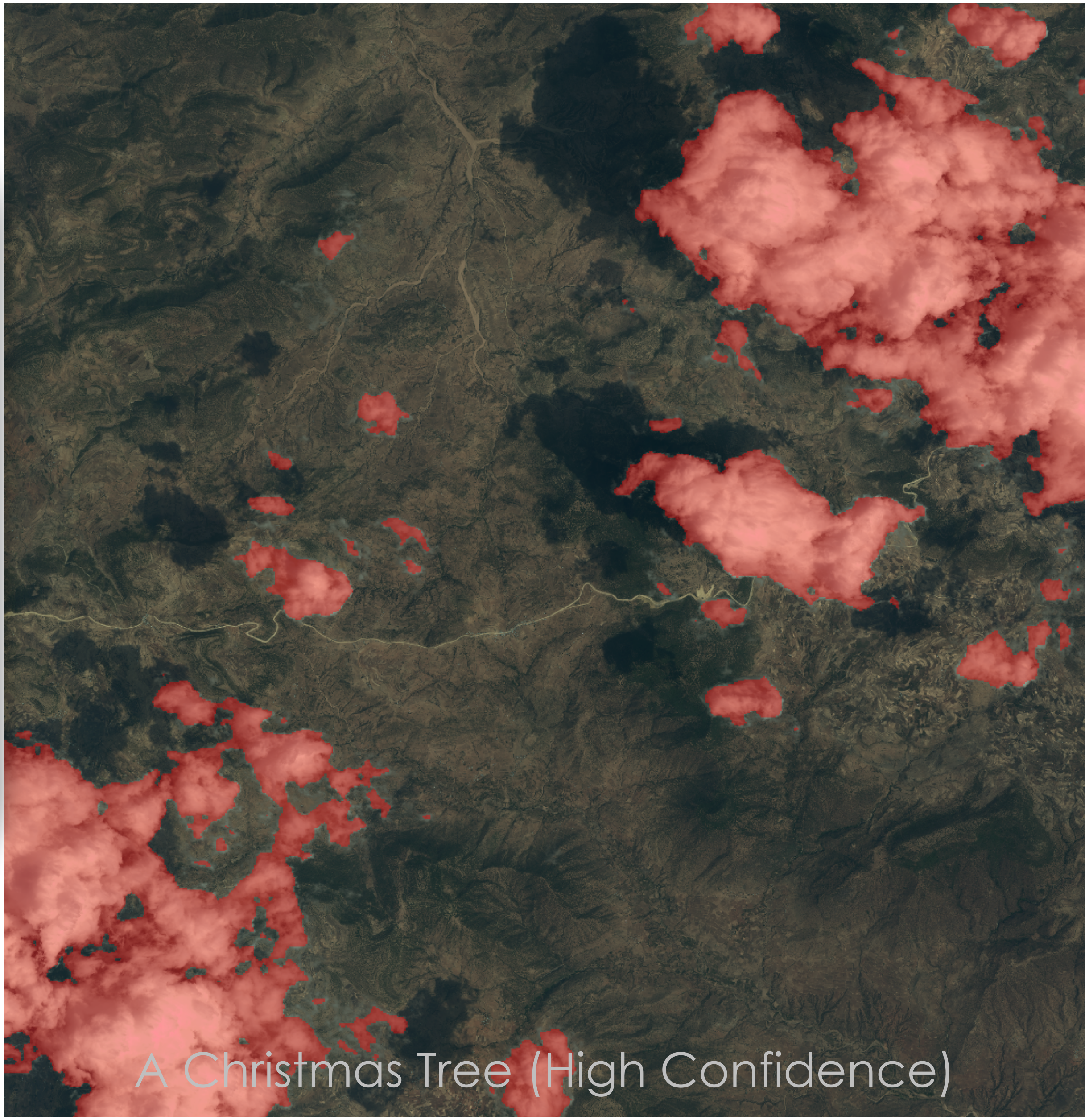


A fabric surface (High Confidence)

Application: Cloud Detection

MENNDL, smart mutation

- 5 hrs on Titan, 1,000 Nodes
- Evaluated 25,000 networks
- 97.5+% Accuracy
  - 200x faster inference
  - 1/10th memory
  - 40% reduction in error over GoogLeNET



Conclusion

- Deep Learning solutions for *commercial* data rarely transfer seamlessly to *scientific* data.
- MENNDL leverages a massively parallel, genetic, asynchronous algorithm on HPC systems to tailor make neural networks when *commercial* solutions fail.
  - Easily achieves performance over 167 Pflops on Summit
  - Evaluates around 2.5 Million neural networks per day
- MENNDL enables custom deep learning for science by removing the time-consuming hand-tuning process of creating custom neural networks.

References

- **SC18 (Gordon Bell Finalist)**
  167-PFlops Deep Learning for Electron Microscopy: From Learning Physics to Atomic Manipulation (Wednesday 4:00-4:30)

- SC17 – MLHPC Workshop (MENNDL)
  Steven R. Young, Derek C. Rose, Travis Johnston, William T. Heller, Thomas P. Karnowski, Thomas E. Potok, Robert M. Patton, Gabriel Perdue, and Jonathan Miller. 2017. Evolving Deep Networks Using HPC. In Proceedings of the Machine Learning on HPC Environments (MLHPC'17). ACM, New York, NY, USA, Article 7, 7 pages. DOI: https://doi.org/10.1145/3146347.3146355

- SC17 – MLHPC Workshop ("Smarter Mutation in MENNDL")
  Travis Johnston, Steven R. Young, David Hughes, Robert M. Patton, and Devin White. 2017. Optimizing Convolutional Neural Networks for Cloud Detection. In Proceedings of the Machine Learning on HPC Environments (MLHPC'17). ACM, New York, NY, USA, Article 4, 9 pages. DOI: https://doi.org/10.1145/3146347.3146352