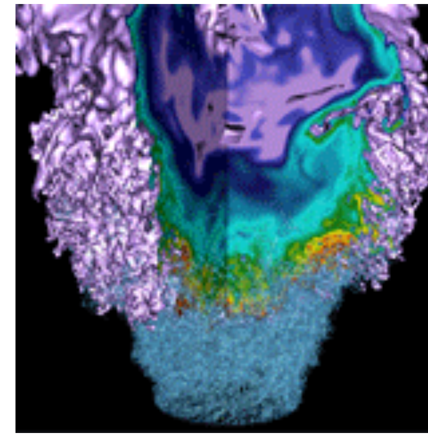
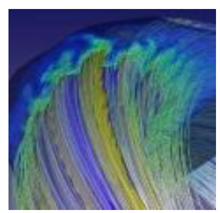
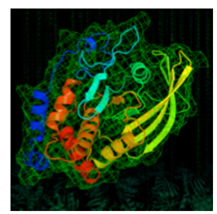
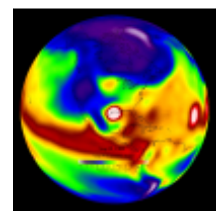
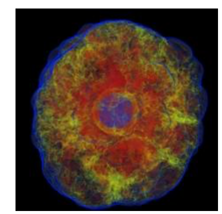
Semi-Supervised Deep Learning for Climate @ Scale

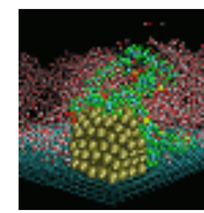












Prabhat
Lawrence Berkeley National Laboratory

QCon
6/28/2017

Outline

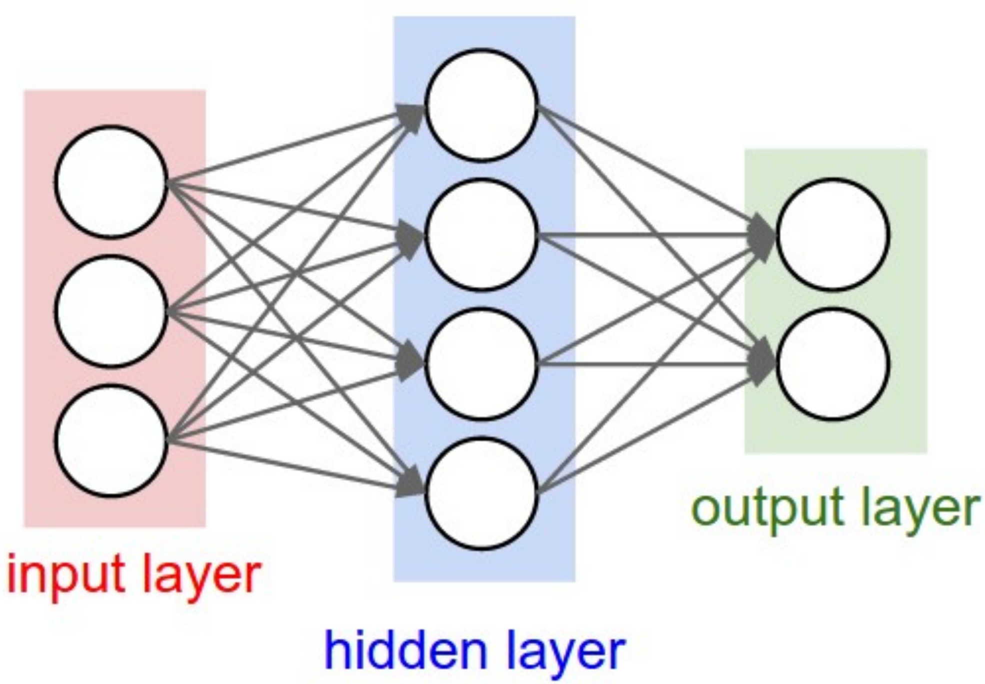
- **Introduction to Deep Learning**
  - History
  - Commercial Applications
  - Convolutional Neural Nets
- **Climate Science**
  - Motivation
  - Representational Challenges
    - Supervised CNNs
    - Semi-Supervised CNNs
  - Scaling Challenges

Outline

- **Introduction to Deep Learning**
  - History
  - Commercial Applications
  - Convolutional Neural Nets
- **Climate Science**
  - Motivation
  - Representational Challenges
    - Supervised CNNs
    - Semi-Supervised CNNs
  - Scaling Challenges

1960's (1^{st} Wave)

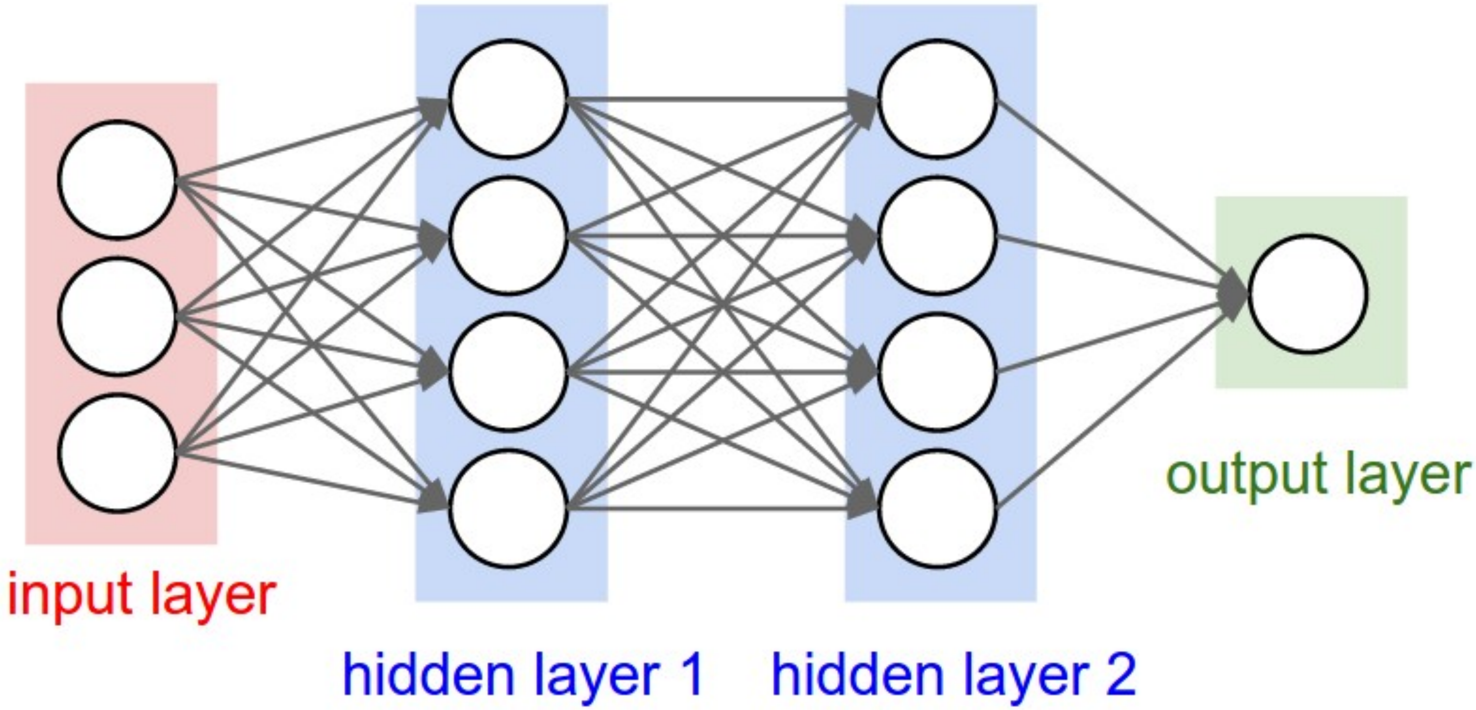
- **Single Layer networks**



- **XOR problem killed research for two decades**

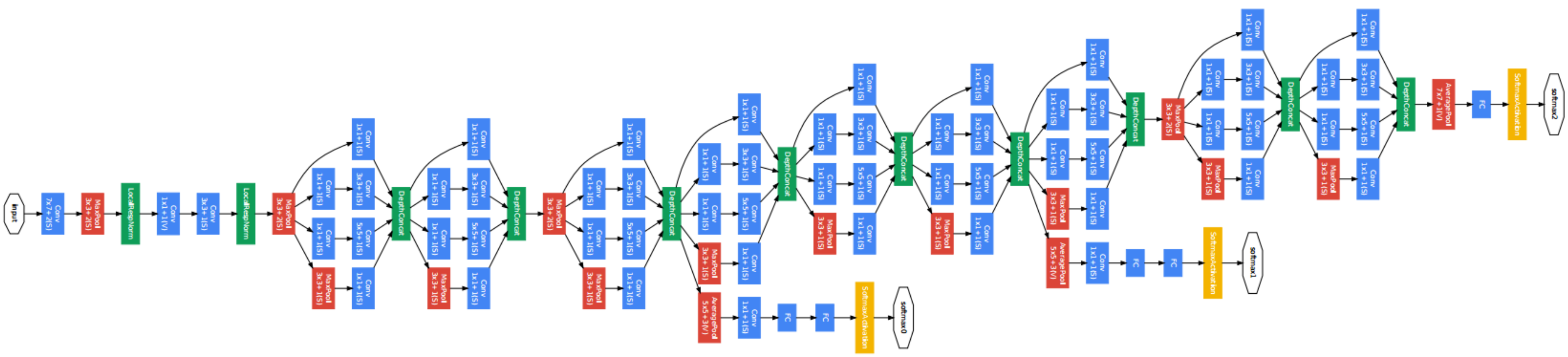
Mid-1980s (2^{nd} Wave)

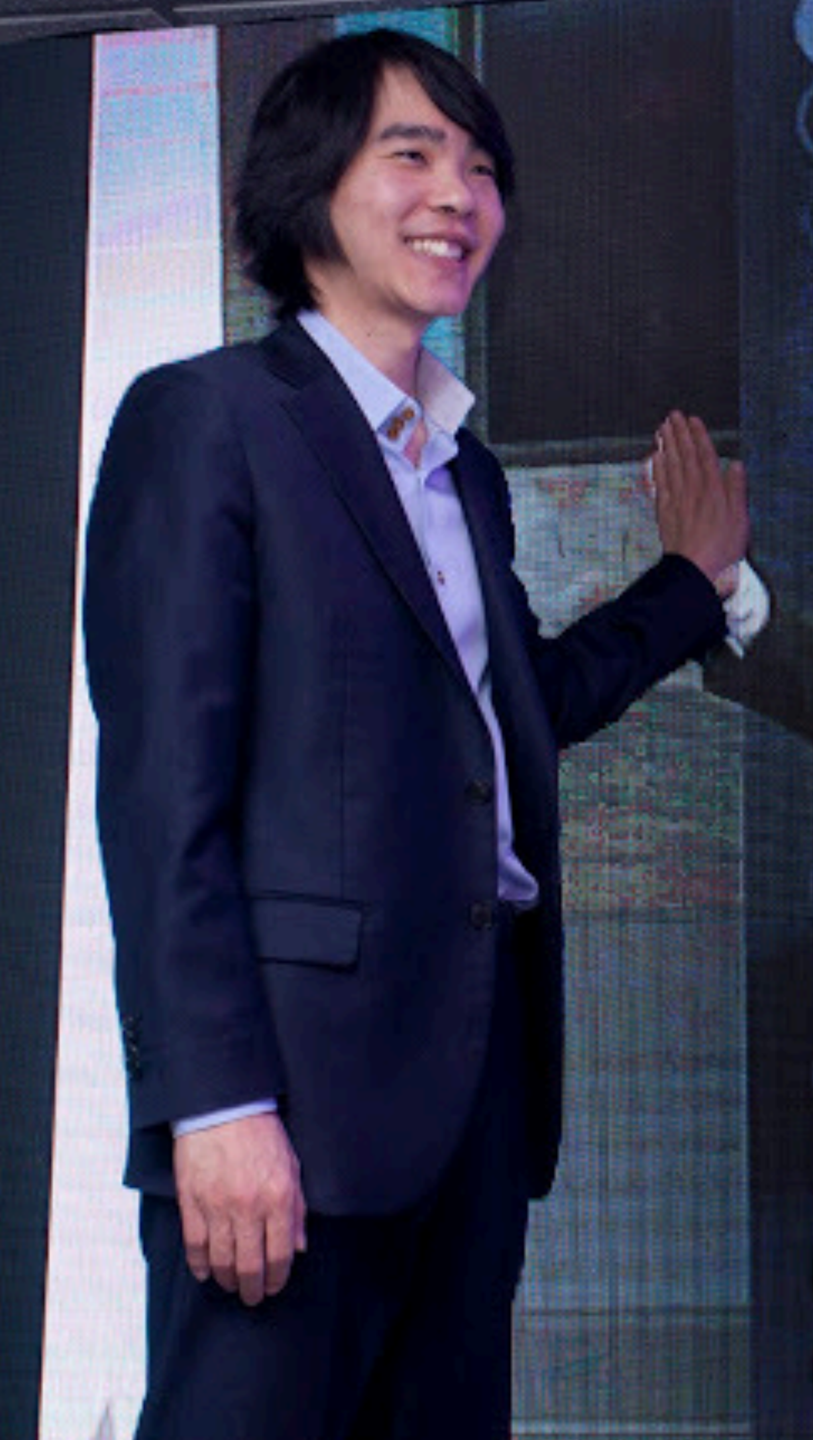
- **Multi-layer networks**
- **Backpropagation algorithm**



2010s (3rd Wave)

- **Big Data**
  - O(M) labeled images
- **Big Compute**
- **'Deep' Learning**



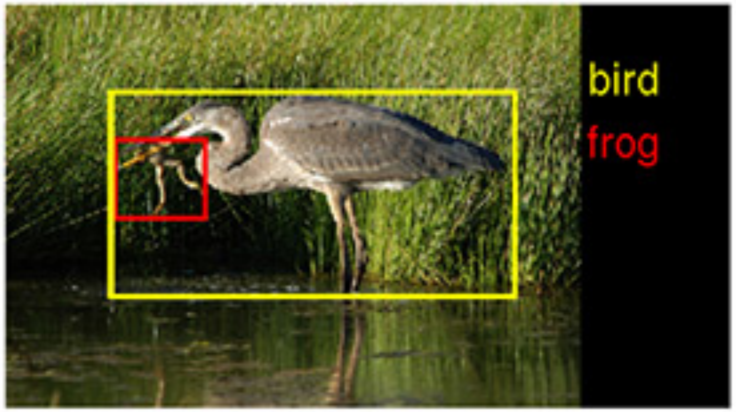


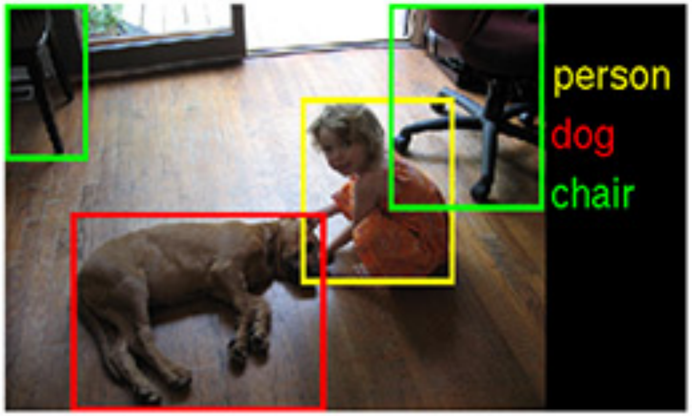

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card 40 years on
PAGE 462
NATUREASIA.COM
28 January 2016
Vol. 529, No. 7587

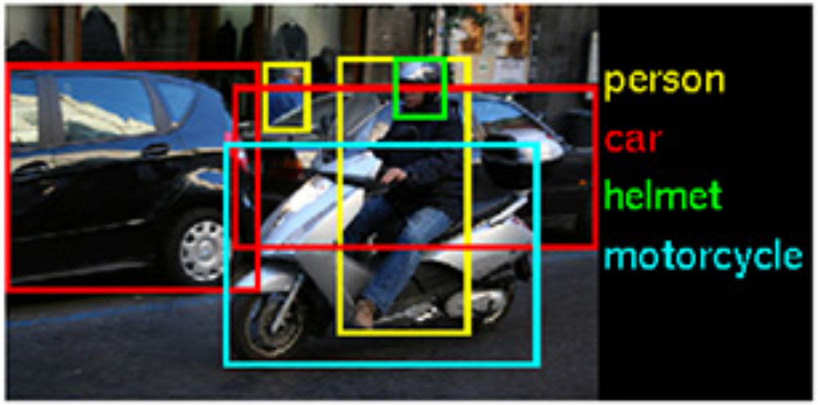
Deep Learning for Self-Driving Cars

Deep Learning for Speech

Deep Learning for Computer Vision

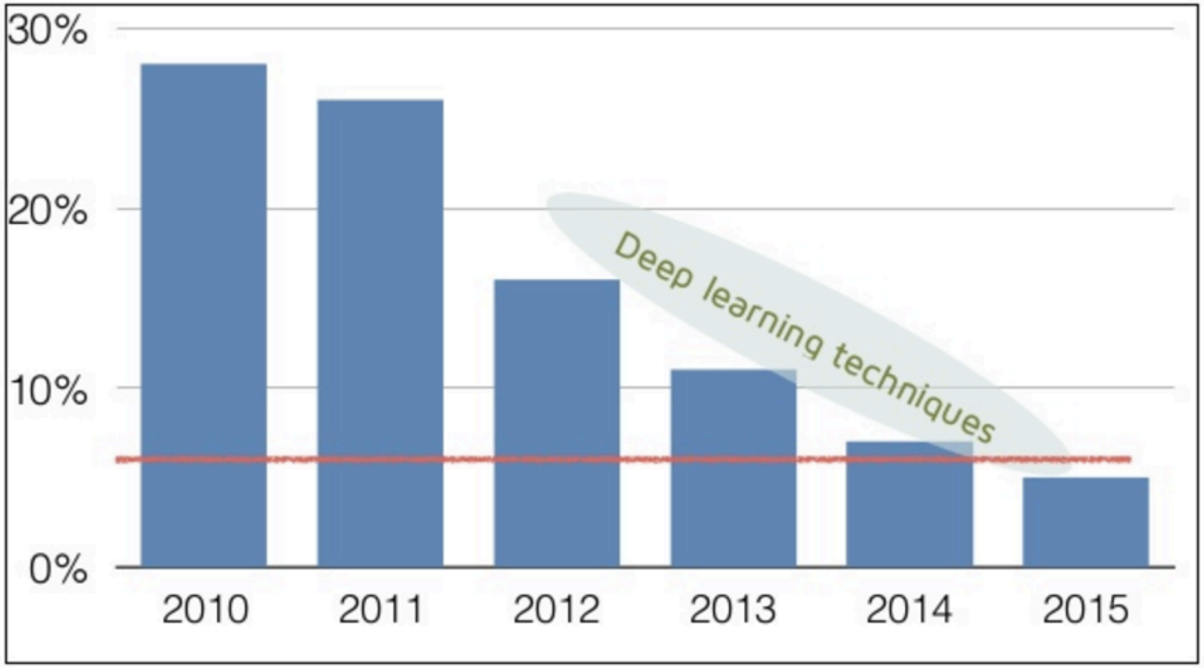






Imagenet ILSVRC Challenge

Error rate[1]

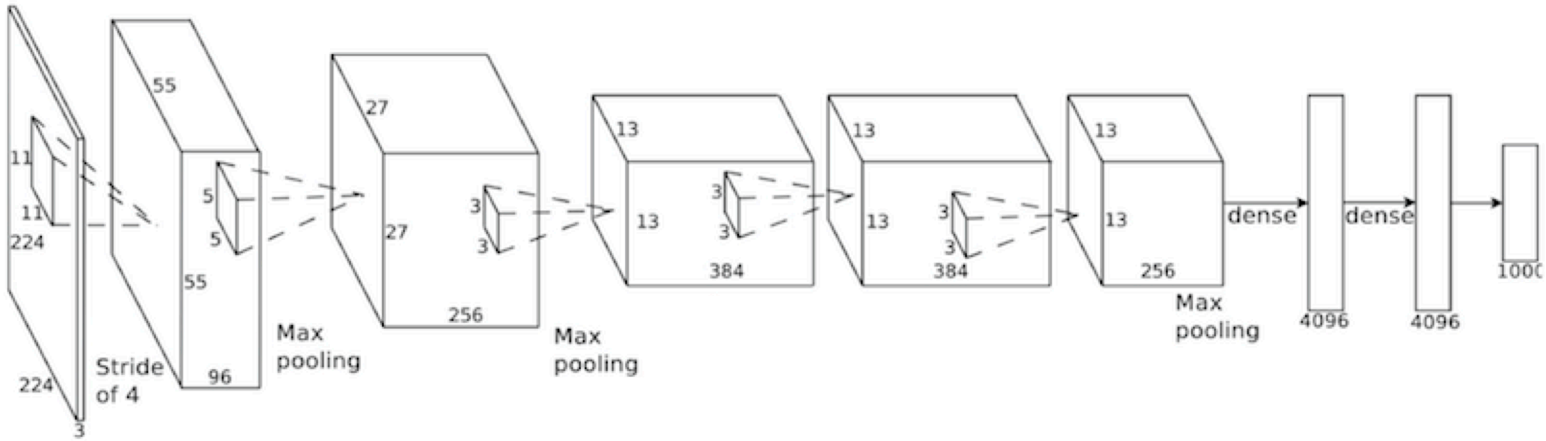


human performance

1: ImageNet top 5 error rate
Source: ImageNet

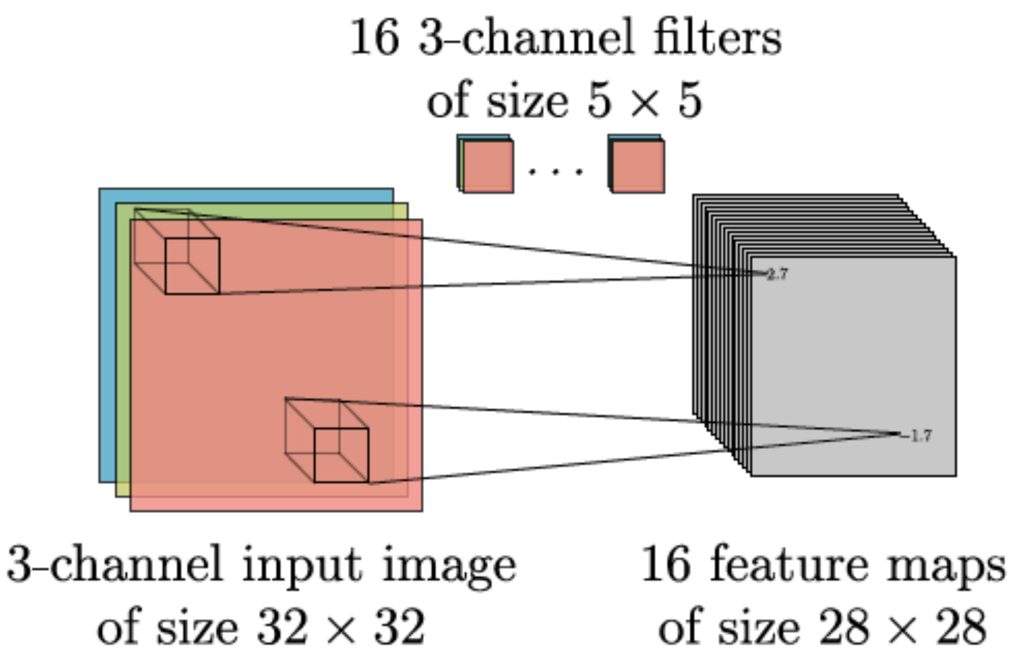
Slide Courtesy of Nervana Systems

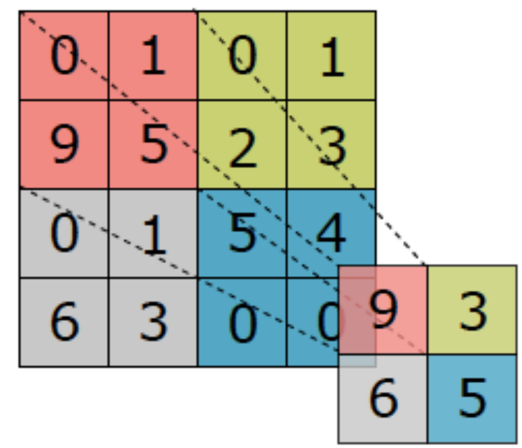
Training Convolutional Networks



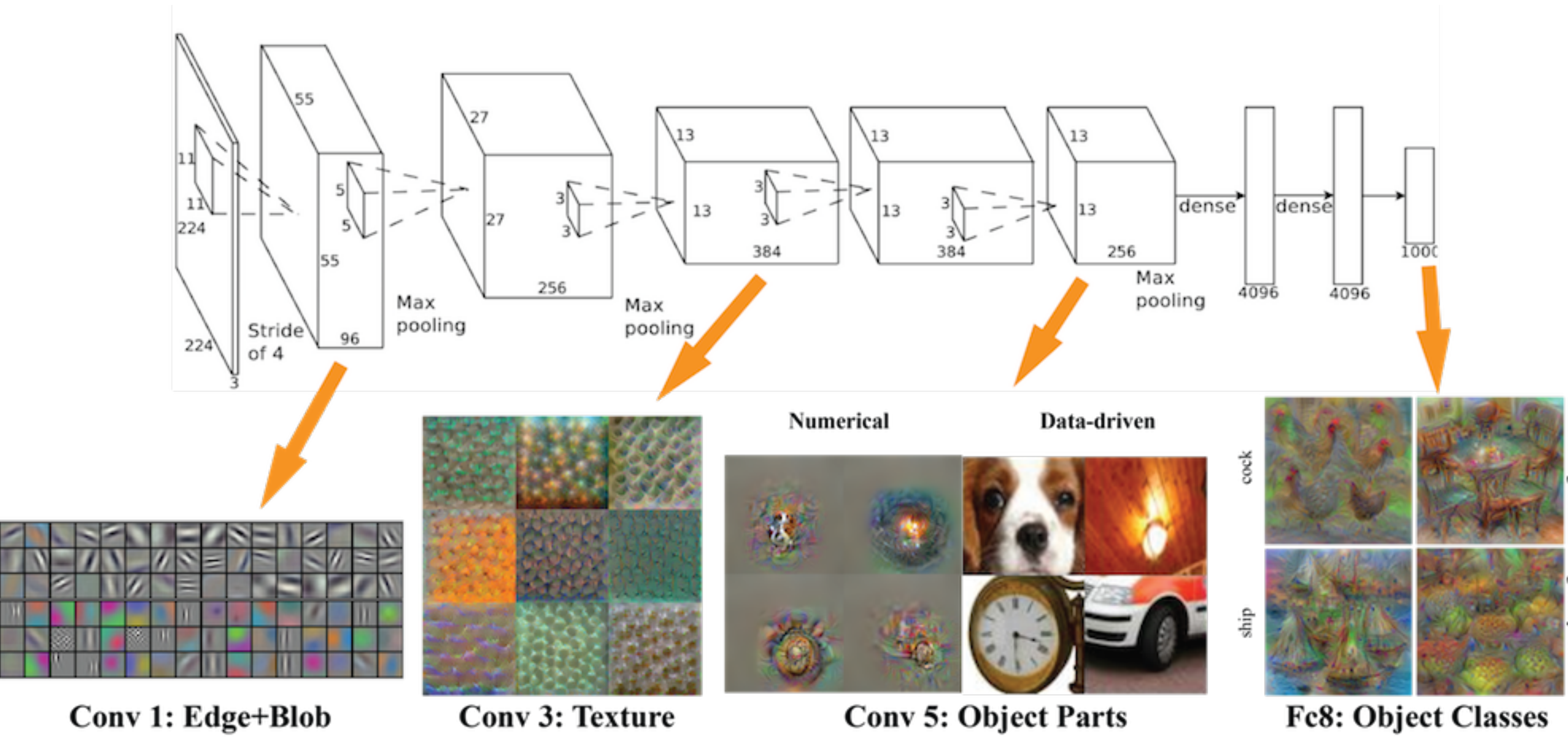
- **Workflow:**
  1. Identify training data (images + labels)
  2. Converge on hyper-parameters (architecture,…)
  3. Random parameter initialization
  4. Forward pass (filter images, make label prediction)
  5. Compute Error
  6. Backward pass (compute gradients, update parameters)

Convolution and Pooling





ImageNet Architecture



Outline

- **Introduction to Deep Learning**
  - History
  - Commercial Applications
  - Convolutional Neural Nets
- **Climate Science**
  - Motivation
  - Representational Challenges
    - Supervised CNNs
    - Semi-Supervised CNNs
  - Scaling Challenges

How will extreme weather change in the future?

How will extreme weather change in the future?

- **Look back in time (Paleoclimate records)**

- **Look forward in time (Climate simulations)**
  - Internal climate system variability
  - External forcings (solar activity, volcanic eruptions)
  - Anthropogenic influence

Climate Simulations

CAM5 hi-resolution simulations (0.25°, prescribed aerosols)

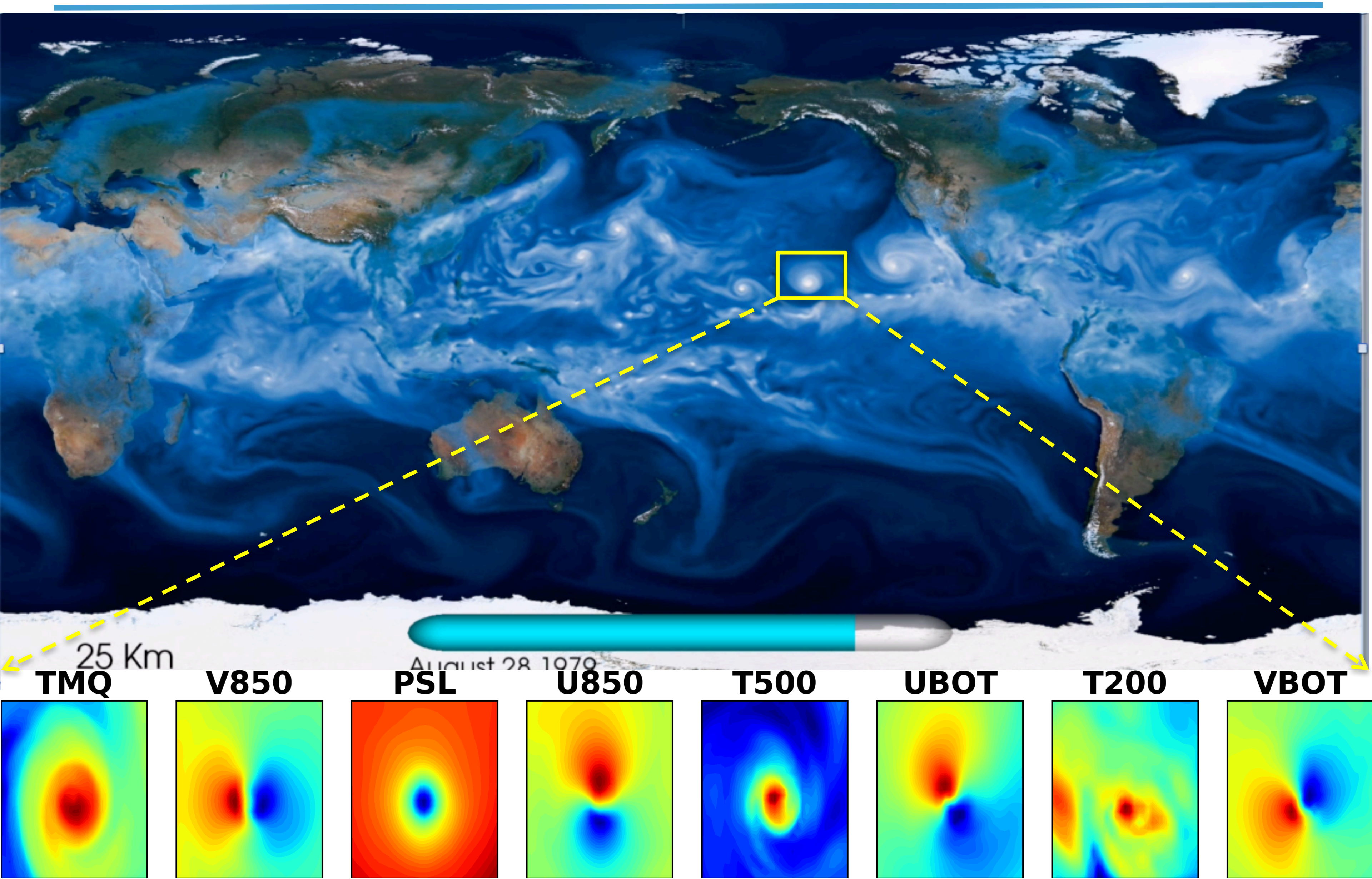
Michael Wehner, Prabhat, Chris Algieri, Fuyu Li, Bill Collins
Lawrence Berkeley National Laboratory

Kevin Reed, University of Michigan

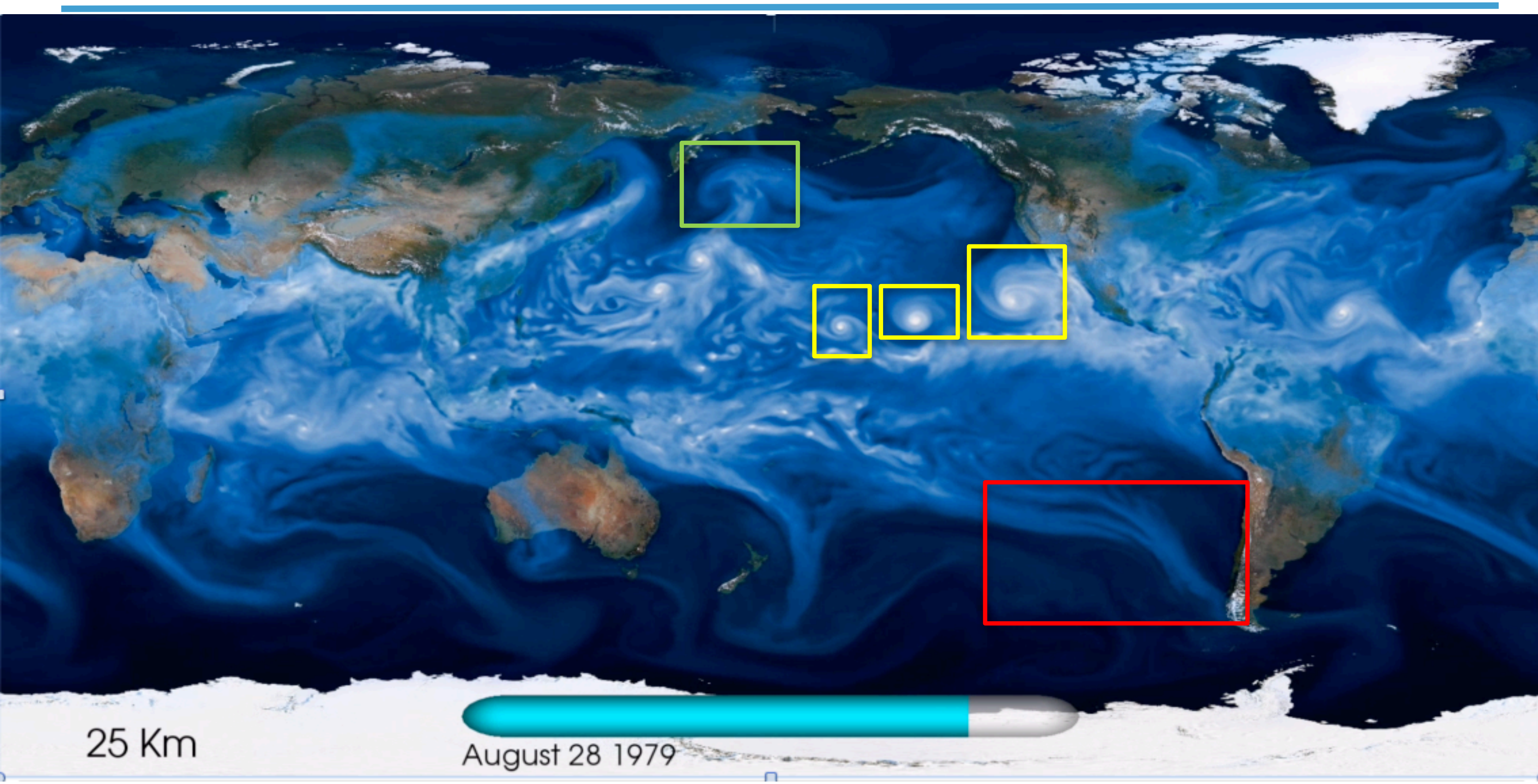
Andrew Gettelman, Julio Bacmeister, Richard Neale
National Center for Atmospheric Research

June 1, 2011

Challenge: Multi-Variate Data



Task: Find Extreme Weather Patterns



Can Deep Learning find Extreme Weather Patterns?

- **Task is analogous to commercial vision applications**
  - Pattern Classification
  - Feature Learning

- **Differences stem from unique attributes of Climate Data**
  - Multi-channel
  - Double precision floating point
  - Statistics are likely different

Outline

- Introduction to Deep Learning
  - History
  - Commercial Applications
  - Convolutional Neural Nets
- **Climate Science**
  - Motivation
  - Representational Challenges
    - Supervised CNNs
    - Semi-Supervised CNNs
  - Scaling Challenges

Supervised Learning

- **Training Input: Cropped, Centered, Multi-variate patches with Labels**
  - Tropical Cyclone (TC)
  - Atmospheric River (AR)
  - Weather Front (WF)

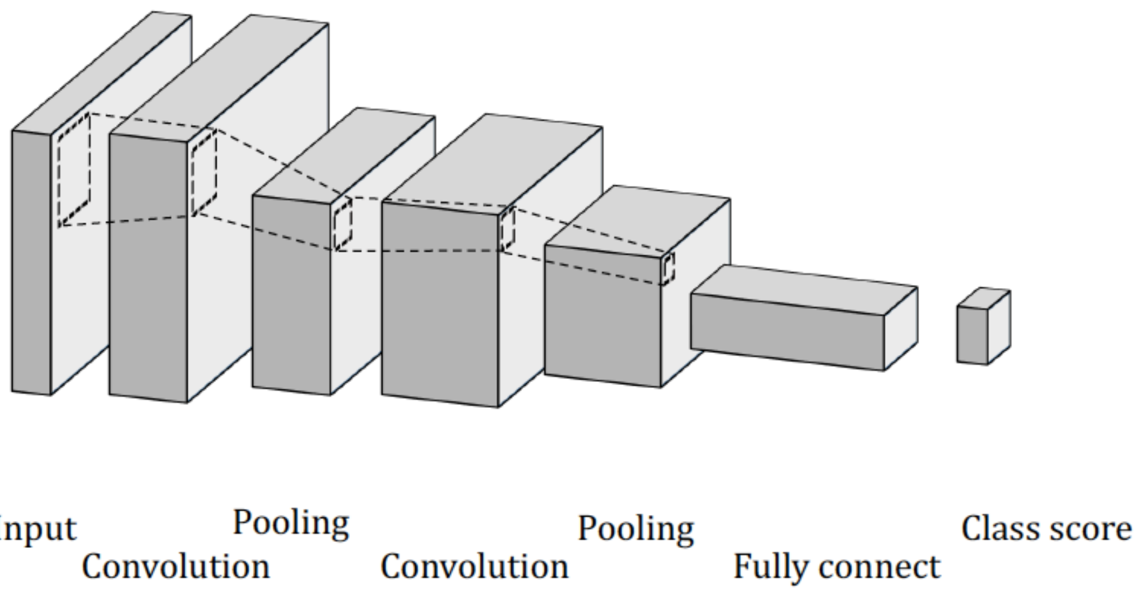
- **Output: Binary (Yes/No) on Test patches**
  - Is there a TC in the patch?
  - Is there an AR in the patch?
  - Is there a WF in the patch?

Training Data

CLASSIFICATION	Image Dimension	Variables	Total Examples (+ve)	(-ve)
Tropical Cyclone	32x32	PSL,UBOT,VBOT,TMQ, U850,V850,T200,T500	10000	10000
Atmospheric Rivers	148x224	TMQ, Land Sea mask	6500	6800
Weather Fronts	27x60	T2m, Precip, PSL	5600	6500

Supervised Convolutional Architecture

CLASSIFICATION	Conv1	Pool1	Conv2	Pool2	Full	Full
Tropical Cyclone	5x5-8	2x2	5x5-16	2x2	50	2
Atmospheric River	12x12-8	3x3	12x12-16	2x2	200	2
Weather Fronts	5x5-16	2x2	5x5-16	2x2	400	2



Supervised Classification Accuracy

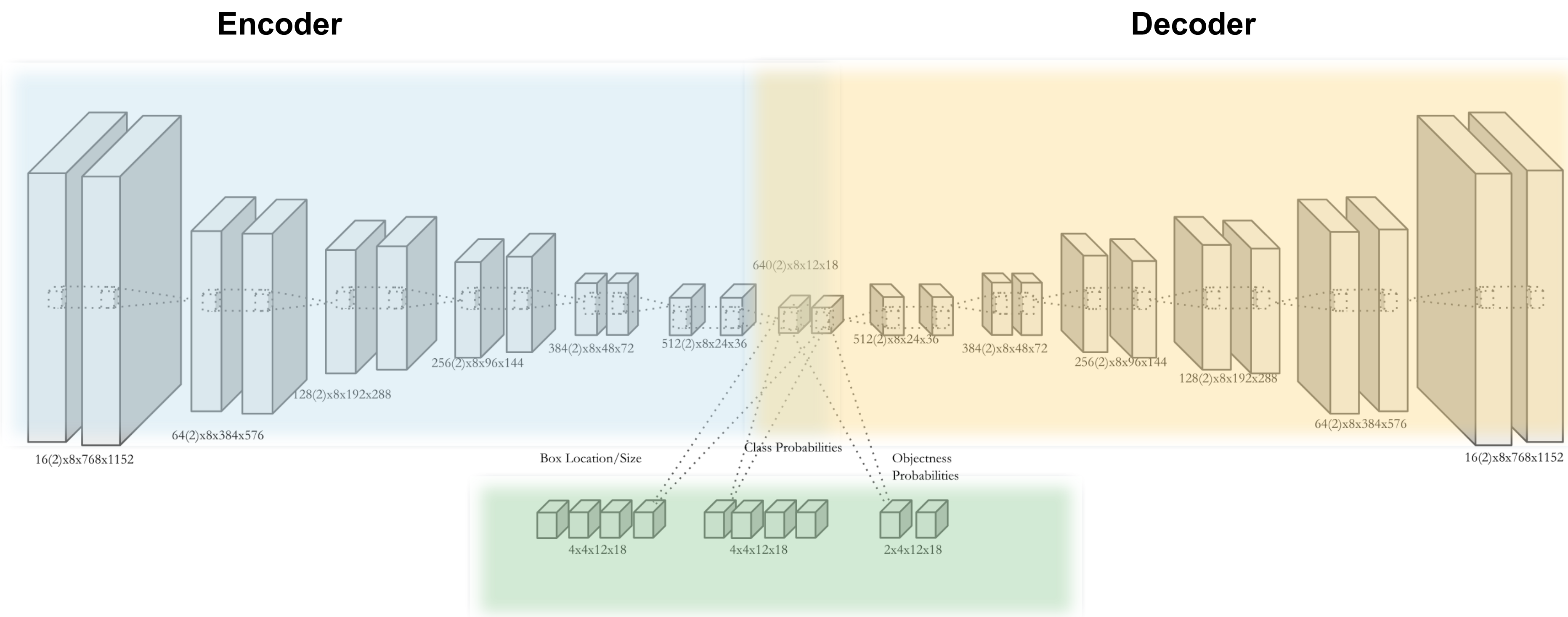
	Logistic Regression		K-Nearest Neighbor		Support Vector Machine		Random Forest		ConvNet	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Tropical Cyclone	96.8	95.85	98.1	97.85	97.0	95.85	99.2	**99.4**	99.3	**99.1**
Atmospheric Rivers	81.97	82.65	79.7	81.7	81.6	83.0	87.9	88.4	90.5	**90.0**
Weather Fronts	84.9	89.8	72.46	76.45	84.35	**90.2**	80.97	87.5	88.7	**89.4**

Hyper-parameter optimization applied with Spearmint for all methods

Semi-Supervised Learning

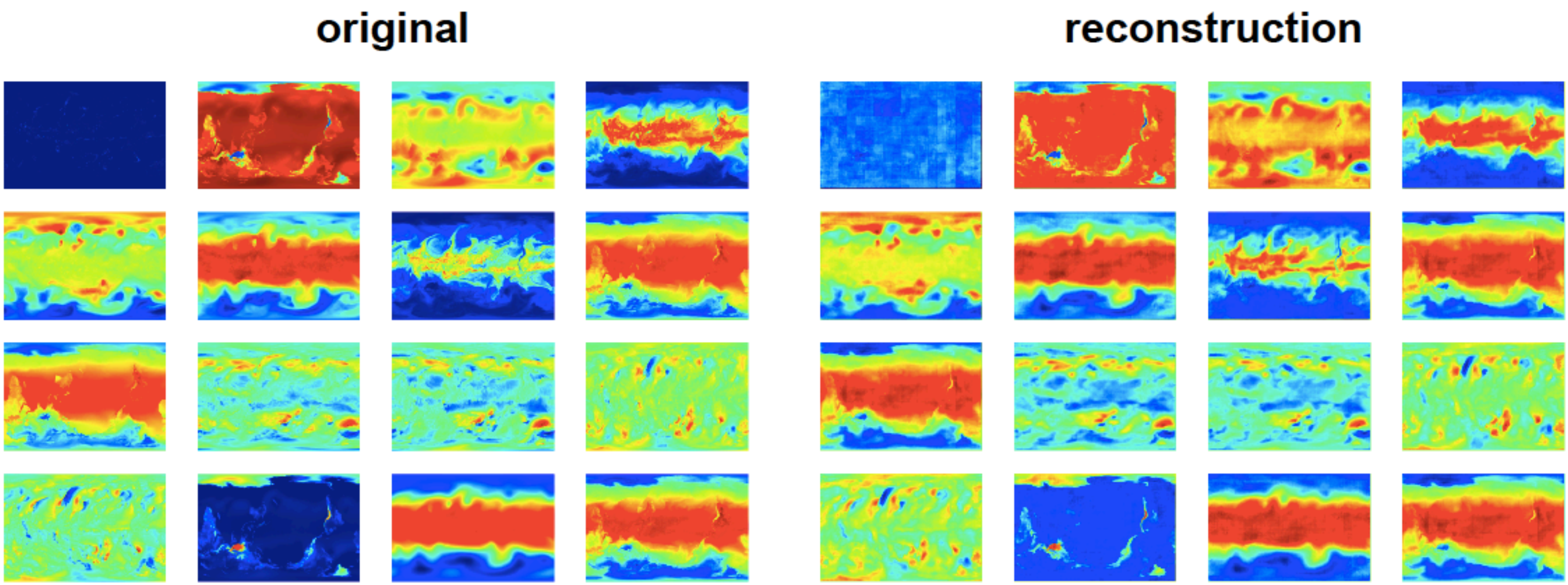
- **Objectives:**
  - Create unified architecture for all weather patterns
  - Predict bounding box location for weather pattern
  - Discover new patterns
    - Might have few/no labels for several weather patterns

Semi-Supervised Convolutional Architecture

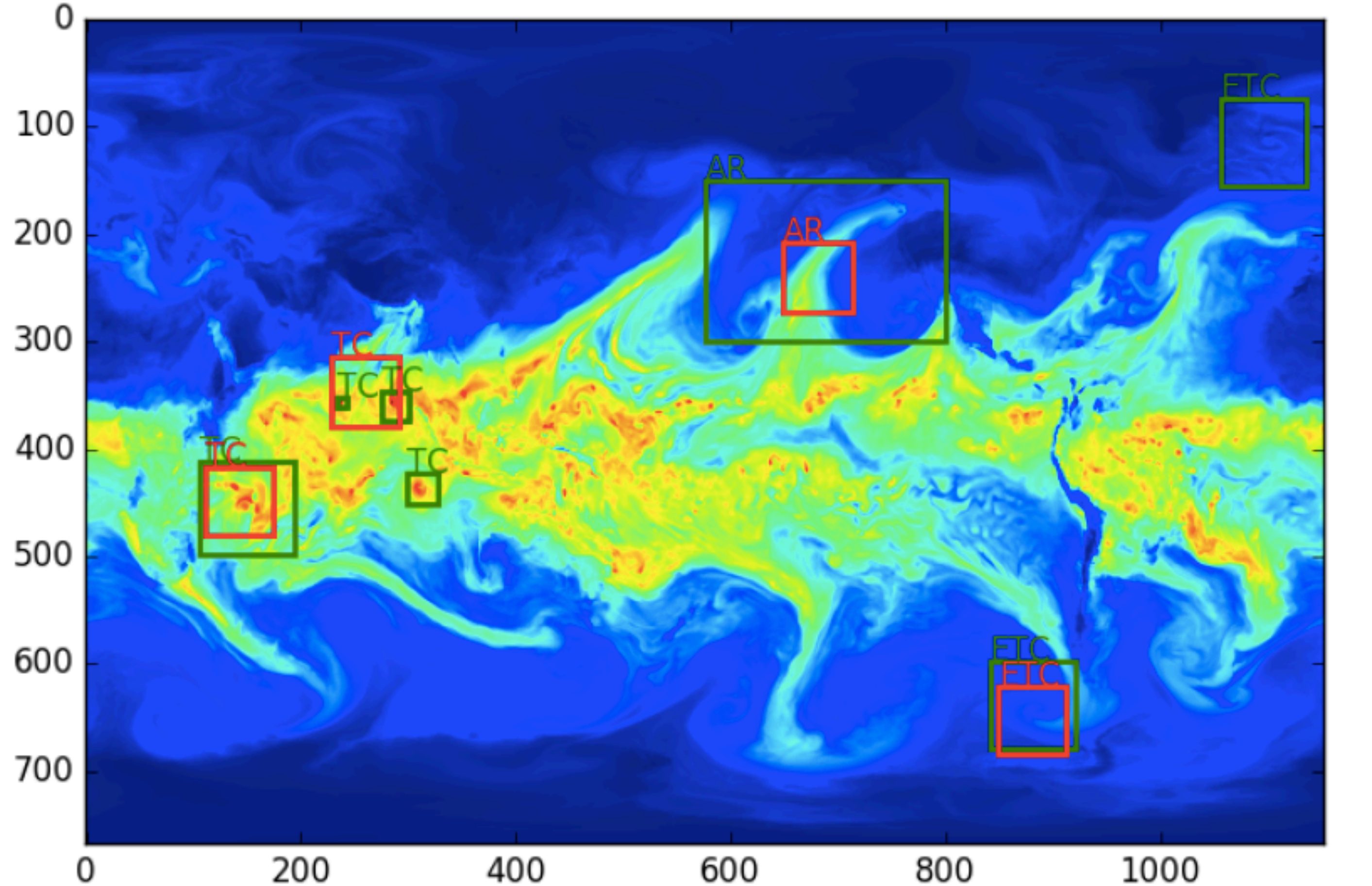


Classification + Bounding Box Regression

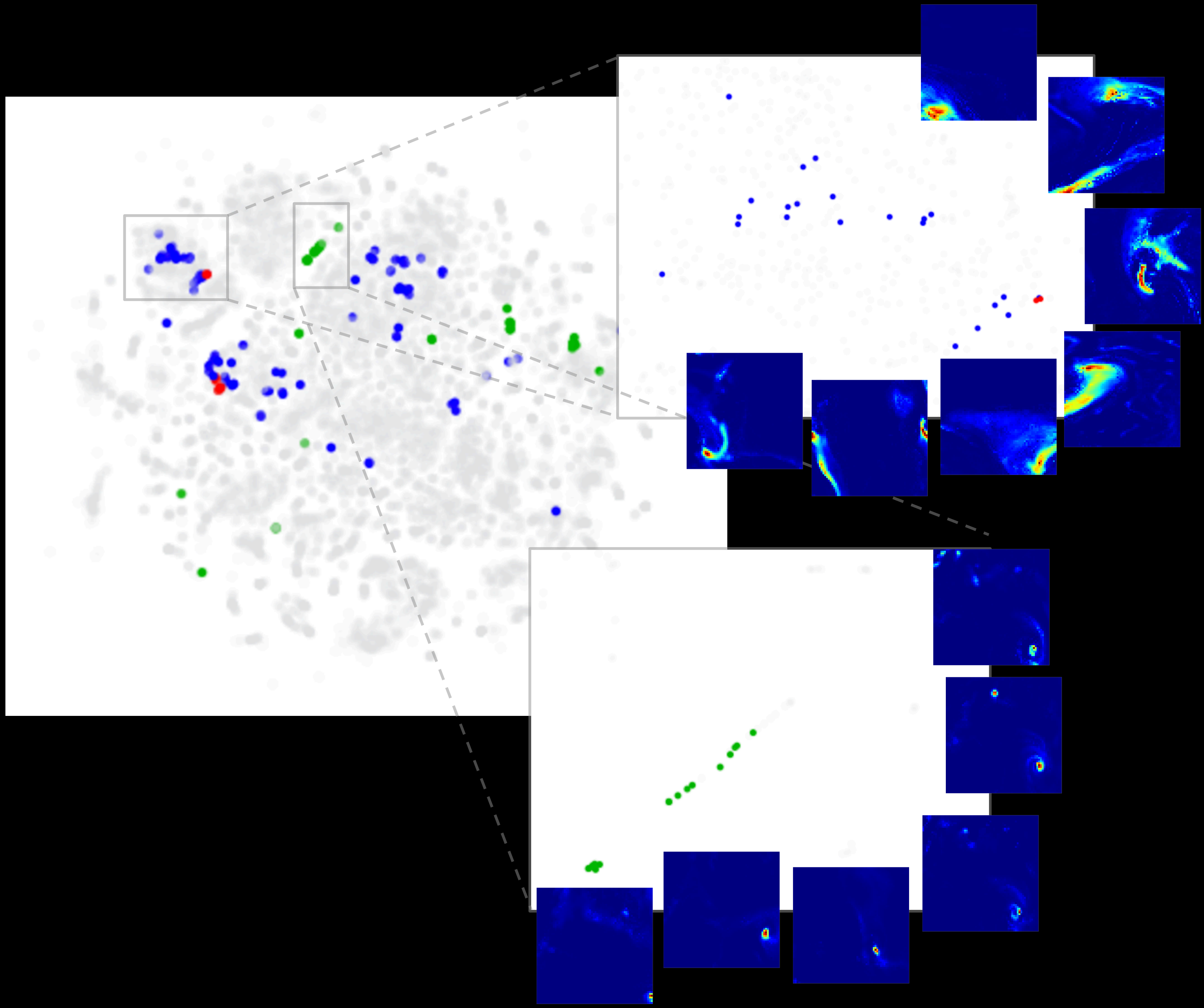
Reconstruction Results



Classification + Regression Results



Ground Truth
Prediction



Outline

- Introduction to Deep Learning
  - History
  - Commercial Applications
  - Convolutional Neural Nets
- Climate Science
  - Motivation
  - Representational Challenges
    - Supervised CNNs
    - Semi-Supervised CNNs
  - Scaling Challenges

Need for Speed

- **Performance and Scaling**
  - Current networks take days to train on O(10) GB datasets
  - We have O(10) TB datasets on hand
- **Quick turnaround is critical for hyper-parameter tuning**
  - # layers, # filters, filter size, stride
  - Pooling operation
  - Learning rates, Learning schedule
  - Optimizers (ADAM, SGD,…)

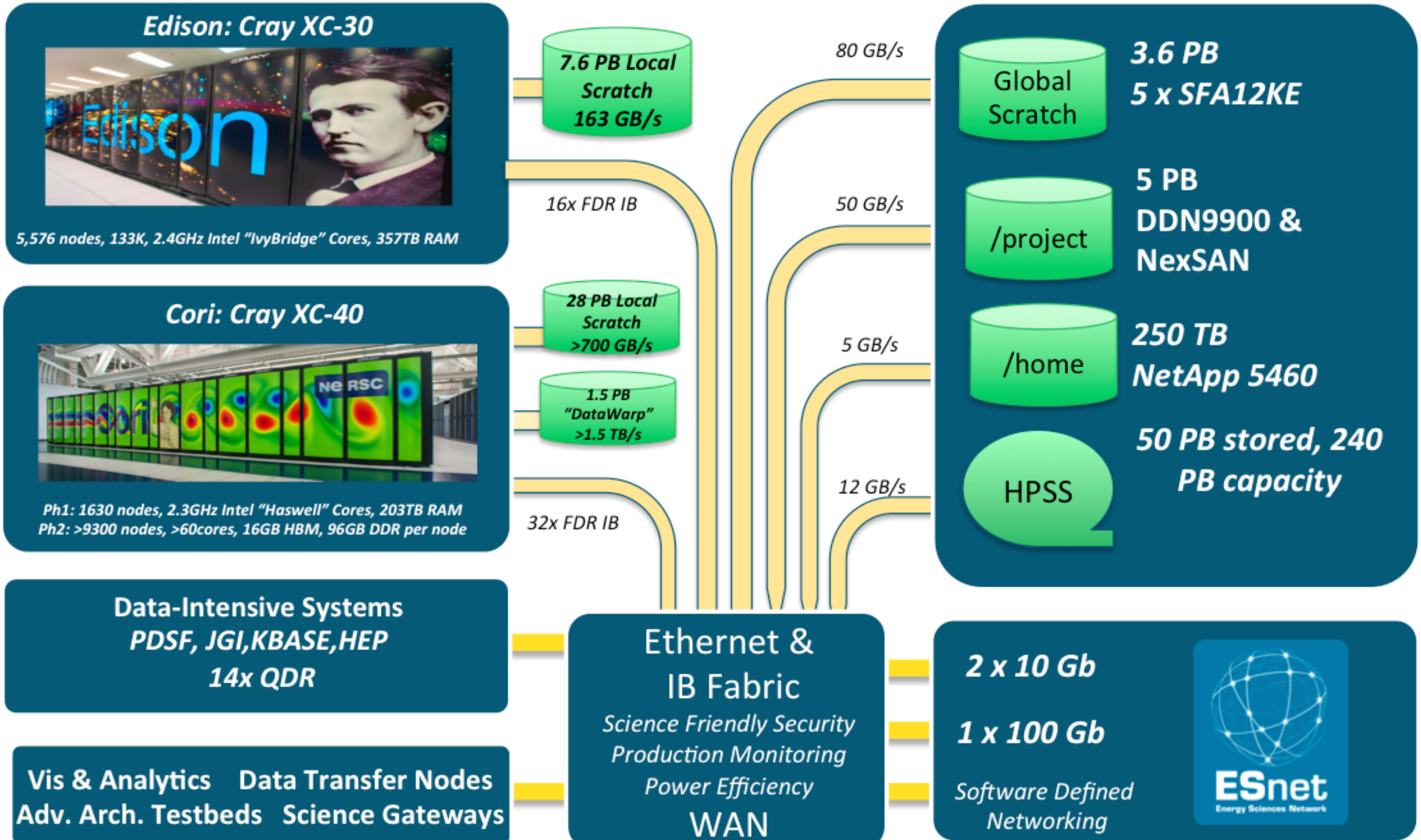
Deep Learning Stack

	Technologies
Deep Learning Frameworks	K, Lasagne, TensorFlow, Caffe, theano, Torch, Neon, CNTK, MXNet, …
Multi Node libraries	MLSL, MPI, GRPC
Single Node libraries	MKL, CuDNN
Hardware	KNL, KNM, GPUs, FPGAs

Deep Learning Stack

	Technologies
Deep Learning Frameworks	K, Lasagne, TensorFlow, Caffe, theano, Torch, Neon, CNTK, MXNet, …
Multi Node libraries	MLSL, MPI, GRPC
Single Node libraries	MKL, CuDNN
Hardware	KNL, KNM, GPUs, FPGAs

Hardware


Edison: Cray XC-30
5,576 nodes, 133K, 2.4GHz Intel "IvyBridge" Cores, 357TB RAM
Cori: Cray XC-40
Ph1: 1630 nodes, 2.3GHz Intel "Haswell" Cores, 203TB RAM
Ph2: >9300 nodes, >60cores, 16GB HBM, 96GB DDR per node
Data-Intensive Systems
PDSF, JGI,KBASE,HEP
14x QDR
Vis & Analytics
Data Transfer Nodes
Adv. Arch. Testbeds
Science Gateways
7.6 PB Local Scratch 163 GB/s
16x FDR IB
28 PB Local Scratch >700 GB/s
1.5 PB "DataWarp" >1.5 TB/s
32x FDR IB
80 GB/s
50 GB/s
5 GB/s
12 GB/s
Global Scratch
3.6 PB
5 x SFA12KE
/project
5 PB
DDN9900 &
NexSAN
/home
250 TB
NetApp 5460
HPSS
50 PB stored, 240 PB capacity
Ethernet &
IB Fabric
Science Friendly Security
Production Monitoring
Power Efficiency
WAN
2 x 10 Gb
1 x 100 Gb
Software Defined Networking
ESnet
Energy Sciences Network

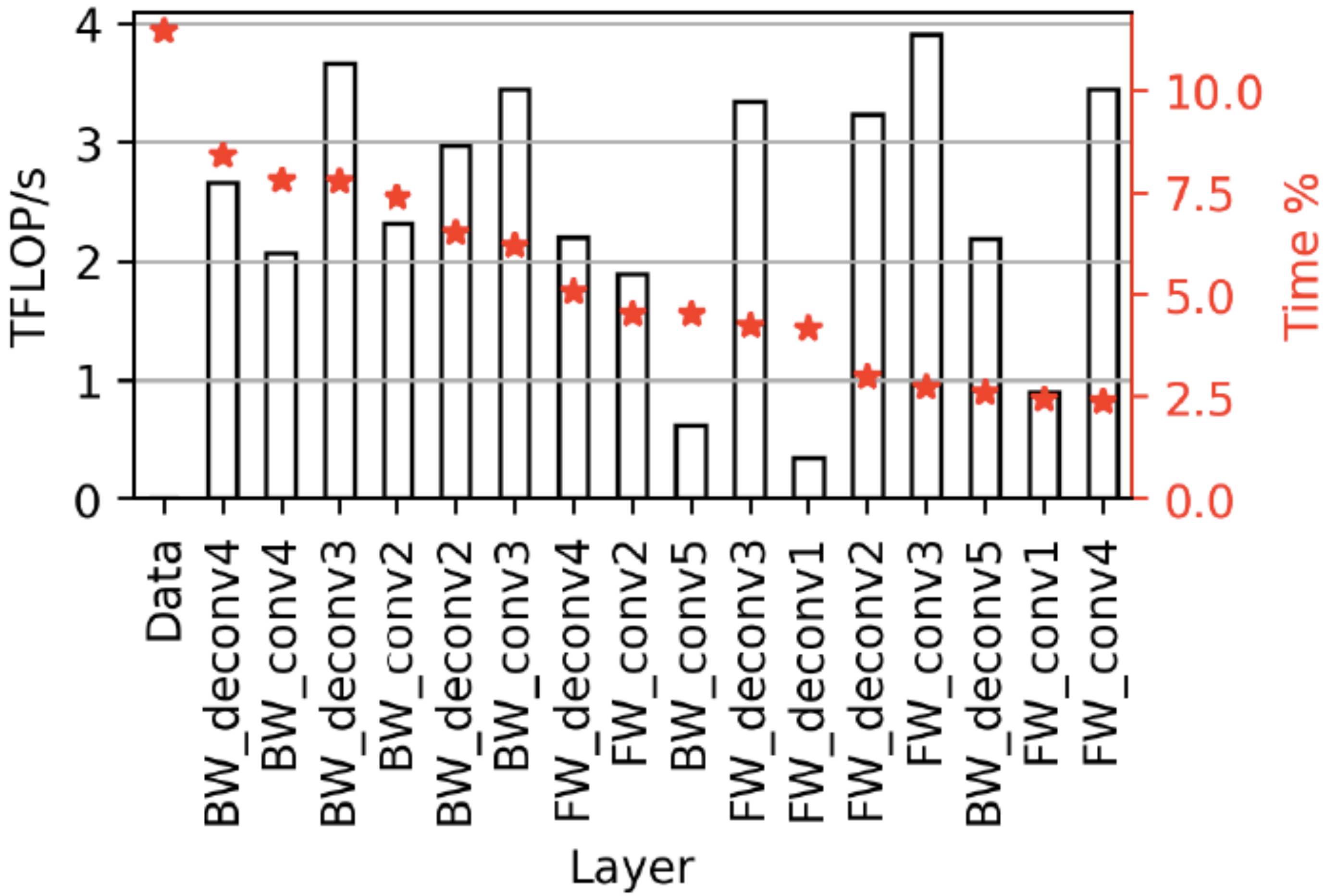
Input Data + Architecture

- **Input Data:**
  - 768x768x16
  - 15 TB
  - 400K images
- **9 convolution, 5 deconvolution layers**
  - 300MB parameters

Single Node Optimizations

- **Target hardware: Intel Xeon Phi (Knights Landing)**
- **Intel Caffe with MKL 2017 library**
  - Optimized DL primitives for KNL
  - Added support for de-convolutions

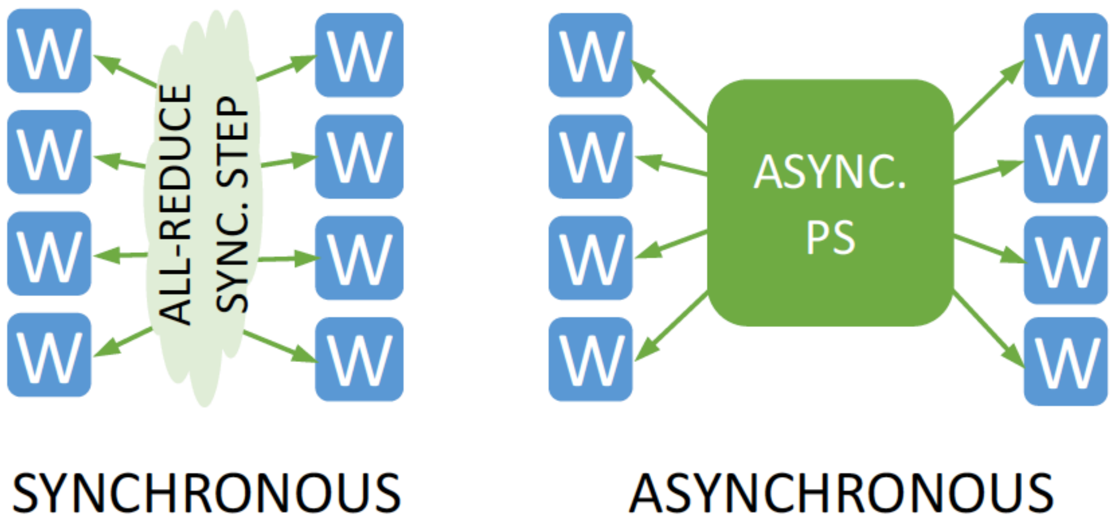
Single Node Performance



Multi-Node strategy

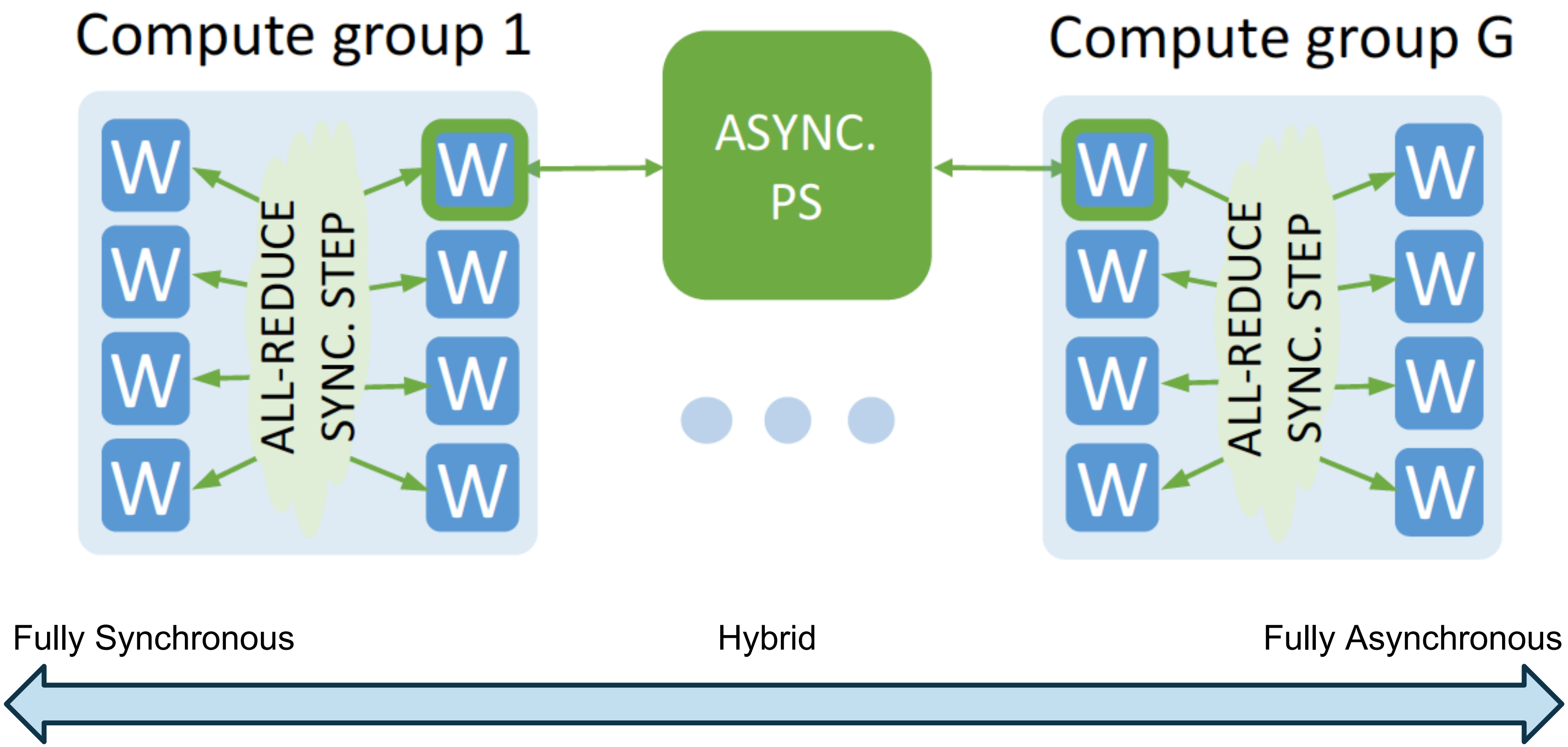
- **Data Parallelism (vs. Model Parallelism)**
- **Optimizations:**
  - Hybrid parameter updates
  - Topology aware placement
  - Dedicated Parameter server per-layer
- **Implementation uses Intel MLSL**
  - Proxy threads/processes drive communication
  - Improvement over vanilla MPI in terms of bandwidth utilization

Multi-Node Scaling Strategy

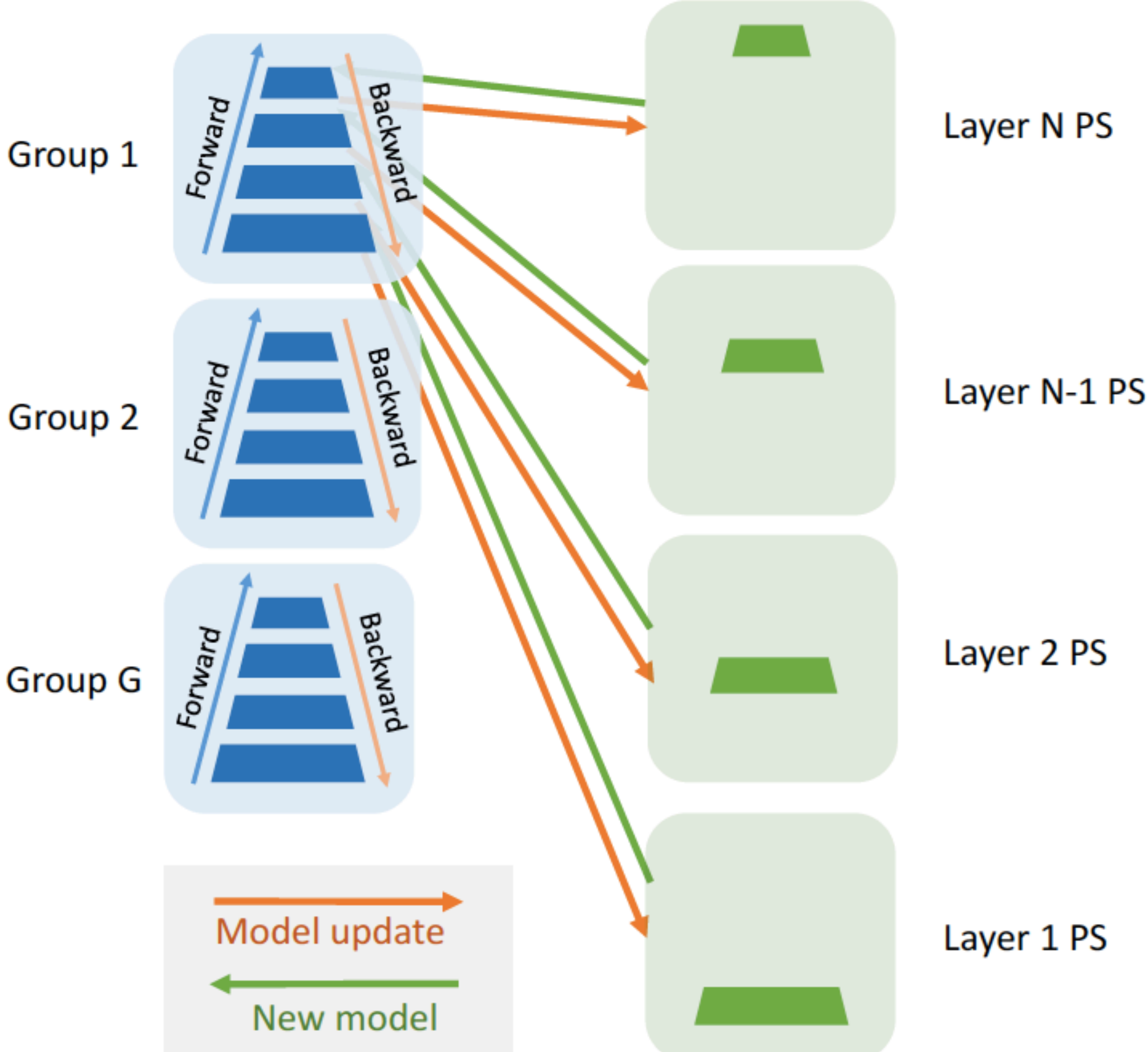


	SYNCHRONOUS	ASYNCHRONOUS
+ve	• **Same # iterations to converge as serial implementation**	• **Faster Iterations** • **Robustness to node failures** • **Better control of batch size**
-ve	• Straggler effect • Susceptible to node failure • Batch size grows with # nodes	• More #iterations to converge as serial implementation

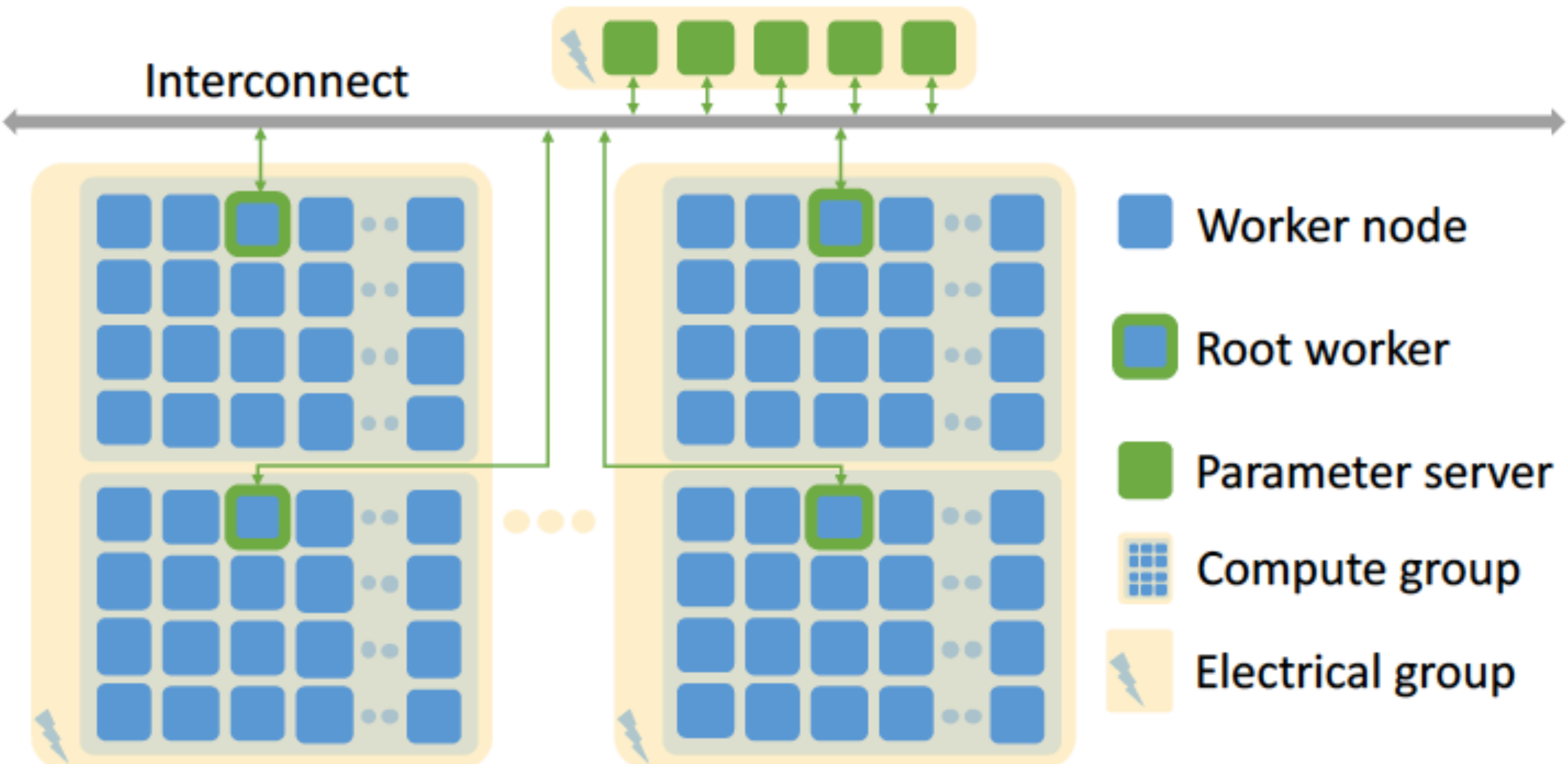
Hybrid Synchronization



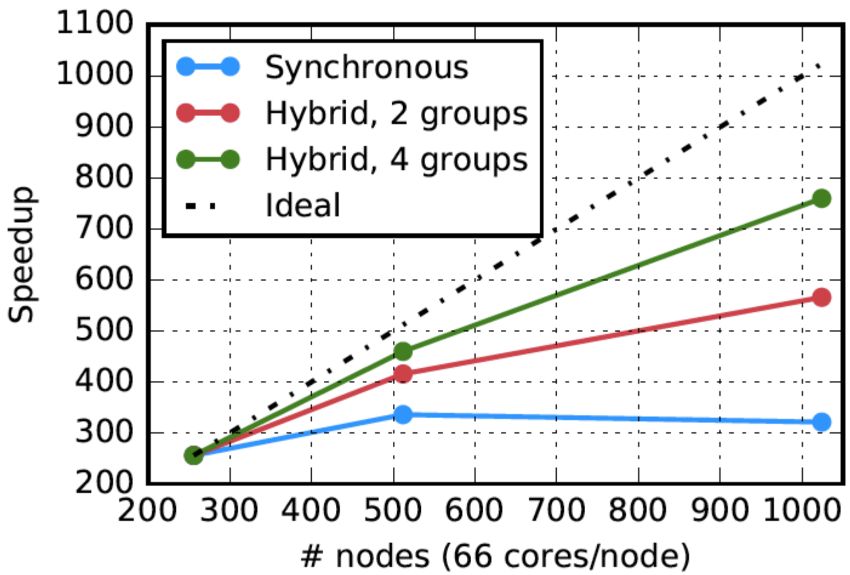
Changing # compute groups controls level of asynchrony


Group 1
Forward
Backward
Group 2
Forward
Backward
Group G
Forward
Backward
Model update
New model
Layer N PS
Layer N-1 PS
Layer 2 PS
Layer 1 PS

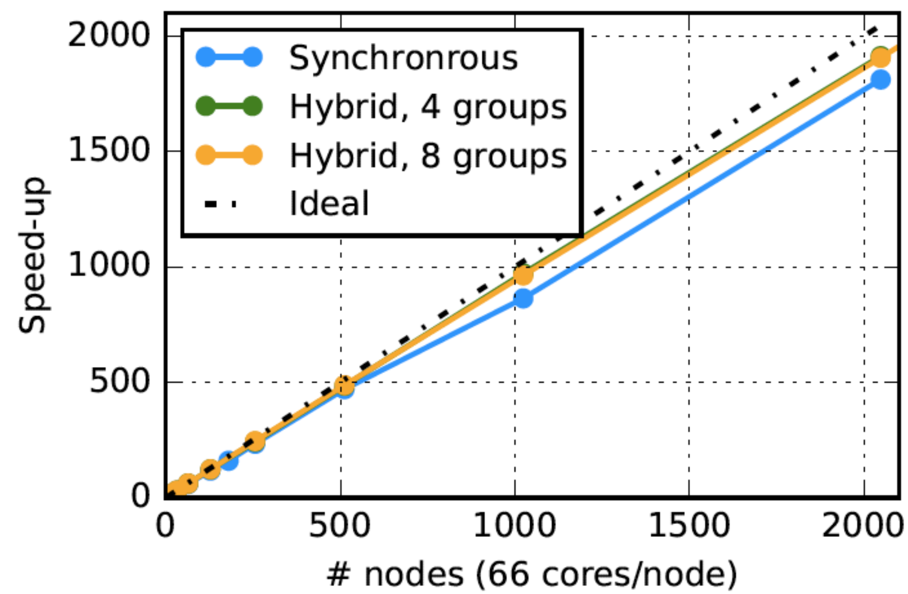
Topology Aware Placement



Multi-Node Scaling Results

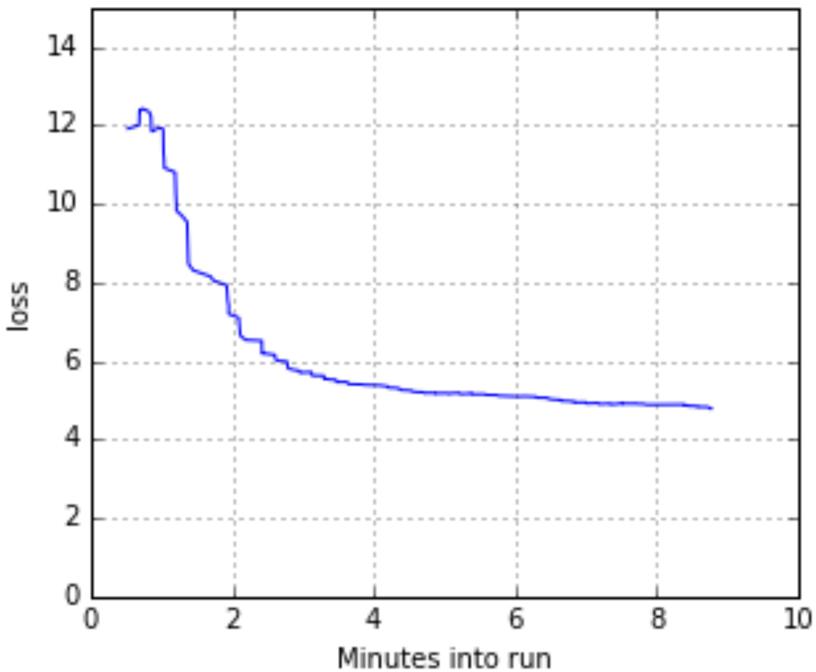


Strong Scaling
Overall Batch size fixed

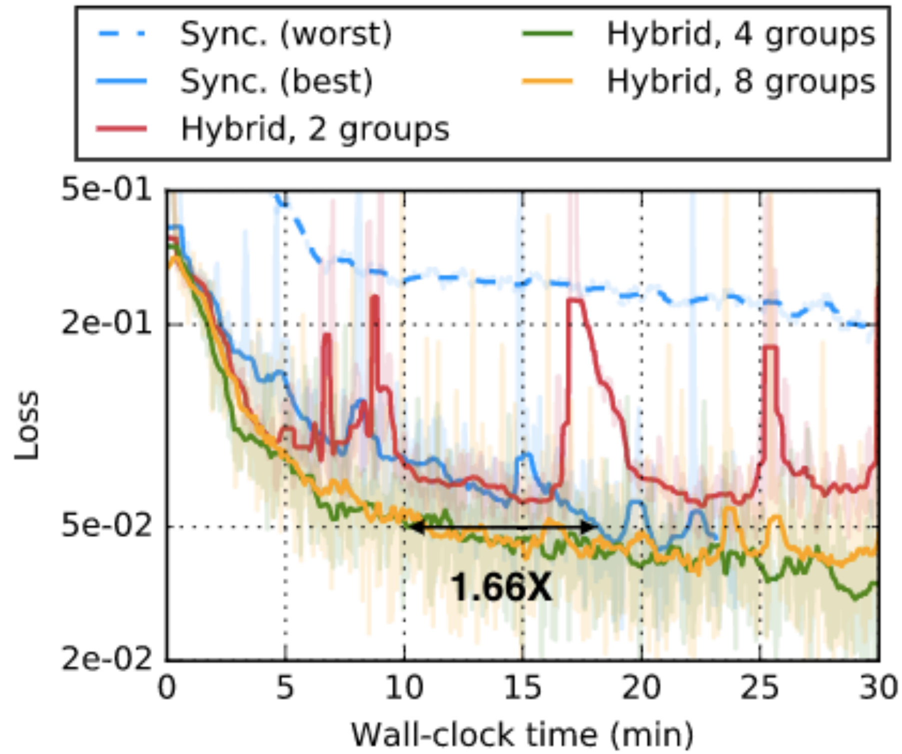


Weak Scaling
Overall batch size increases with #nodes
Batch size/node fixed

Statistical Convergence..



Climate



HEP

Overall Performance

- **Single KNL node (66 cores)**
  - 1-4 TFLOP/s
- **9600 KNL nodes (633,600 cores)**
  - 15 PFLOP/s peak
  - 13 PFLOP/s sustained
  - 12.6 seconds/iteration; 7200x speedup over single node runtime

Conclusions

- **Genuine excitement in the field of AI and DL**
  - Commercial applications spanning vision, speech and control
- **Deep Learning is viable tool for find extreme weather patterns**
  - Helps in characterizing changes in the future
- **Representational Challenges:**
  - Supervised architectures can match hand-tuned criteria
  - Semi-supervised architectures can potentially discover new patterns
- **Computational Challenges:**
  - Single node performance on KNL: 1-4 TF
  - Multi-node scaling on 9600 nodes: 15 PF

Acknowledgements

- **Intel Research**: Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Amir Khosrowshahi, Pradeep Dubey
- **Stanford**: Ioannis Mitliagkas, Jian Yang, Chris Re (**Stanford**)
- **U. Montreal**: Chris Beckham, Tegan Maharaj, Chris Pal
- **Berkeley Lab**: Thorsten Kurth, Evan Racah, Wahid Bhimji, Yunjie Liu, Michael Wehner, William Collins

Questions?

prabhat@lbl.gov