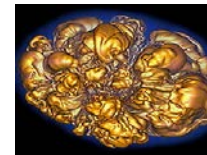
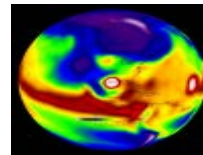
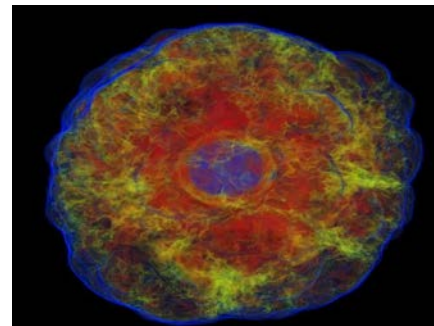
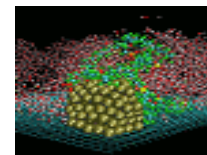
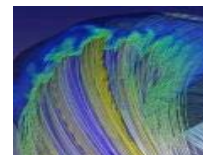
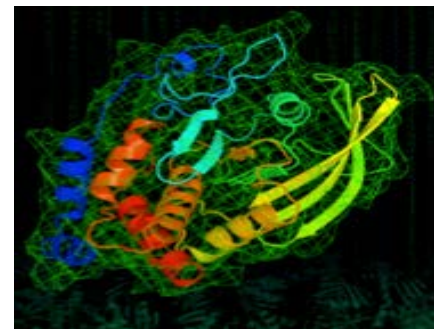
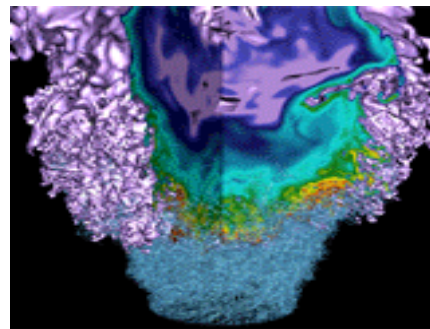


# Machine Learning At Scale

IXPUG BoF SC18  
Nov 15 2018



Wahid Bhimji  
Berkeley Lab

## Mission HPC center for DoE Science:

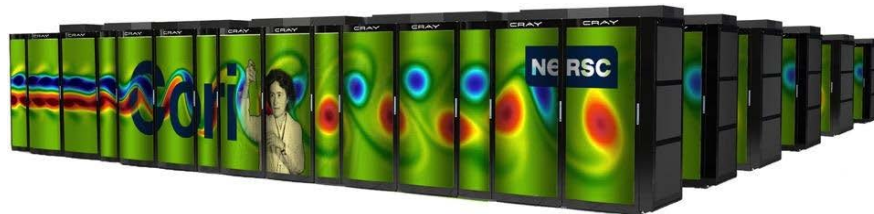
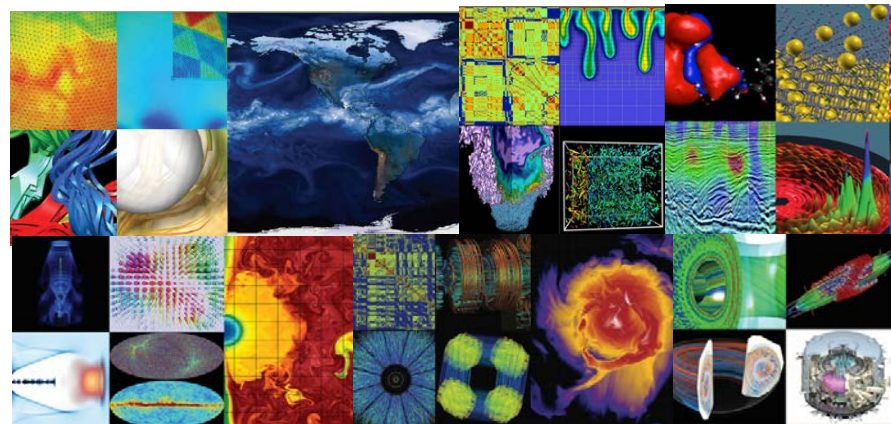
>7000 users; 100s of projects

## Cori: 31.4 PF Peak

- 2388 Haswell
- 9668 KNL XeonPhi
- 28 PB Lustre FS: 700 GB/s peak
- 1.8 PB Flash Burst Buffer: 1.7 TB/s

## NERSC Machine/Deep Learning:

- ML training and tools
- Optimize for hardware and scale
- Collaborative Projects (e.g. with Intel):  
advance methods and applications



# Deep learning can accelerate science

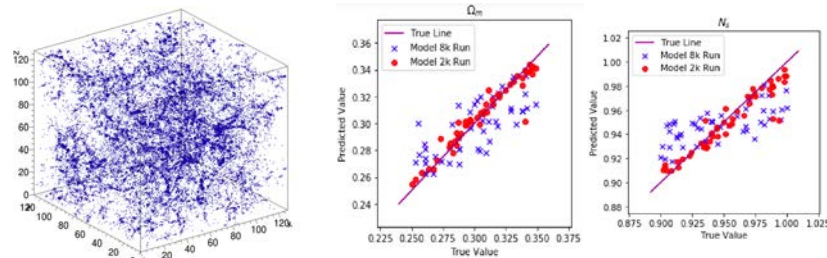


Exploit high-dimensional ‘raw’ data for e.g.:

- **Classification** and segmentation of phenomena
- **Regression** of fundamental parameters
- **Unsupervised clustering** of underlying structure
- **Generation** to augment science simulations

Mathuriya, Bard et. al <http://arxiv.org/abs/1808.04728>

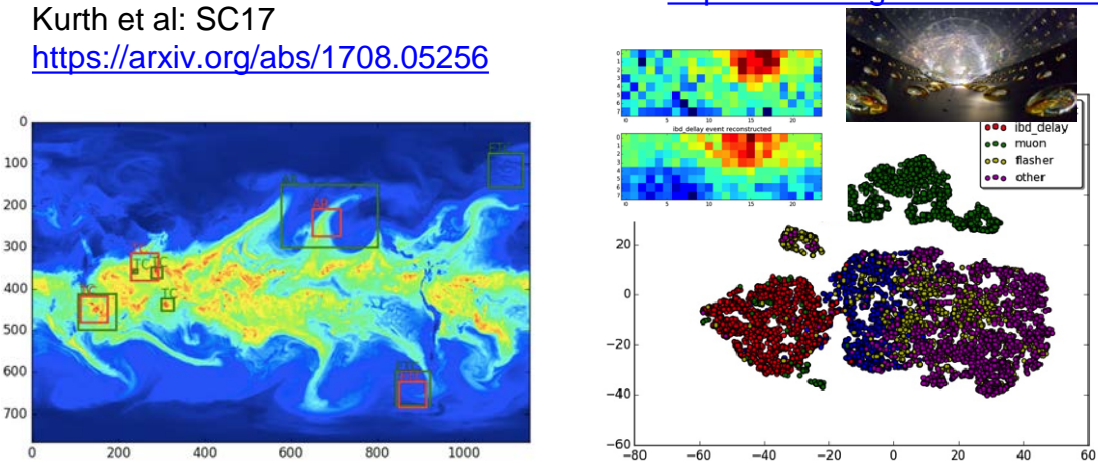
CosmoFlow SC18 – Thu 2pm



Kurth et al: SC17

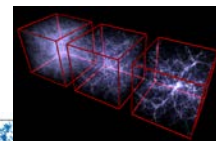
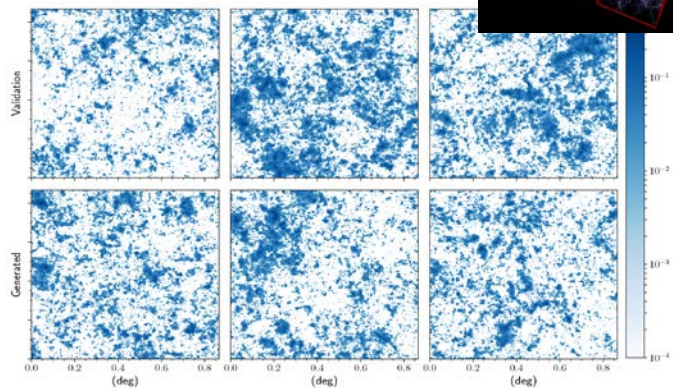
<https://arxiv.org/abs/1708.05256>

Racah et. al. <https://arxiv.org/abs/1601.07621>



CosmoGAN Mustafa Mustafa et. al

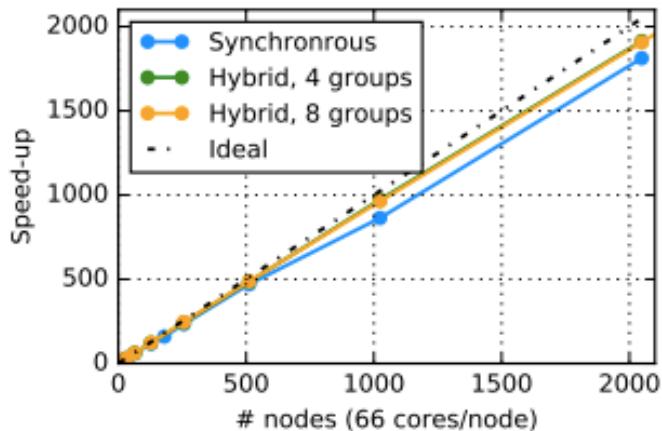
<https://arxiv.org/abs/1706.02390>



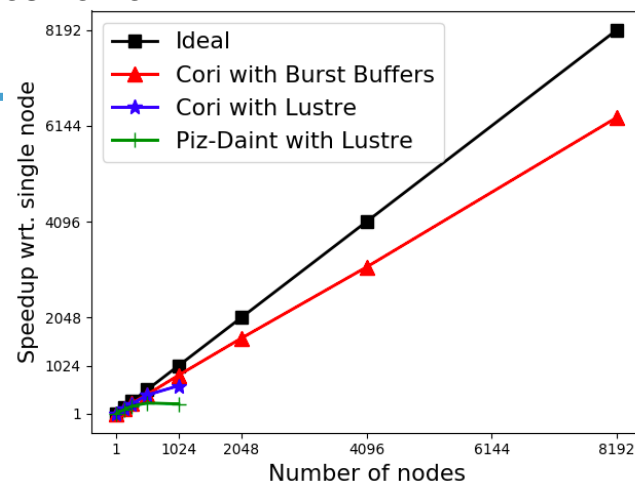
# Some successes

- Scale to thousands of Cori KNL nodes (data parallel): achieve good (weak) scaling
- Frameworks and MPI libraries available and continue to get easier to use
- Great collaborations between Scientists, NERSC, Industry, ML researchers and others

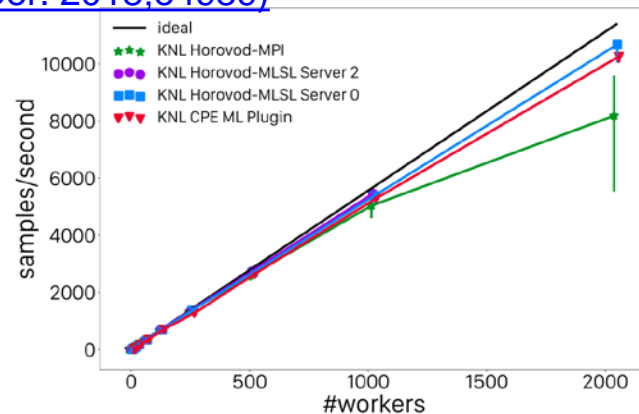
Climate Classification (Kurth et al: SC17):



CosmoFlow:

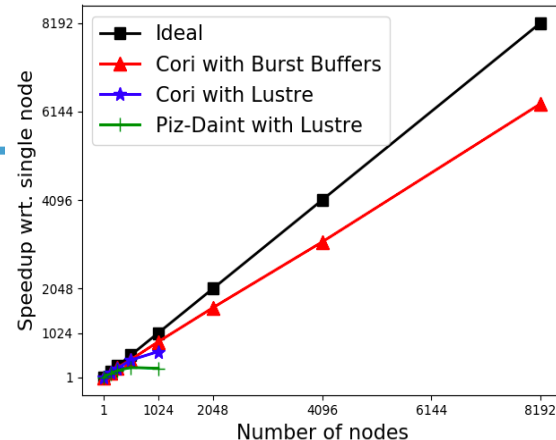
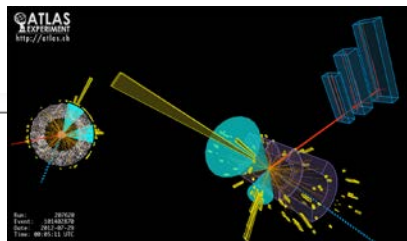
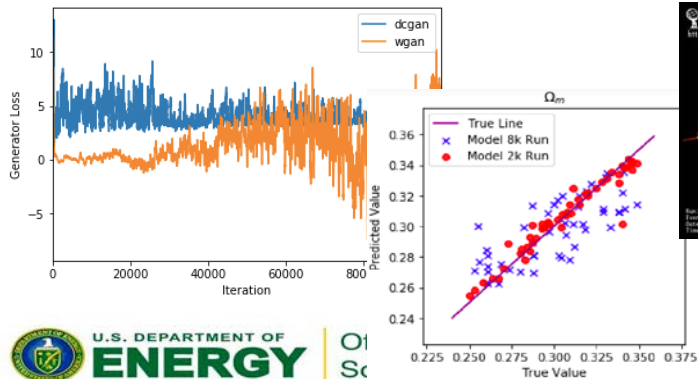


CosmoGAN: (Kurth et al. [Concurrency Computat Pract Exper. 2018:e4989](#))



# Some open challenges

- I/O and data pipelines / structures
- Convergence (at scale)
  - Painful tuning – difficult to provide practical user guidelines
- Interface to real science applications
  - Interpretability; Complex pipelines and sw stacks; Combining best of ML and other approaches
- **Move Proof-of-concept -> Production**



Current / Future work  
(Oxford/UBC/NYU/Intel/NERSC...  
collaboration

Etalumis: Efficient / interpretable Bayesian  
inference with existing science simulators

<https://arxiv.org/abs/1807.07706>

