Overview of the Argonne Data Science Program & Science Highlight

Elise Jennings

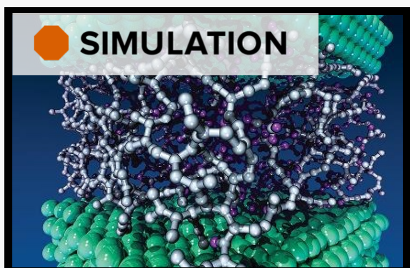
ejennings@anl.gov

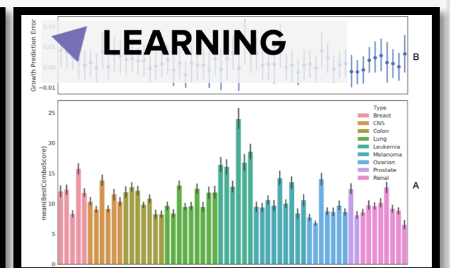
Argonne
NATIONAL LABORATORY

Argonne Leadership Computing Facility

The Argonne Leadership Computing Facility provides world-class computing resources to the scientific community.

- Users pursue scientific challenges
- Resources fully dedicated to open science
- In-house experts to help maximize results





ALCF offers different pipelines based on your computational readiness. Apply to the allocation program that fits your needs.

https://www.alcf.anl.gov

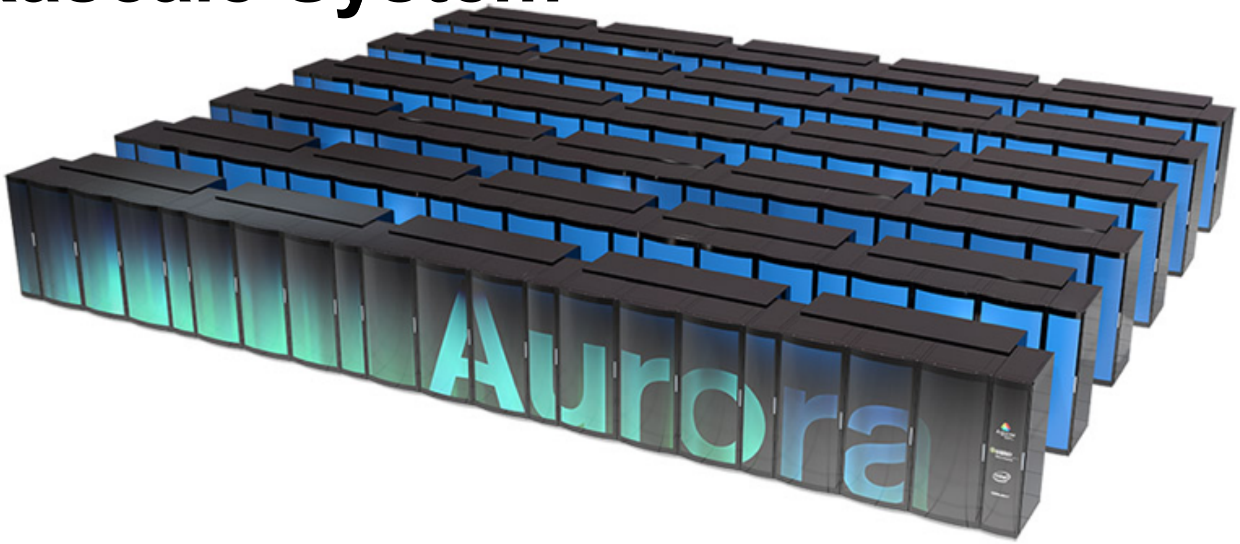
Theta Intel/Cray

4,392 nodes
281,088 cores
69 TiB MCDRAM
824 TiB DDR4
549 TB SSD
Peak flop rate: 11.69 PF

Cooley

NVIDIA system
126 nodes
126 K80 GPUs
Peak flop rate: 293 TF

Aurora 2021 (A21) The first US Exascale System



Architecture supports three ways of computing

- **Large-scale Simulation (PDEs, traditional HPC)**
- **Data Intensive Applications (scalable science pipelines)**
- **Deep Learning and Emerging Science AI (training and inferencing)**

Tom Uram

Venkat Vishwanath

Taylor Childers

Murat Keceli

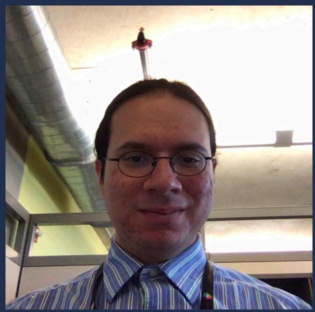
Elise Jennings

Alvaro Vazquez Mayagoitia

Adrian Pope

Misha Salim

Prasanna Balaprakash



Antonio Villarreal

Bethany Lusch

Huihuo Zheng

datascience@alcf.anl.gov

Himanshu Sharma

Xiao-Yong Jin

Ganesh Sivaraman

Corey Adams

ALCF Data Science Program (ADSP)

The ADSP program started in 2016 and is now in its 4th year.

ADSP's goal is to support "Big Data" science that require the scale and performance of leadership computing.

Successful projects have

- high potential impact
- data scale readiness
- diversity of science domains and algorithms
- can fully exploit the architectural features of Theta

Two-year proposal period. Yearly call for proposals.

ALCF Data Science Program (ADSP)

Two main targets for development

Science applications

Tools

To date the majority of proposals have been science applications.

Tool development and support is becoming a major requirement.

ADSP projects

- span a diverse set of science domains (Materials, Imaging, Neuroscience, Engineering, Combustion/CFD, Cosmology).
- involve large science collaborations (multiple APS, LSST, DESC, LIGO, DES, ATLAS) and smaller research groups developing ML at scale.
- have used nearly 300M core hours on Theta (26% as capability runs)

Emerging trends

We now have a mix of applications at scale

- HPC simulations, Big Data analytics and ML

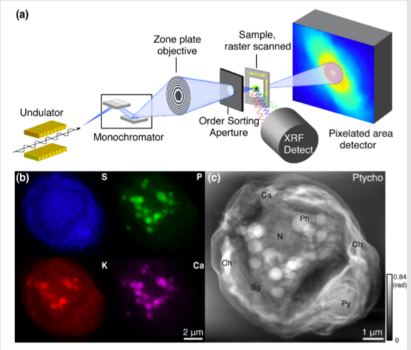
Deep integration of HPC simulations and Machine/Deep Learning

- Augment training data, provide supervised labels for training
- ML model can be embedded into the simulation
- Speed/accuracy trade off in replacing first principal model with ML

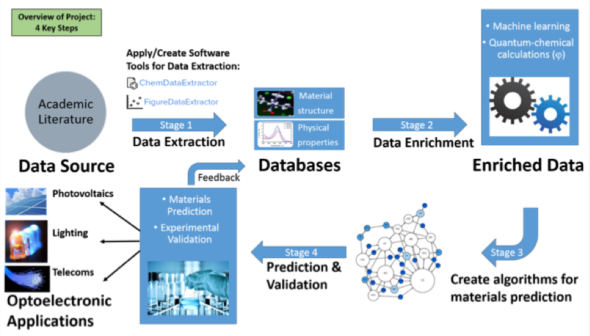
Big drive for

- Scientific techniques (Uncertainty Quantification, reproducibility etc.)
- Image processing, in situ analysis and visualization
- Complex and interactive workflows with performance capabilities
- Smart configuration space sampling
- Tools to couple simulations with analysis and ML

X-ray microscopy of extended 3D objects: scaling towards the future
PIs: Chris Jacobsen, Wild, Nashed (ANL, Northwestern)



Molecular Engineering of Solar-Powered Windows
PI: Cole, Cambridge University, ANL



Argonne NATIONAL LABORATORY

Keeping up with the pace of Machine Learning is challenging

The pace of Machine Learning is very different to traditional HPC.

ML/DL software

- Updates occur every few weeks (TensorFlow, Keras, PyTorch, Horovod, etc.)
- Stack must enable performance libraries (Intel MKL, MKL-DNN, LibXSMM)
- Must work seamlessly with simulation and data frameworks
- Development does not always prioritize backwards compatibility

Keeping up requires

- Dedicated team members to track and update software regularly
- Containers which can provide portable customized software stack
- Regular training/workshops to update the scientific community

Machine Learning, Deep Learning & Workflow software

ML/DL:

TensorFlow, Keras, Neon, MXNet, Caffe2, Theano, CNTK, PyTorch, Sci-kit Learn, Graph Analytics (Cray Graph Engine), Horovod

With performance libraries e.g. Intel MKL, MKL-DNN, LibXSMM enabled

Intel optimized Tensorflow

Conda package on Theta

Intel Distribution for Python’s optimized numpy

Uncertainty Quantification: Edward, Tensorflow Probability

Containers: Singularity

Data Analysis:MongoDB, Apache Spark, R, Python, Jupyter Hub

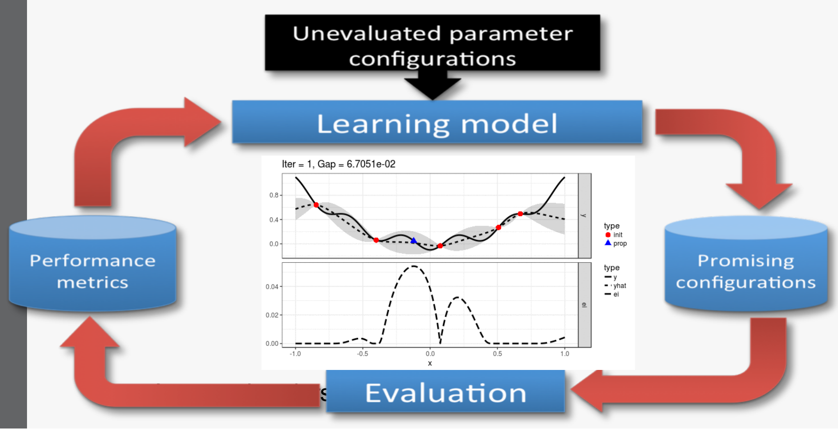
High impact Data Science software

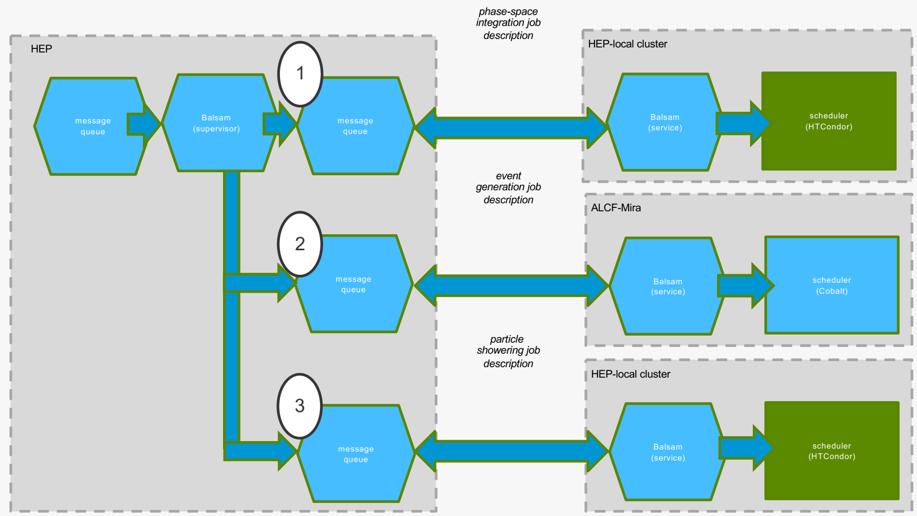
Balsam workflow manager

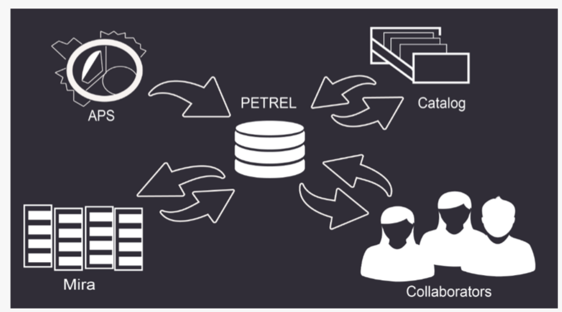
ATLAS experiment used Balsam to run ~100's million compute hours of jobs across ALCF and NERSC systems. Balsam is used by ADSP, ESP, ALCC and ECP applications.
https://www.alcf.anl.gov/balsam

Deep Hyper

ALCF is currently conducting hyperparameter optimization for DL on thousands of nodes of Theta
https://github.com/deephyper/deephyper







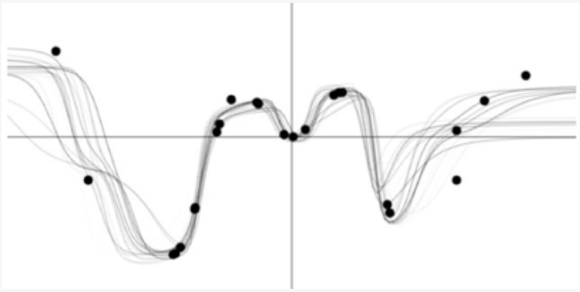
PETREL
Data Management and Sharing Pilot

Petrel

Petrel leverages storage and infrastructure provided by ALCF and Globus Transfer and Sharing services. 100TB allocation per project.

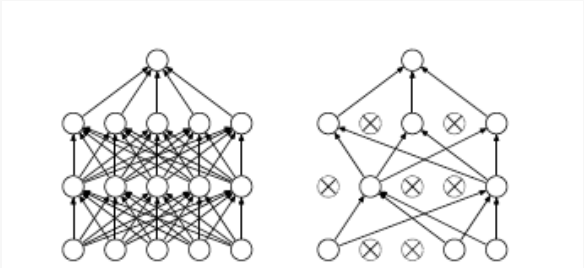
http://petrel.alcf.anl.gov

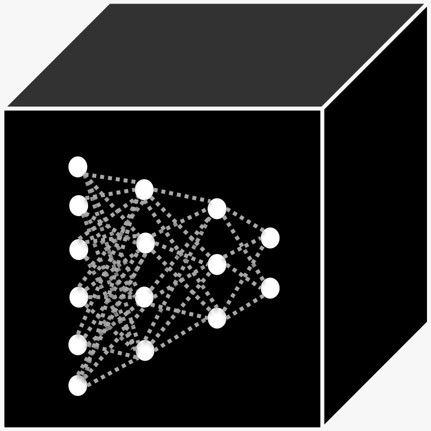
Scientific Machine Learning



Uncertainty Quantification

Software and method development for
- Bayesian Neural Networks at scale
- Drop out methods





Interpretability

Data Science team provide support on
- visualization of neural network
- information theoretic model of neural networks
- Domain knowledge to develop specific network,loss function

Science highlight: Deep Learning in astrophysics

Argonne
NATIONAL LABORATORY

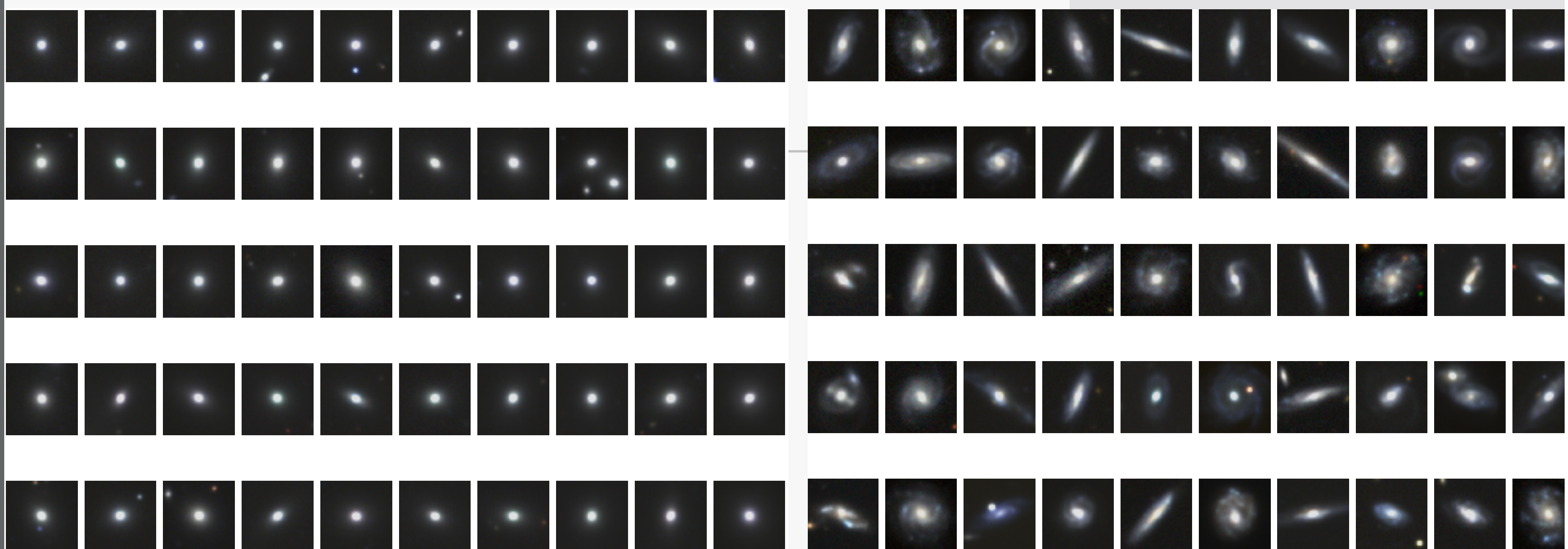
Deep Learning in Cosmology/Astronomy

Applications:

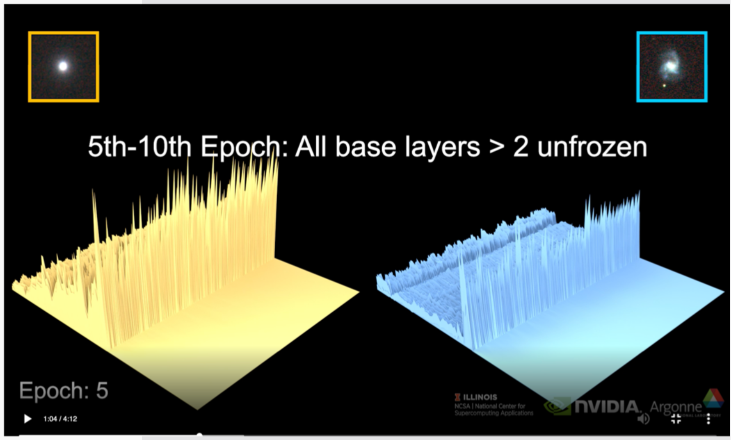
- Mock catalogue creation
- Augment N-body simulations
- Detect gravitational lenses
- Detect transients in real time
- Detect gravitational waves in real time
- Classify galaxy images
- Parameter estimation from time series data

Multimessenger Astrophysics through the NCSA-Argonne Collaboration
PIs: Huerta, Zhao, Haas, Saxton (NCSA)

Deep Transfer Learning to classify galaxy images



Deep Transfer Learning to classify galaxy images



Network:

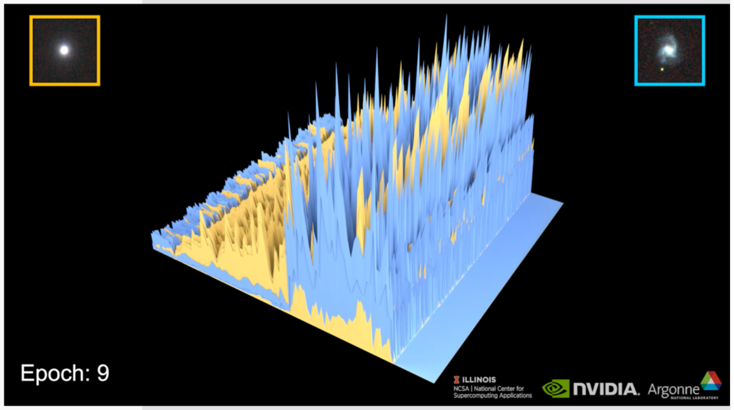
Xception + a few custom defined fully connected layers

Weights:

pretrained weights with the ImageNet dataset

Data:

resized all the galaxy images 299 × 299 pixels



Method:

Progressively unfreeze earlier layers of the whole network
Fine tune their weights for a few epochs of training

Retain earlier layers of a trained network: versatile filters for features like lines and edges

Using the network as a feature extractor

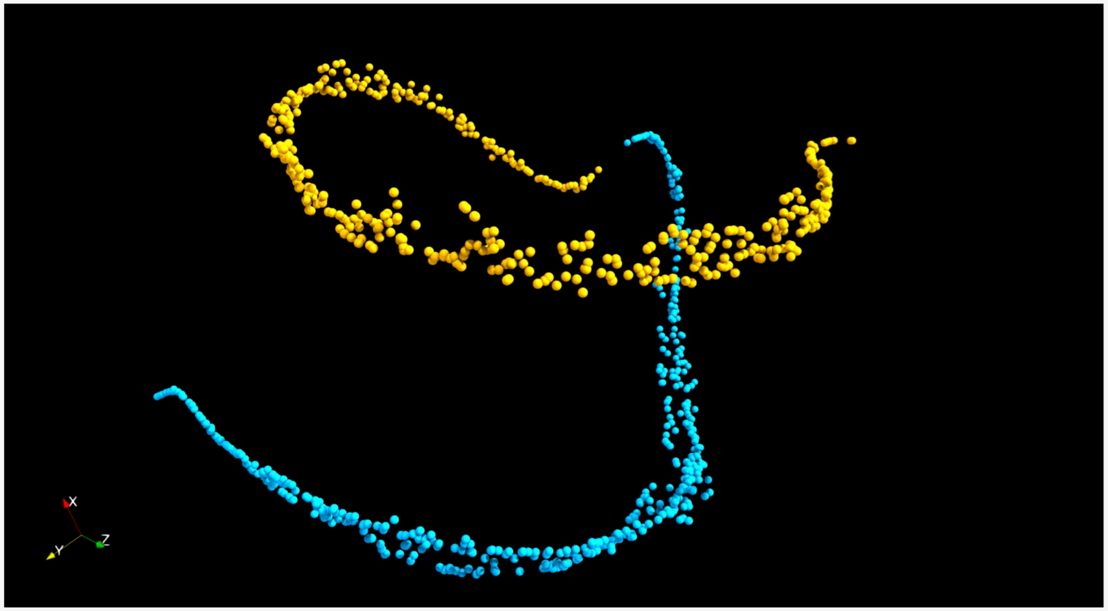
Network output:

output activation values from second-to-last layer

3D representation

t-Distributed Stochastic Neighbor Embedding (t-SNE)

Addresses common problem of large unlabeled datasets



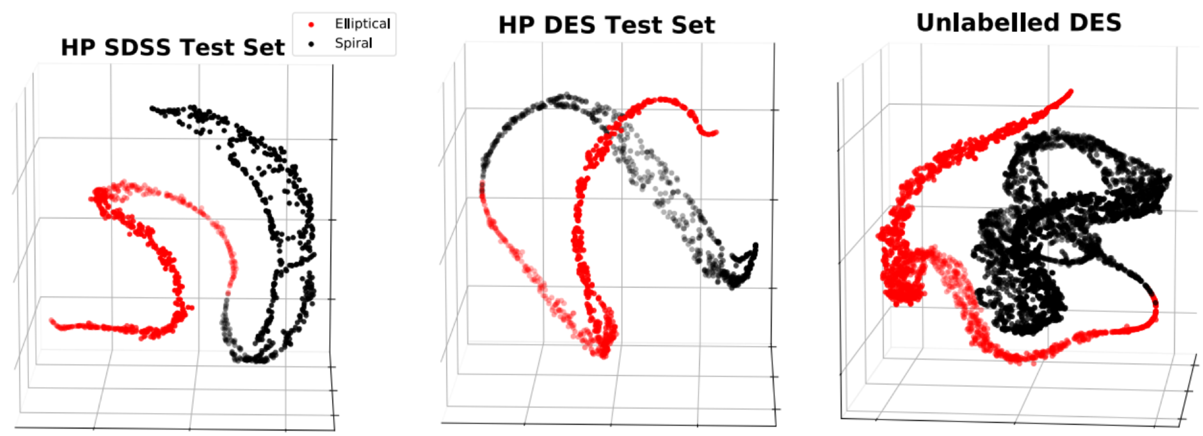
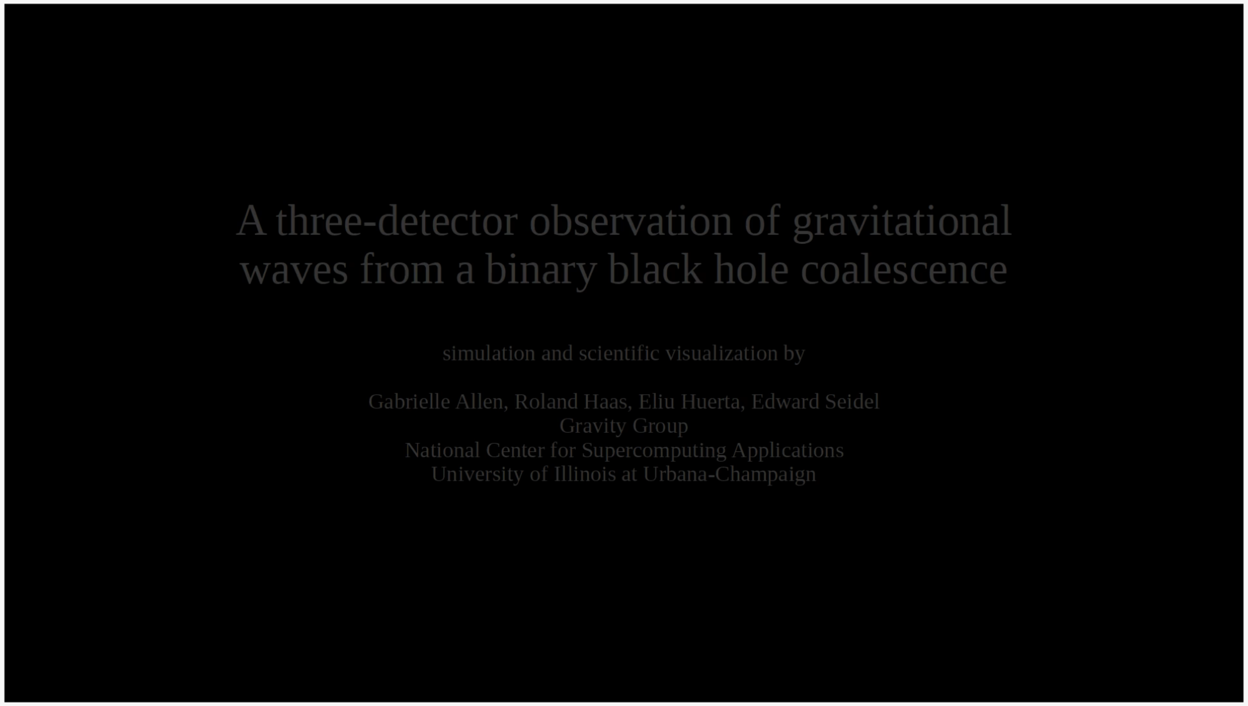


FIG. 4: t-SNE visualization of the clustering of HP SDSS and DES test sets, and unlabelled DES test.

ADSP mmaADSP
Khan et al 2019

Gravitational Wave detection and parameter estimation



Novel data-parallel deep learning fusing HPC and AI for MultiMessenger Astrophysics (MMA).

Huge potential for scientific discovery

- Convergence of all-sky GW observations (LIGO) with deep, high-cadence electromagnetic observations (LSST)
- Novel visualization of Neural Networks

Deep Learning for Multi-Messenger Astrophysics. A Gateway for Discovery in the Big Data Era, Huerta et al., Nature Review Physics

Argonne NATIONAL LABORATORY

Bayesian Neural Networks at scale

- Deep Learning at scale for parameter estimation of Binary Black Hole (BH) mergers
  (spins are aligned or anti-aligned, evolve on quasi-circular orbits)

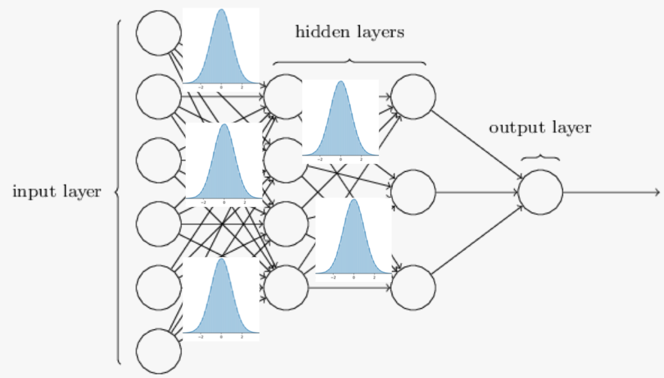
- L2loss re-defined to be negative Evidence Lower Bound (ELBO) loss.

  ELBO loss = expected negative log-likelihood +
  Kullback-Leibler divergence

- Variational Inference method

  Posterior distribution parameter fit by network
  Prior distributions for all network parameters

$$w_i^1 \sim \mathcal{N}(0, \epsilon^1)$$
$$w_i^2 \sim \mathcal{N}(0, \epsilon^2)$$



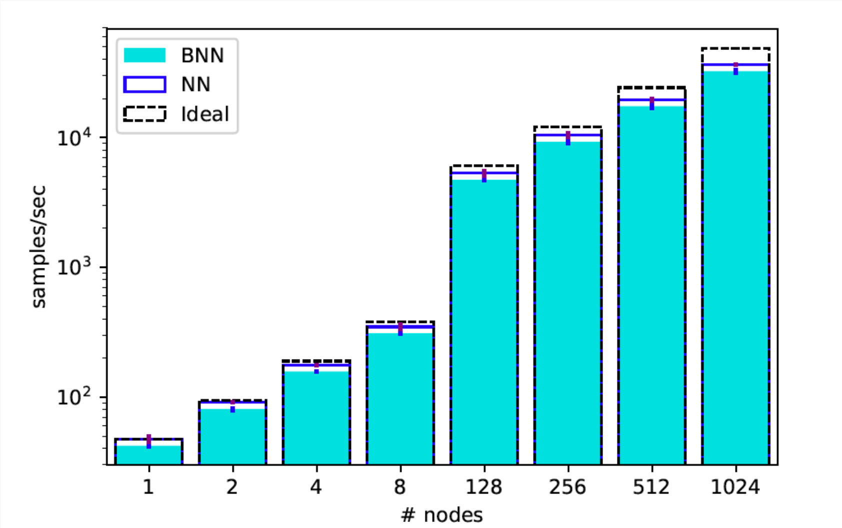
Bayesian Neural Networks at scale

Inference is very costly

- Negative Log Likelihood is approximated via Monte Carlo
- Validation, test steps only carried out when needed

Scaling on Theta with Horovod.

- Horovod and Tensorflow timeline show increased time in AllReduce due to increased parameter set.



Tensorflow Probability on Theta

Thank you !

datascience@alcf.anl.gov