ANL HEP AI projects

Introduction

- Aim is to collect ideas for short-term (call coming in March 2020) AI ideas.
- Long-term ideas are also welcome.
- Consider overarching themes, if there are any (no need to force any).
  - Possible theme: ML experience already present within HEP.
  - Possible theme: common algorithm expertise.
- Connect ideas with ANL resources: A21 and CELS.
  - There may already be existing collaborations.

DOE AI for Science

- Several "AI town hall" meetings held at several labs and DC.
  - ANL, ORNL, LBNL, DC
- Very broad: many disciplines with cross cutting themes.

DOE uniquness (vs industry)

- Techniques for robust uncertainty quantification (UQ) -- neglected by non-scientific applications. Without UQ methodology, many ML techniques will not be used in critical areas.
- Interpretabilty -- scientists need to understand what the methods are doing.
- Incorporation of physical models/constraints -- this is a key issue and in early stages in ML.

DOE AI for Science: DC HEP summary

- Usable tools for large-scale distributed training and optimization of ML models to enable scaling up the complexity of models to orders of magnitude above the current state-of-the-art
- Training methodologies that are able to detect rare features in high-dimensional spaces while being robust against systematic effects
- Tools to quantify the impact of systematic effects of the accuracy and stability of complex ML models
- High-quality generative models satisfying physical constraints and symmetries
- Fast methods for solving high-dimensional statistical inverse problems

Should attempt to be aligned with this message.

PSE AI for Science

- Effort within PSE to come up with AI strategy within the lab.
- Weekly "town hall" meetings with presentation of current AI work.
- Culminated in a workshop and a white paper that summarizes AI plans/desires within PSE.

Many practical/organizational aspects discussed will high light short-term strategy.

PSE AI for Science: short term goals summary

- Methods and software tools for using active learning in observations/experiments.
- Methods, software tools, and workflows for data curation in preparation for AI applications.
- Tools that leverage robotics expertise to integrate AI in experimental apparatuses.
- AI techniques for rare event identification.
- Developing tools to model performance scaling in training neural networks, and deployment of neural networks in science problems.

Should make sure we aren't in conflict with PSE AI strategy.

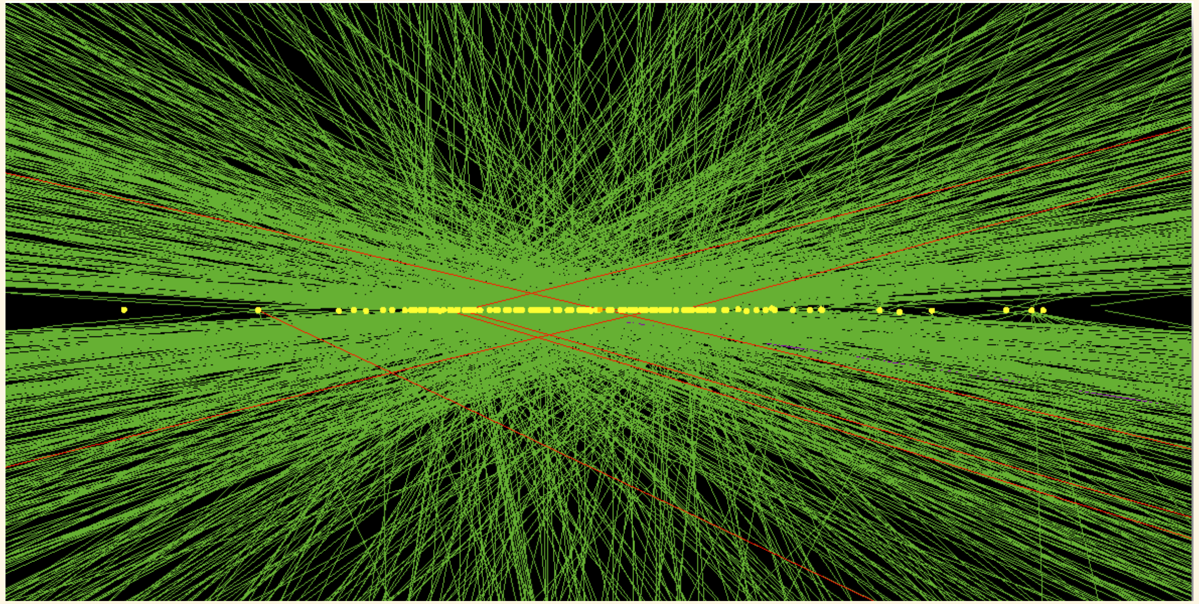
Summary of efforts

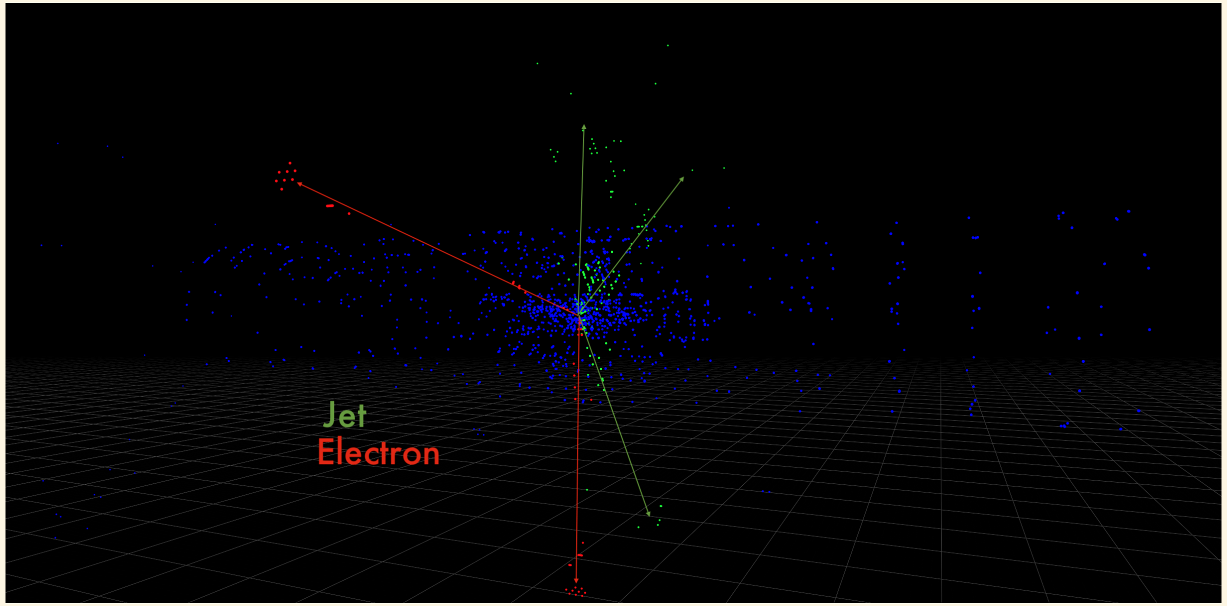
- Divided efforts into (arbitrary) categories:
  - Object classification/reconstruction.
  - AI/ML in analysis.
  - AI/ML as a surrogate/dimensional reduction.
  - Testing/exploring new AI/ML techniques.
- Division of projects is somewhat based on application.
- Alternative would be to use DOE AI themes: UQ, interpretability, physics constrained models.
  - I could only really see UQ commonality... but that's what the discussion is for.

Object/channel classification

Semantic segmentation for LHC trigger/reco

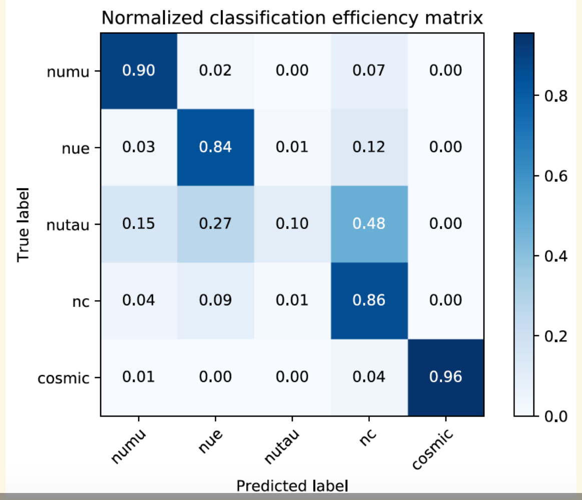
- HL-LHC will cause significant increase in detector occupancy → need to discern pileup background from interesting events.
- Semantic segmentation: channel-by-channel categorization.
  - Each input channel is associated with a physics object (jet, tau, etc).
- Semantic segmentation could identify only relevant detector channels.
  - Drastically reduce input data for trigger/reconstruction → large speedup

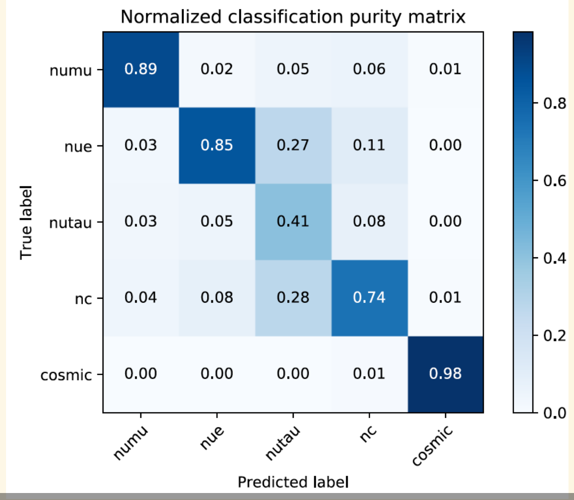




NOvA/DUNE

- Need to identify flavor of neutrinos and individual particles.
- Used MobileNetv2 convolutional neural network (CNN) to identify/classify flavor of neutrino based on NOvA event images.
- Built infrastructure to use ANL HPCs (Cooley) and ANL hyperparameter scan (DeepHyper) for training.
- NOvA experience useful for DUNE:
  - Classify events/particles using raw waveform and event geometry data from ProtoDUNE-SP.

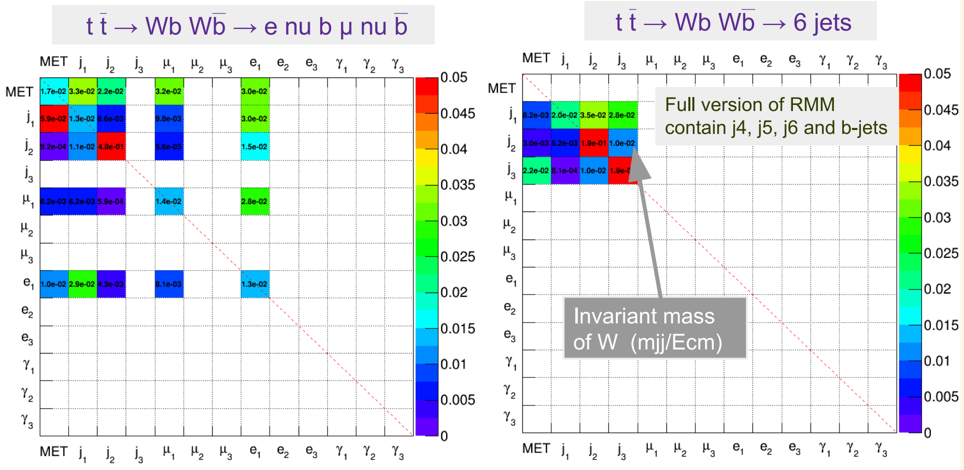




AI/ML in analysis

Generic inputs for ML-based LHC BSM searches

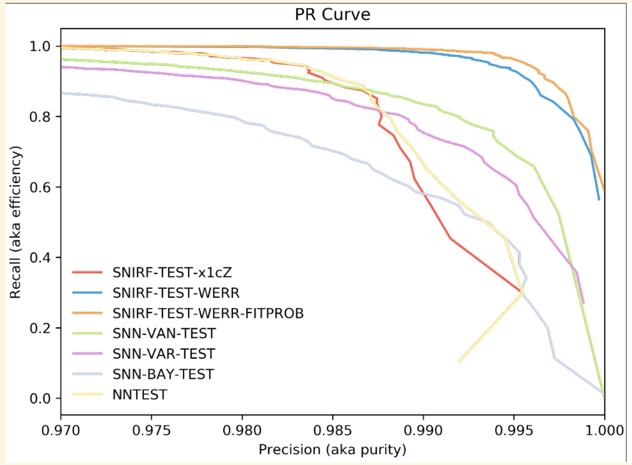
- Instead of building complex variables manually use ML to find optimal signal/background variable combination.
- Rapidity Mass Matrix (RMM) combines combinations of objects to form masses and rapidity differences.
  - Could be processed fully-connected neural network or more advanced techniques like CNNs.
- Promising result for ATLAS dijets+lepton search.
- Plan: apply this to for next dijets+lepton search.



Cell with MET, μ and e leptons activated

Many jets, no MET and leptons

Supernova ID with random forest (SNIRF)

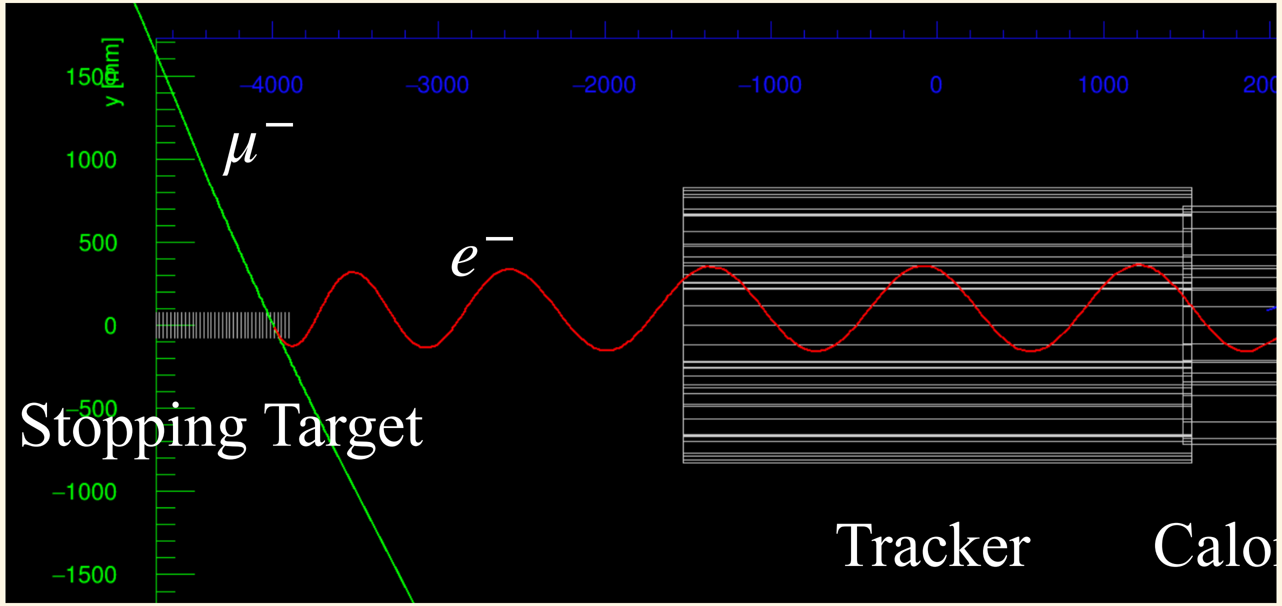


- Current/future imaging surveys discover too many SNe to classify them all spectroscopically.
- Need photometric classification to select cosmologically useful Type Ia Sne.
- Random Forest: classic ML algorithm that is fast and accurate for relatively small training sets.
  - Well suited for systematic studies.
- Many issues in common with other AI applications:
  - Training with simulations versus real data.
  - Feature optimization.
  - Uncertainty quantification.
  - Calibration of probabilities

Mu2e: cosmic ray veto and beam induced noise

- Mu2e ($\mu \rightarrow e$ conversion) probes new physics mass scales of 10^4 TeV, 10^4 improvement from past experiments.
- Cosmic Ray Veto (CRV) subdetector needs to suppress cosmic background by 10^4.
- Current cut-based muon ID algorithm can't achieve 10^4 reduction.

Cosmic background event

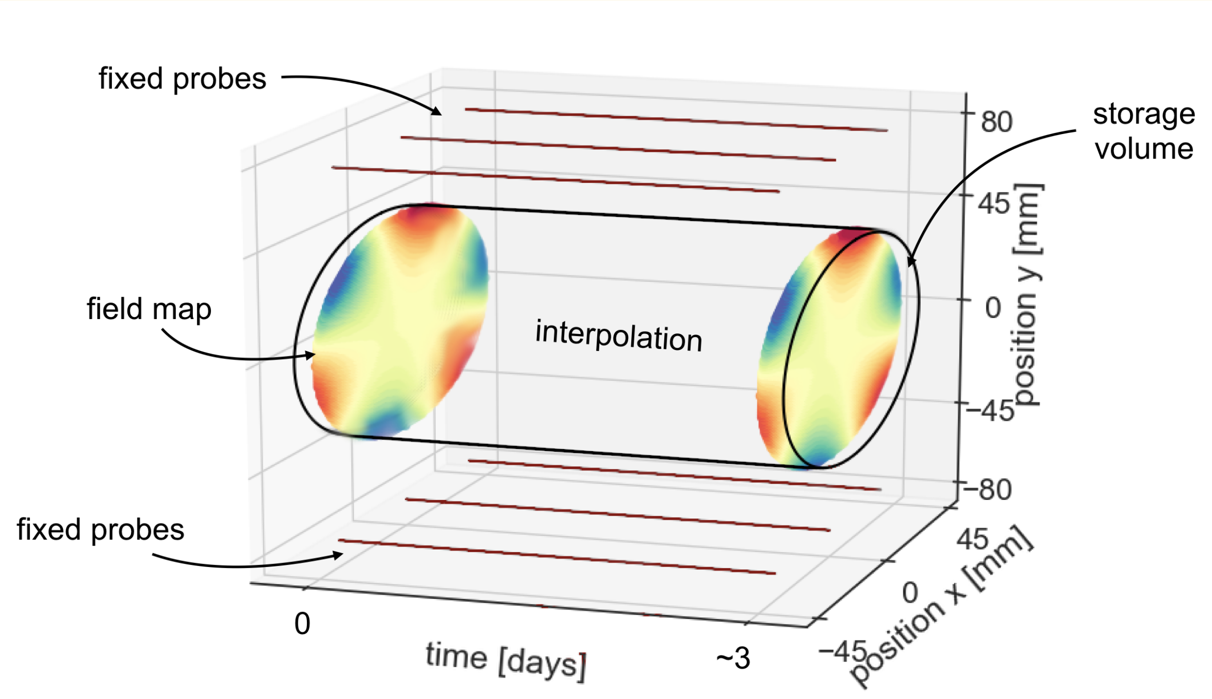


Mu2e: ML for cosmic ray veto and noise reduction

- ML can exploit correlations across CRV channels and full multi-dimensional space for beam induced noise reduction.
- ML requires large data sets for training.
  - Currently simulating electronic noise and cosmic ray background datasets on Theta.
- Plan to test algorithm on pilot CRV modules and eventually on fully commissioned CRV.
- What is learned for CRV ML rejection can be used in other subdetectors.

g-2: magnetic field interpolation with ML

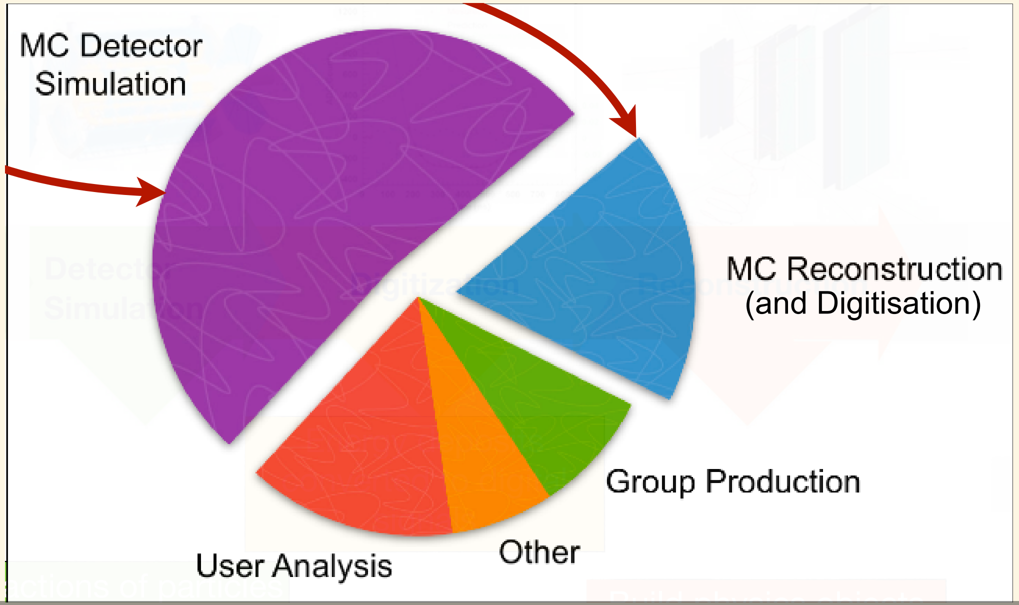
- Magnetic field measurement is essential for g-2 precision (to resolve previous >3σ discrepancy).
- Currently the field map is interpolated to get field inside storage region.
  - This is the dominant source of uncertainty in field measurement.
- ML can probe non-linear dynamic high dimensional parameter space to find optimal interpolation.

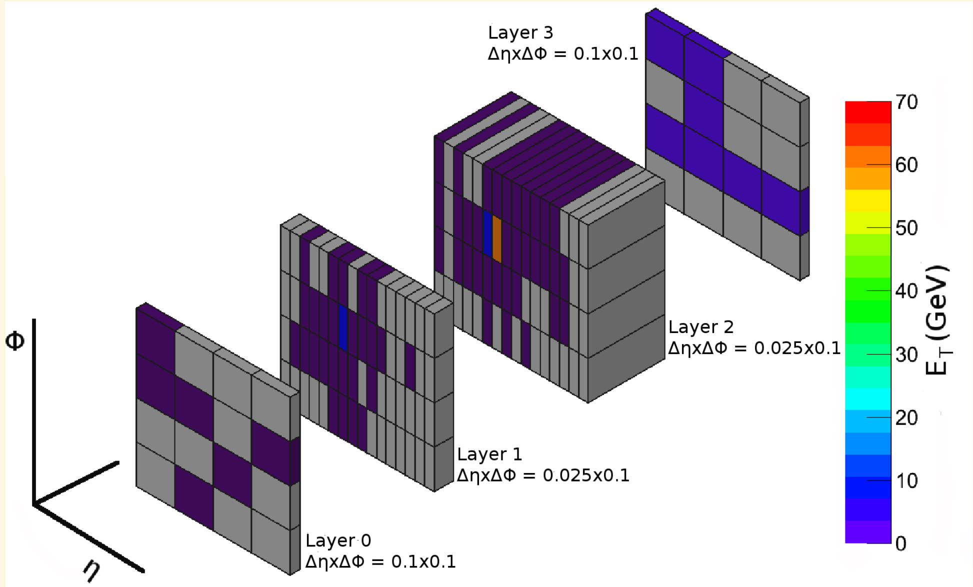


AI/ML as a surrogate/dimensional reduction

ML for LHC detector simulation

- Need: fast/accurate transformation of particle (~10 parameters) to detector quantities (~100s, with detector geometry convolved).
- Learn only what is needed with least amount parameters: hyperparameter scan/reinforcement learning.
  - **Make use of ALCF human and computing resources.**
- ML gives access to hardware agnostic backends → fast inference on any type of resource.



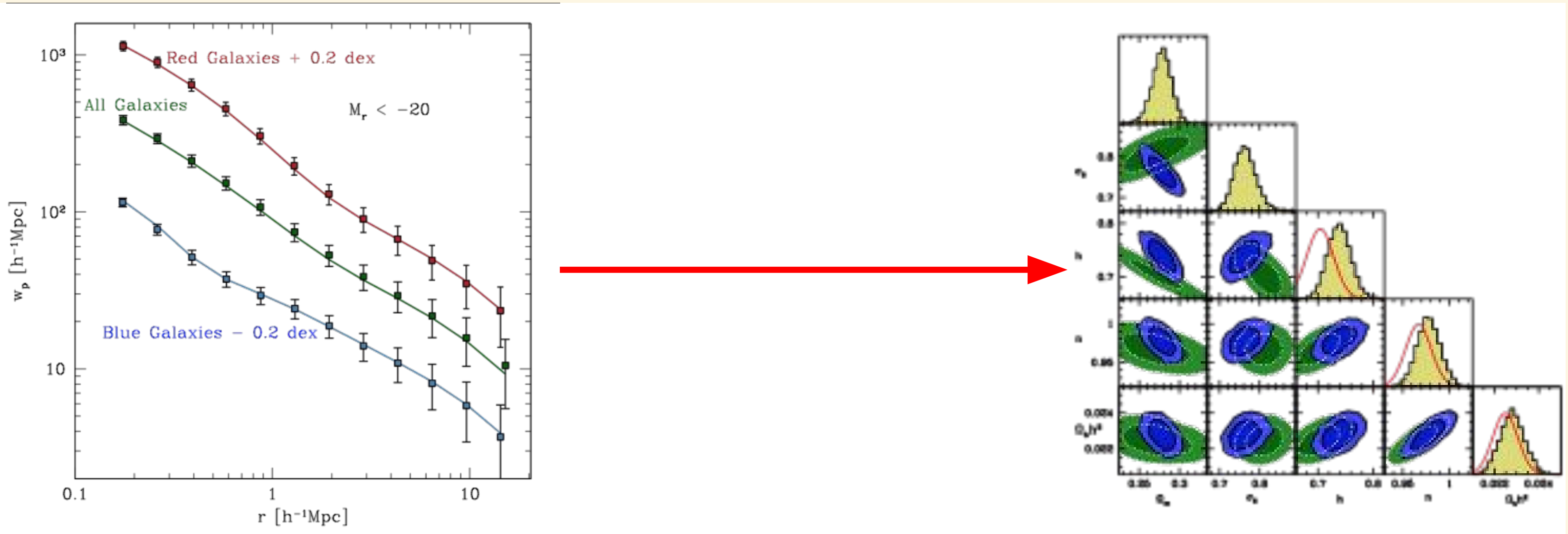


AI-accelerated forward-modeling in cosmology

- Workflow: **cosmological simulation** $\rightarrow$ compute summary statistic $\rightarrow$ compare Λ_{CDM} and other parameters to real observed ones.
- Galaxy models can be moderately expensive to evaluate.
- Explore AI & classical ML techniques to build fast and accurate surrogate model.
  - Surrogate models could easily run on CPU and GPUs due to AI software backends.
- Possible solutions for galaxy models:
  - Autoencoders to reduce parameters.
  - NN approximation to galaxy model

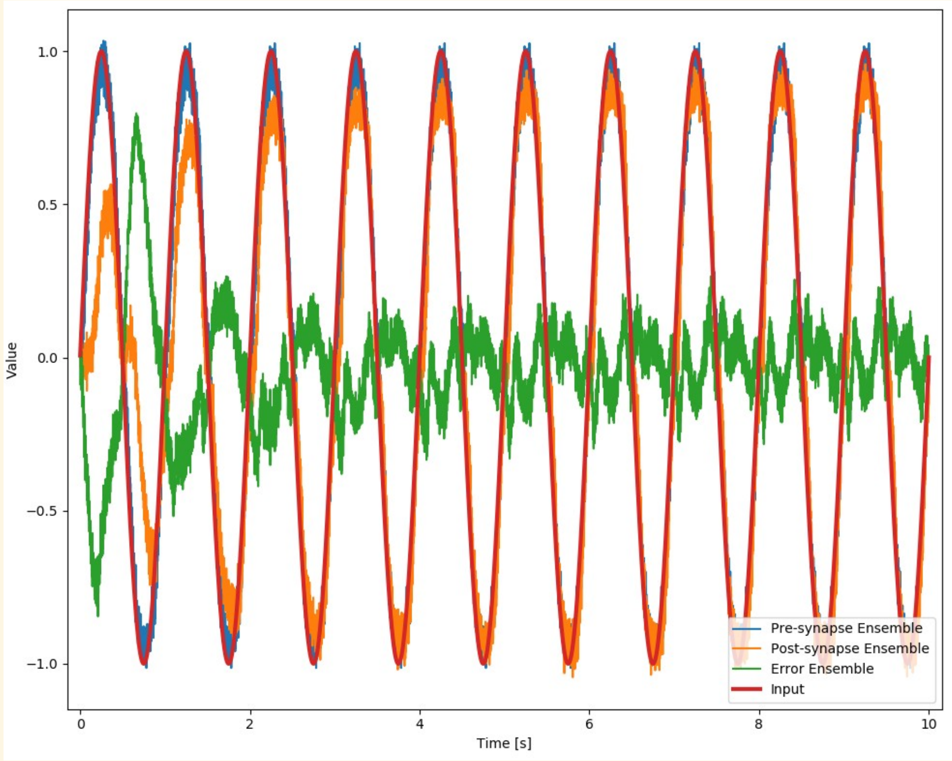
AI-approximated posterior inference in cosmology

- Workflow: cosmological simulation → **compute summary statistic** → compute Λ_{CDM} and other parameters to real observed ones.
- Brute force MCMC rapidly becomes impractical.
- AI for inference: given observables estimate Λ_{CDM} and other parameters.
  - Variational inference to approximate posterior PDF.
  - AI surrogate for the full likelihood function.



Testing/exploring new ML techniques

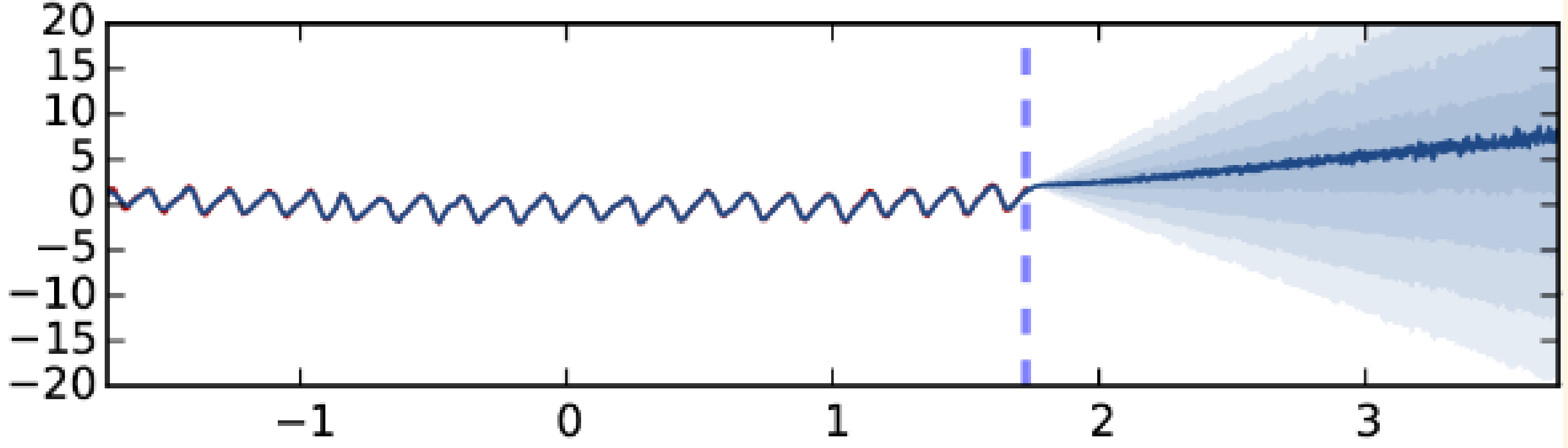
Neuromorphic computing



- Neuromorphic Computing - Computing algorithms and architectures based on applications of the biological behavior of neurons and synapses.
- Neuromorphic hardware systems are extremely low power and exhibit lifelong learning: good candidate for on-detector learning like Smart Sensors.
- Neuromorphic systems process time dependent streams of data. To make use must re-encode digitized data into spatio-temporal correlated spike trains.
- ALCF Expeditionary LDRD to investigate how to optimally encode ATLAS Data.

Uncertainty quantification using MC Dropout

- Dropout stochastically removes NN nodes for overtraining robustness.
  - Traditionally Dropout is enabled for training, disabled for evaluation.
  - Enabling Dropout during evaluation approximates the Bayesian posterior PDF for the NN response (arXiv:1506.02142).



(c) MC dropout with ReLU non-linearities

Uncertainty quantification using MC Dropout

- Project with Graduate Student to quantify performance of Dropout based UQ for ATLAS Flavor Tagging algorithms.
  - ATLAS Deep Learning models are already trained with Dropout.
  - Compare uncertainties from Dropout with traditional techniques.
  - Excellent test case as input and output uncertainties, and well validated data driven calibration techniques already exist.
- ALCF CANDLE group using Dropout to evaluate model uncertainties.
  - Studying ways to speed up the evaluation.
  - Train NN to learn posterior PDF.

Discussion