Deep Learning Applications in HEP

Suyong Choi

Korea University

Outline

- Computing resources at KISTI
- Introduction
- Properties of deep learning in top-jet tagging
- Interesting directions
- Outlook

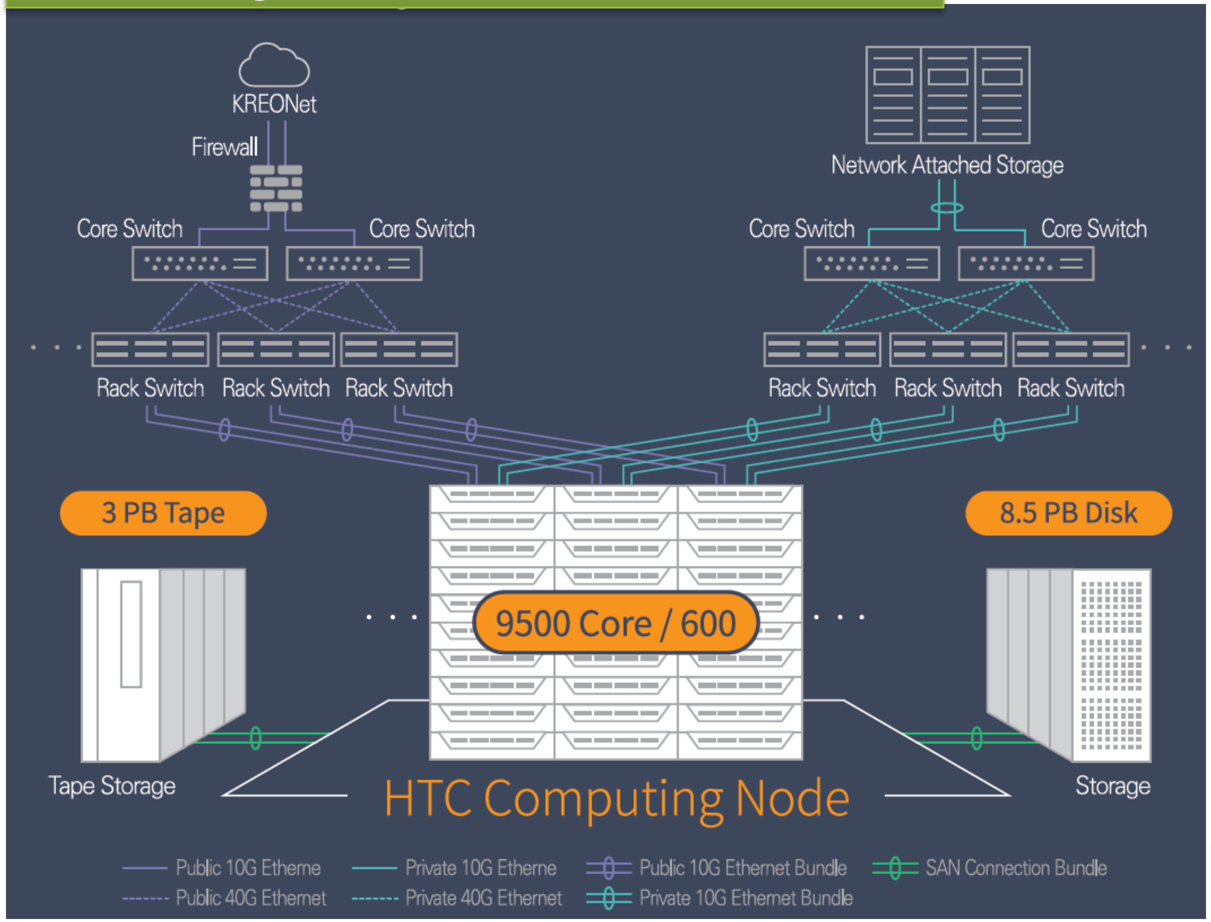
  - Disclaimer: Personal preferences

Infrastructure @ KISTI-GSDC

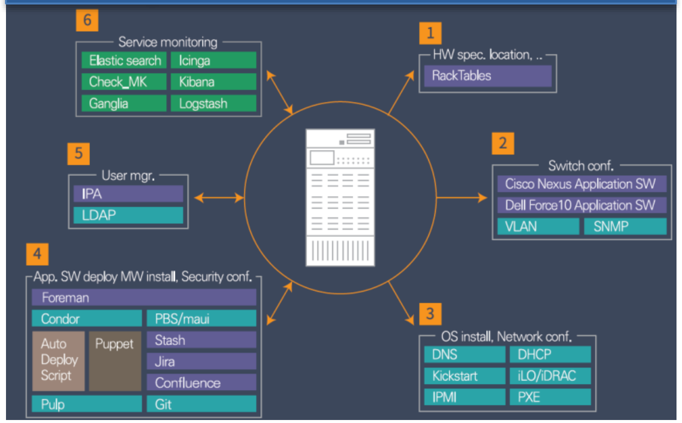
KISTI Provides Tier 1 center for ALICE, Tier 2 center for CMS

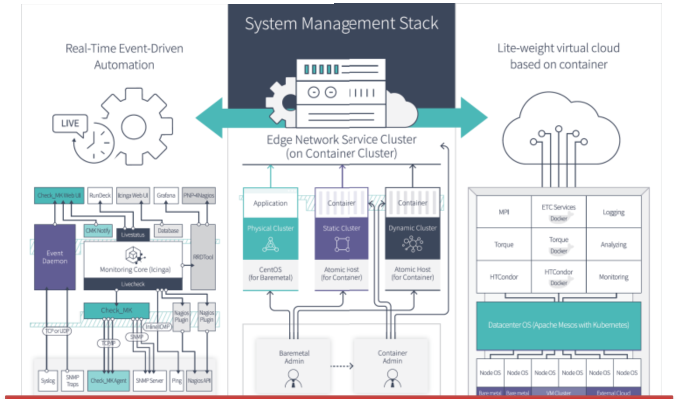
It is impossible without expertise.

25 storage racks with 5 different vendors



Six steps to enable basic functions





Automatic System Management Stack for Service

About GPU machine

Server Spec

- Server Product : Dell R730 (2U)
- CPU : 2x Xeon 2.6GHz 14Core
- RAM : 384GB
- GPU : NVIDIA P100
  - Double-Precision : 4.7TFLOPS
  - Single-Precision : 9.3TFLOPS

System environment

- OS : CentOS7
- Driver Version: 396.26
  - CUDA 9.2 + cuDNN v7.14
- Compute Mode
  - ⇒ Exclusive Process Mode

Deep Learning in High-Energy Physics

- Many efforts with deep learning in a variety of settings
  - Classification in physics object reconstruction
    - Track reconstruction – complicated hit patterns
    - Jet flavor tagging – particle kinematic information
  - Distinguishing new physics”from Standard Model signatures with complicated event topologies

- Original and creative new ways
  - Keeping abreast with new developments
  - Modifying to our data and our needs

- Are we learning about properties of these algorithms?

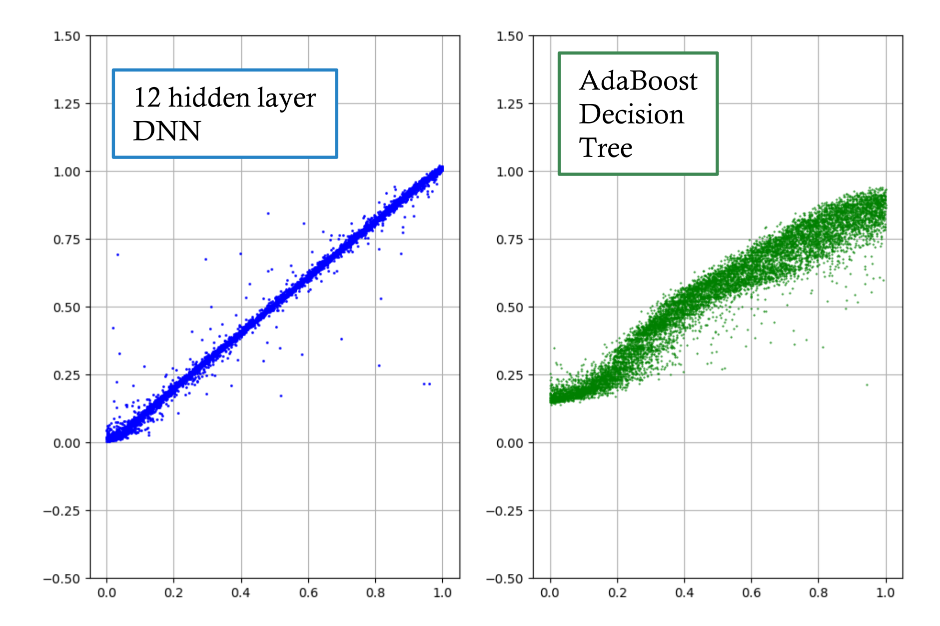
Deep Learning

- Deep neural network as a universal function approximator

- Classification: $f: R^n \to \{1, \dots, k\}$
  - Decide among k-classes

- Regression: $f: R^n \to R^m$
  - Predict real-valued outputs

- For a DNN, crafting sensitive variables (feature inputs) is less crucial



Mass reconstruction using 4-vectors in a two-body decay

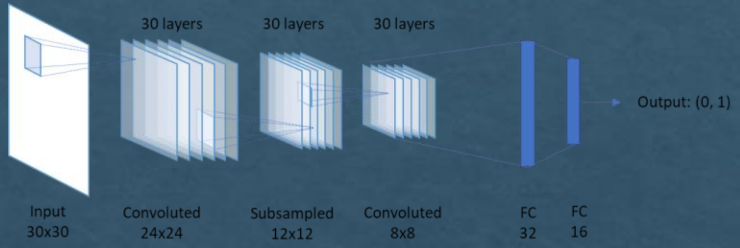
Properties of a DNN
Top-jet Tagging

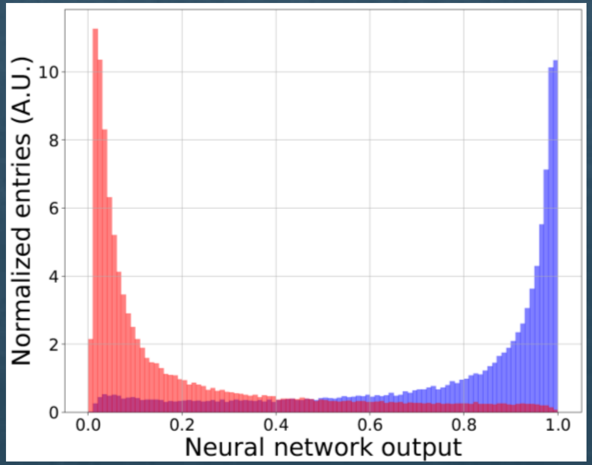
Top-jet tagging

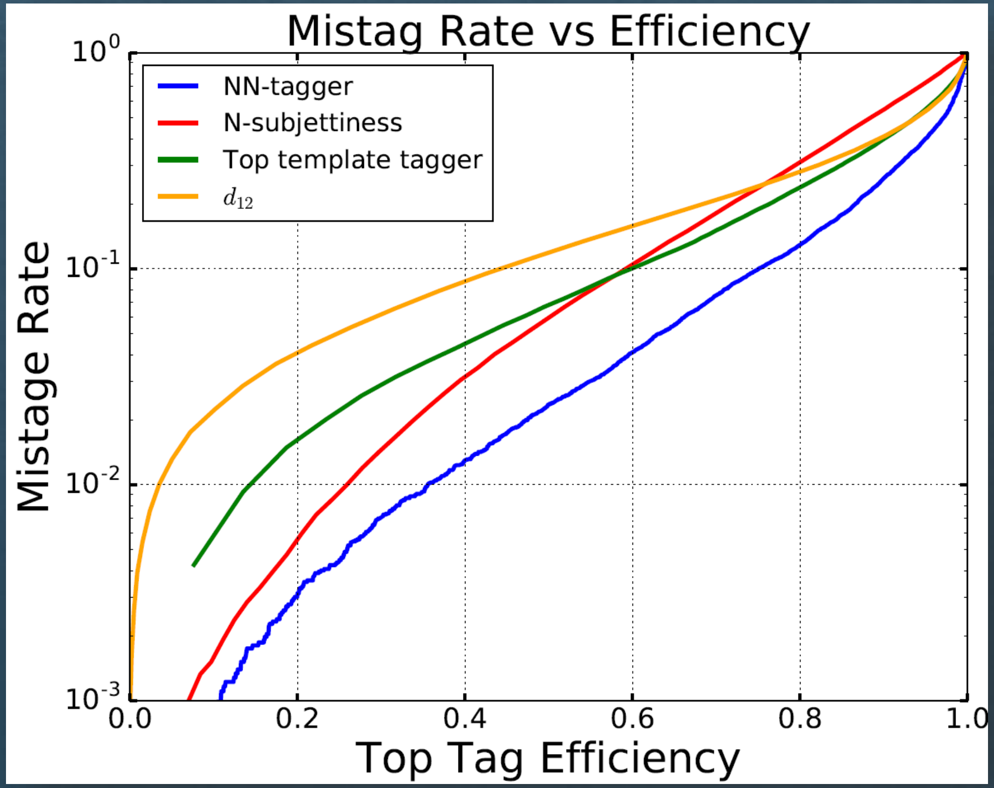
- For a high momentum top quark, three quarks from the top decay can fall within a single jet cone – a "top-jet"
- Top-jet tagging algorithm
  - Distinguish between top-jet, W-jet and QCD jet
  - Makes use of energy distribution inside the jet
- N-subjettiness is a theoretically well-motivated infrared safe observable
- What about algorithms using deep neural network?

Top-jet tagging with DNN

- CNN Network architecture

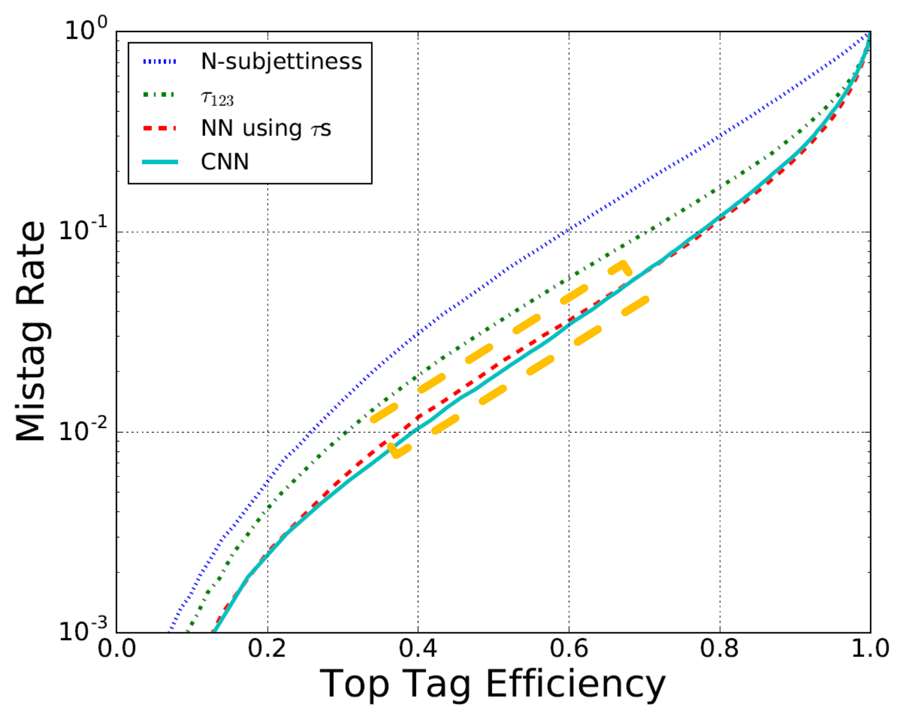






Deep NN with N-subjettiness

- Setup
  - PYTHIA6 generated top quark to hadronic decays
  - $800 < p_T < 900$ GeV, $|\eta| < 2.0$,
    $130 < m_J < 210$ GeV
- Performance comparable to using jet energy depositions
  - 24 inputs - τ_1, τ_2, τ_3, τ_4 with $\beta =$ 0.1, 0.25,0.5,1,1.5,2
  - 3 hidden layers (48, 48, 16) used
- Performances comparable, but, there is crucial difference



Infrared safe property of a DNN top-jet tagging algorithm

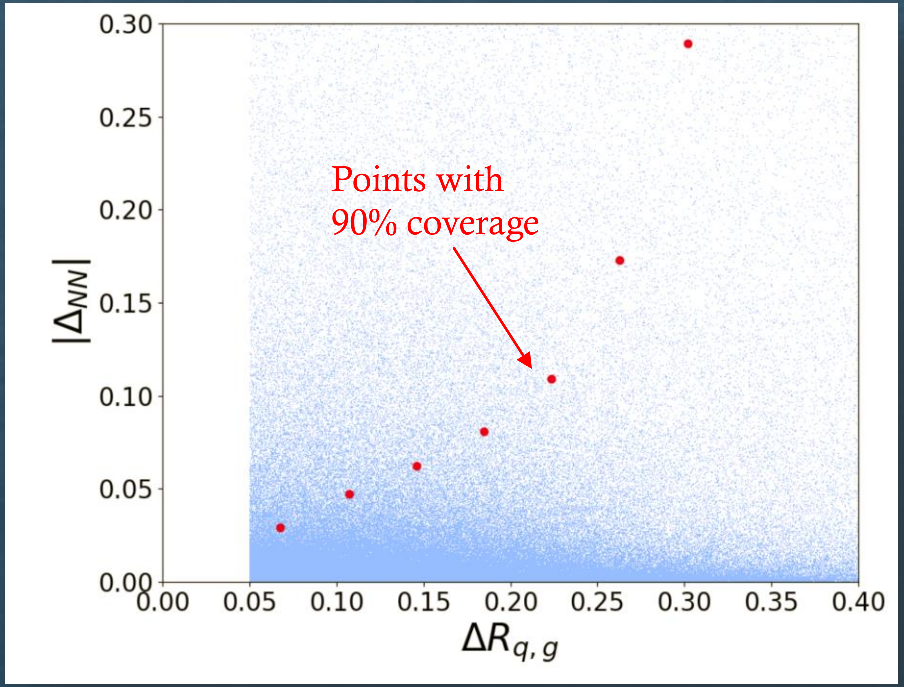
- A DNN based on using N-subjettiness will be infrared-safe
  - The feature inputs τ_i variables are infrared-safe

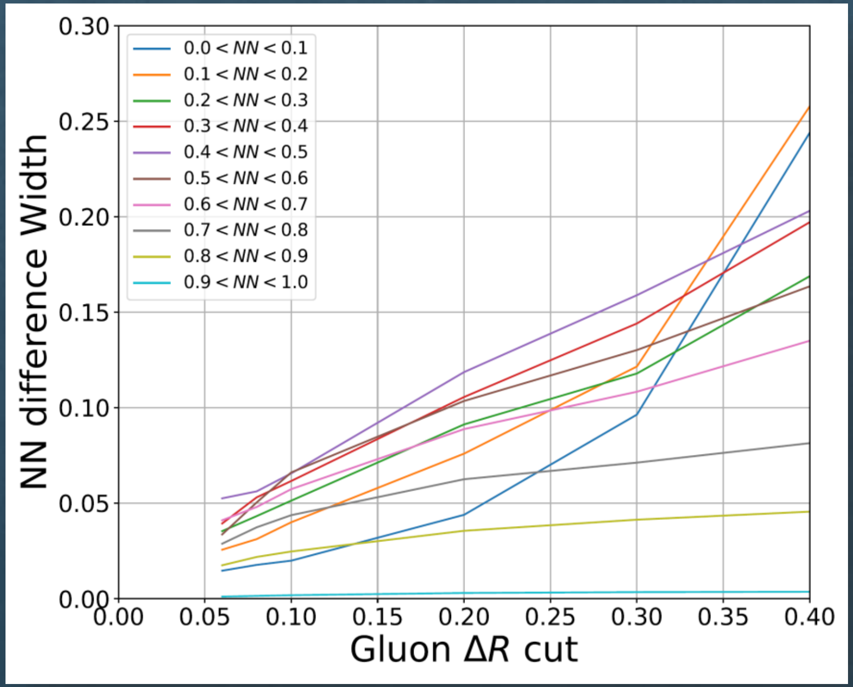
- Is a DNN based on energy distribution infrared safe?
  - Important property for a physically meaningful result
  - Collinear gluon emission
  - Soft gluon emission

Infrared safety property of a DNN Top-jet tagger

- $|\Delta_{NN}| = |NN_{qqq+g} - NN_{qqq_{merged}}|$

arXiv:1806.01263
S. Choi, S. Lee, M. Perelstein

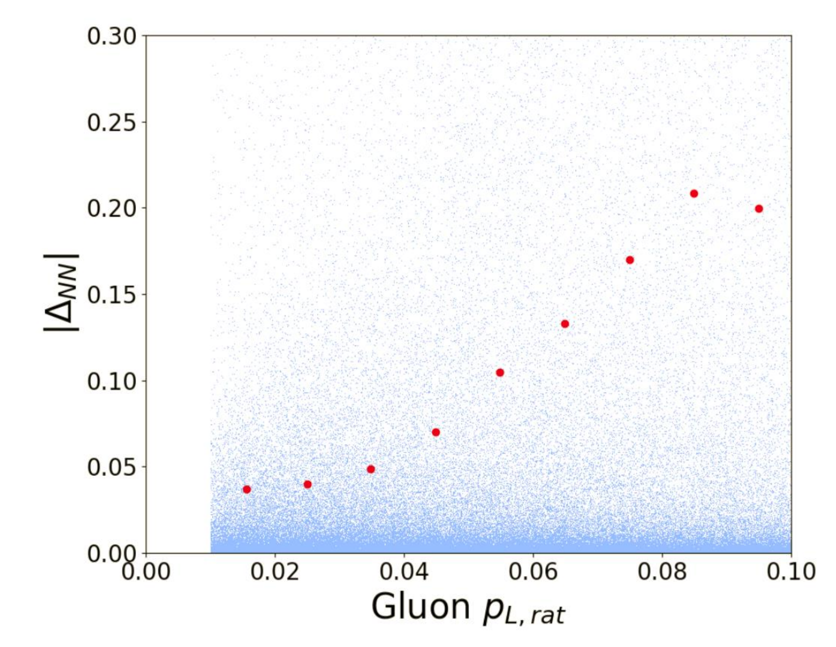


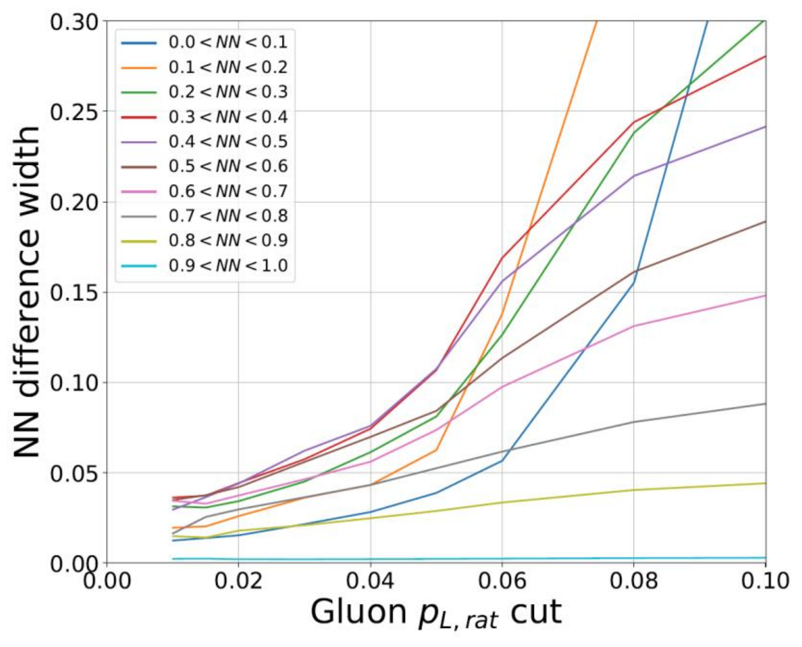


Collinear

Infrared safety property of a DNN

- $|\Delta_{NN}| = |NN_{qqq+g} - NN_{qqq_{merged}}|$





$$p_{L,rat} = \frac{\vec{p}_q \cdot \vec{p}_g}{\vec{p}_q \cdot \vec{p}_q}$$

Soft

Directions of deep learning

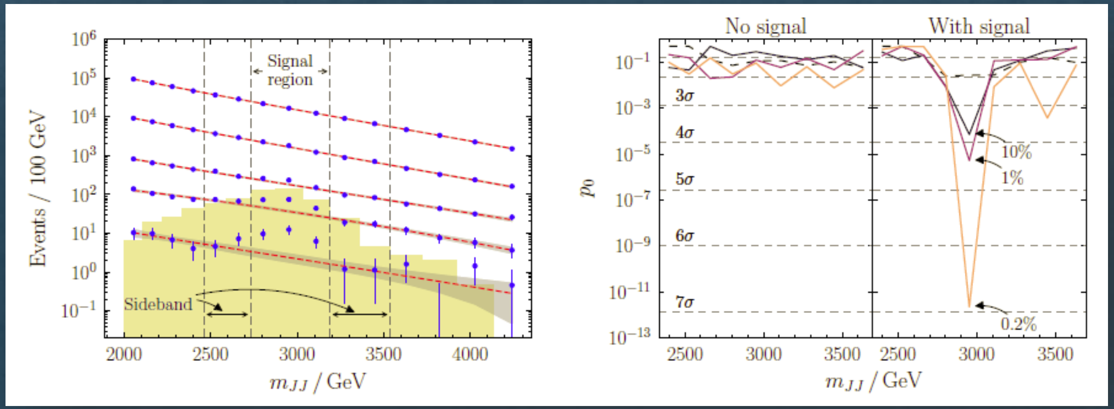
- Learning something from the data itself

- Anomaly detection without specific NP model
  - So called “model independent”
  - Without assuming shape of excess

- Generating fake data mimicking real ones

Searching for Anomaly

- Predict backgrounds from side-band data
  - No assumption on the shape of signal except width
  - Backgrounds: side band events. Signal: events in signal mass region
  - Train classifier using variables nearly uncorrelated with mass
- In $pp \to W' \to W + X(\to WW)$
  - For a massive W', boosted W-jets: Jet substructure variables



J. H. Collins, K. Howe, B. Nachman
arXiv:1805.02664

Searching for anomaly

R. D'Agnolo and A. Wulzer
arXiv:1806.02350

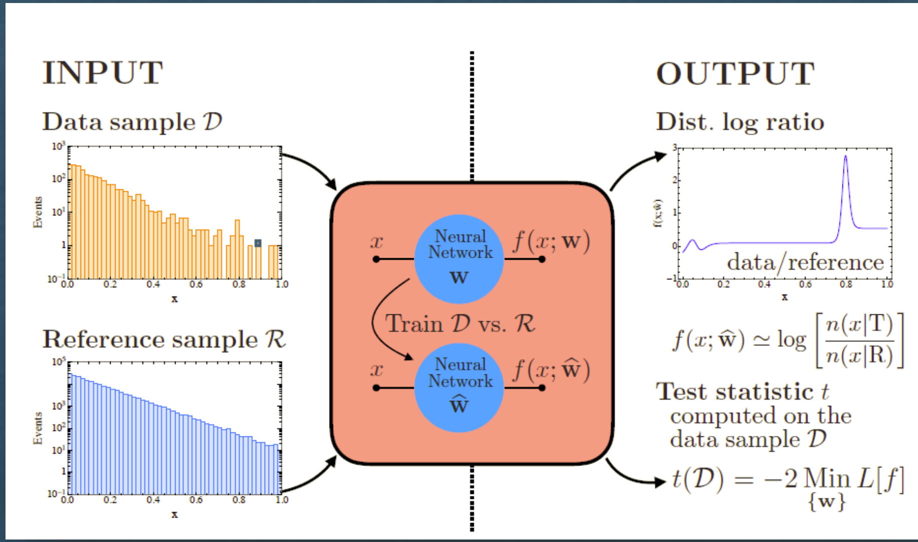
- New physics signature could take on many different shapes, not always a gaussian bump
- Parametrize the difference between reference data (null hypothesis) and observed data using a NN

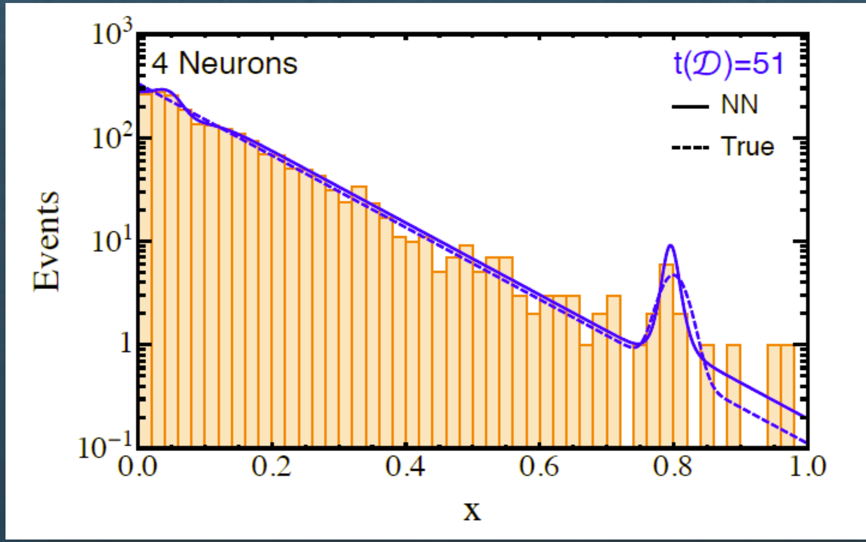
$$n(\vec{x}|w) = n(\vec{x}|R)\, e^{f(\vec{x};w)}$$

- Using a test-statistic based on maximum-likelihood principle, observed p-value calculated from

$$p_{obs} = \int_{t_{obs}}^{\infty} P(t|R)\, dt$$

Searching for anomaly





R. D'Agnolo and A. Wulzer
arXiv:1806.02350

Directions in Deep Learning in HEP

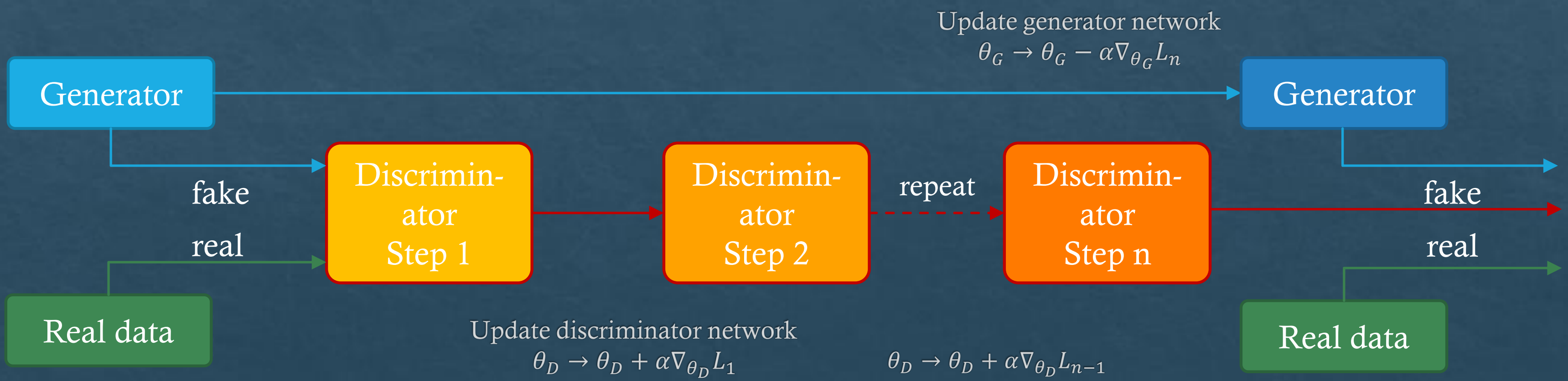
- Learning from data itself
  - Approximating observed probability density function
  - Anomaly detection

- Why do we want to do that?
  - Some processes are copious yet difficult to get it right or difficult (or time-consuming) to simulate correctly
  - Model-independent way of finding new physics

Generative methods

- Learn from data and generates fake data that mimics the original
  - Generative adversarial network (GAN)
  - Variational auto encoder (VAE)

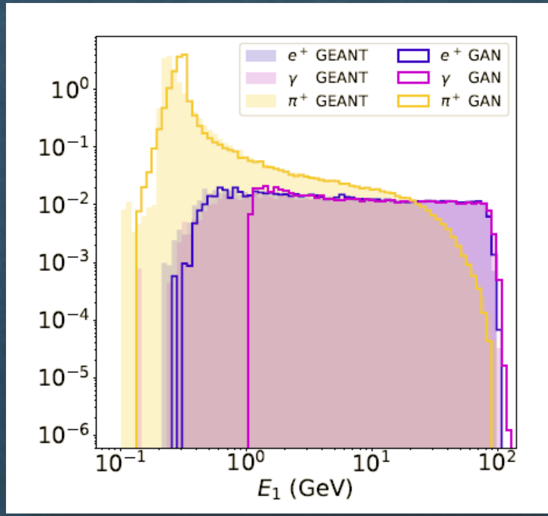
- GAN



Applications of generative methods

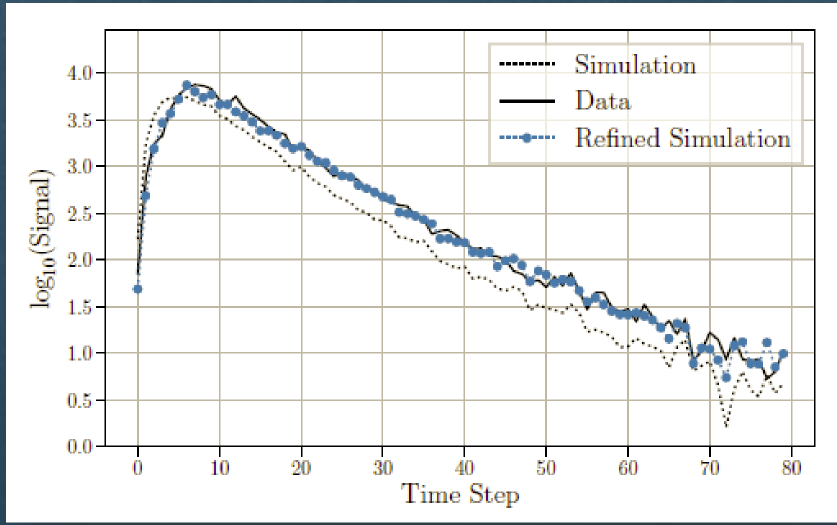
- Simulating and Augmenting MC data

3-D energy deposition in calorimeter



M. Paganini, L. Oliveira, B. Nachman
Phys.Rev. D97 (2018) 014021
arXiv:1712.10321

69 EeV airshower



M. Erdmann, L. Geiger, J. Glombitza, D. Schmidt
arXiv:1802.03325

Outlook

- Progress in applying deep learning methods in classification problem in HEP
- Understanding properties of deep learning algorithm is important in making sense of its outcome
- How we can make use of the big data
  - Model-independent searches
  - Faster simulated data generation
  - Better modeling
- How can we make it assist us?