Jet Representations for Machine Learning:

An Overview

David Shih
NHETC, Rutgers University

Fermilab ML4Jets Workshop 2018

ML @ LHC

These are very exciting times for machine learning and LHC physics!!!

Recent breakthroughs in deep learning have spurred much activity in our field.

Many different ideas are being explored for how to use these breakthroughs to improve the analysis of LHC data.

We are seeing huge gains in performance resulting from the ability to harness much lower-level information than was previously possible.

An important issue is how to represent the LHC data in a way that can be easily fed to "off-the-shelf" deep learning algorithms.

Also, whether physics motivates particular representations that lend themselves to particular NN architectures, or even lead to new ones.

ML @ LHC

Jet representation ⟷ ML architecture

4-vectors
Images
Sequences
Trees
Graphs
Kinematic invariants
Latent spaces
…

DNN
CNN
RNN
GRU
LSTM
RecNN
GNN
GAN
…

Application

Jet tagging
Pileup removal
Event generation
Parton shower
Triggering
Anomaly detection
…

Jets as…

Lists of four vectors

Images

Sequences

Trees

Graphs

Kinematic invariants

Latent spaces

…

Jets as…

Lists of four vectors

Images

Sequences

Trees

Graphs

Kinematic invariants

Latent spaces

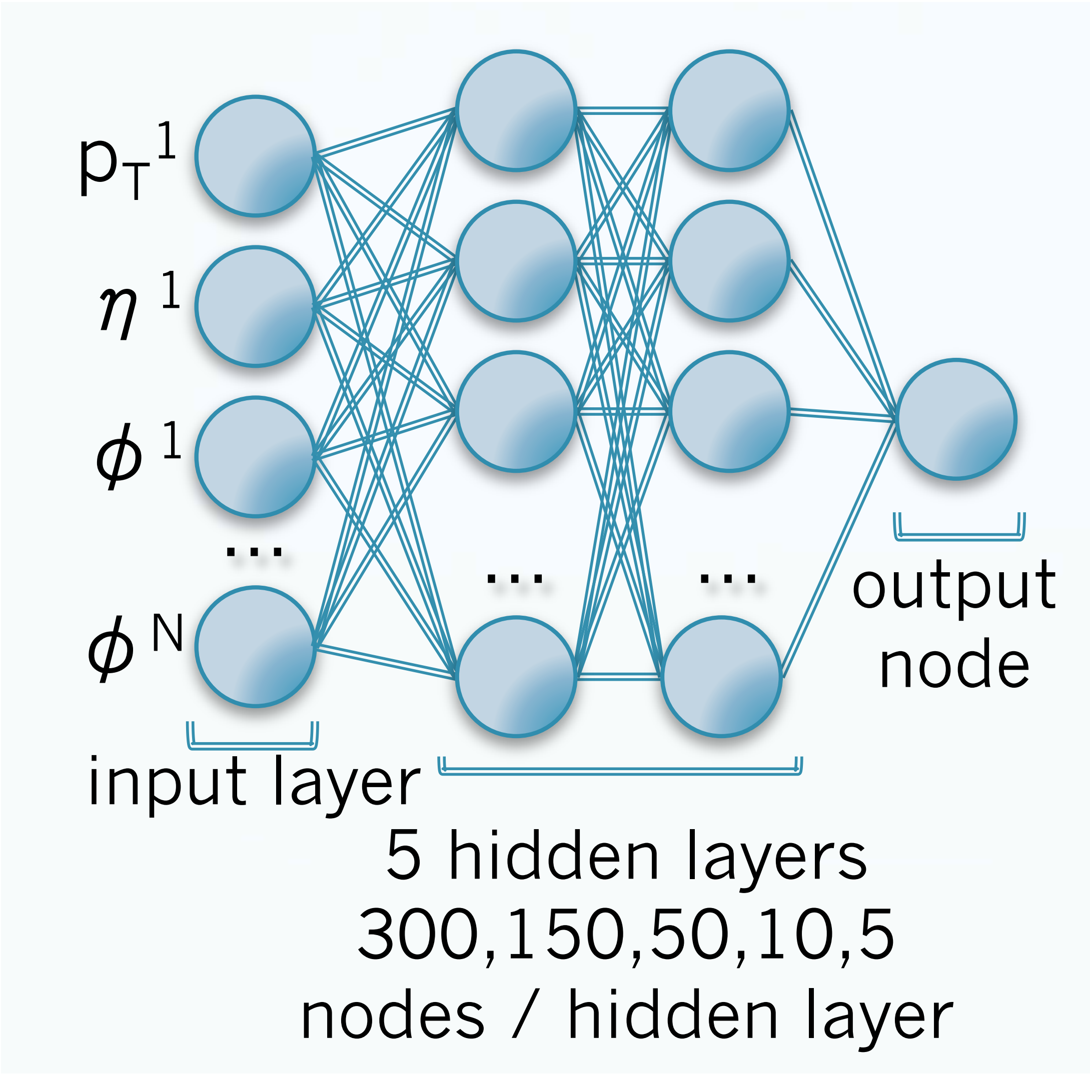
…

In this talk, I will attempt to set the stage for the rest of the session:

- Review last year's meeting
- Some selected developments since then

[Apologies if I left you out!!]

Lists of four vectors



Pros:

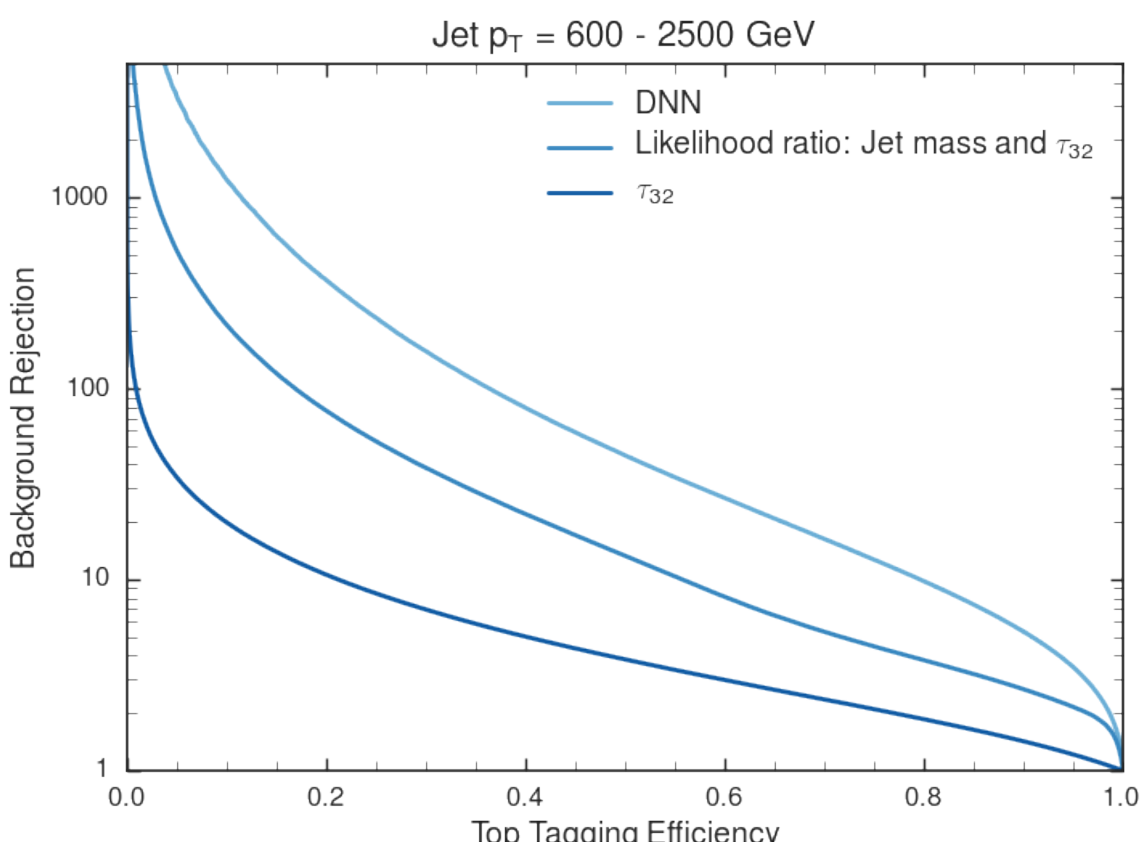
- no information loss

Cons:

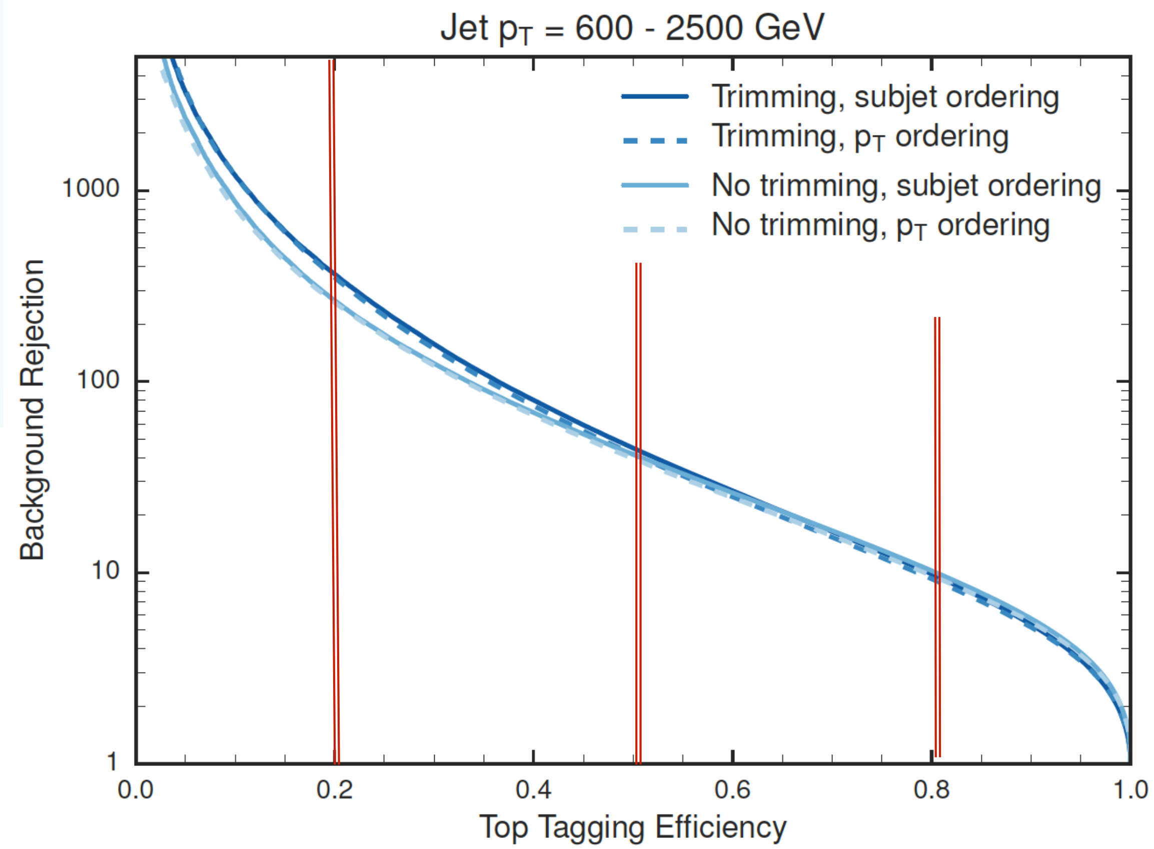
- no canonical ordering
- spatial correlations obscured
- large number of NN parameters

- Input width fixed: up to 120 constituents, 0-padded

W. Fedorko, ML4Jets 2017
1704.02124



Outperforms high-level taggers



robust against different orderings

W. Fedorko, ML4Jets 2017
1704.02124

Images

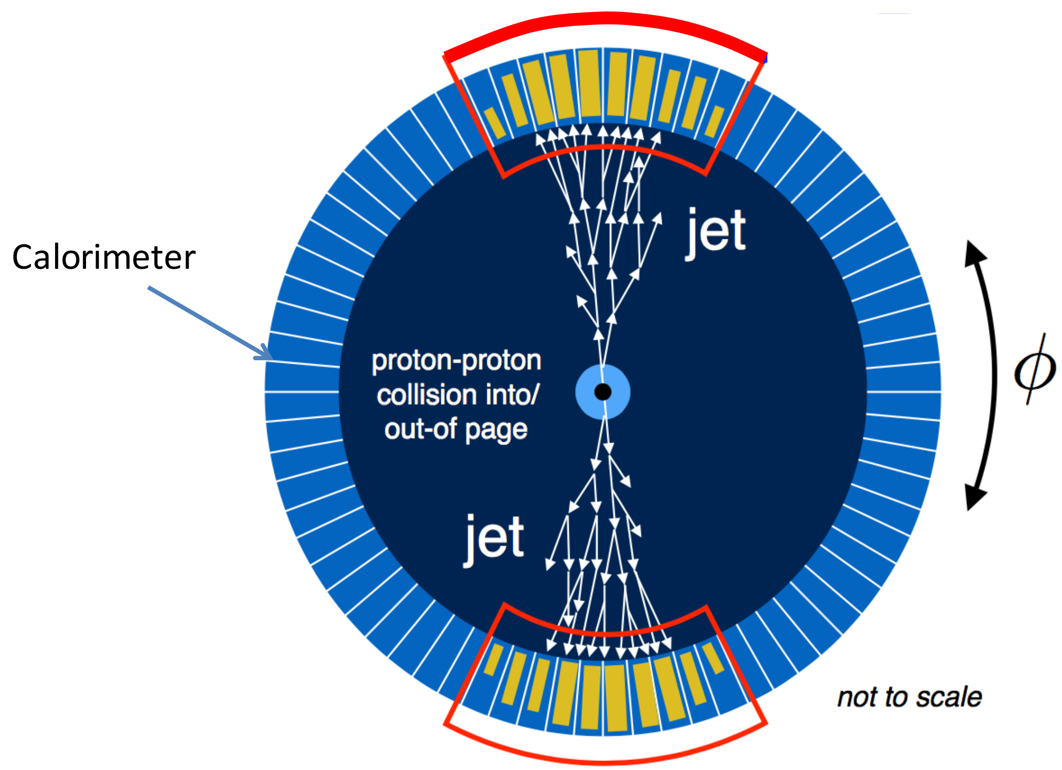
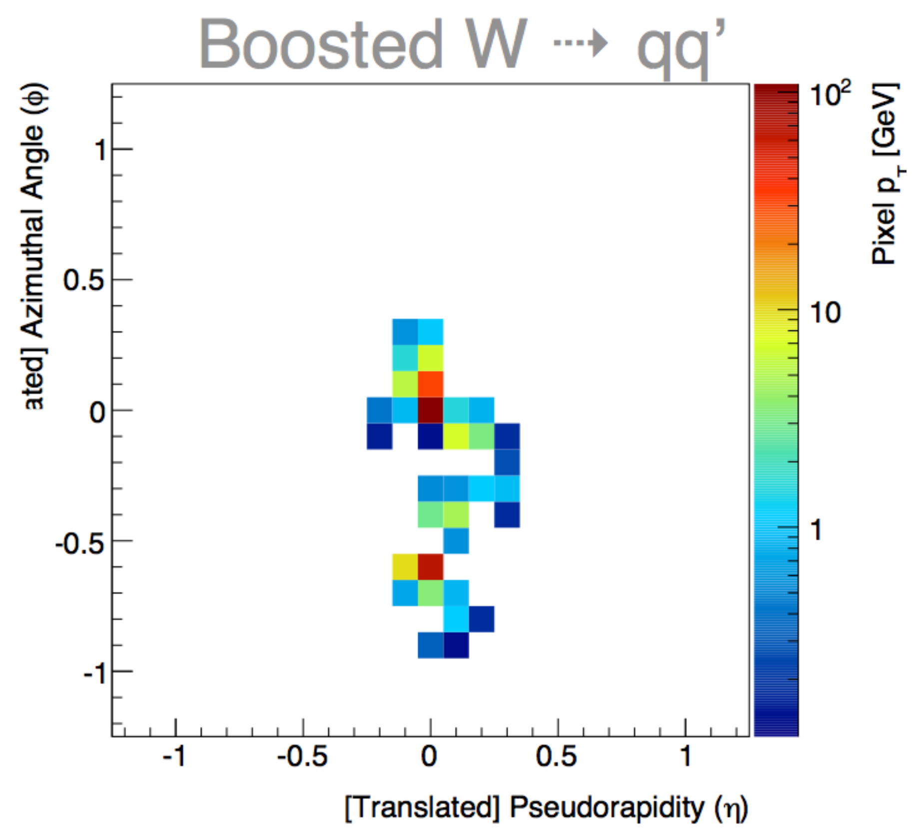


Figure credit:
B. Nachman



Can think of a jet as an image in eta and phi, with

- Pixelation provided by calorimeter towers
- Pixel intensity = pT recorded by each tower

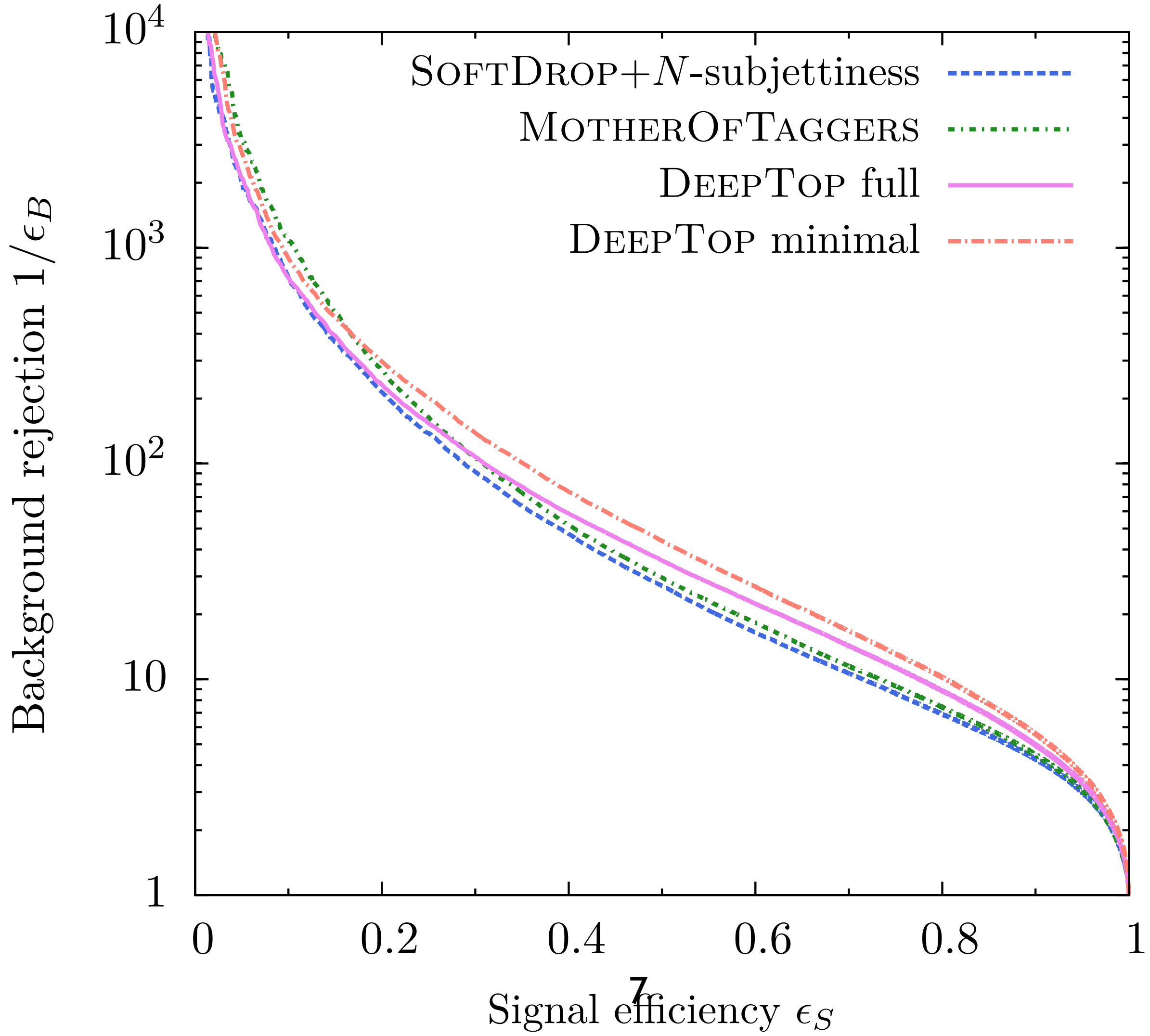
Cogan et al 1407.5675
de Oliveira et al 1511.05190

Pros:

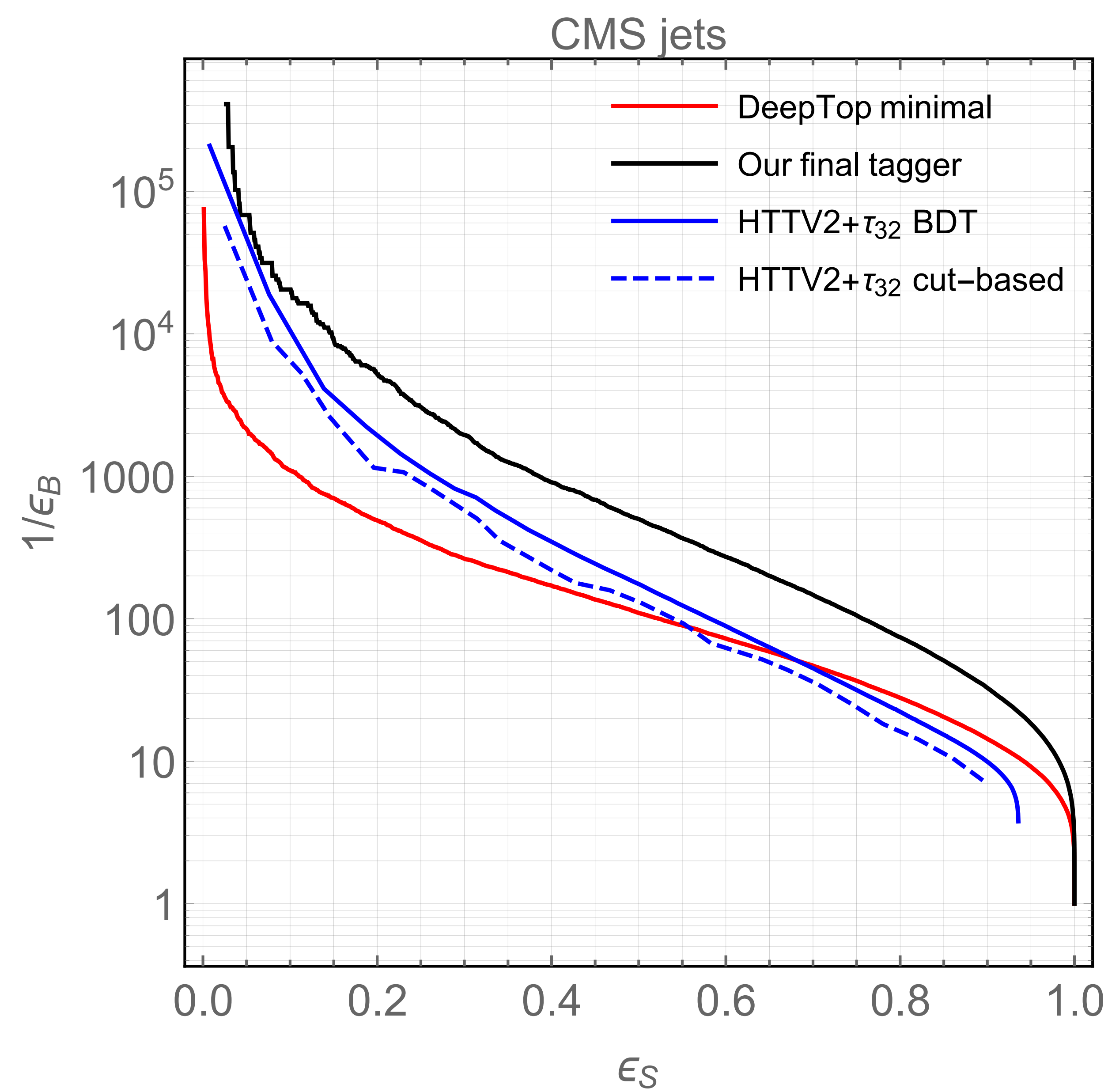
- captures spatial correlations
- fewer parameters

Cons:

- information loss (pixelation)

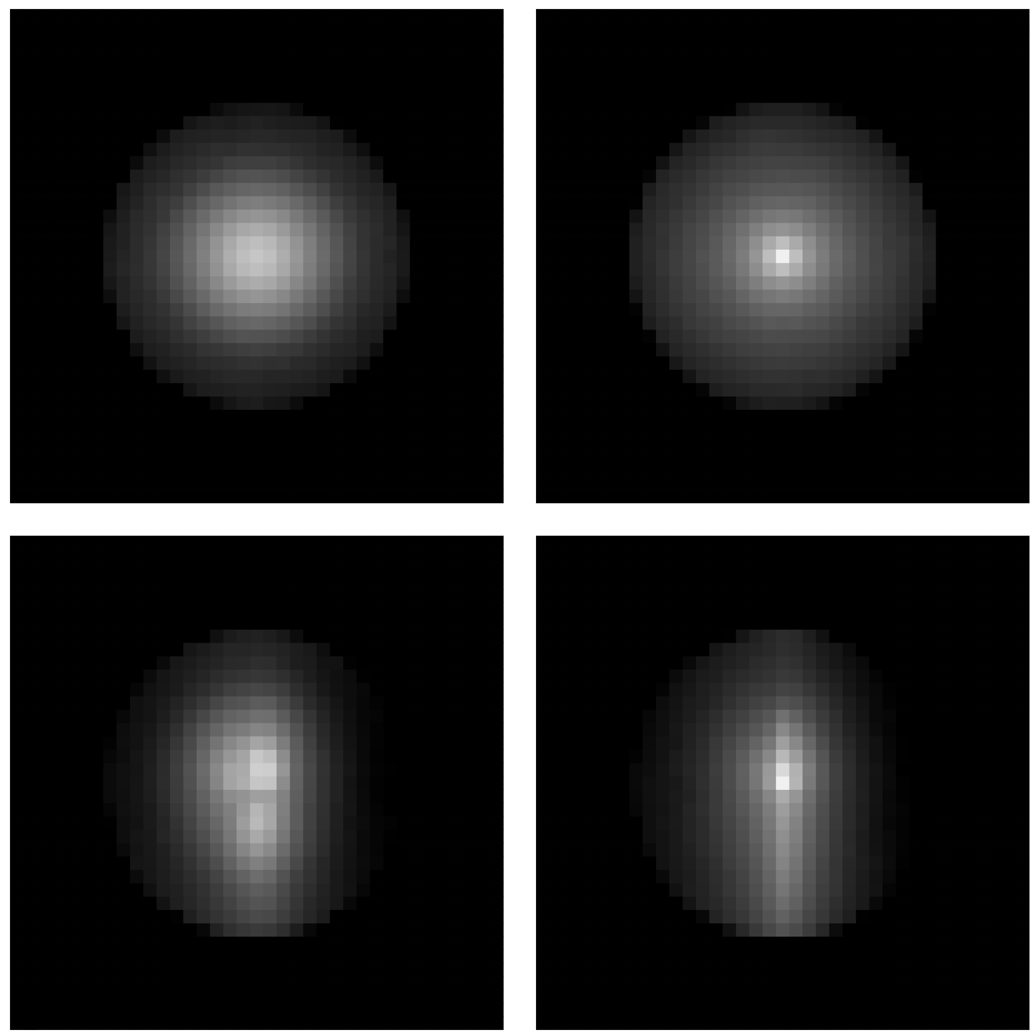


G. Kasieczka MLJets 2017
1701.08784



Big improvements since 2017
Macaluso & DS 1803.00107
see Sebastian's talk today for details

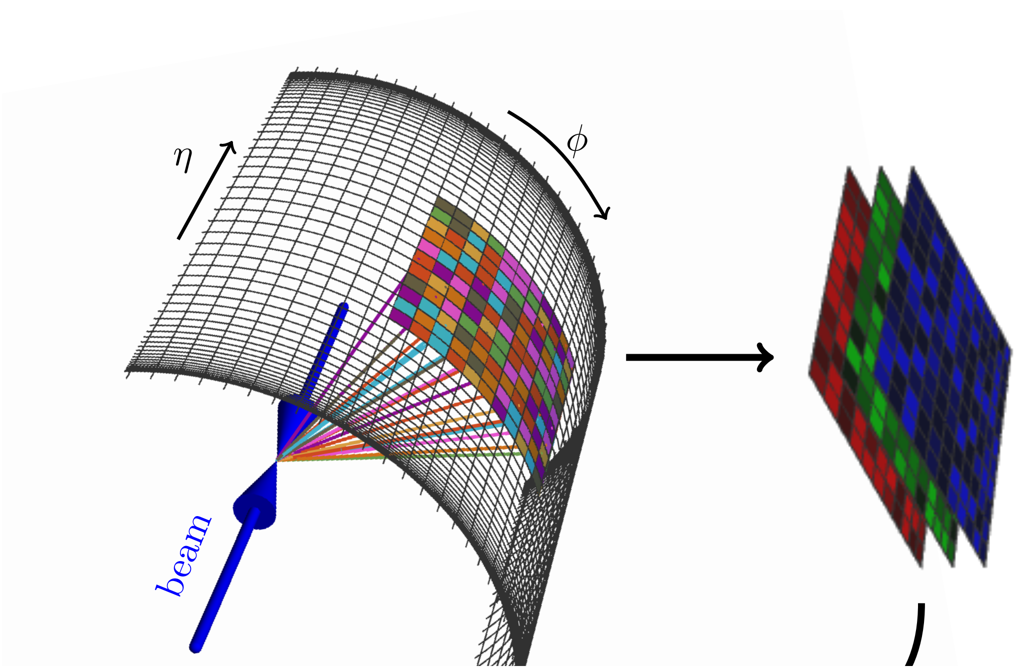
Importance of preprocessing



Average of 100k jet images

Importance of color

Can feed multiple pixel-wise features to CNN as different "colors" of the image



red = transverse momenta of charged particles

green = the transverse momenta of neutral particles

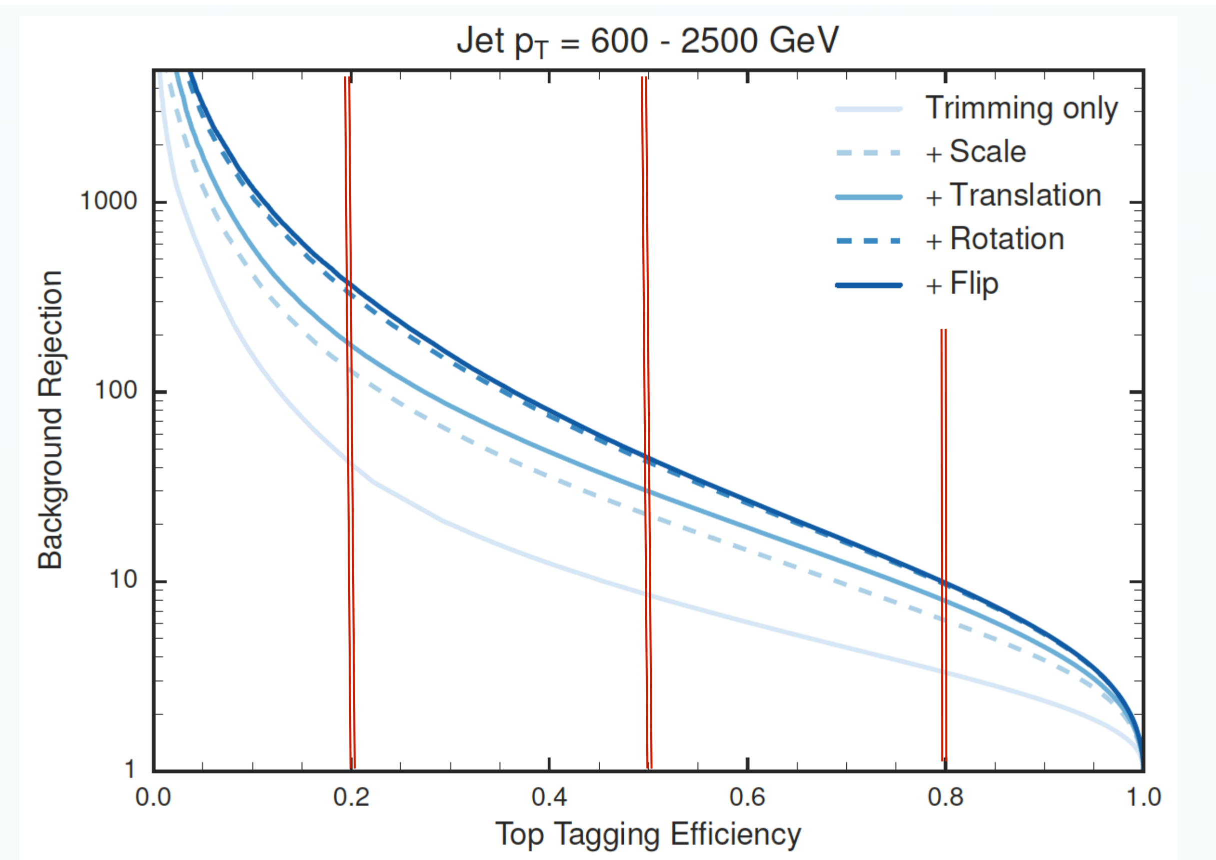
blue = charged particle multiplicity

Komiske, Metodiev & Schwartz 1612.01551
for q/g tagging

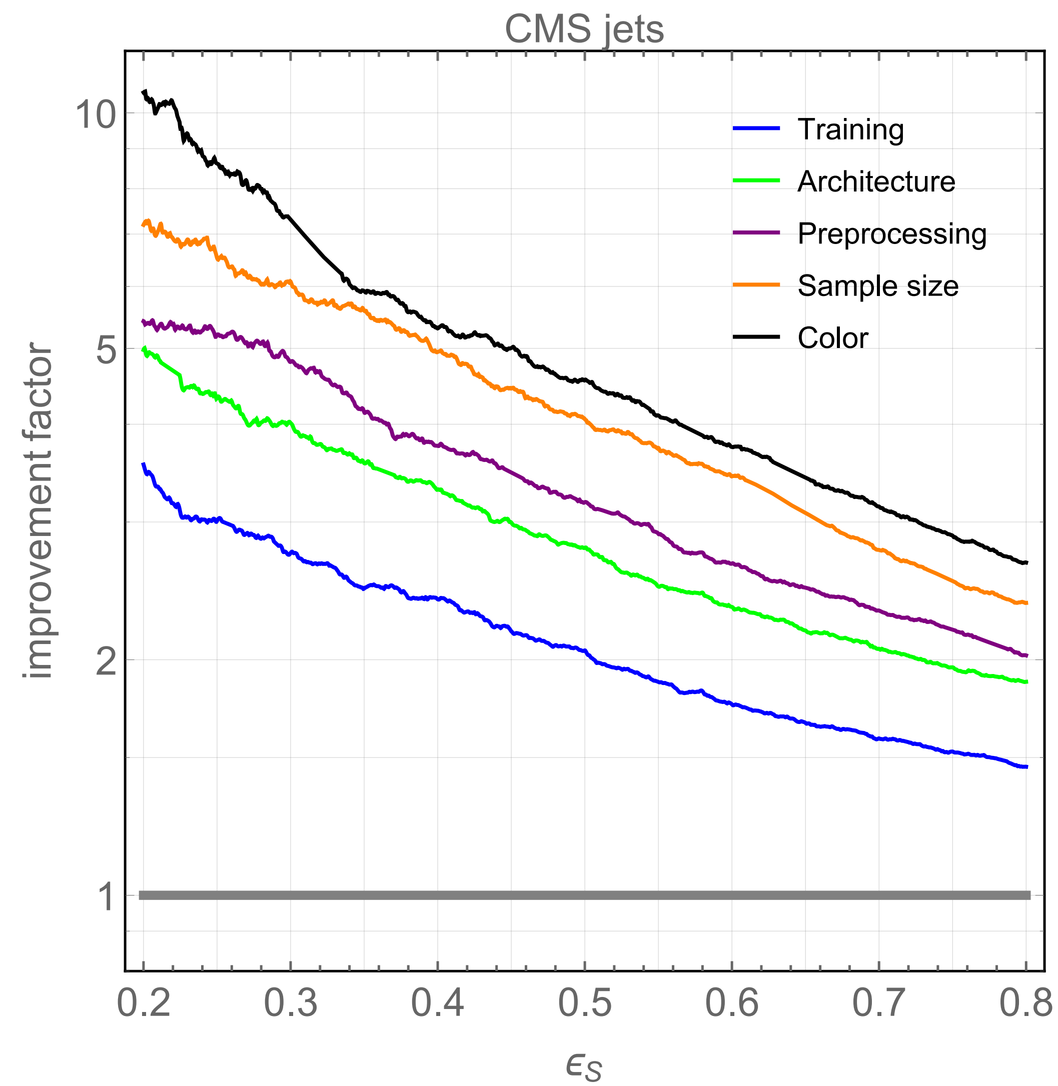
	DeepTop	CMS
Jet sample	14 TeV $p_T \in (350, 450)$ GeV, $\lvert\eta\rvert < 1$ $R = 1.5$ anti-k_T calo-only match: $\Delta R(t, j) < 1.2$ merge: NONE	13 TeV $p_T \in (800, 900)$ GeV, $\lvert\eta\rvert < 1$ $R = 1$ anti-k_T particle-flow match: $\Delta R(t, j) < 0.6$ merge: $\Delta R(t, q) < 0.6$
Image	40×40 $\Delta\eta = 4,\ \Delta\phi = \frac{10}{9}\pi$	37×37 $\Delta\eta = \Delta\phi = 3.2$
Colors	p_T^{calo}	$(p_T^{neutral},\ p_T^{track},\ N_{track},\ N_{muon})$

Table 1: The two jet image samples used in this work.

Macaluso & DS 1803.00107
top tagging



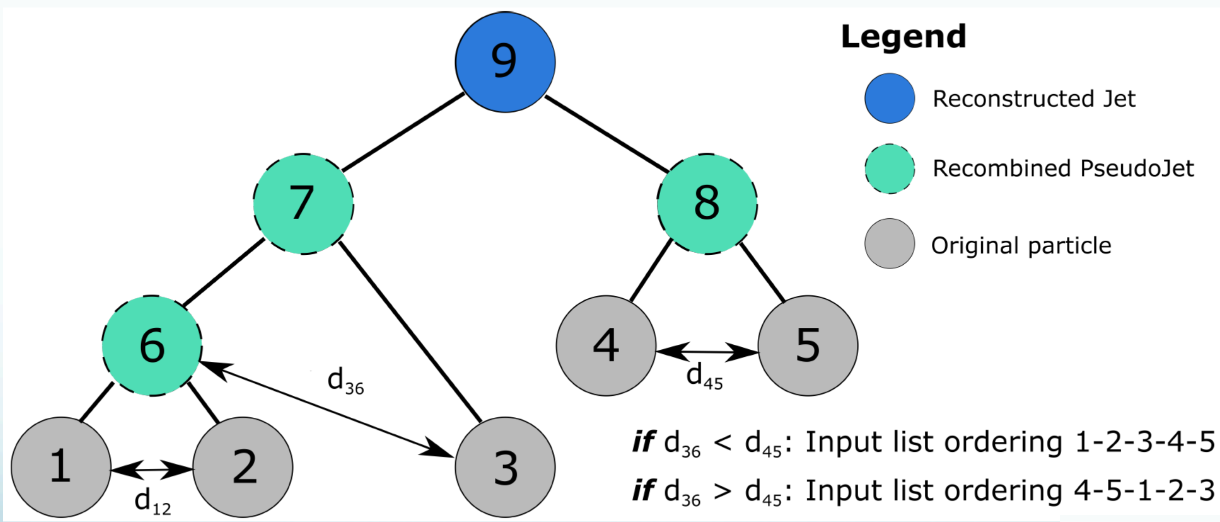
W. Fedorko, ML4Jets 2017
1704.02124



Macaluso & DS 1803.00107

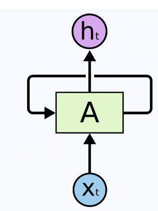
Sequences

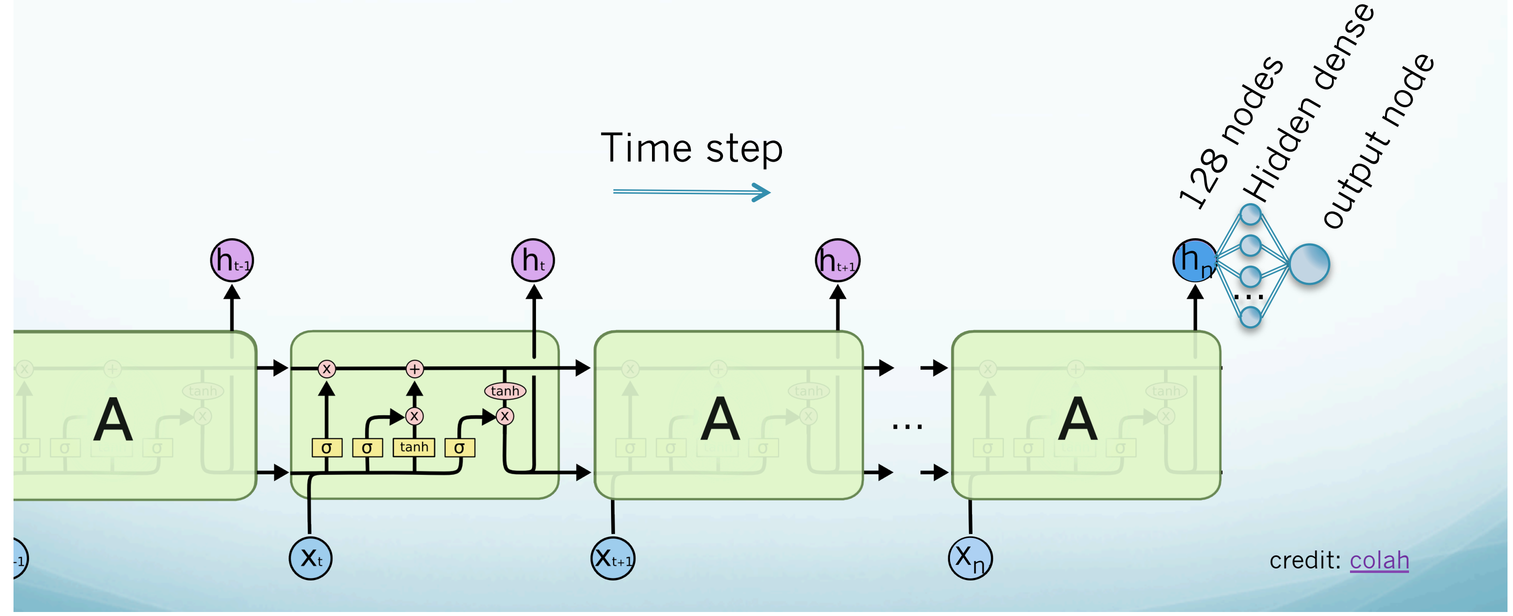
- View anti-k_T sequence as a binary tree
- Order using depth-first traversal prioritizing jets with ‘parents’ whose d_{ij} is smaller



W. Fedorko, ML4Jets 2017
1711.09059

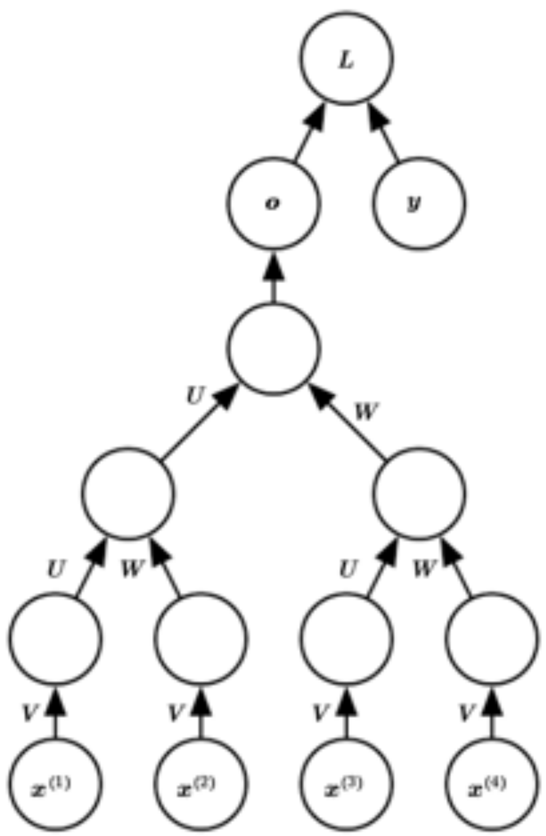
- Use a Recurrent Neural Network:
  - Long Short-Term Memory Network



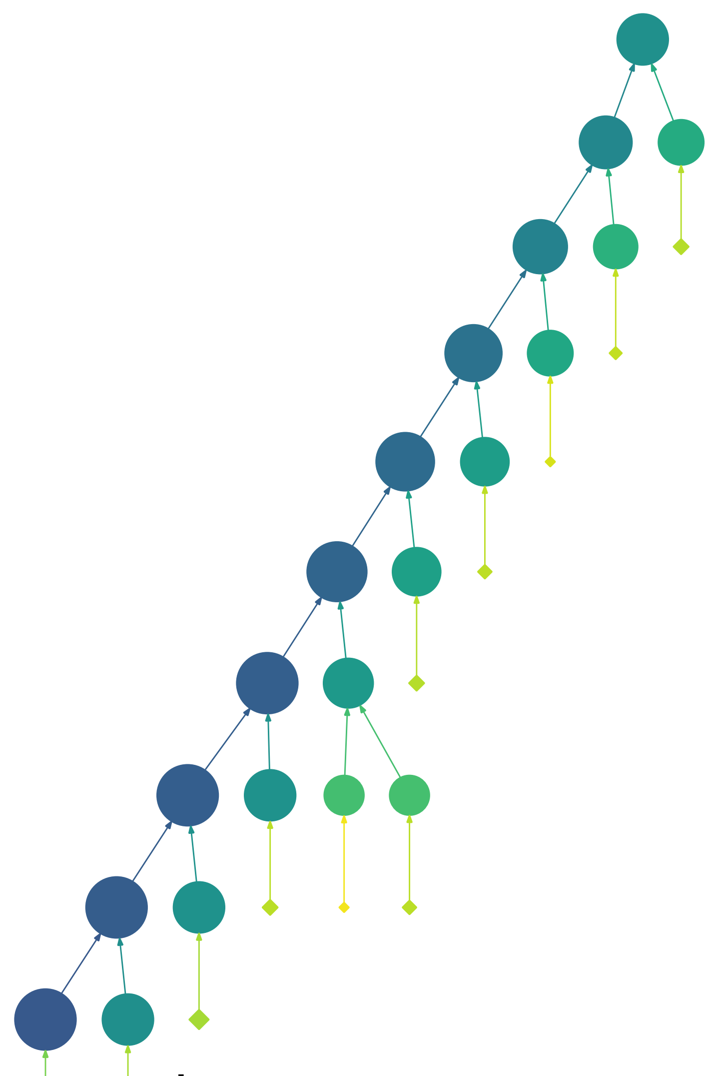


See also Andreassen & Frye’s ML4Jets 2017 talks: RNNs for parton showers 1804.09720

Trees



RecNN



jet tree structure

Use jet clustering history to build binary tree

Train recursive neural net to turn jet tree into "embedding"

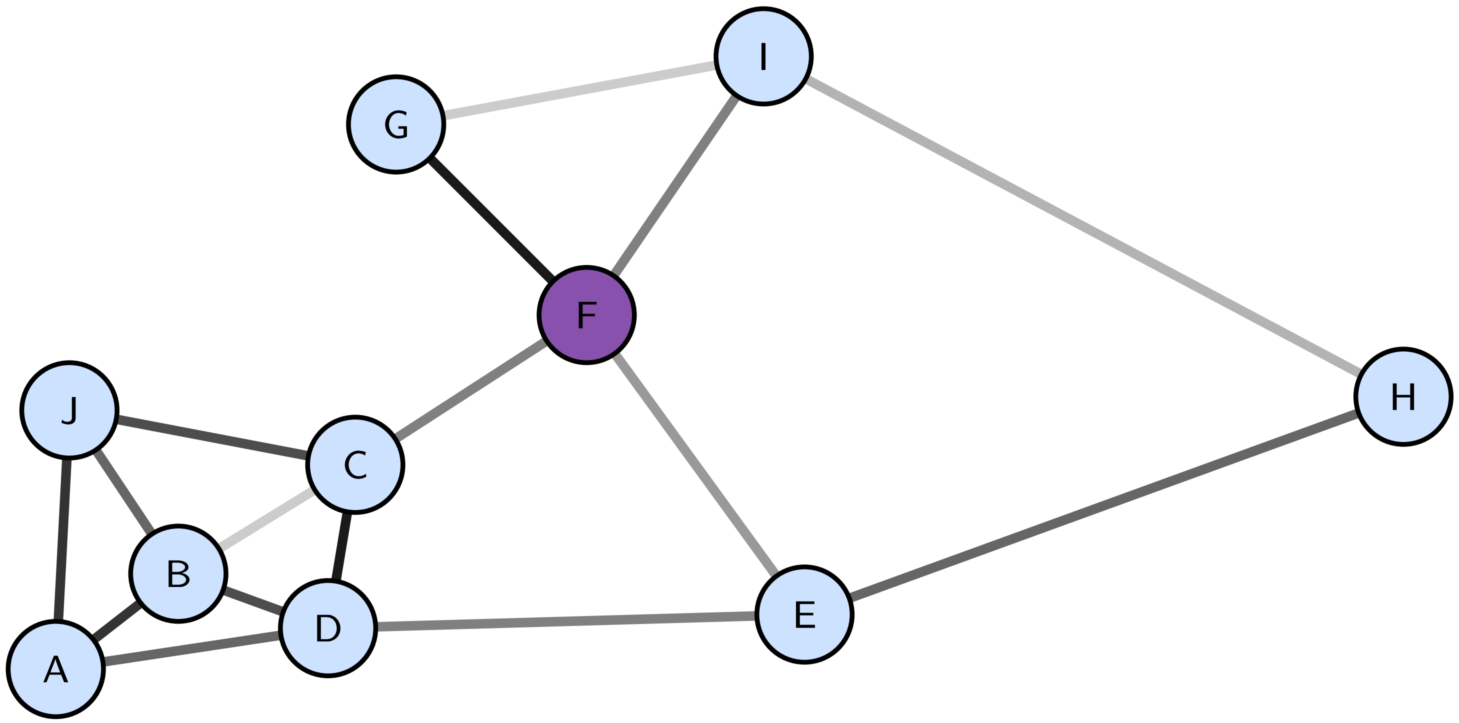
Use jet embedding for classifier, etc

Can perform comparable to other architectures with far fewer weights!

Louppe et al
1702.00748
RecNN for W tagging

T. Cheng MLJets 2017
1711.02633
applied to q/g tagging

Graphs

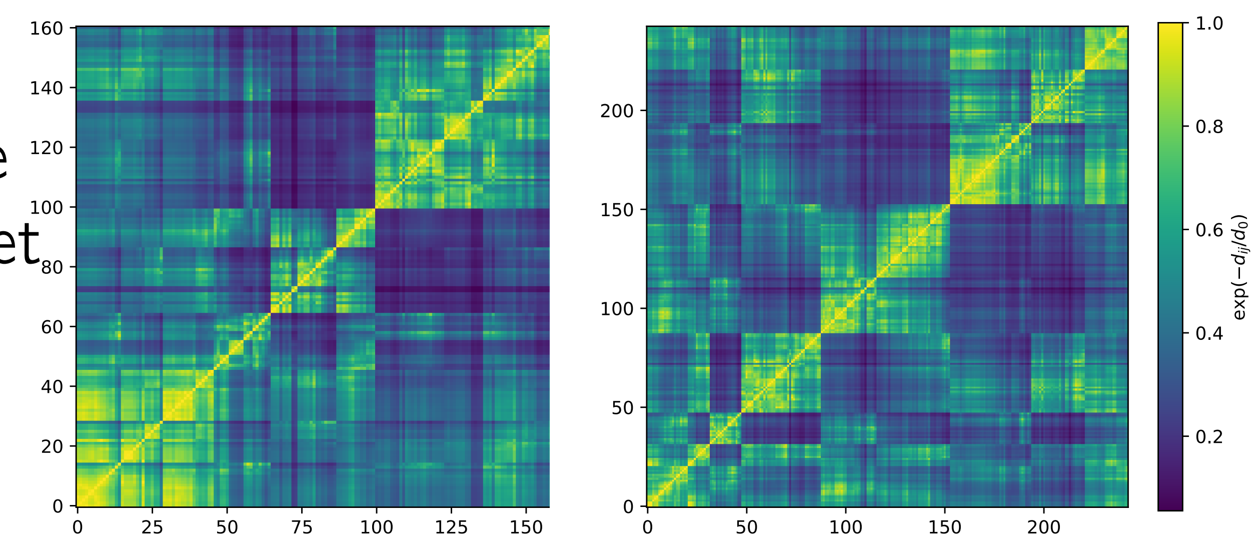


I. Henrion ML4Jets 2017

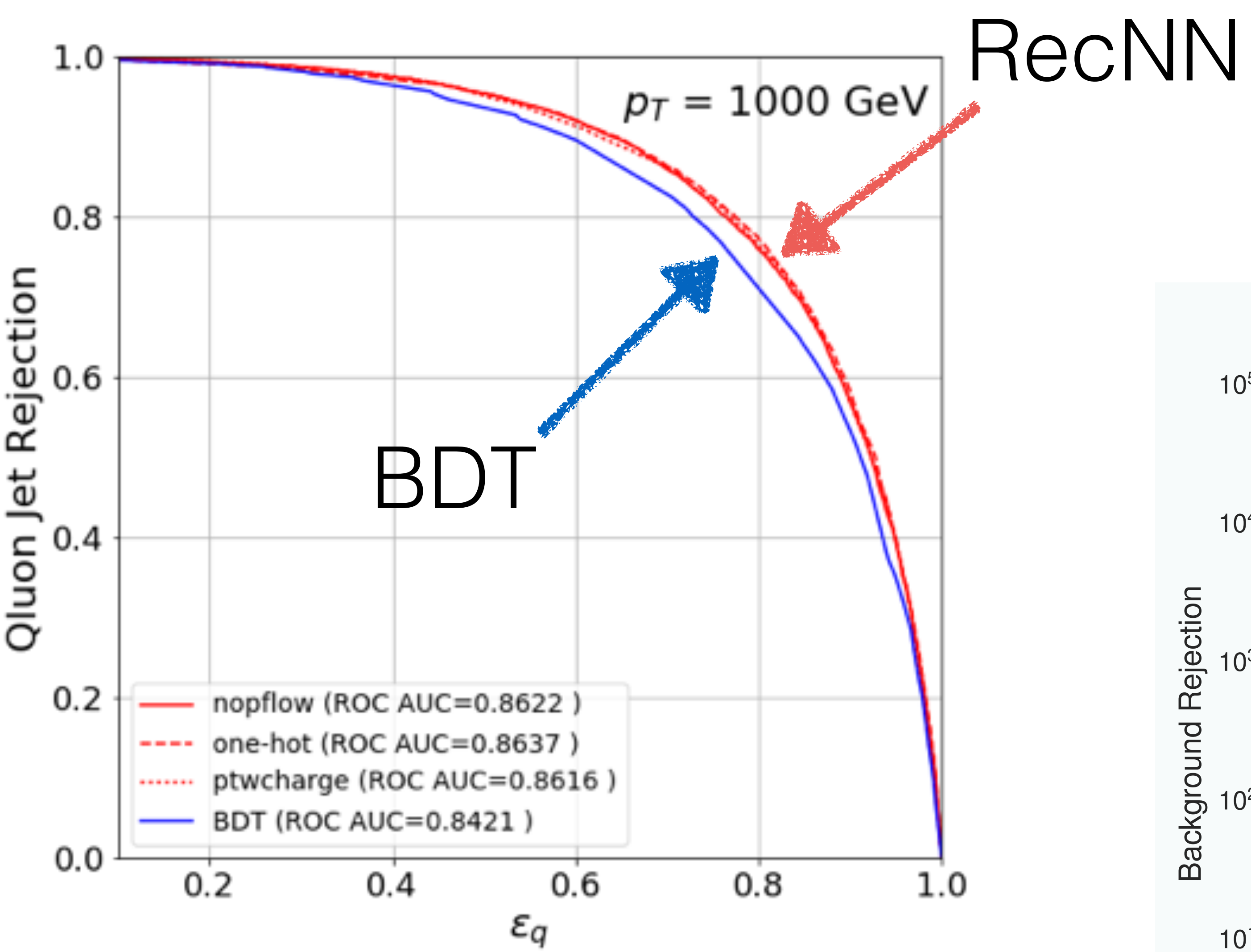
Can learn clustering history via graph adjacency matrix

$$d_{ij}^{\alpha} = \min(p_{ti}^{2\alpha}, p_{tj}^{2\alpha}) \frac{\Delta R_{ij}^2}{R^2}$$

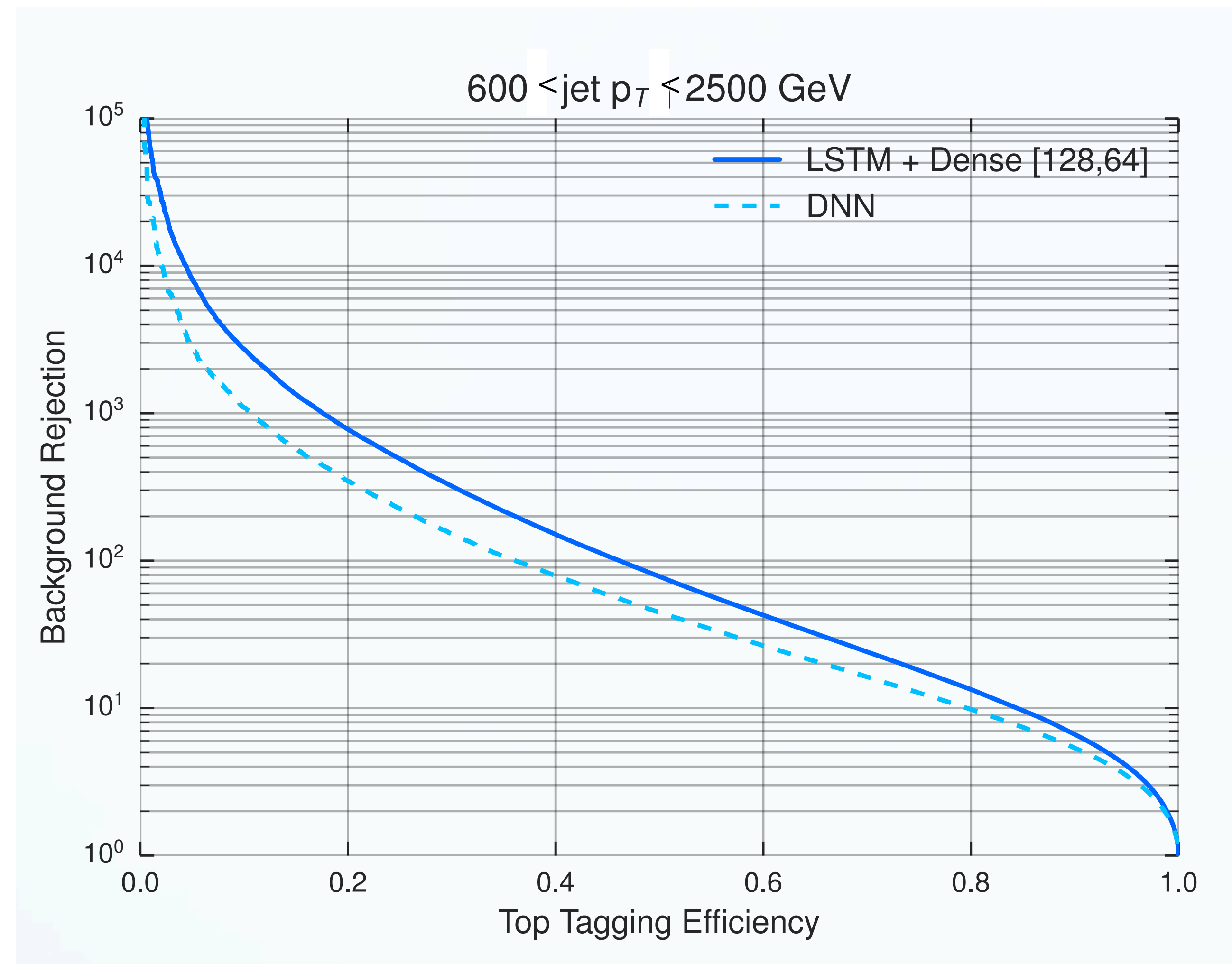
Single QCD jet



Single W jet



T. Cheng MLJets 2017
1711.02633



W. Fedorko, ML4Jets 2017
1711.09059

Kinematic invariants: LoLa

Input is a p_T sorted list of Lorentz four-vectors: (calo towers or particle flow objects)

$$k_{\mu,i} = \begin{pmatrix} E_0 & E_1 & \cdots & E_N \\ p_{x,0} & p_{x,1} & \cdots & p_{x,N} \\ p_{y,0} & p_{y,1} & \cdots & p_{y,N} \\ p_{z,0} & p_{z,1} & \cdots & p_{z,N} \end{pmatrix}$$

Combination Layer (**CoLa**): create linear combinations: $k_{\mu,i} \stackrel{\text{CoLa}}{\longrightarrow} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$

Lorentz Layer (**LoLa**): Use resulting matrix to extract physics features.
Main assumption is the Minkowski metric

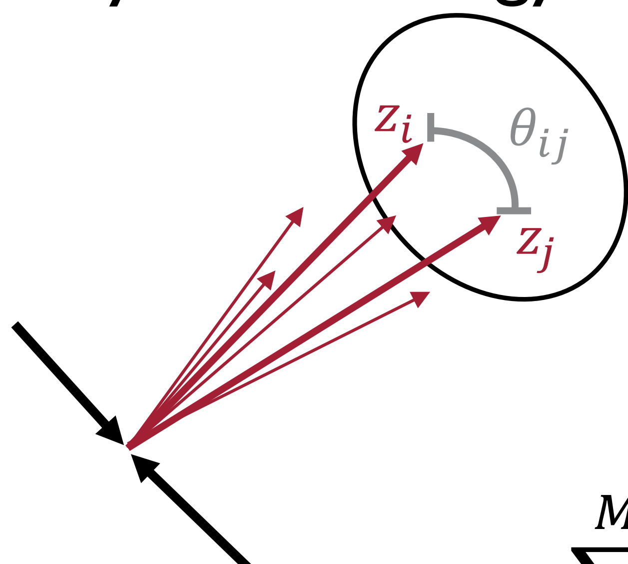
Fully connected layers for final output

$$\tilde{k}_j \stackrel{\text{LoLa}}{\longrightarrow} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w^{(E)}_{jm}\, E(\tilde{k}_m) \\ w^{(d)}_{jm}\, d^2_{jm} \end{pmatrix}$$

G. Kasieczka, ML4Jets 2017
1707.08966

Kinematic invariants: EFP

Anatomy of an Energy Flow Polynomial:



Energy Fraction | Pairwise Angular Distance

$$e^+e^-: \; z_i = \frac{E_j}{\sum_k E_k}, \quad \theta_{ij} = \left(\frac{2p_i^\mu p_{j\mu}}{E_i E_j}\right)^{\frac{\beta}{2}}$$

$$\text{Hadronic:} \; z_i = \frac{p_{Tj}}{\sum_k p_{Tk}}, \quad \theta_{ij} = \left(\Delta y_{ij}^2 + \Delta\phi_{ij}^2\right)^{\frac{\beta}{2}}$$

In equations:

$$\mathrm{EFP}_{\mathrm{G}} = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} z_{i_2} \cdots z_{i_N} \prod_{(k,l)\in \mathrm{G}} \theta_{i_k i_l}$$

multigraph

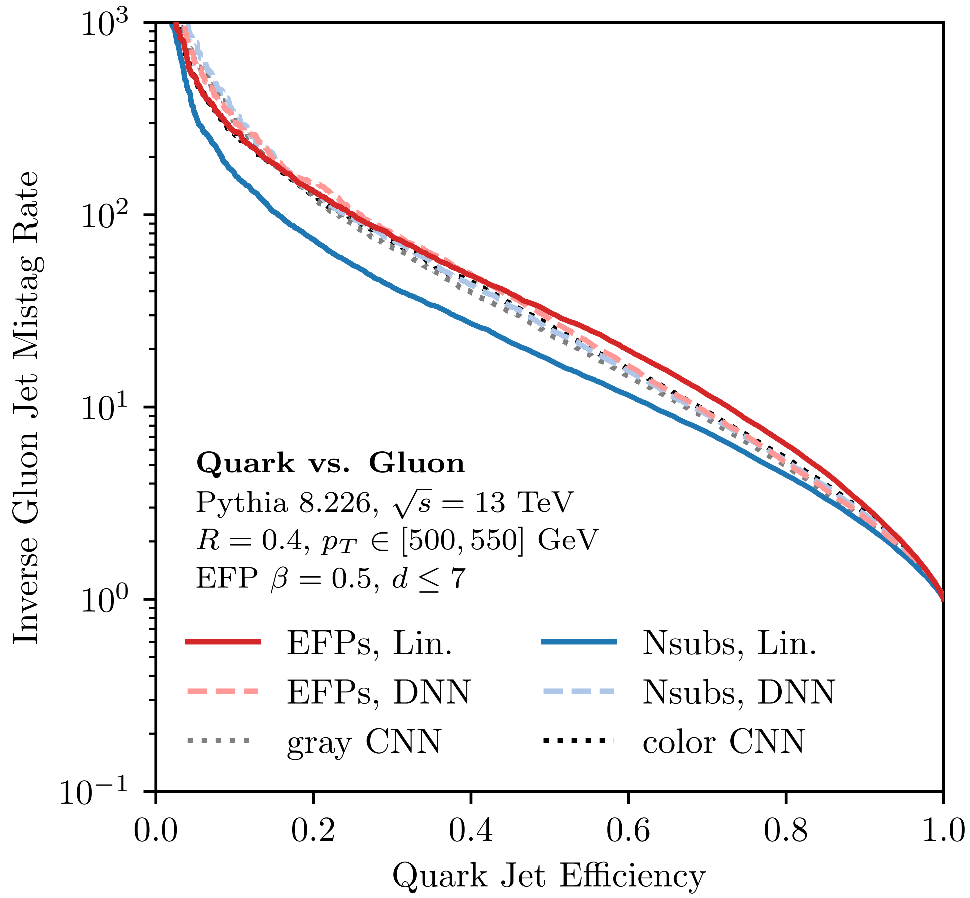
In words: **Correlator** of **Energies** and **Angles**

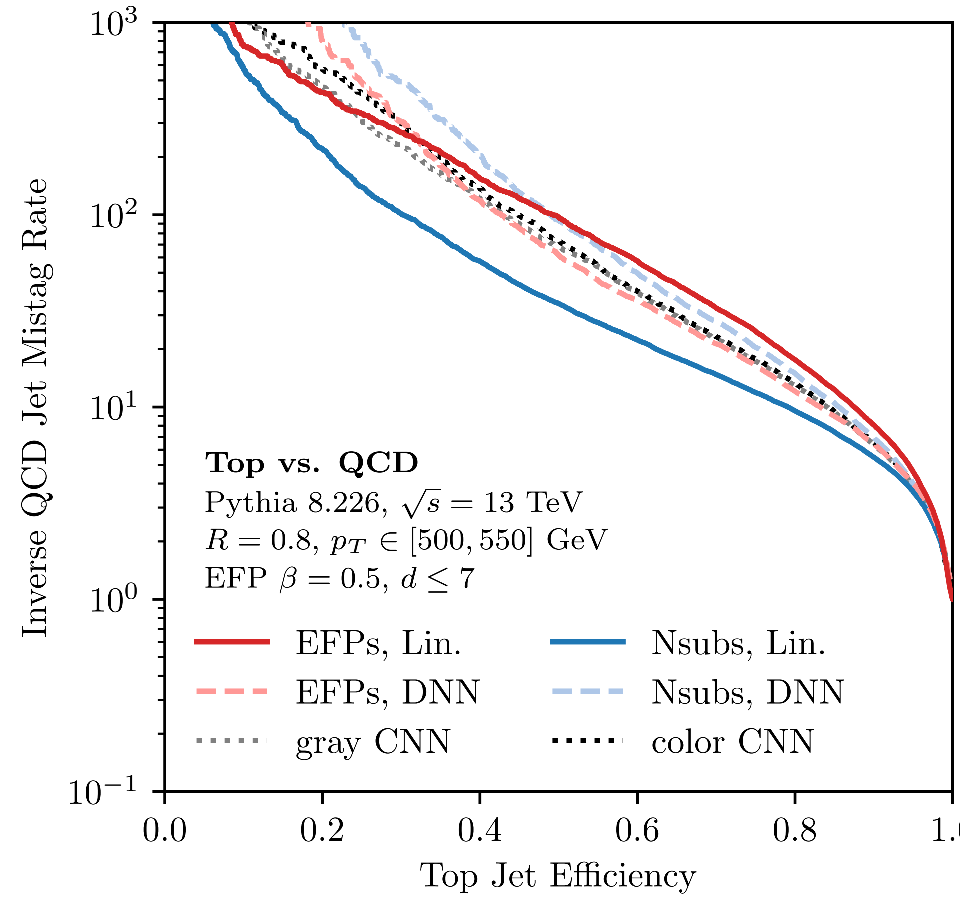
Correlator: Sum over all N-tuples of particle in the event

Energies: Product of the N energy fractions

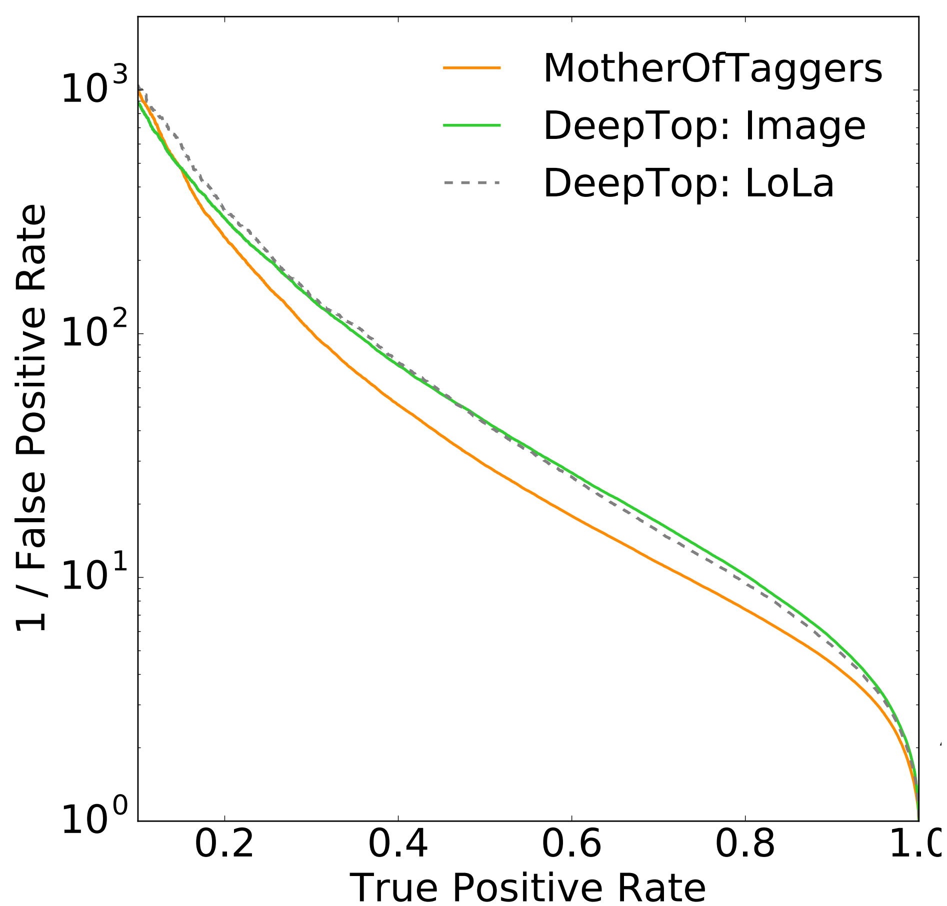
Angles: One $\theta_{i_k i_l}$ for each edge in $(k,l) \in G$

E. Metodiev ML4Jets 2017
1712.07124



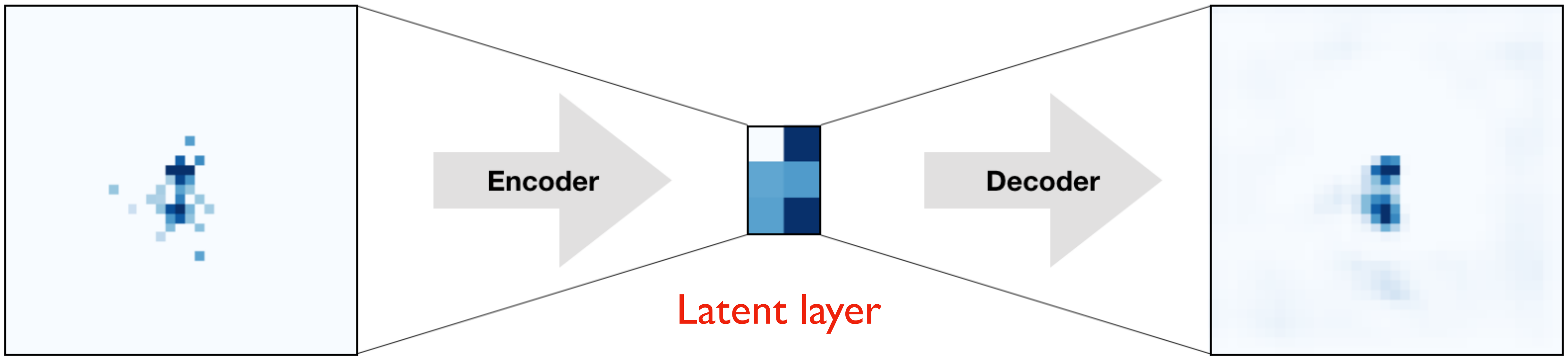


P. Komiske ML4Jets 2017
1712.07124



G. Kasieczka, ML4Jets 2017
1707.08966

Latent space representation



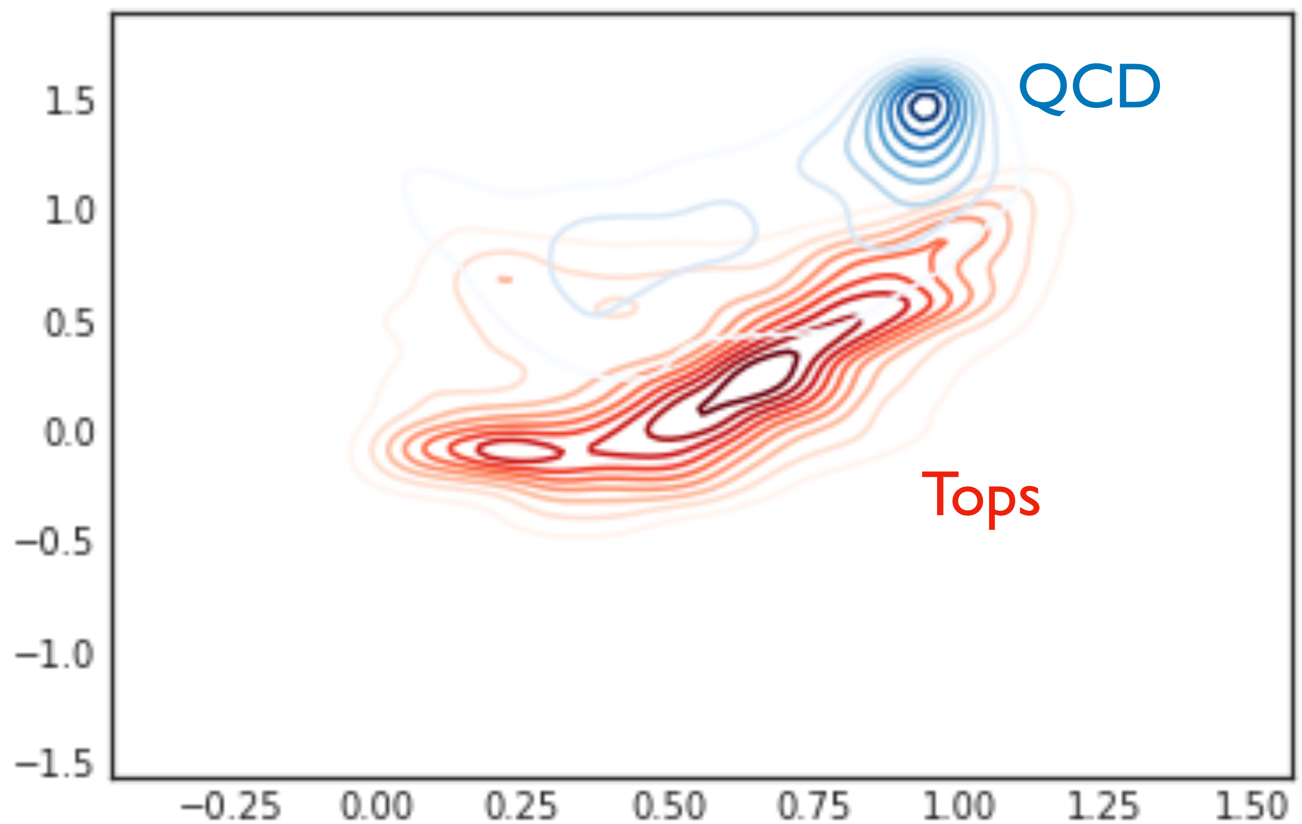
An autoencoder compresses an input into a "latent representation" and then attempts to reconstruct the original input.

(See talks by Gregor, Marco and Tao on Friday for details!)

Can explore the meaning of this latent representation…

Hajer et al 1807.10261
Heimel et al 1808.08979
Farina, Nakai & DS 1808.08992

Latent space representation



Signal and background cluster in latent space!

Simple autoencoder — no metric on latent space…meaning of latent space variables far from clear…

Can explore variational AE, adversarial AE, …

Does the representation even matter???

Notion: all these representations have comparable performance, if you optimize architecture and train enough.

Not a huge amount of hard evidence for this. Need more apples-to-apples comparisons!

A good start: a community top-tagger comparison

(see Gregor's talk yesterday; https://goo.gl/XGYju3)

Approach	AUC	Acc.	1/eB (@ eS=0.3)	Contact	Comments
LoLa	**0.979**	**0.928**		**GK / SImon Leiss**	**Preliminary number, based on LoLa**
LBN	**0.981**	**0.931**	**863**	**Marcel Rieger**	**Preliminary number**
CNN	**0.981**	**0.93**	**780**	**David Shih**	**Model from *Pulling Out All the Tops with Computer Vision and Deep Learning (1803.00107)***
P-CNN (1D CNN)	**0.980**	**0.930**	**782**	**Huilin Qu, Loukas Gouskos**	**Preliminary, use kinematic info only (https://indico.physics.lbl.gov/indico/event/546/contributions/1270/)**
6-body N-subjettiness (+mass and pT) NN	**0.979**	**0.922**	**856**	**Karl Nordstrom**	**Based on 1807.04769 (*Reports of My Demise Are Greatly Exaggerated: N-subjettiness Taggers Take On Jet Images*)**
8-body N-subjettiness (+mass and pT) NN	0.980	**0.928**	**795**	**Karl Nordstrom**	**Based on 1807.04769 (*Reports of My Demise Are Greatly Exaggerated: N-subjettiness Taggers Take On Jet Images*)**
Linear EFPs	0.980	0.932	**380**	Patrick Komiske, Eric Metodiev	d<= 7, chi <= 3 EFPs with FLD. Based on 1712.07124: *Energy Flow Polynomials: A complete linear basis for jet substructure.*
Particle Flow Network (PFN)	0.982	0.932	888	Patrick Komiske, Eric Metodiev	Median over ten trainings. Based on Table 5 in 1810.05165: *Energy Flow Networks: Deep Sets for Particle Jets.*
Energy Flow Network (EFN)	0.979	0.927	619	Patrick Komiske, Eric Metodiev	Median over ten trainings. Based on Table 5 in 1810.05165: *Energy Flow Networks: Deep Sets for Particle Jets.*

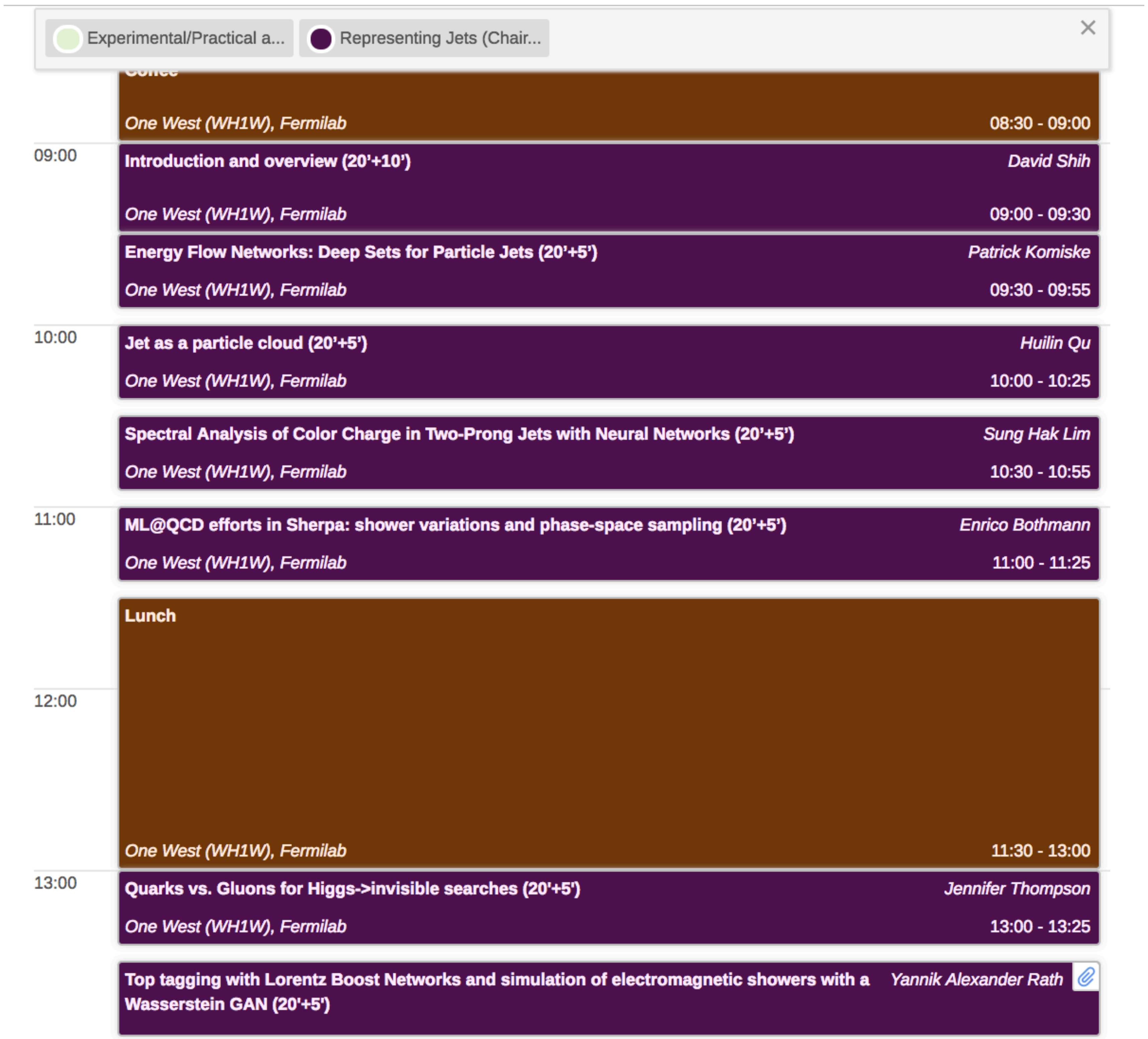
Does the representation even matter???

I think it is too soon to conclude that all representations are equivalent.

We are just starting to explore their relative strengths and weaknesses.

- Is top tagging too easy?
- ROC curve is not everything!!
- Correlations with jet mass and other variables?
- Does classifier sculpt jet mass distribution?
- Scale factors in actual data?
- Systematic uncertainties in actual data?

Outlook



Many new ideas being explored at this years workshop.

Even more exciting times to come!

Thanks for your attention!