Introduction

- Many measurements of the Standard Model and searches for new physics signatures involve studying top quarks and vector bosons in hadronic final states
- Quarks from the top & boson hadronic decays form showers of particles – jets
- Jets from decay more collimated at high energy – angular separation $\approx \frac{1}{\eta}$
- Reconstruct jets with large radius ($R$) to capture decay products

> Boosted tagging – use of large-$R$ jet substructure to discriminate tops, W’s, etc. from multijet background – light-quark/gluon jets
- Energy patterns in the jet, multi-prong structure
- Top(W) jets: 2(2) prongs
- Light-quark/gluon jets: 1 prong with (typically) soft, wide-angle emissions
- Tagging algorithms optimizing using Monte Carlo (MC) simulation.

Data/MC comparisons of jet substructure

Data/MC measurement of signal efficiency and background rejection

Tagging algorithms investigated

- "Simple" taggers – selection on two substructure observables (high-level features)
- Likelihood-based approach – Shower Deconstruction (matrix element method)
- Probability of signal or background scenario based on a simplified showering model
- Shape-based boosted decision tree & deep neural network – extending approach of "simple" taggers, using multiple high-level features and their correlations
- Topocluster-based deep neural network – using fixed number of low-level features; momentum vectors of 10 highest-$p_T$ jet constituent clusters

Boosted decision tree (BDT)

Deep neural network (DNN)

- Gradient boosting technique with bagging
- 500 trees in forest used, depth $< 20$
- 10(15) inputs for top(W) tagging
- Using TMVA

- Fully-connected feed-forward network, up to 5 hidden layers, up to 18 nodes in layer
- Adam optimizer, rectified linear unit activation function
- 13(12) inputs for top(W) tagging
- Using Keras with Theano as backend

Training sample of jets flat in $p_T$ for both ML techniques

Topocluster-based deep neural network (TopoDNN)

- $(p_T, \eta, \phi)$ of 10 leading-$p_T$ clusters, reduced dimensionality by rotating and flipping cluster vectors (rotational symmetry of the jet constituents)
- 4 hidden layers with up to 300 nodes
- Training sample jet $p_T$ distribution signal-like (sub-sampling for light-quark/gluon jets)

Performance in MC simulation

- Machine learning (ML)-based algorithms outperform other investigated taggers
- TopoDNN at high $p_T$ outperforms ML taggers using high-level features
- Larger performance gains by ML taggers for top jets – more distinct decay structure

Signal efficiency measurement

- Using boosted $R$ events with single lepton in final state
- Tag-and-probe approach, very pure $Jt$ sample using leptonic top signature selection, probe the hadronically-decaying top

Top jet ($W$) jet selection – require small-$R$ $b$-jet inside (outside) large-$R$ jet cone

Signal efficiency definition in MC and data:

- Data/MC comparisons of jet substructure
- Many measurements of the Standard Model and searches for new physics signatures

Background rejection measurement

- Gluon-enriched mixture of jets at low $p_T$
- Light-quark jets dominant contribution

Background rejection in MC and data:

- 1/Background rejection measurement in MC (top) – $W$ (top)
- Background rejection measurement in data (top) – $W$ (top)

Dijet events

- Background rejection in MC and data:

References