A Boosted Event Shape Tagger for Heavy Object Classification

Justin Pilot, John Conway, Robin Erbacher, Ramya Bhaskar, Maia Paddock,
on behalf of the CMS Collaboration
University of California, Davis

APS Meeting of the Division of Particles and Fields
Fermilab
1 August 2017

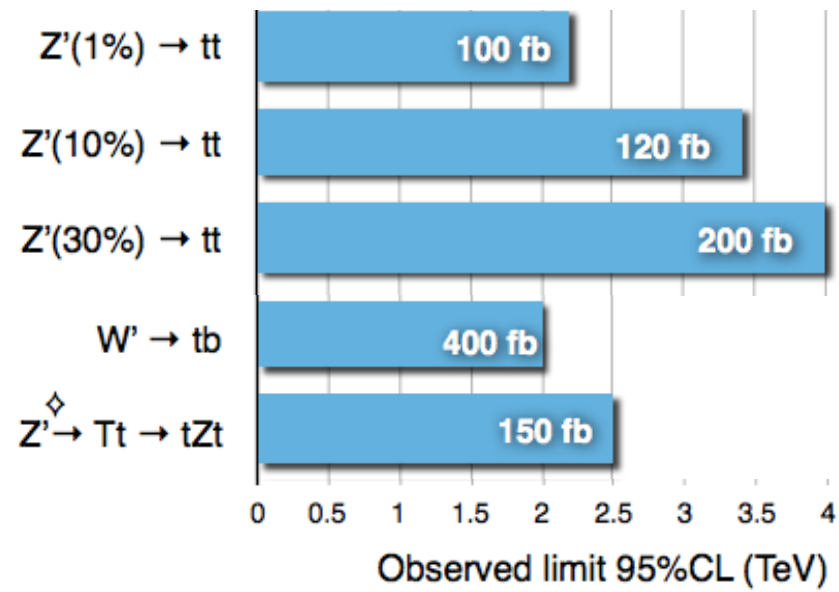
Introduction

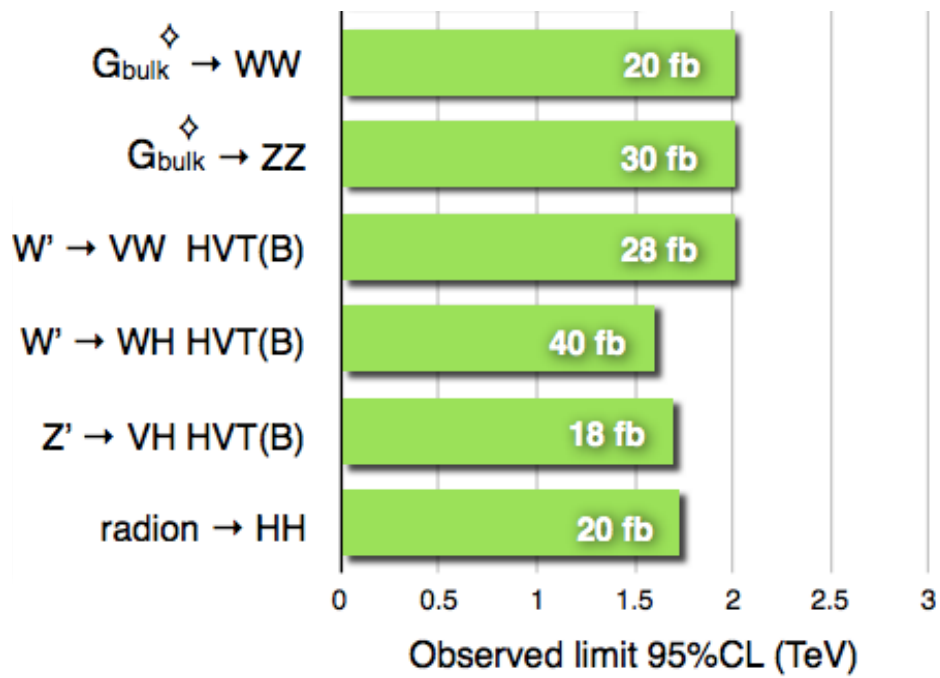
- Searches for new physics at the LHC continue to push exclusion limits higher and higher

- High resonance mass → high energy, collimated SM decay products

- **Jet substructure** techniques critical to maintain search sensitivity to these unique signatures
  - Identify hadronic decays of high-p_T, heavy SM objects
    - Top quarks, W, Z, H bosons

  - Already prevalent in CMS analyses, becoming more widespread
    - Also powerful for pileup mitigation

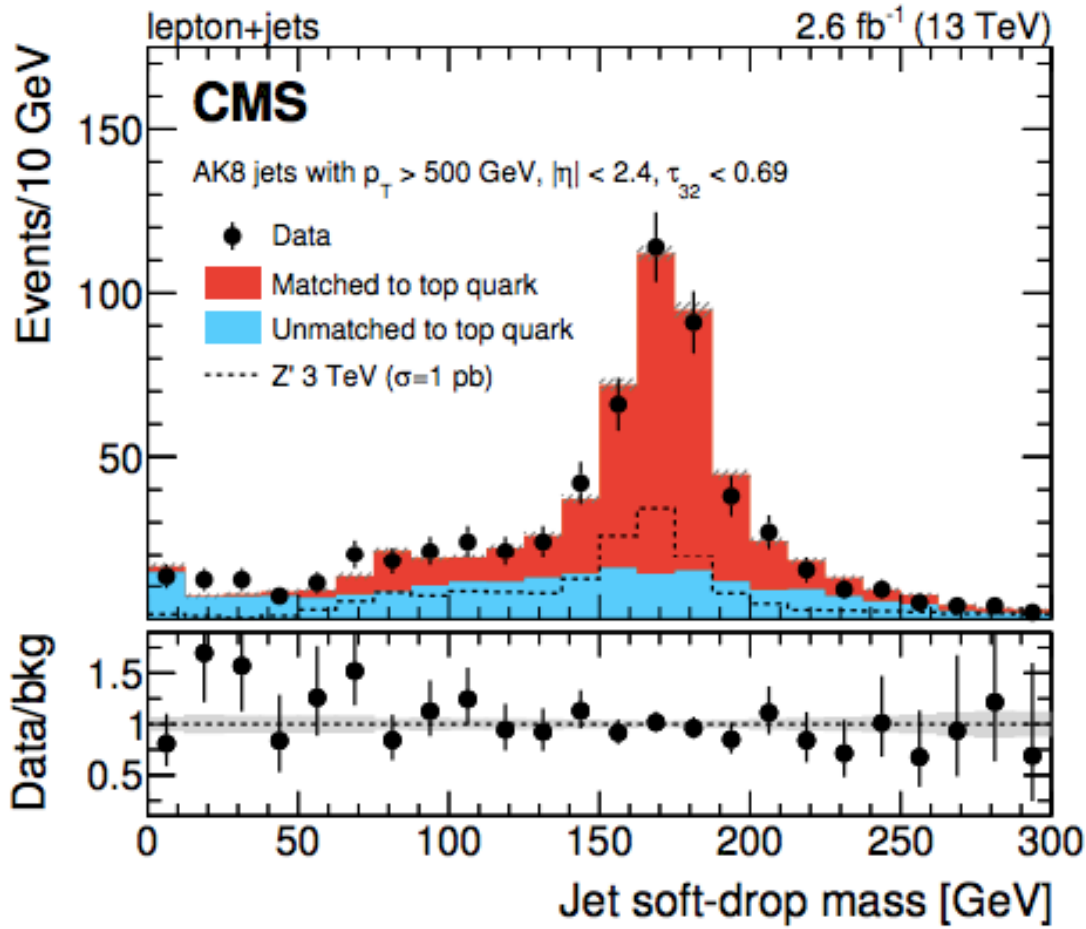




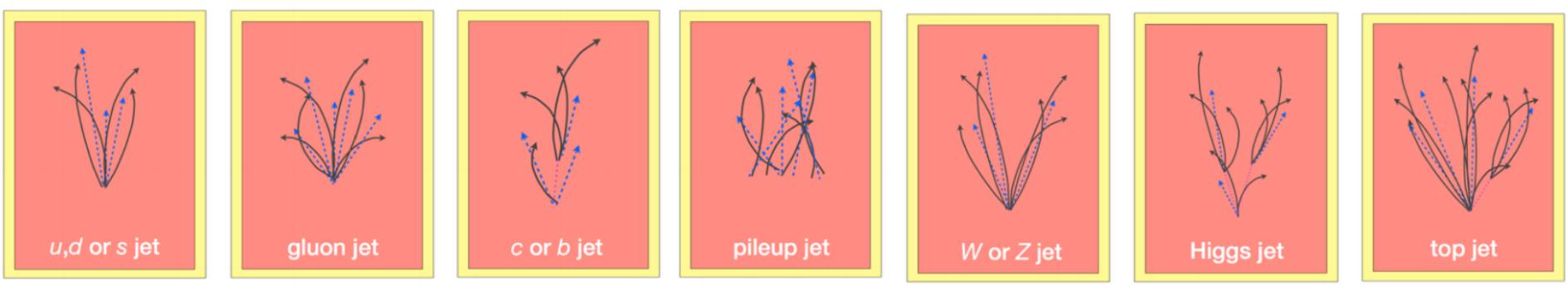
Jet Substructure

JHEP 07 (2017) 001 / B2G-16-015

- **Jet substructure signatures are discovery signatures!**
  - Use information within a jet to identify hadronic decays of heavy objects
    - Choose large jet cone size to capture all decay products
    - R = 0.8 →
      - Merged W at p_T = 200 GeV
      - Merged top at p_T = 400 GeV
    - "Top jet", "W jet", etc.



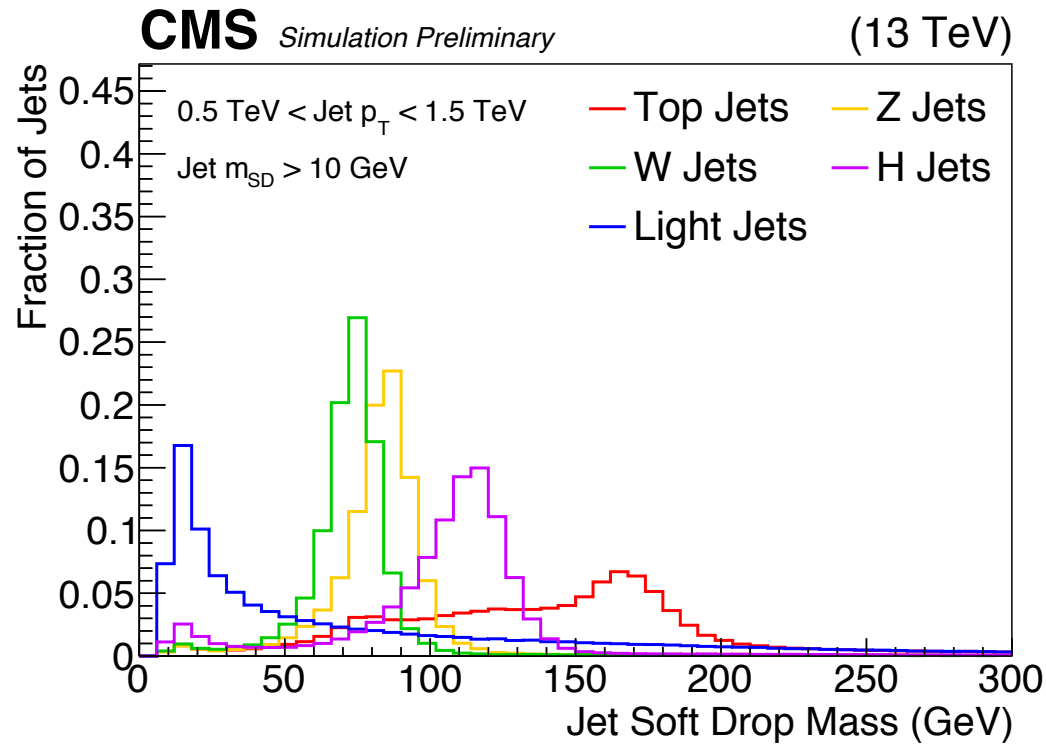
- Several algorithms in use in CMS targeting individual particles
  - Well-validated in data



N. Tran

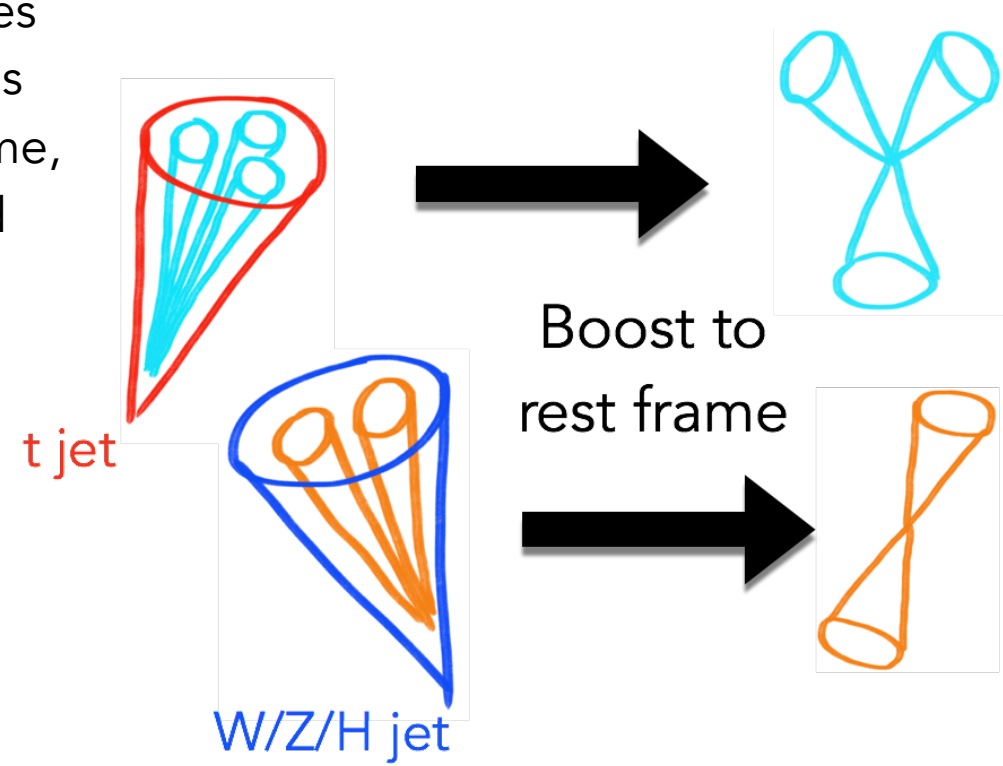
New Idea

- We present a new algorithm to **simultaneously** classify jets according to heavy object type
- **Boosted Event Shape Tagger (BEST)**
  - Consider top quarks, W, Z, H bosons
  - Light jets from QCD processes as background



- Use different hypothesized reference frames corresponding to the heavy particle masses
  - When boosting to the 'correct' rest frame, jet constituents should be isotropic and show the expected N-prong structure

- Based on fall 2016 phenomenology paper [1], now implemented using full CMS simulation and reconstruction



[1] PRD 94 094027 (2016)

Methodology

Public Material in CMS Detector Performance Note CMS-DP-2017/027

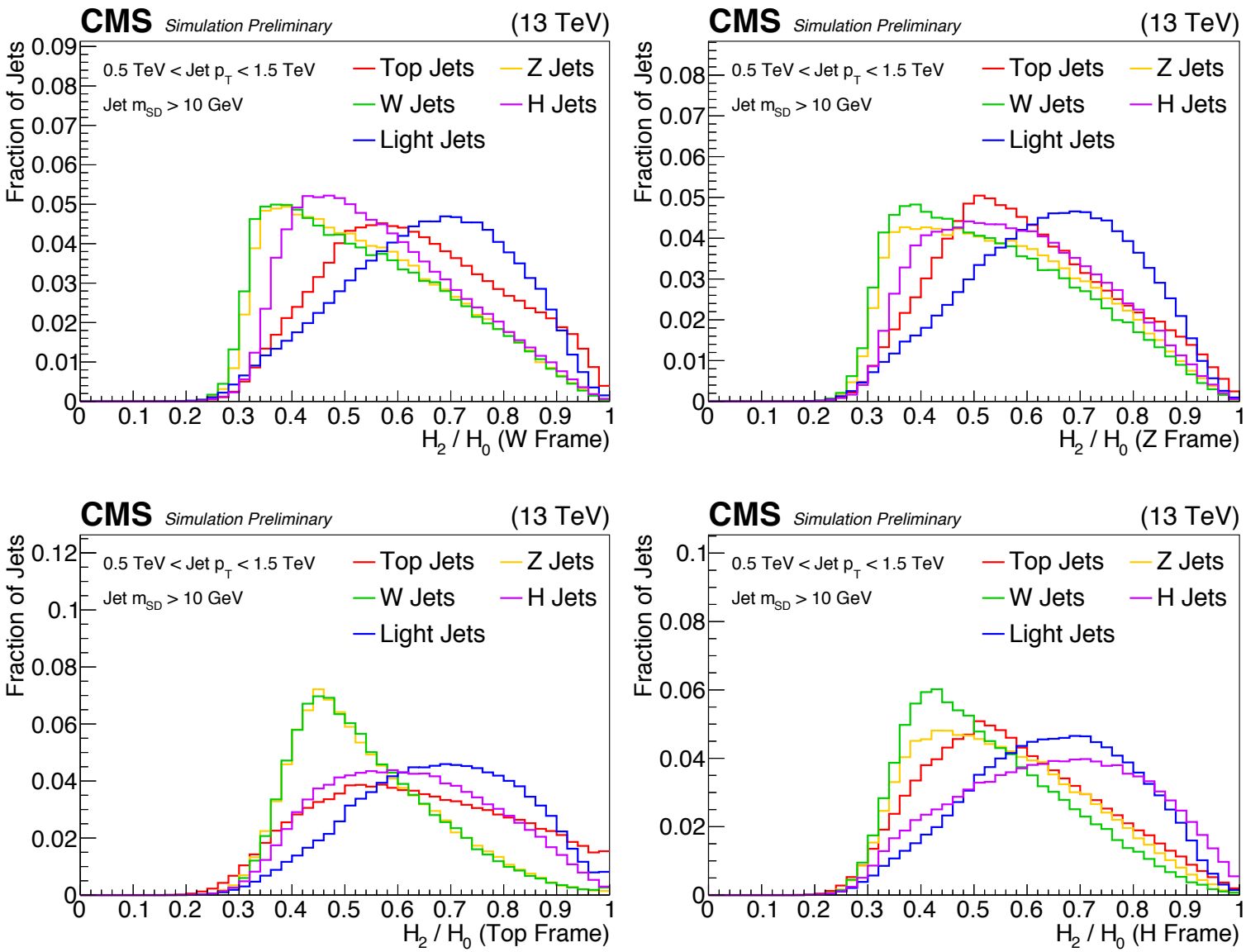
- Select an anti-kT R = 0.8 (AK8) jet with $p_T > 500$ GeV, $|\eta| < 2.4$
- Define four reference frames based on hypothesized particle origins of jet
  - Assume mass of top, Z, W, H; with same jet momentum
- Boost all jet constituents into each of the 4 hypothesized reference frames
- Compute angular and kinematic quantities in each frame
  - Four distributions of same quantity for each jet in respective frames
  - Recluster constituents in boosted frame (relative to boost axis) with smaller distance parameter R = 0.4 (AK4 jets)
- Train a neural network using these observables
- Obtain discrimination between heavy object jets and light jets; particle origin classification from NN outputs

NN Input Quantities	
Sphericity (t, W, Z, H)	Jet Soft-Drop Mass
Isotropy (t, W, Z, H)	Jet p_T (flattened)
Aplanarity (t, W, Z, H)	Jet η
Thrust (t, W, Z, H)	Jet τ_{21}
Jet Asymmetry A_L (t, W, Z, H)	Jet τ_{32}
Fox-Wolfram H_1/H_0 (t, W, Z, H)	Fox-Wolfram H_3/H_0 (t, W, Z, H)
Fox-Wolfram H_2/H_0 (t, W, Z, H)	Fox-Wolfram H_4/H_0 (t, W, Z, H)

Inputs - Fox-Wolfram Moments

- Related to Legendre polynomials
  - Angular moments of the distribution of particles

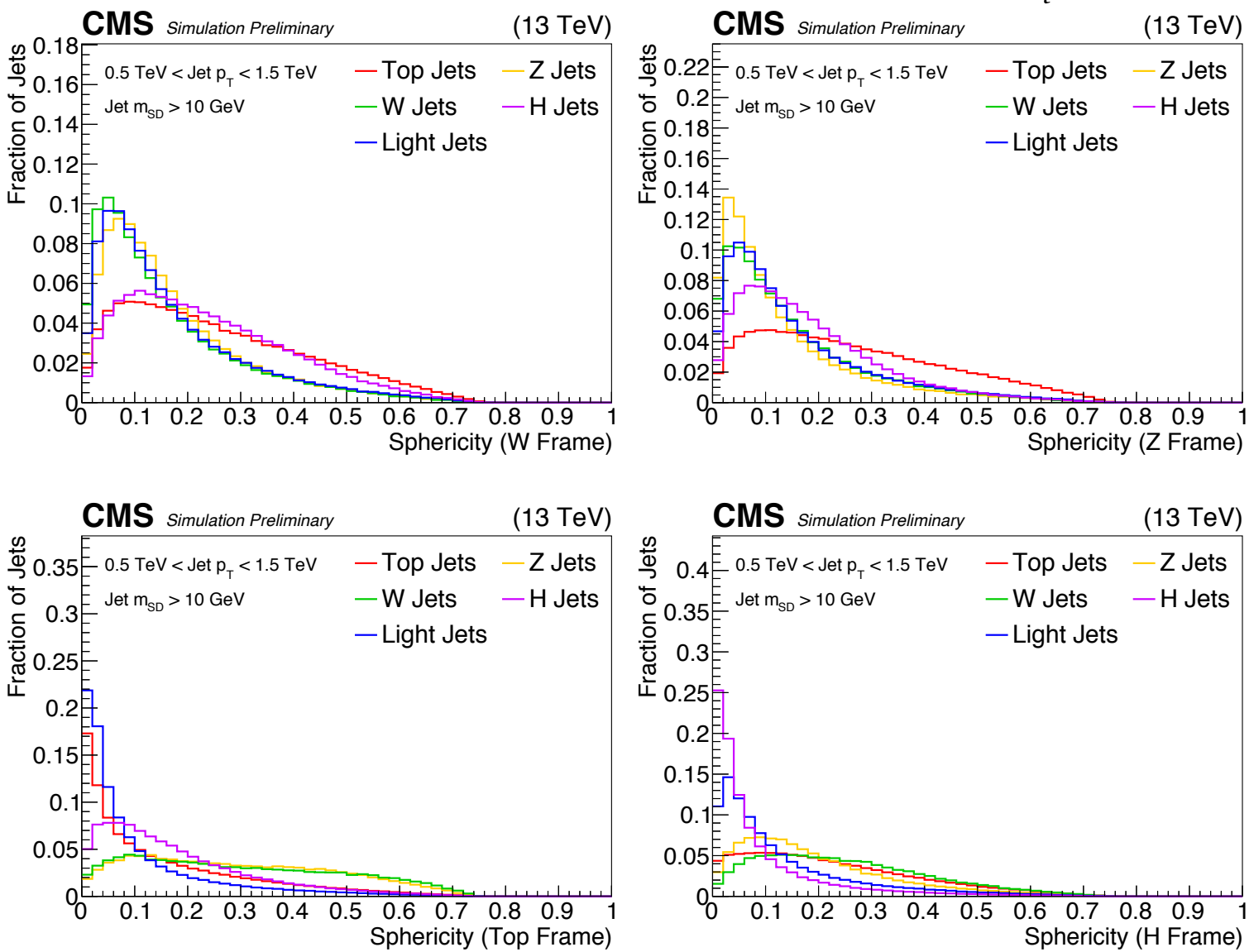
$$H_\ell = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{s} \cdot P_\ell(\cos(\phi_{i,j}))$$



Inputs - Sphericity Tensor

- Measure of how uniformly distributed a set of particles is
  - 0 → Spherical, 1 → Collimated

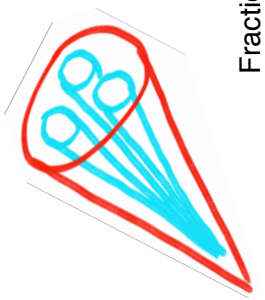
$$S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}$$

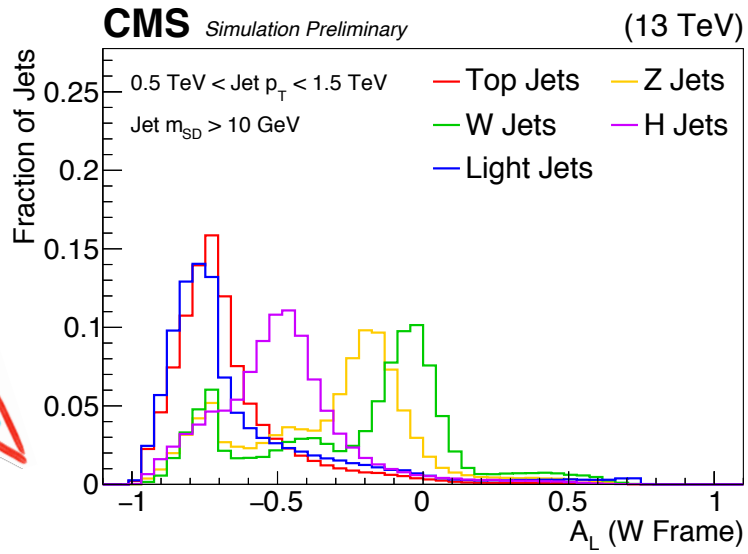


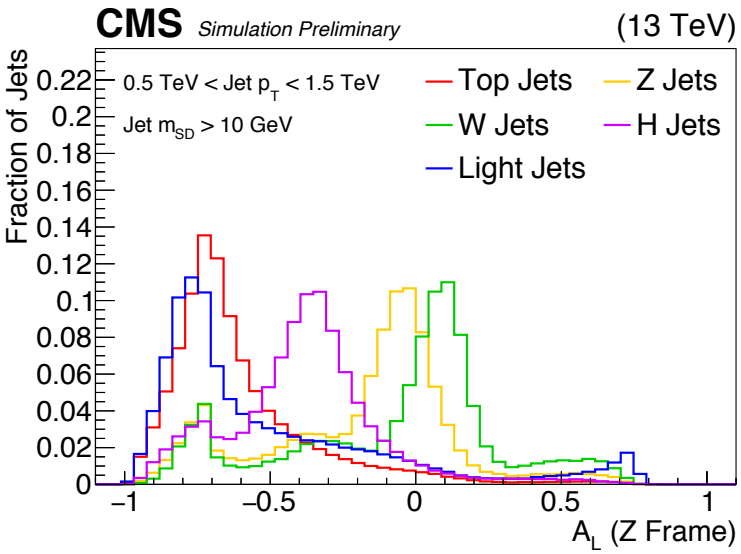
Inputs - Jet Asymmetry

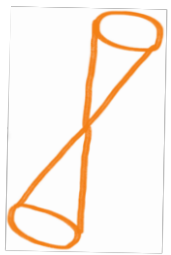
- Use the AK4 jets reclustered in boosted frames
  - In 'correct' frame, $A_L \rightarrow 0$

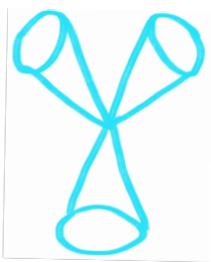
$$A_L = \frac{\sum_{jet} p_z^{jet}}{\sum_{jet} p^{jet}}$$

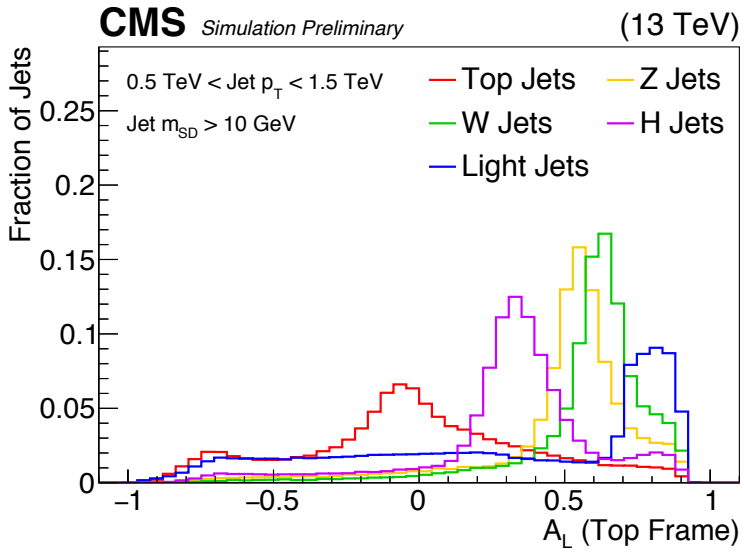


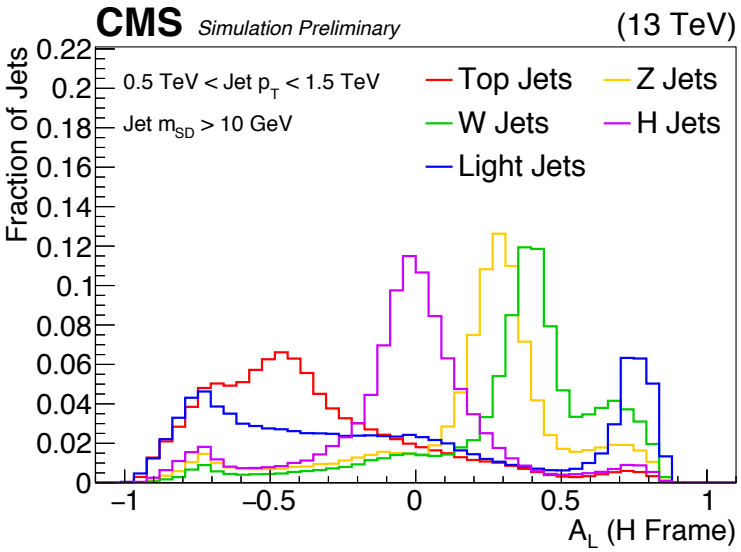


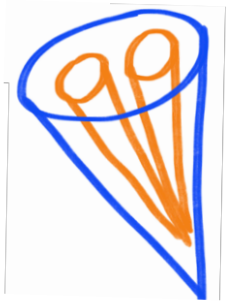






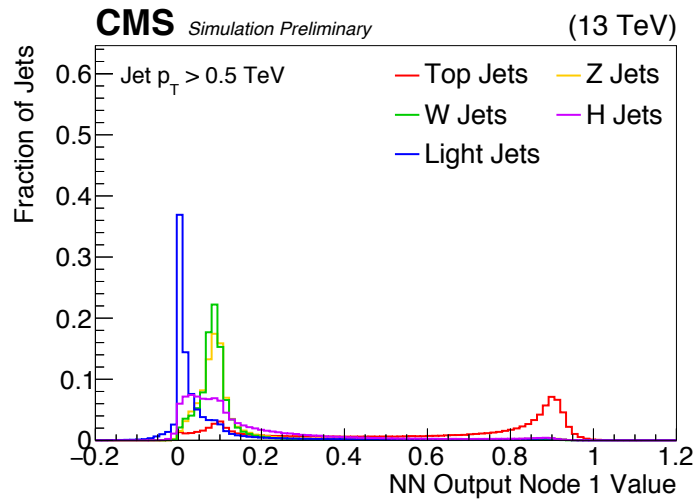


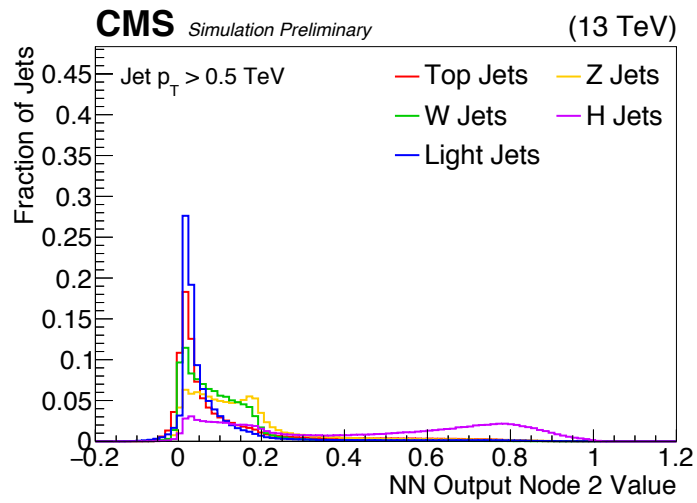


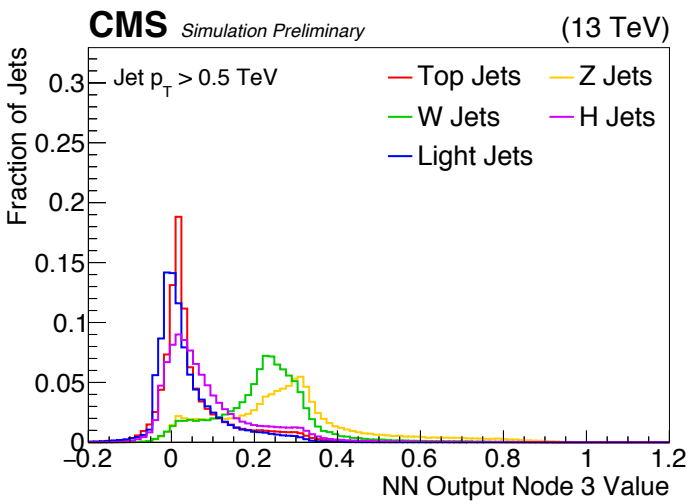


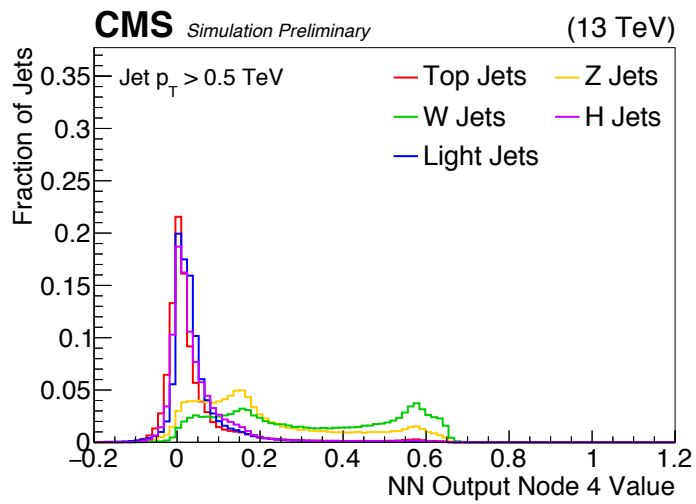
Neural Network

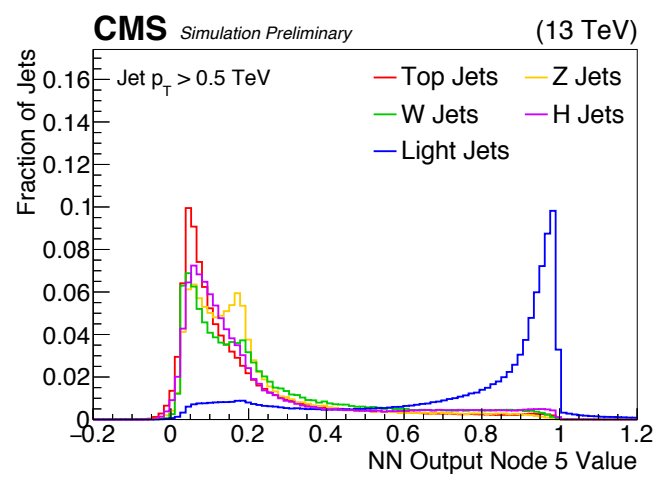
- We train a NN using the TMVA software package
  - 41 input nodes for the input distributions
  - 2 x 20-node hidden layers
  - 5 output nodes
    - One target each for t, W, Z, H jets and light-flavor jets from QCD
- 100k individual jets used for training
  - 20k from each particle species
- Outputs show good separation!





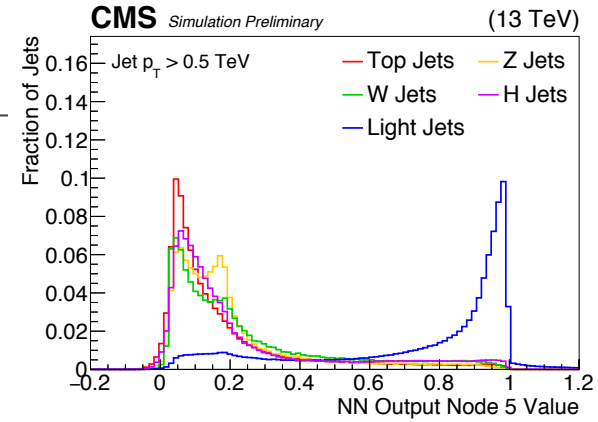


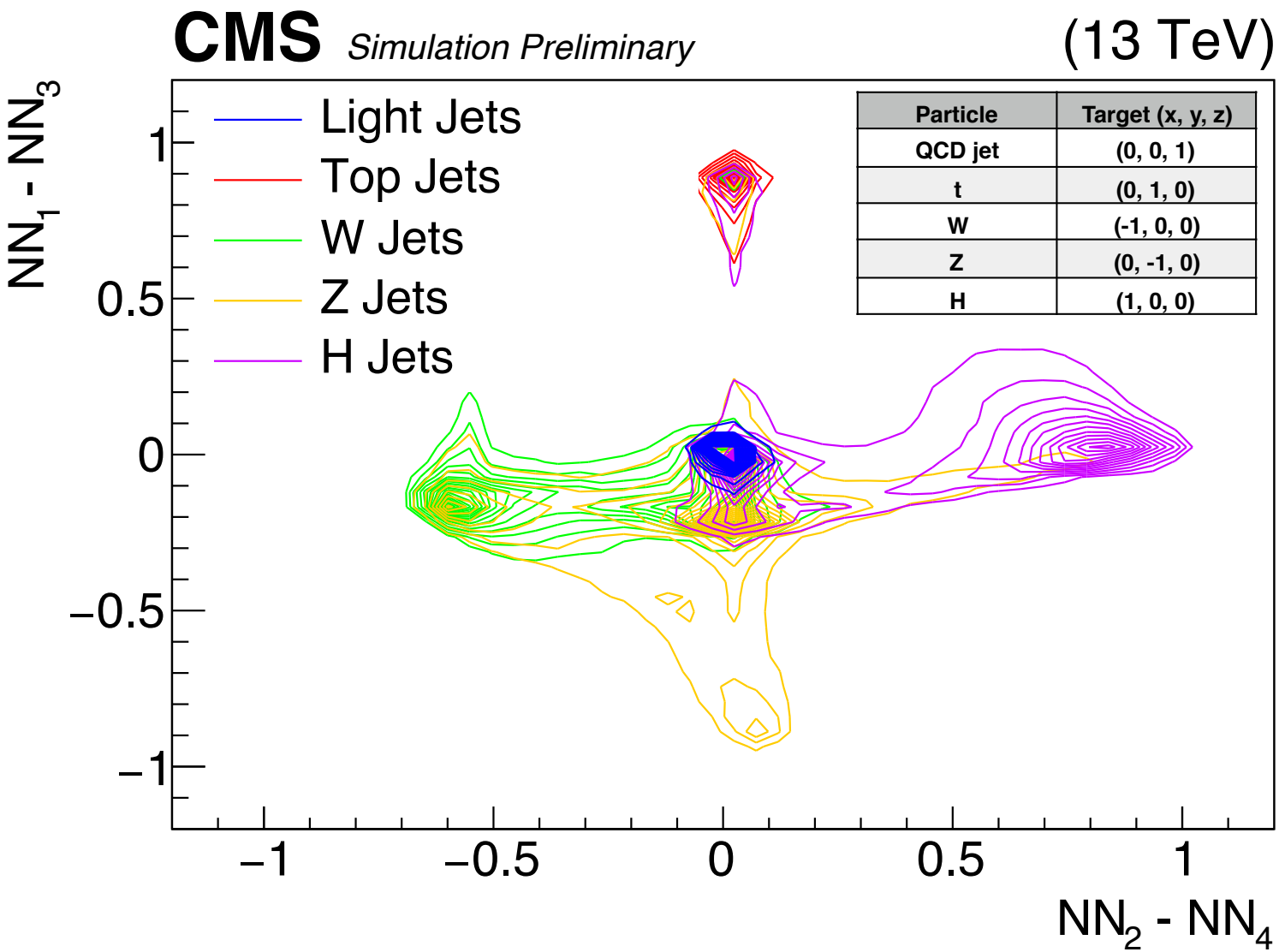




Neural Network

- Two-dimensional visualization of the 5D space shows simultaneous separation of the species
  - QCD will fall out of the page

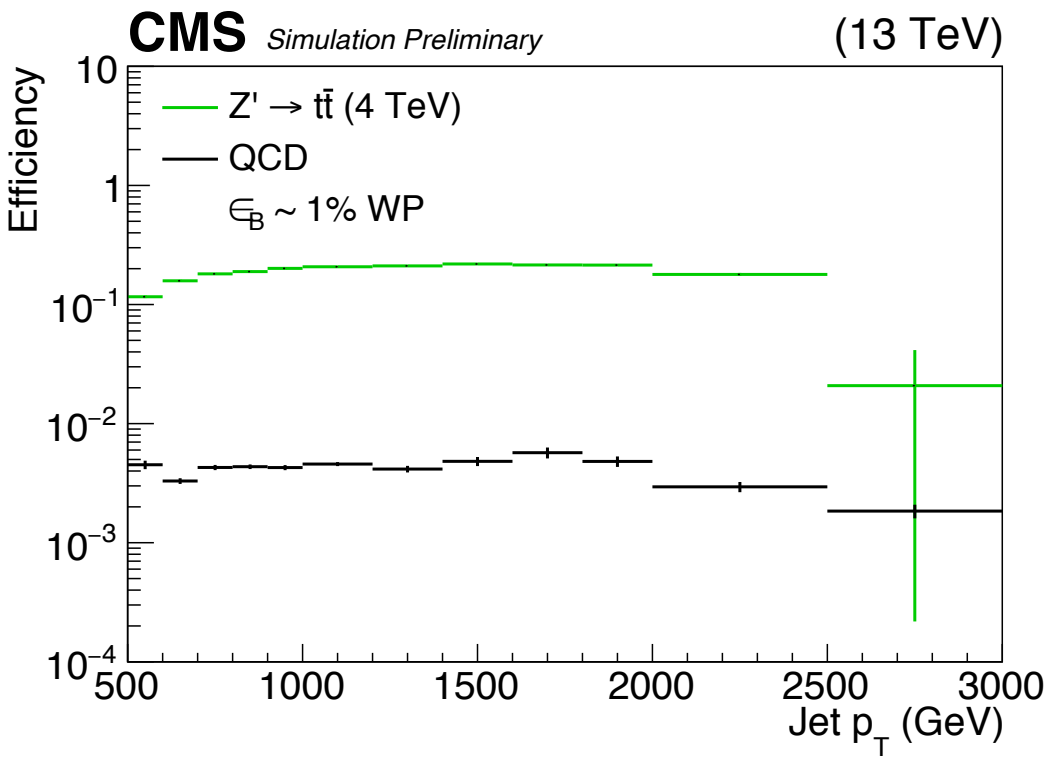


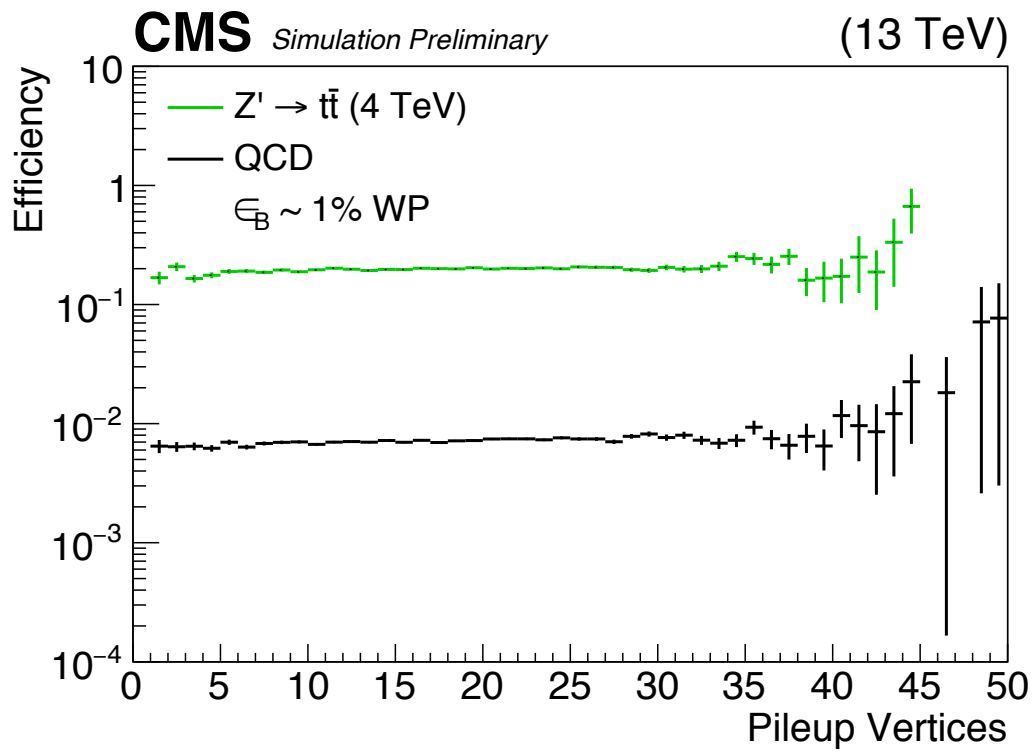


Particle	Target (x, y, z)
QCD jet	(0, 0, 1)
t	(0, 1, 0)
W	(-1, 0, 0)
Z	(0, -1, 0)
H	(1, 0, 0)

Extension to Analysis

- To use in an analysis, we can create working points for identifying the different objects, e.g. top-tagging
  - Combining cuts on the NN outputs with existing methods (jet soft-drop mass, N-subjettiness) gains performance
- Efficiency is stable as a function of p_T and pileup activity
  - ~25% top-tagging efficiency for <1% background efficiency

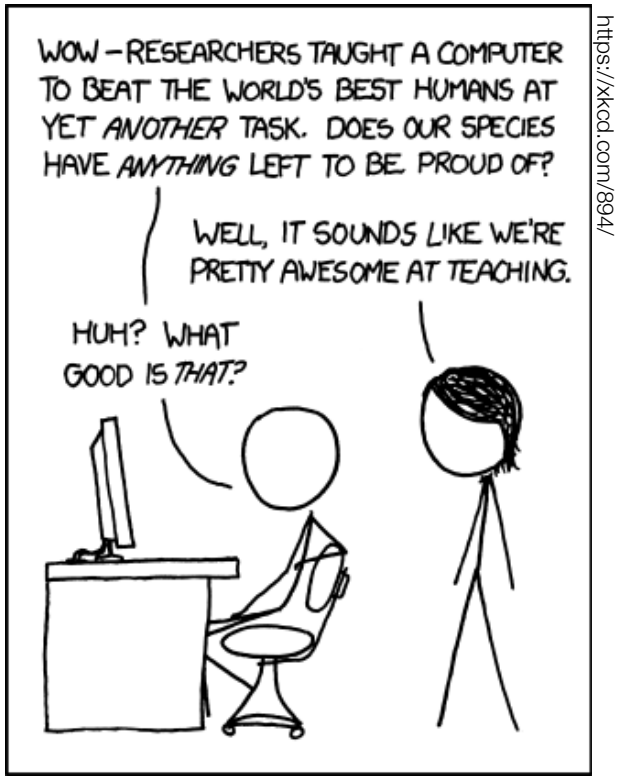




Conclusions

PRD 94 094027 (2016)
Public Material in CMS Detector
Performance Note CMS-DP-2017/027

- We have demonstrated the Boosted Event Shape Tagger (BEST) in CMS simulation/reconstruction
  - Achieved simultaneous classification of t/W/Z/H jets and discrimination of light-flavor jet background
- Useful for high-multiplicity final states, e.g. vector-like quark searches (TT→tZtH, bWtZ, etc.)
- Can also be used for single-particle identification through cuts on NN outputs
  - Improve top-pair resonance searches, diboson resonance searches, e.g.
- Next steps
  - Optimize input variables to remove redundant information
  - Validation in data events
  - Performance comparisons to existing tools
  - Commission in analysis scenario



https://xkcd.com/894/