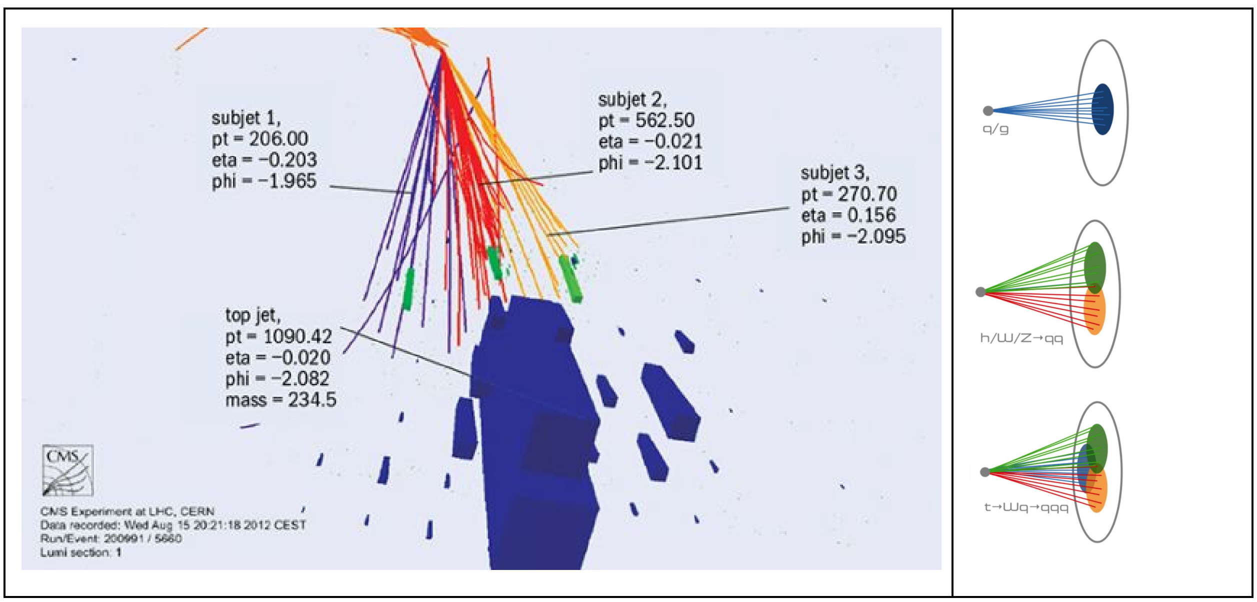
Goal: feature studies of jet tagging samples

One of the most difficult challenges at the LHC is to identify and classify hadronic decays of highly Lorentz-boosted W/Z bosons and top quarks from quark and gluons. A jet is a group of collimated particles, which can be used as a proxy of a fundamental particle, e.g. up quark, down quark, strange quark, gluon, W boson, Z boson or top quark.

A light quark (q) jet or gluon (g) jet is usually composed of one subjet. The light quark is either up quark, down quark or strange quark. W-jet and Z-jet are composed of two subjects because hadronic decay of W or Z bosin is quark and ant-quark final state. A top jet is composed of three subjets because of the intrinsic decay structure (top -> W b -> qq’ b). The figure below ([link]()) shows an example of a top jet. The blue lego bricks represent the Calorimeter clusters; while purple, red, yellow lines represent charged Tracks. Calorimeter clusters and/or Charged Tracks can be used as consistent to reconstruct Jets.



Samples:

[processed-pythia82-lhc13-all-pt1-50k-r1_h022_e0175_t220_nonu_withPars_truth.z]()

This sample allows developers to explore modern Machine Learning methods. The samples are used in the JEDI-net paper ([1908.05318]()). We will follow this JEDI-net paper closely for the whole exercise. The detailed description of the features can be found in [arXiv:2004.08262]().

Task1: Count number of jets in each truth jet label.

Create a counting table:

Label	Number of jets	Description
Total		Total statistics
q		Light (up, down, strange)-quark jet
g		Gluon jet
w		W-boson jet
z		Z-boson jet
t		Top-quark jet
undef		Undefined jet

Task2: Study High Level features

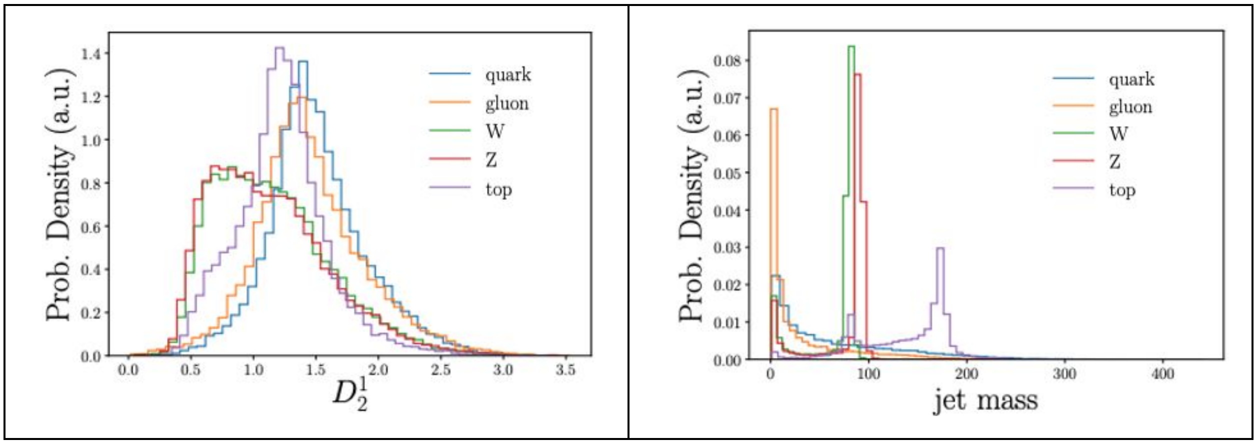
These high level features (convention: j_*) are computed from the constituents of each jet. Let's plot 1D distribution of each high level feature and compare distributions among different truth labels. Let's focus on a few selected variables after the Modified Mass Drop Tagger (mmdt).

Jet kinematics
- 'j_mass_mmdt', 'j_pt', 'j_eta'

Jet substructures:
- 'j_zlogz', 'j_multiplicity'

Energy-correlation functions
- 'j_c1_b0_mmdt', 'j_c1_b1_mmdt', 'j_c1_b2_mmdt', 'j_c2_b1_mmdt', 'j_c2_b2_mmdt',
- 'j_d2_b1_mmdt','j_d2_b2_mmdt', 'j_d2_a1_b1_mmdt', 'j_d2_a1_b2_mmdt',
- 'j_m2_b1_mmdt', 'j_m2_b2_mmdt',
- 'j_n2_b1_mmdt', 'j_n2_b2_mmdt',

Example results:



What is Energy-Correlation Function ([1901.10342](https://arxiv.org/abs/1901.10342))?

Energy-correlation functions (ECFs) achieve essentially the same objective than N-subjettiness without requiring the selection of N reference axes.

$$ {}_k e_N^{(\beta)} = \sum_{i_1<\ldots<i_N\in \mathrm{jet}} \left(\prod_{j=1}^{N} z_{i_j}\right)\left(\prod_{\ell=1}^{k} \overset{\ell}{\min_{u<v\in\{i_1,\ldots,i_N\}}} \Delta R_{uv}^{\beta}\right), \qquad (5.23) $$

where $\overset{\ell}{\min}$ denotes the ℓ-th smallest number.

with $z_i = p_{t,i} / \mathrm{Sum}\,(p_{t,j})$.

Two-prong taggers:

$$ C_2^{(\beta)} = \frac{{}_3e_3^{(\beta)}}{\left({}_1e_2^{(\beta)}\right)^2} \equiv \frac{e_3^{(\beta)}}{\left(e_2^{(\beta)}\right)^2}, \qquad D_2^{(\beta)} = \frac{e_3^{(\beta)}}{\left(e_2^{(\beta)}\right)^3}, \qquad (5.24) $$

$$ N_2^{(\beta)} = \frac{{}_2e_3^{(\beta)}}{\left(e_2^{(\beta)}\right)^2}, \qquad M_2^{(\beta)} = \frac{{}_1e_3^{(\beta)}}{e_2^{(\beta)}}, $$

Task3: Study Low Level features

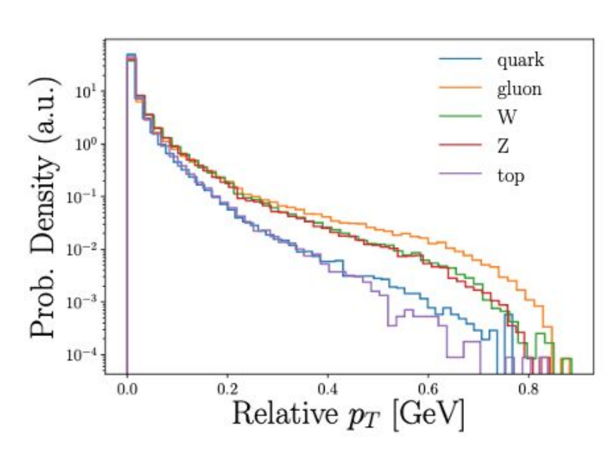
The low level features (convention: j1_*) are constituents of each jet.

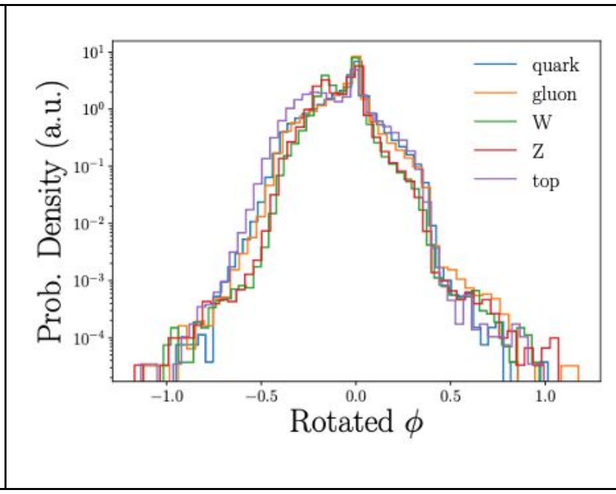
1. Draw 1D distributions
   Loop each constituent of each jet and categorize in each label
   'j1_px', 'j1_py', 'j1_pz', 'j1_e', 'j1_pdgid', 'j1_erel', 'j1_pt', 'j1_ptrel', 'j1_eta', 'j1_etarel', 'j1_etarot', 'j1_phi', 'j1_phirel', 'j1_phirot', 'j1_deltaR', 'j1_costheta', 'j1_costhetarel', 'j1_e1mcosthetarel'
2. Draw 2D weighted distributions

    a. (jet_etarot, j1_phirot) weighted by j1_ptrel
    b. (jet_eta, j1_phi) weighted by j1_ptrel

Example plot:

a. 1D distributions





b. 2D temperate maps