Quark gluon tag in HL-LHC Constituent-Based Tagger

Sabine Elles, Jessica Leveque and Florencia Castillo

Our goals

Test Constituent-Based Transformers for HL-LHC

- Focus on both the **central** and **forward regions**.

Integrate HGTD

- Apply to the **forward region**.

Integrate HGTD and Topo Towers

- Extend to the forward region for enhanced tagging.

Evaluate the Tagger

- Different (y,pT)

Test with Varying Pile-Up Conditions

- Use samples with different **pile-up values**

Publication!

Working on

Test Constituent-Based Transformers for HL-LHC

- Focus on both the **central** and **forward regions**.

Integrate HGTD

- Apply to the **forward region**.

Integrate HGTD and Topo Towers

- Extend to the forward region for enhanced tagging.

Evaluate the Tagger

- Different (y,pT)

Test with Varying Pile-Up Conditions

- Use samples with different **pile-up values**

Publication!

Derivation request

- JEMT 2 currently includes the constituents and TT, but we need to incorporate the HGTD.
  - I am in contact with the conveners and organizers to coordinate this update.
    - We need to check the impact on the overall size due to this inclusion.

Derivation request: Option 1

- HGTD information is added InDetTrackParticles in JETM2SlimmingHelper.AllVariables
  - Currenly only SmartCollection are included in JETM2
  - https://acode-browser1.usatlas.bnl.gov/lxr/source/athena/PhysicsAnalysis/DerivationFramework/DerivationFrameworkJetEtMiss/python/JETM2.py#0131

Comparision size:

Adding All variables the size of InDetTrackParticles is (1 Event):

151.888 kb 38.195 kb 38.195 kb 3.977 1 InDetTrackParticles

And without is:

70.382 kb 16.510 kb 16.510 kb 4.263 1 InDetTrackParticle

Derivation request: Option 2

- Only adding the HGTD variables I need:
  - TrackingVariablesHGTD = ["HGTD_cluster_time", "HGTD_extrap_x", "HGTD_extrap_y"]
  - In the job option:
  - if flags.GeoModel.Run == LHCPeriod.Run4:

    JETM2SlimmingHelper.ExtraVariables += [".".join(["InDetTrackParticles"] + TrackingVariablesHGTD)]

Comparison size:

Only adding the variables I need (1 Event):

73.348 kb 17.268 kb 17.268 kb 4.248 1 InDetTrackParticles

And without is:

70.382 kb 16.510 kb 16.510 kb 4.263 1 InDetTrackParticle

Successfully added TT+Track(HGTD)+Constituents

```
info_index = []
for i, const_type in enumerate(self.const_types):
    m_const.append(sample[f'{self.jet_type}_{const_type}_m'])
    pt_const.append(sample[f'{self.jet_type}_{const_type}_pt'])
    eta_const.append(sample[f'{self.jet_type}_{const_type}_eta'])
    phi_const.append(sample[f'{self.jet_type}_{const_type}_phi'])
    if const_type == "Track":
        hgtd_x_const.append(sample[f'{self.jet_type}_{const_type}_HGTD_extrap_x'])
        hgtd_y_const.append(sample[f'{self.jet_type}_{const_type}_HGTD_extrap_y'])
    else:
        # Append tensor filled with -1 (float32) matching the shape of pt_const[-1]
        hgtd_x_const.append(tf.fill(tf.shape(pt_const[-1]), -1.0))  # -1.0 as float32
        hgtd_y_const.append(tf.fill(tf.shape(pt_const[-1]), -1.0))  # -1.0 as float32

    info_index.append(i * tf.ones_like(pt_const[-1]))
```

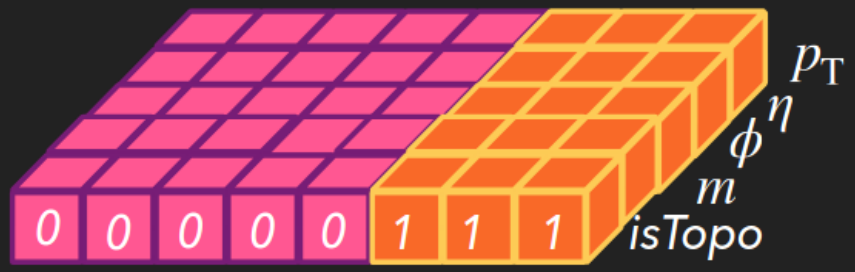
N pfos, M topotowers and L tracks the input would have shape (N+M+L, K+2), where K is the normal nariables like eta, phi, logpT, etc. and 2 variables: isTopo, isTrack

From Samuel slides

INPUTS

- **UFO (PFO)** 4-momenta as an input
- additionally in high η region - **TopoTowers**
- concatenate to UFOs (PFOs) + add *isTopo* variable
- use only **relative variables**, w/o overall jet information

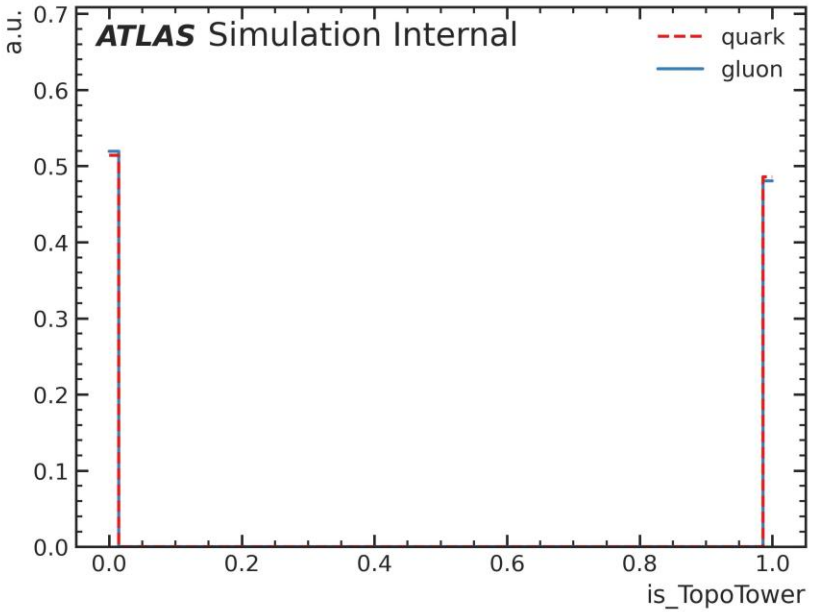
Constituent Variables
$\Delta\eta = \eta - \eta^{\text{jet}}$
$\Delta\phi = \phi - \phi^{\text{jet}}$
$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$
~~$\log p_{\text{T}}$~~
~~$\log E$~~
$\log \frac{p_{\text{T}}}{p_{\text{T}}^{\text{jet}}}$
$\log \frac{E}{E^{\text{jet}}}$
m

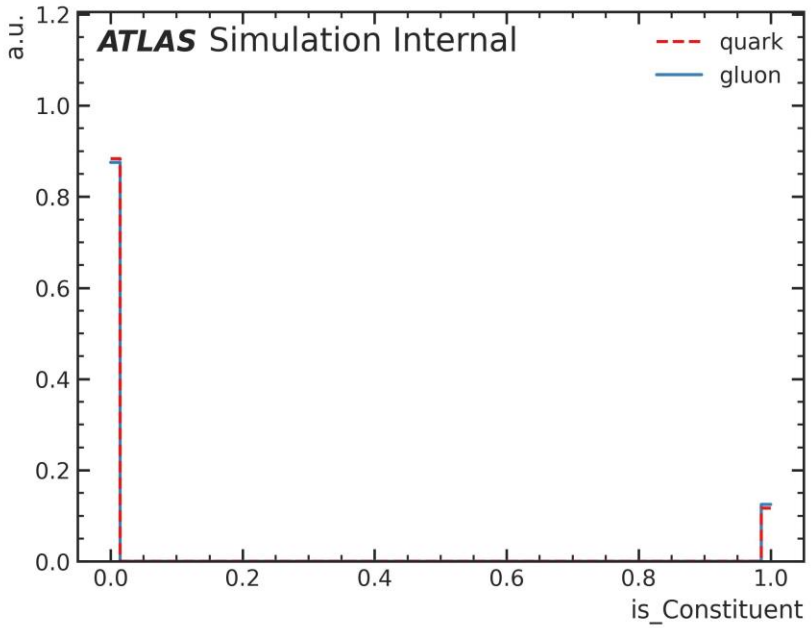


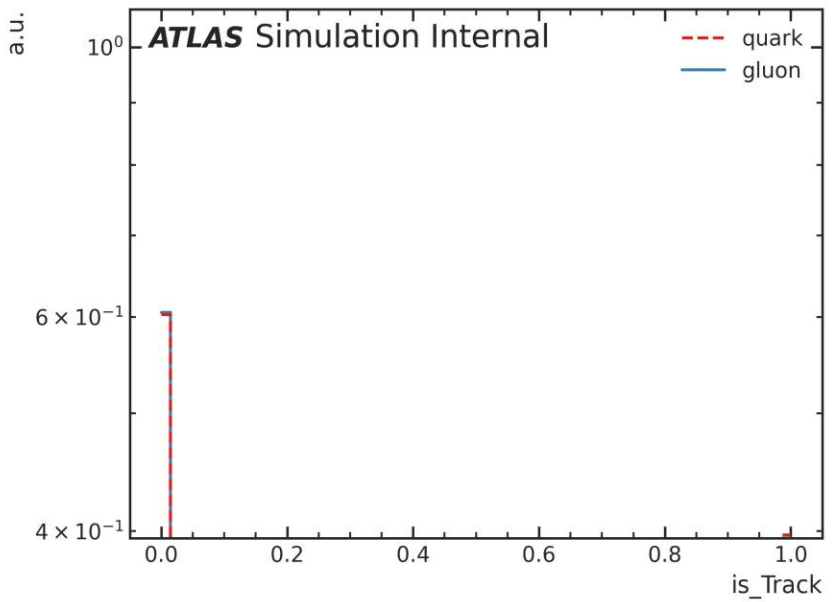
Constituent Interaction Variables
$\log \Delta^{ab} = \log \sqrt{(\eta^a - \eta^b)^2 + (\phi^a - \phi^b)^2}$
$\log k_{\text{T}}^{ab} = \log (\min (p_{\text{T}}^a, p_{\text{T}}^b)\Delta^{ab})$
$z^{ab} = \min (p_{\text{T}}^a, p_{\text{T}}^b)/(p_{\text{T}}^a + p_{\text{T}}^b)$
$\log m^{2,ab} = \log (p^{\mu,a} + p^{\mu,b})^2$

https://indico.cern.ch/event/1380446/contributions/6084131/attachments/2918147/5122658/QGTagging_lightening_talk_HCW_2024.pdf

Successfully added TT+Track(HGTD)+Constituents







Machinery working! Tested in a small dataset!

Working on

Test Constituent-Based Transformers for HL-LHC

- Focus on both the **central** and **forward regions**.

Integrate HGTD

- Apply to the **forward region**.

Integrate HGTD and Topo Towers

- Extend to the forward region for enhanced tagging.

Moving forward! ---> Next

Evaluate the Tagger

- Different (y,pT)

Test with Varying Pile-Up Conditions

- Use samples with different **pile-up values**

Publication!

Need to prepare a bigger sample, while preparing the code for the evaluation in y,pT regions