arnab.laha@cern.ch

ATLAS EXPERIMENT

Search for Emerging Jets at ATLAS

Arnab Laha, Rachid Mazini

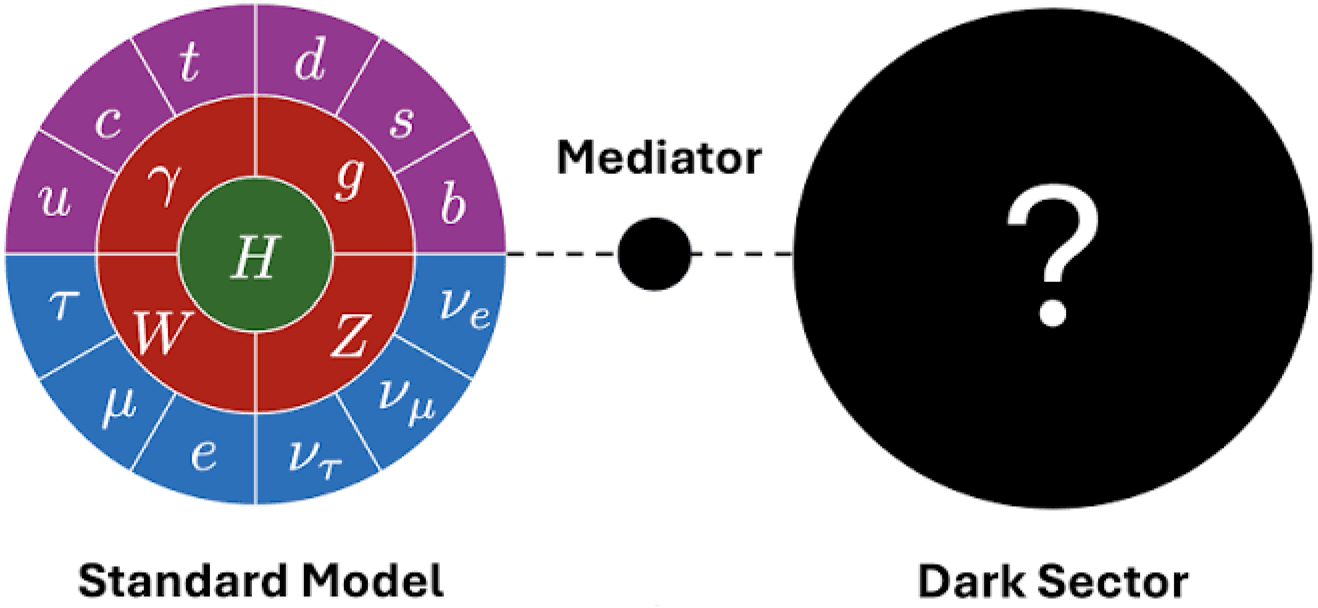
70th ANNUAL CONFERENCE OF THE SOUTH AFRICAN INSTITUTE OF PHYSICS
6 - 10 July 2026
University of the Western Cape

NRF National Research Foundation

iThemba LABS Laboratory for Accelerator Based Sciences

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

A Hidden World: Beyond the Standard Model

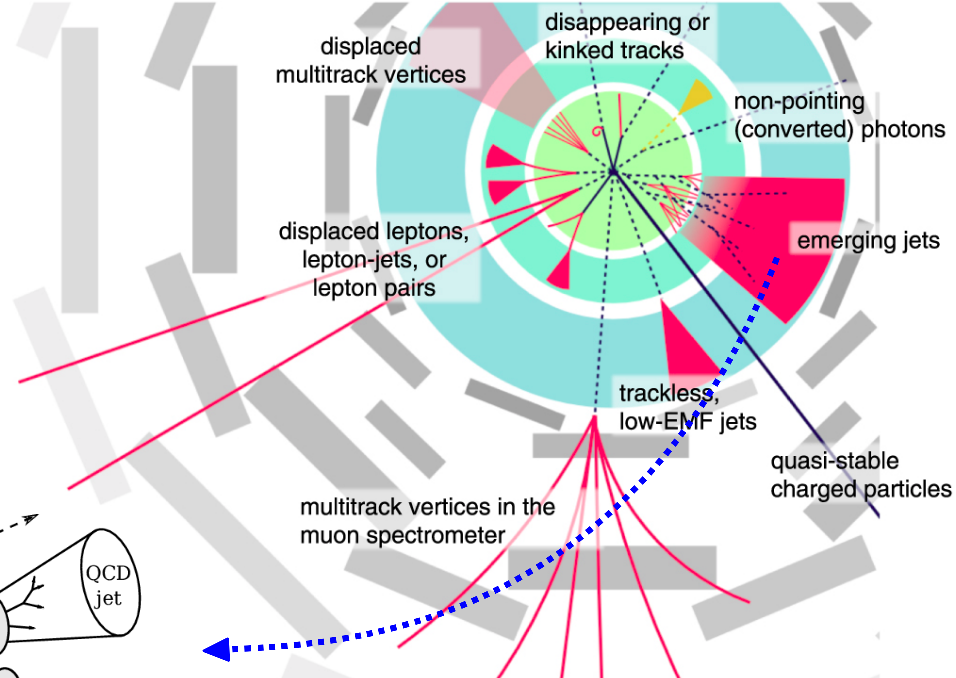


- Standard Model is incomplete; We don't know the particle nature of dark matter yet!
- **Hidden World?** A new sector of physics beyond the Standard Model which is connected to Dark Matter
  - The mediator particle(s) bridges the SM and Dark Sector

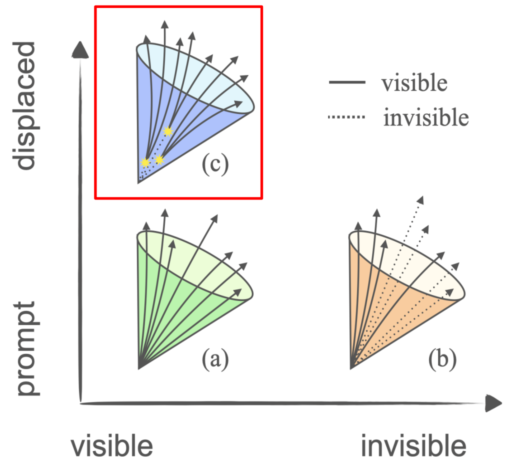
Dark QCD and Emerging Jets

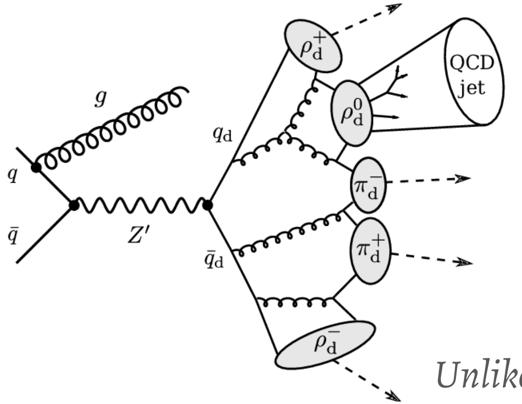
Dark QCD is a popular BSM framework which extends the SM with “**dark quarks**” charged under a new $SU(N)_D$

- Dark quarks will shower and hadronize into dark meson in the dark sector before decaying back to SM quarks
- Novel signatures than QCD jets
- Emerging jets are a part of different types of **dark jets** predicted in various hidden sector models



https://arxiv.org/abs/1903.04497



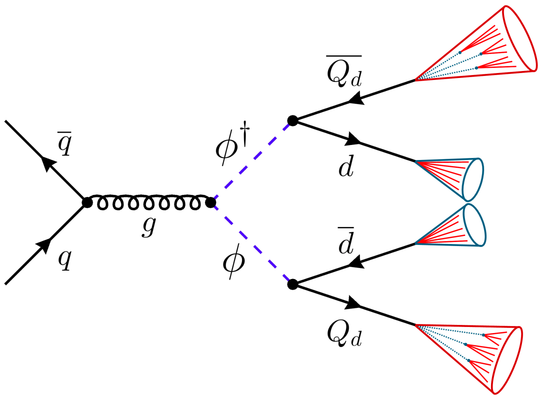


Unlike SM jets, the emerging jets develop gradually with many displaced tracks and secondary vertices.

Benchmark models

based on two ATLAS EJ searches

Run-2 search

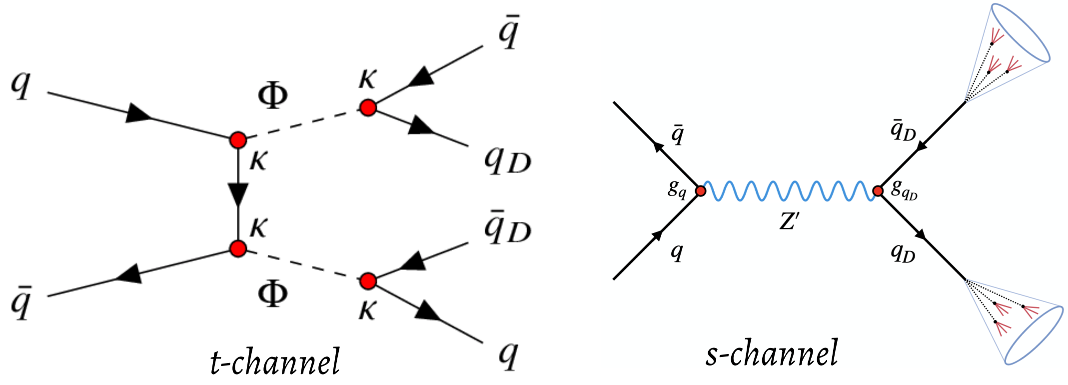


Mediator: Scalar ϕ

Production: ϕ pair production

Final State: **2 QCD Jets + 2 EJs**

Run-3 search



Mediator: Scalar ϕ / Vector Z'

Production: t-channel (scalar mediated)/ s-channel

Final State: **2 Ejs (+ QCD jets in t-channel)**

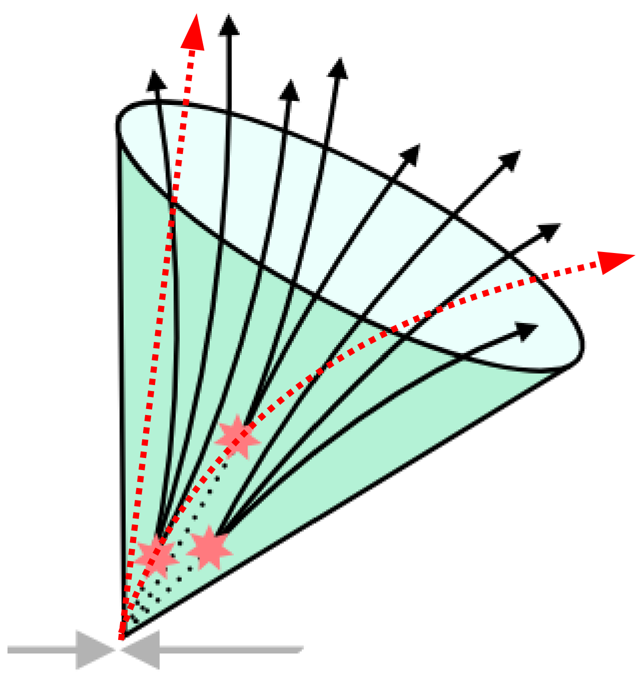
Model parameters

$(m_{\pi_D}, \Lambda_D, m_{\rho_D})$ [GeV]	(5, 10, 20), (10, 20, 40), (20, 40, 80)
$m_{Z'}$ [TeV] m_Φ [TeV]	0.6, 0.8, 1.0, 1.2, 1.5, 1.8, 2.2, 2.6, 3.0, 3.5 0.6, 1.0, 1.5, 2.0
$c\tau_{\pi_D}$ [mm]	1, **5**, 10, **50**, 100, **500**, 1000

$\mathrm{N_c^{dark}} = 3$, $\mathrm{N_f^{dark}} = 7$

$m_{\rho_d} = 2\Lambda_d = 2m_{q_d} = 4m_{\pi_d}$

Anatomy of Emerging Jets

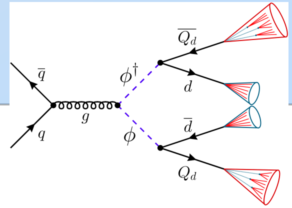


- **Presence of Displaced Vertices**
- **More displaced tracks than prompt tracks** experimentally
  - Prompt track fraction
- **Jets substructure**
  - Energy concentration inside the jet-cone is distinguishable than normal QCD jets
  - Energy correlation functions, subjettiness could probe the pronginess of the jets

- Special Large-Radius Tracking to reconstruct highly displaced tracks
- Secondery Vertices considering LRT tracks

https://arxiv.org/abs/2304.12867

Run 2 search



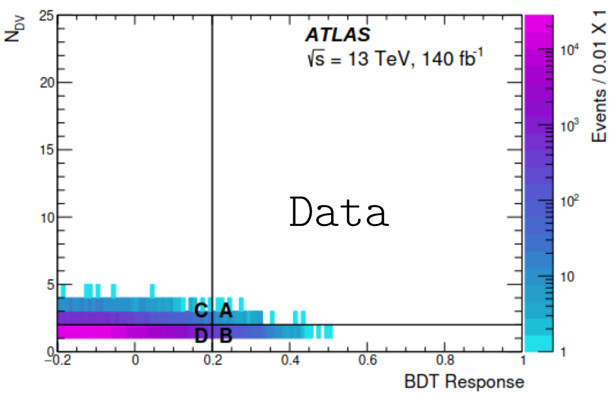
- **Trigger:** atleast 4 jets with $P_T > 120$ GeV
- Events are required to have **at least one displaced vertex** matched to a reconstructed jet

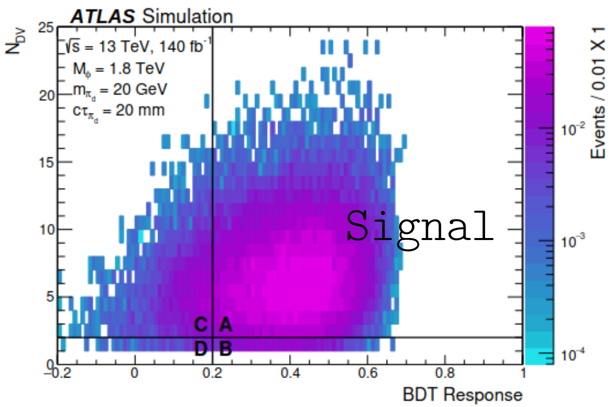
- Two stage BDT classifier:
  - **Jet-level-BDT:** Identify emerging jets
  - **Event-level-BDT:** Reduce QCD background

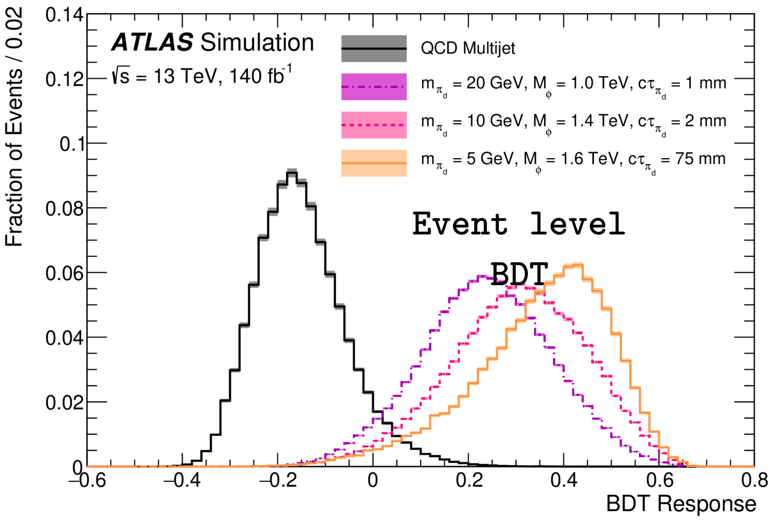
- **Data-driven** background estimation strategy (ABCD method)
  - Variables: N_{DV} vs Event-level-BDT score

Jet-level-BDT

Observable	Description
E_{Jet}	Jet energy, measured directly from calorimeter deposits
$m_{\text{Jet}} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2}$	Jet mass, invariant mass from the sum of topological clusters
$W_{\text{Jet}} = \frac{1}{p_T^{\text{Jet}}} \sum_i p_{T,i} \Delta R_i$	Jet width, p_T-weighted angular spread of topological clusters
η_{Jet}	Jet pseudorapidity

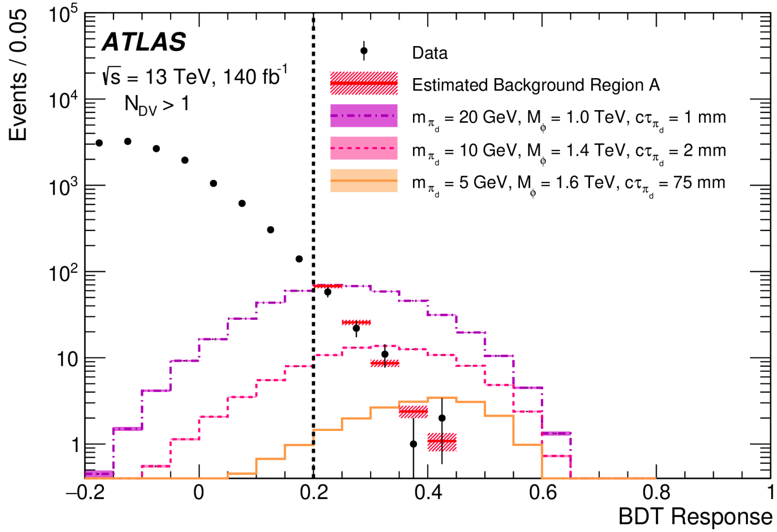


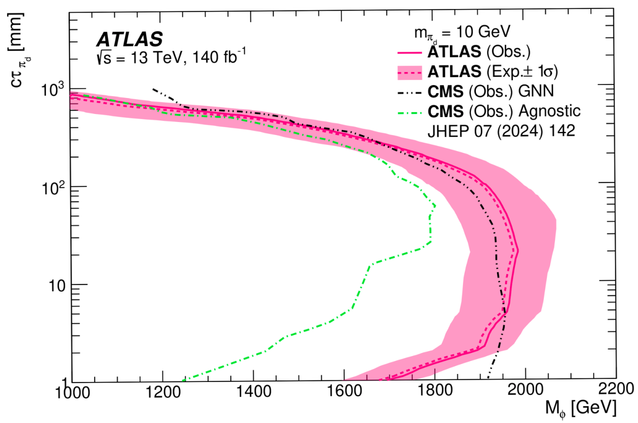


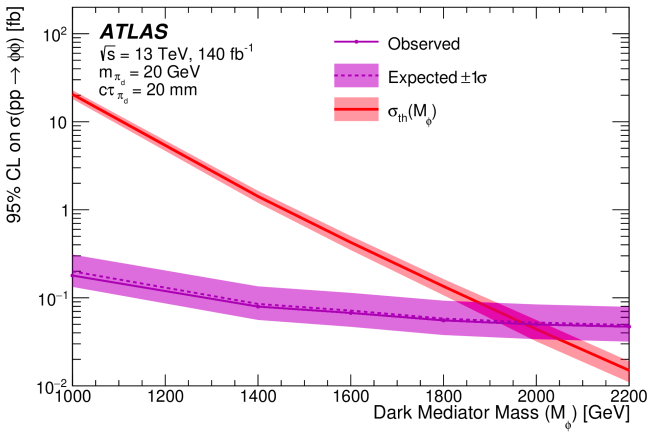


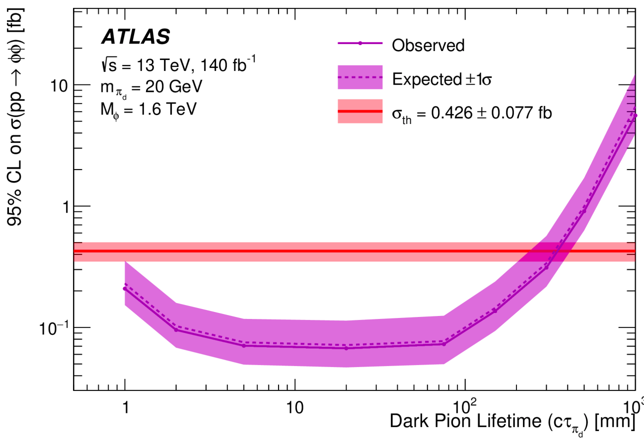
Run 2 result

- Data agree with SM prediction
- Limits in 2D plane of ϕ Mass vs cτ
- 1D limits in different model parameters









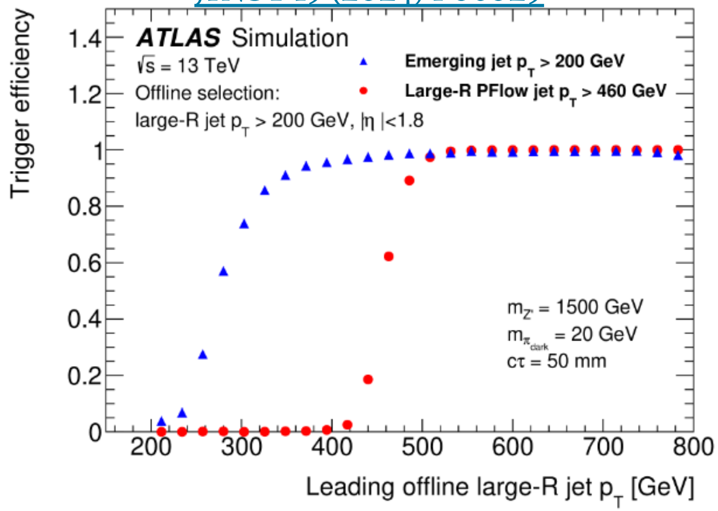
Run 3: Improved trigger and analysis strategy

- New dedicated trigger in Run 3 allows to probe lower Z' masses

Two different triggers:

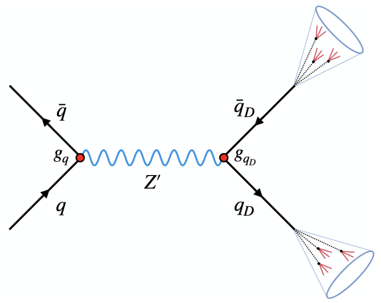
- **Standard single-jet trigger:**
  - Selects events containing a large- jet with $P_T > 460$ GeV
  - Fully efficient offline for jets with $P_T > 520$ GeV
- **Emerging Jet (EJ) trigger:**
  - Uses `Prompt Track Fraction (PTF)` as a discriminant.
  - Events selected with at least one jet having PTF < 0.08
  - Efficient down to much lower jet P_T than the single large-R jet trigger

JINST 19 (2024) P06029



Analysis Strategy

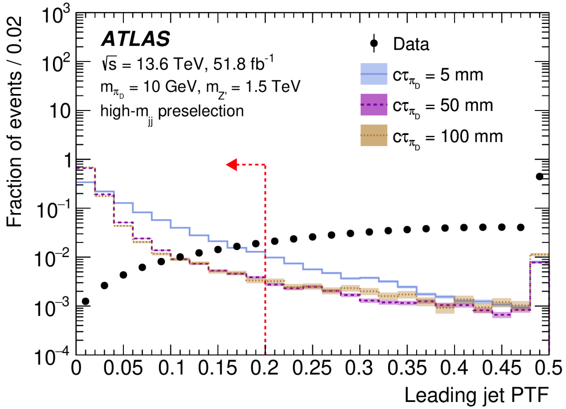
	low-m_{jj} EJ trigger & mjj < 1 TeV	high-m_{jj} jet trigger & mjj > 1 TeV
Cut-based	Easier Reinterpretation	
ML-based	Higher Sensitivity	



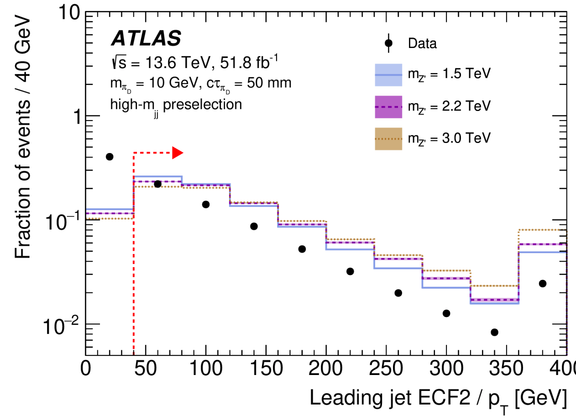
Cut Based Strategy

- Key discriminating variables to exploit emerging jets feature

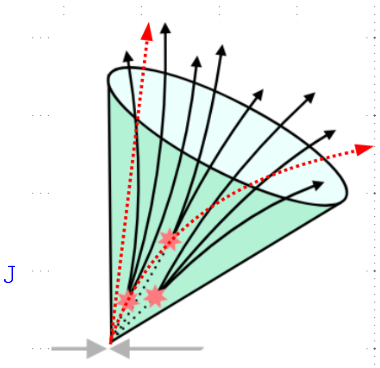
Displacement Properties

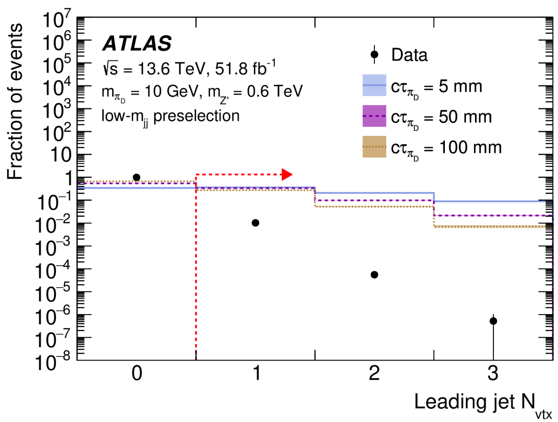


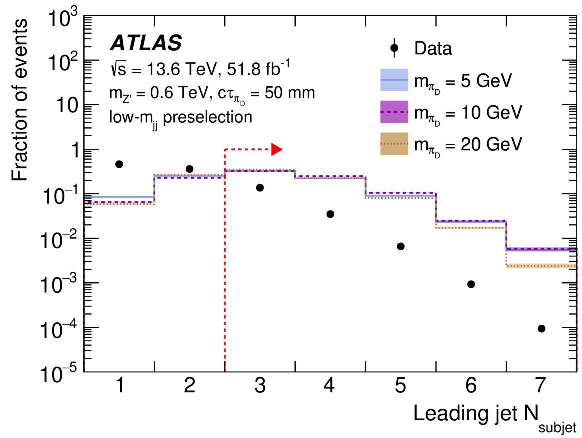
Jet substructure/ pronginess



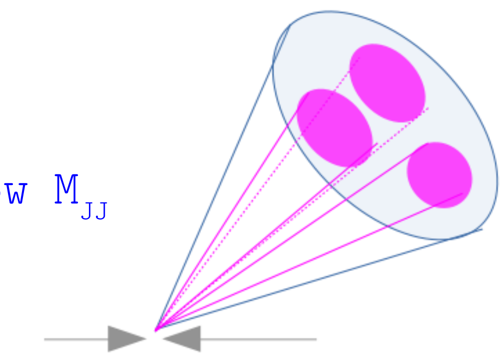
High M_{JJ}







Low M_{JJ}

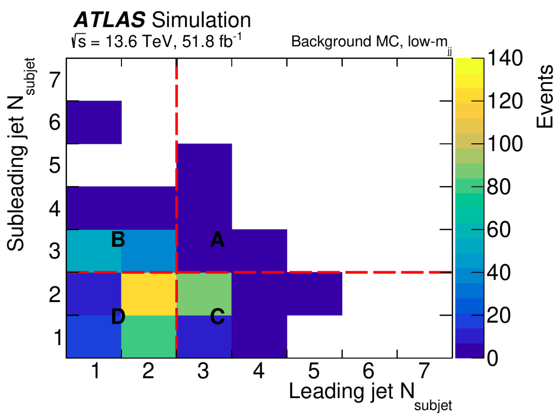


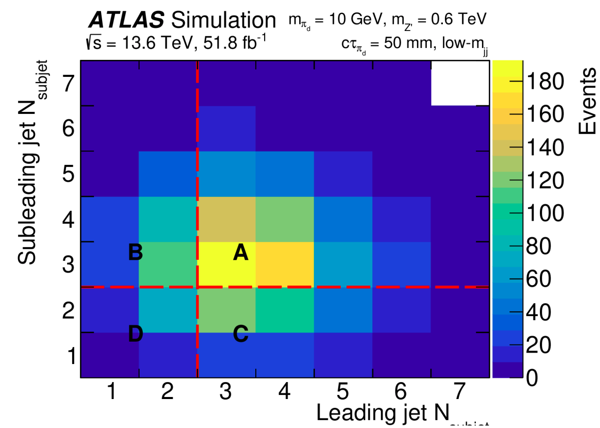
Background Estimation : ABCD method

- QCD MC is not reliable to predict background events in Signal Region
- Data-driven ABCD method is adopted in both regions

Low M_{JJ}

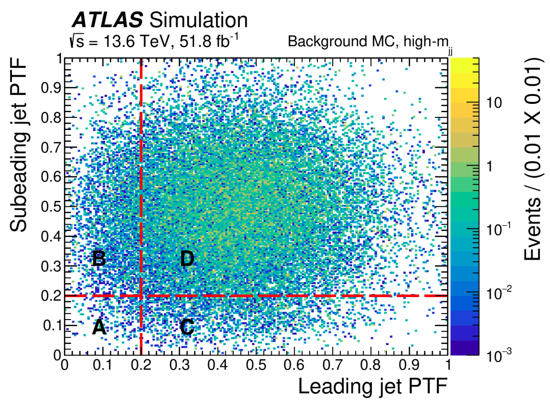
number of small-R jet constituents contained in the two leading jets

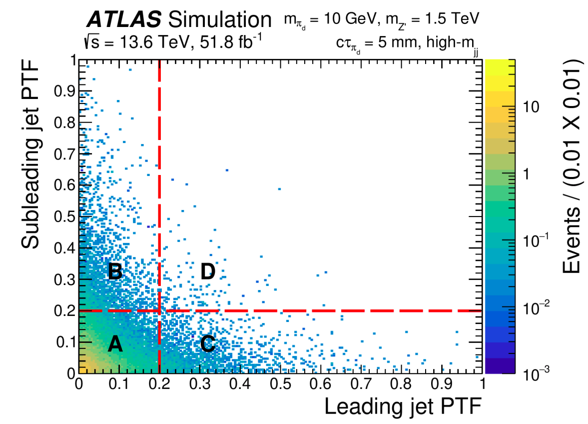




High M_{JJ}

prompt track fraction of the two leading jets

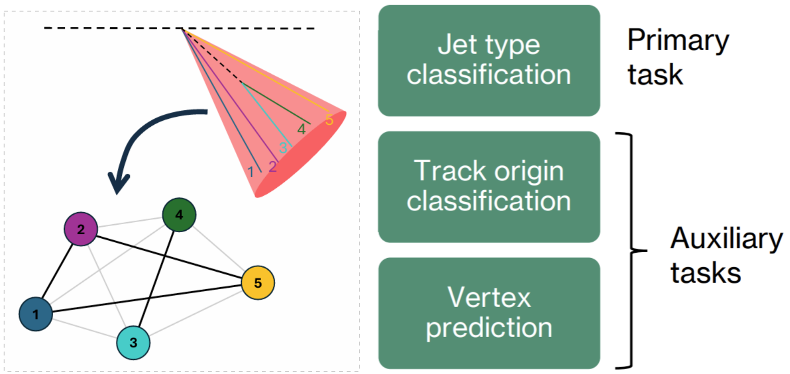




ML Strategy: Graph Neural Network

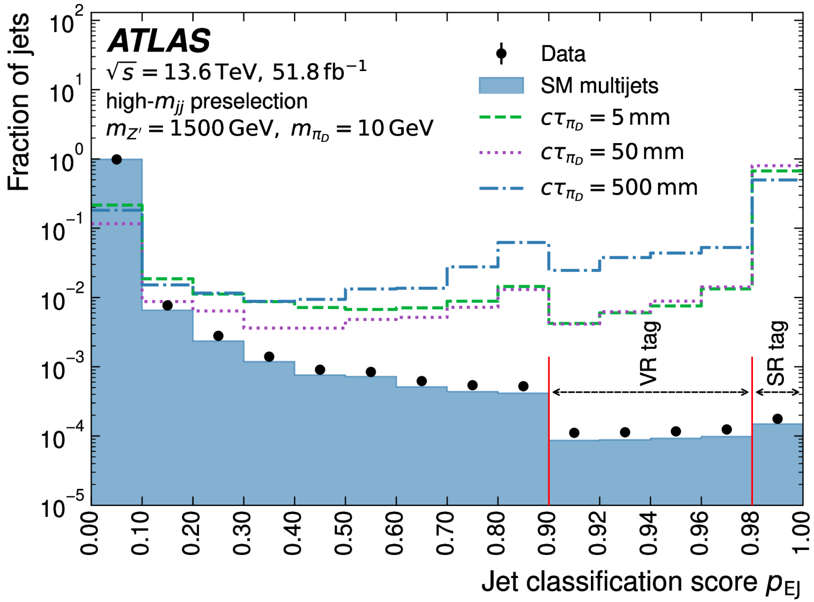
Graph Attention Network (GAT), based on GN2 b-tagging algorithm

- Track as nodes and track level features as node features
- **Total Loss = $L_{jet} + \alpha L_{trk} + \beta L_{vertex}$**



Visual representation of how jets are represented as graphs: the tracks in the jet are represented as nodes, and the tracks originating from the same vertices as the edges connecting nodes

GNN Score for identifying EJ



Bkg. Rejection = 2600, Signal Eff = 97.5 %

$p_{EJ} > 0.98$

	CR	SR
	Events with $N_{EJ} < 2$	Events with $N_{EJ} \geq 2$
	Used to estimate background	Search for EJ signal

Background Estimation: Mistag Map Method

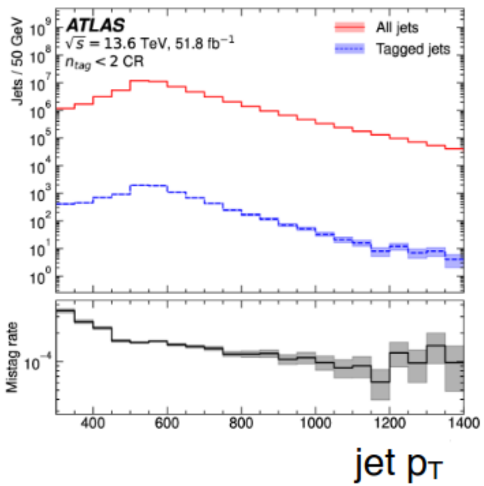
Data-driven mistag rate approach is adopted

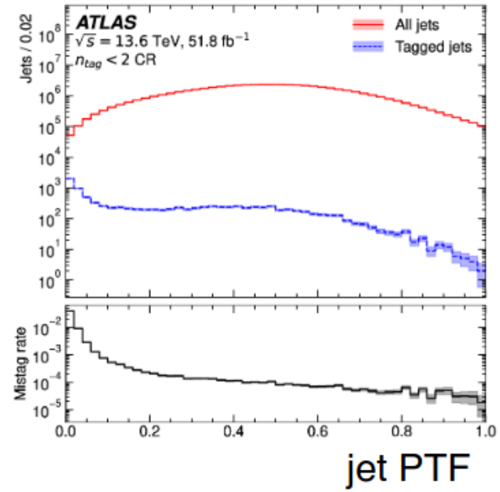
- Probability that a bkg jet is misclassified as signal
- mistag rate determined in CR
- 2D efficiency maps using jet-level variables correlated with the GNN score

$$P(0\ \text{tag}|\text{event}) = \prod_{i=1}^{n_{\text{jet}}} (1 - P(\text{tag}|j_i))$$

$$P(1\ \text{tag}|\text{event}) = \sum_{i=1}^{n_{\text{jet}}} P(\text{tag}|j_i) \times \prod_{k\neq i}^{n_{\text{jet}}} (1 - P(\text{tag}|j_k))$$

$$P(\geq 2\ \text{tag}|\text{event}) = 1 - P(1\ \text{tag}|\text{event}) - P(0\ \text{tag}|\text{event})$$





Key Limitation

→ Difficult to model the correlation with tagger score in a few variables

→ varying mistag rate parameterisation (nbjets, ntrk, nDV) → 23% (60%)

Results

No excess was found in any of the defined region

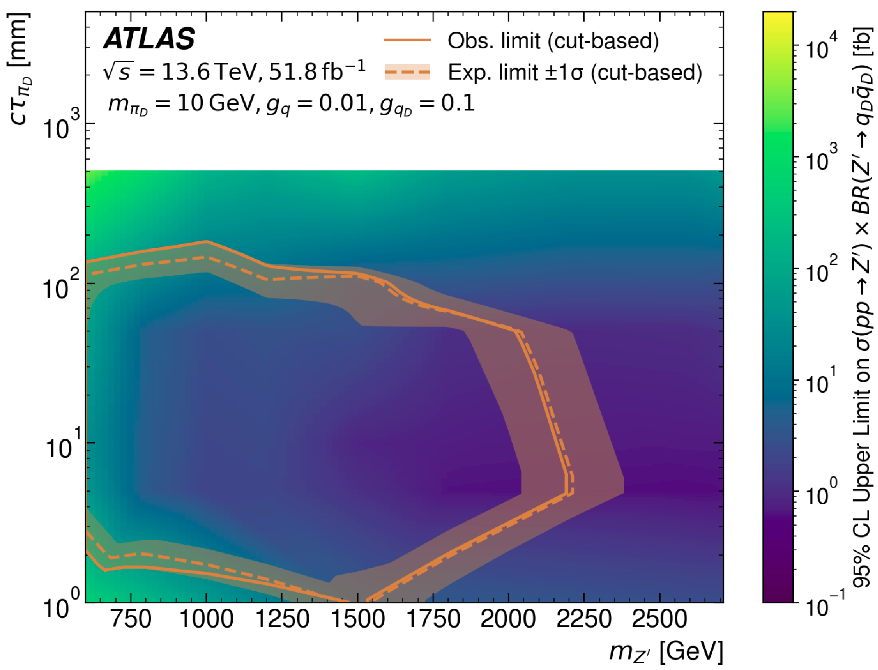
Strategy	Region	Prediction (± stat ± syst)			Observed yield
Cut-based	High-m_{jj}	7.5	±1.1	±1.1	8
	Low-m_{jj}	16.6	±4.8	±4.8	10
ML-based	High-m_{jj}	4.5	±0.3	±2.8	3
	Low-m_{jj}	31.8	±0.8	±7.5	24

Simultaneous fit of low and high mjj regions for each approach in the limit interpretation

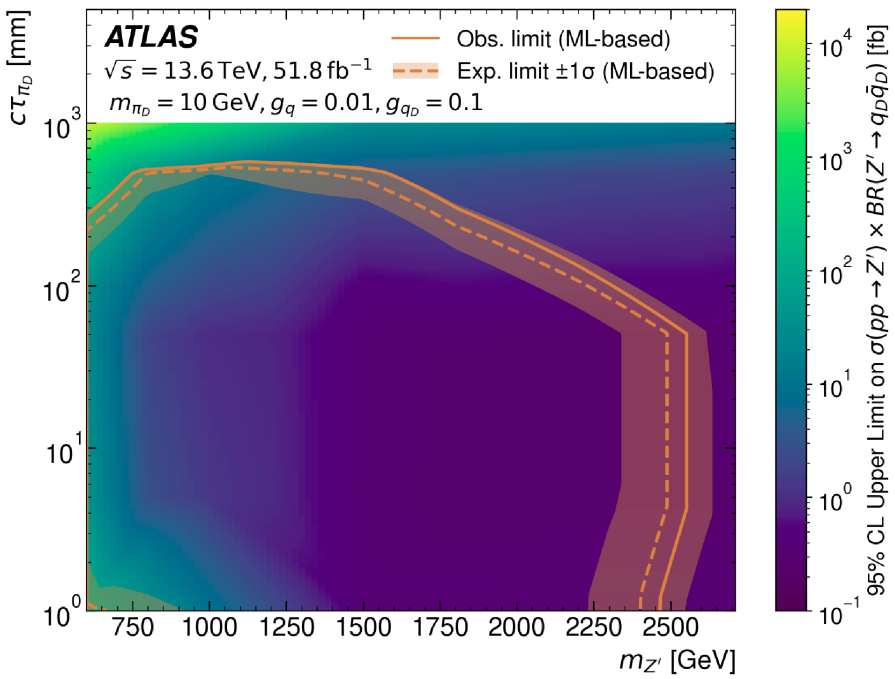
Results : s-channel

- Results are interpreted as cross-section x BR limits vs proper decay length and Z' mass
  - ML-based set limits about a order of magnitude more stringent than cut-based

Cut based



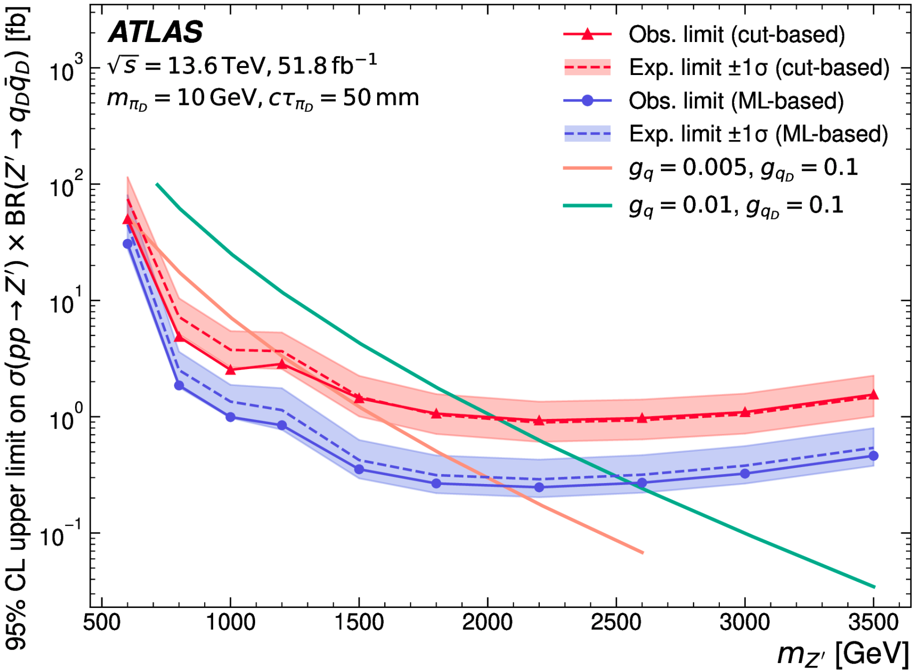
ML based



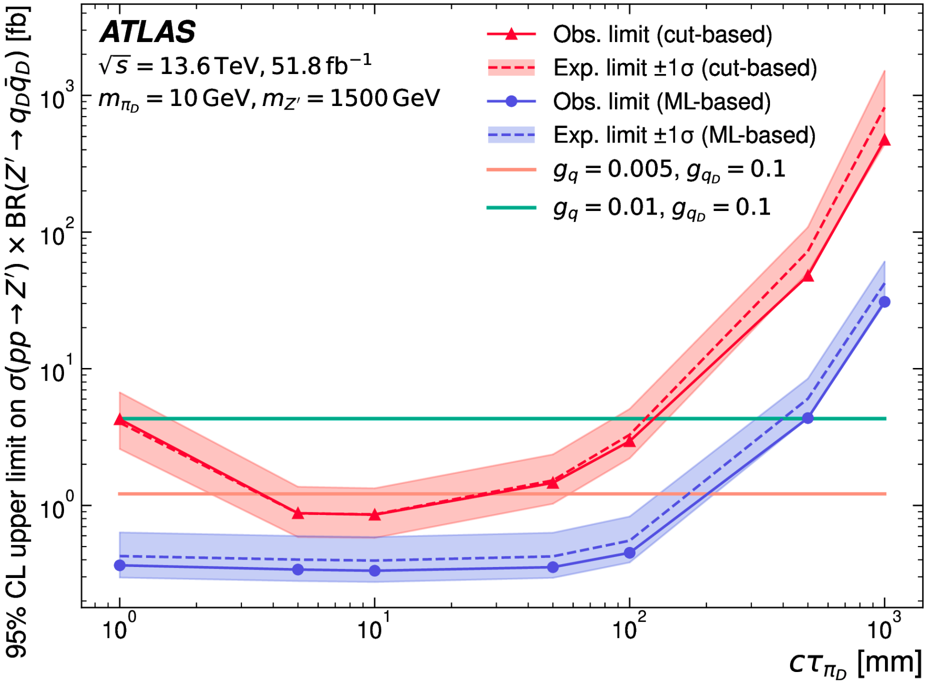
Results: s-channel

- EMJ trigger allows to probe low Z' mass phase-space
- Track reconstruction limitations at boosted regime

Limit on cross-section as $M_{Z'}$ varies



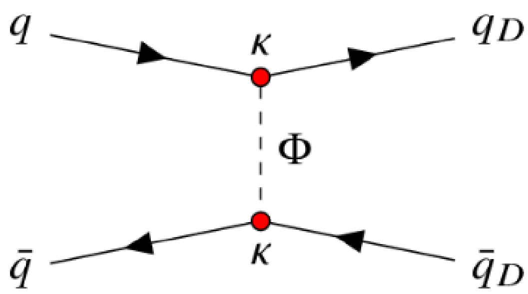
Limit on cross-section as cτ varies

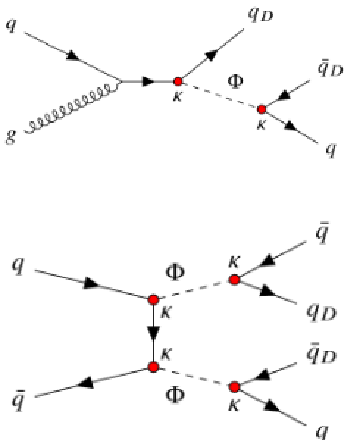


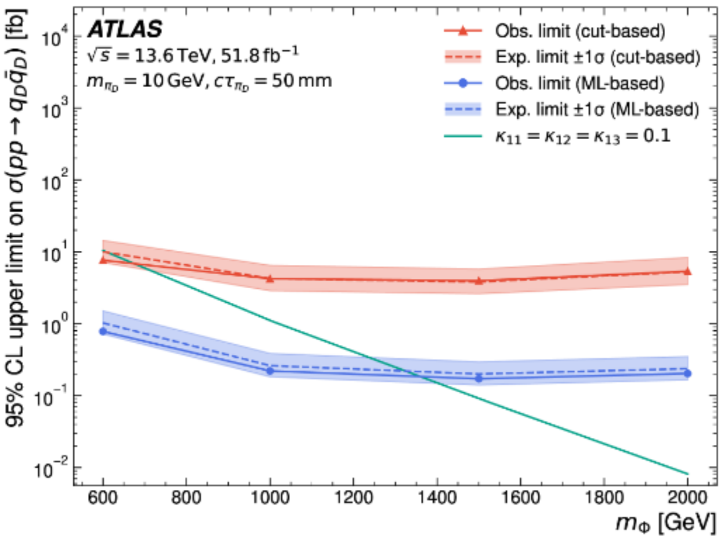
Results: t-channel

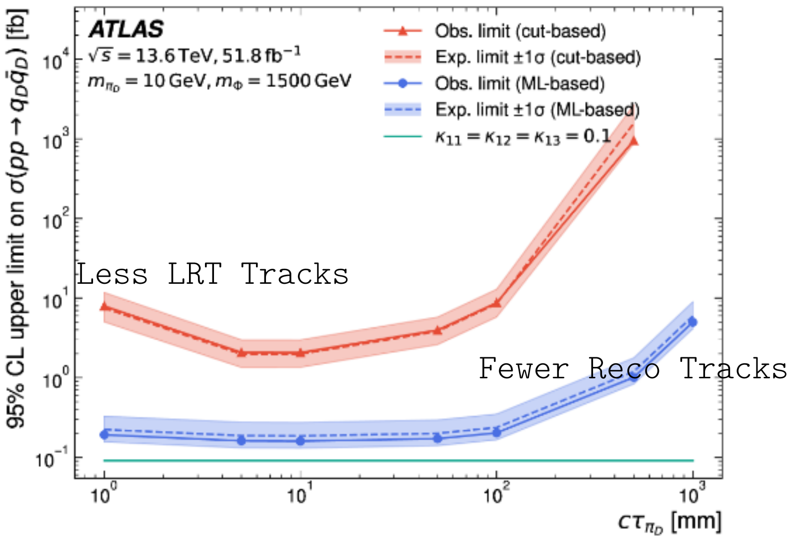
- Low sensitivity in cut-based strategy due to presence of extra SM jets
  - ABCD method based on leading and subleading jet of the event

- ML-based approach is magnitude more sensitive









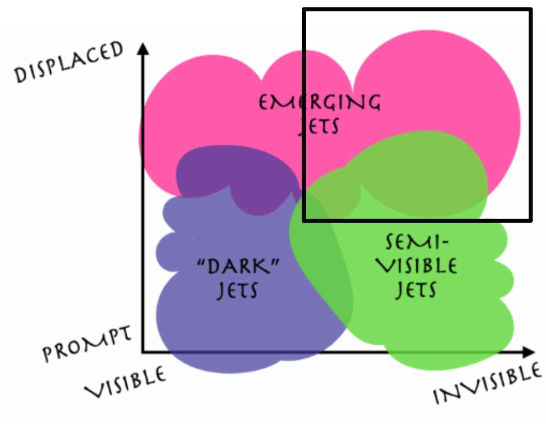
o

Looking ahead: towards Full Run 3

- Full Run 3 analysis is ongoing to improve on several aspects of the previous analysis beyond luminosity benefits

Stay tuned!

Benchmark Model
- Consistent model parameters across experiments
- Simplified model
- Semivisible Emerging Jets?



Cut-based tagger
- Going away from ABCD method to tackle stat limitation issue
- Improvement in both models
- Relying on jet-level properties

What's new for Full Run 3 Analysis?

- Harmonize background estimation
- Improvement in Mistag Map method
  - BDT-based mistag rate

Anomaly Detection Channel to be sensitive to a wide range of models

Re-interpretation
- Challening due to extensive model parameters
- ML-models to map truth level - reco level selection efficiency?

Back up

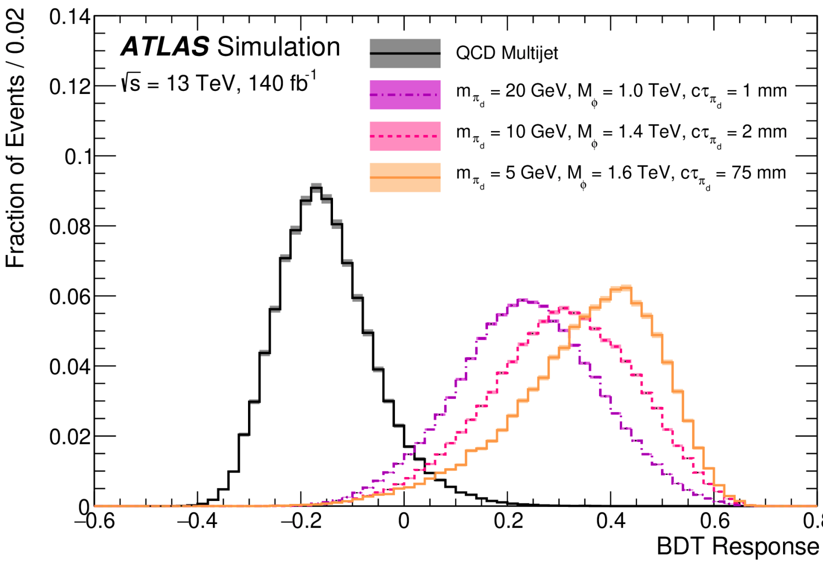
Run-2 search

Jet-level-ML

Observable	Description
E_{Jet}	Jet energy, measured directly from calorimeter deposits
$m_{\mathrm{Jet}} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2}$	Jet mass, invariant mass from the sum of topological clusters
$W_{\mathrm{Jet}} = \frac{1}{p_{\mathrm{T}}^{\mathrm{Jet}}} \sum_i p_{\mathrm{T},i} \Delta R_i$	Jet width, p_{T}-weighted angular spread of topological clusters
η_{Jet}	Jet pseudorapidity

Event-level-ML

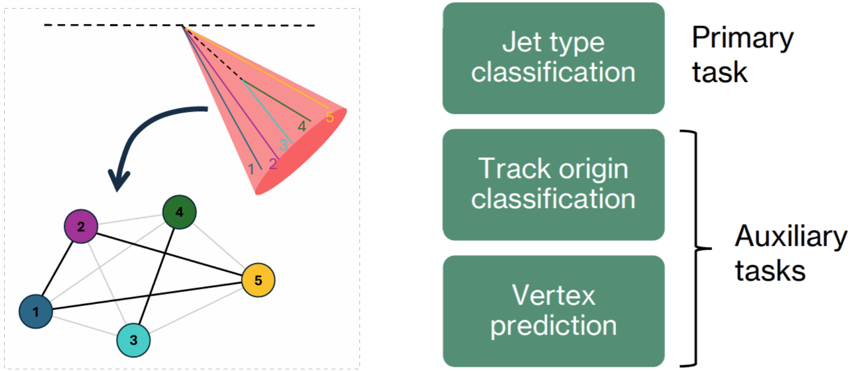
Observable	Description
R_{Jet}^{1-4}	BDT response of the four highest p_{T} jets, individually
$H_{\mathrm{T}} = \sum_{j=1}^{4} p_{\mathrm{T},j}$	Scalar sum of the four highest p_{T} jet transverse momenta
minPTF	The minimum PTF value in the four highest p_{T} jets
N_{Jet}	Reconstructed jet multiplicity
S_{T}	Transverse Sphericity



ML Strategy: Graph Neural Network

Graph Attention Network (GAT), based on GN2 b-tagging algorithm

- Leveraging ATLAS GN2 flavour tagging architecture based on transformers
- Each node (track hit) receives weighted inputs from neighbors
- Weights (α_{tt}) determine feature importance for LLP tagging.
- **Total Loss = $L_{jet} + \alpha L_{trk} + \beta L_{vertex}$**



Visual representation of how jets are represented as graphs: the tracks in the jet are represented as nodes, and the tracks originating from the same vertices as the edges connecting nodes

G.Gustavino's LHC Seminar

low-level track inputs

up to 200 tracks per jet from QCD and EJ

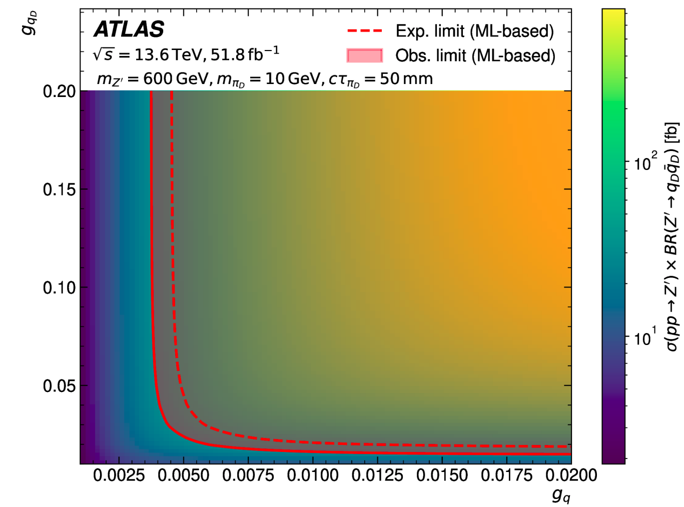
jet: Jet η; tracks: d_0 through $N_{\mathrm{SCT\ shared}}$

Input	Description
Jet η	Jet pseudorapidity
d_0	Track closest distance to PV in transverse plane
$z_0\sin(\theta)$	Track closest distance to PV in longitudinal plane
$\Delta\phi$	Azimuthal angle of the track, relative to the jet ϕ
$\Delta\eta$	Track pseudorapidity, relative to jet η
q/p	Track charge over momentum
$\sigma(\phi)$	Uncertainty in track ϕ
$\sigma(\theta)$	Uncertainty in track θ
$\sigma(q/p)$	Uncertainty in track q/p
$d_0/\sigma(d_0)$	signed d_0 significance
$z_0/\sigma(z_0)$	signed z_0 significance
$N_{\mathrm{PIX\ hits}}$	Number of Pixel hits per track
$N_{\mathrm{SCT\ hits}}$	Number of SCT hits per track
$N_{\mathrm{IBL\ hits}}$	Number of innermost pixel layer hits
$N_{\mathrm{PIX\ shared}}$	Number of Pixel shared hits
$N_{\mathrm{SCT\ shared}}$	Number of SCT shared hits

Run 3 search

Event Selections

<table>
<tr><th>Preselection</th><th>Low-m_{jj}</th><th>High-m_{jj}</th></tr>
<tr><td>leading jet p_T</td><td>–</td><td>$>$ 520 GeV</td></tr>
<tr><td>subleading jet p_T</td><td>–</td><td>$>$ 300 GeV</td></tr>
<tr><td>trigger-matched jet</td><td>$p_T > 250$ GeV, PTF < 0.04</td><td>–</td></tr>
<tr><td>leading and subleading jet PTF</td><td colspan="2">< 1.0</td></tr>
<tr><td>leading and subleading jet ΔPTF</td><td colspan="2">< 0.4</td></tr>
<tr><td>leading and subleading jet N_{vtx}</td><td colspan="2">≥ 1</td></tr>
<tr><td>leading and subleading jet ECF2/p_T</td><td>–</td><td>$>$ 40 GeV</td></tr>
<tr><td>m_{jj}</td><td>$<$ 1 TeV</td><td>$>$ 1 TeV</td></tr>
<tr><td>number of jets</td><td colspan="2">≥ 2</td></tr>
<tr><th>Signal Region</th><th></th><th></th></tr>
<tr><td>leading and subleading jet N_{subjet}</td><td>≥ 3</td><td>–</td></tr>
<tr><td>leading and subleading jet PTF</td><td>–</td><td>< 0.2</td></tr>
</table>



Limit on g_q vs g_{qd}