HEPData for the ATLAS Dark Jet Resonance Search

Danielle Wilson-Edwards

On behalf of the analysis team: Marcella Bona, Joe Davies, Caterina Doglioni, Alison Elliot, Marie-Hélène Genest, Nikolina Ilić, Nathan Lalloué, Dilia María Portillo-Quintero, Jia Jian Teoh

Dark Jet Resonance Search Analysis Strategy:

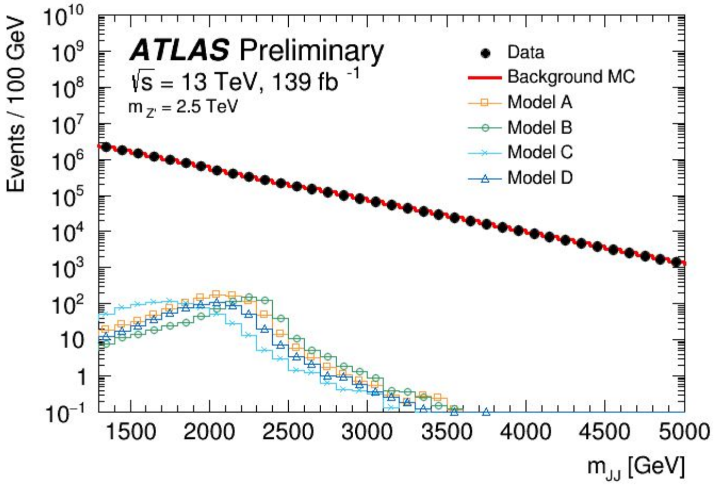
ATLAS-CONF-2023-047

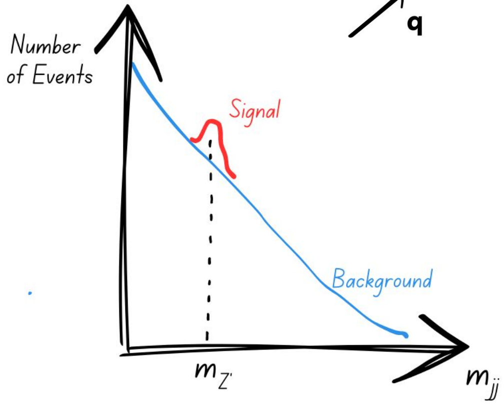
See Dilia's talk for more information!

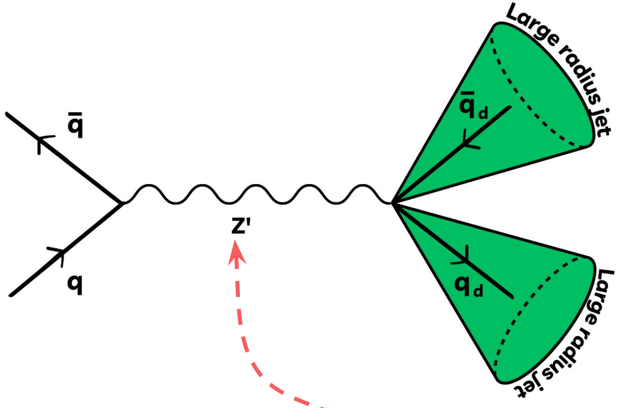
- **Reconstructed Objects:** Two Large radius jets (Trimmed LCTopo jets)
- **Trigger on Large-R jets**
- **Preselection cuts:** $|\eta(j_{1,2})| < 2.0$ $p_{\mathrm{T}}(j_1) > 500$ GeV , $p_{\mathrm{T}}(j_2) > 400$ GeV $m(j_{1,2}) > 50$ GeV $m_{\mathrm{JJ}} > 1.3$ TeV

Signal scenario: Two dark jets

1) Tag them using substructure information
2) Look for a resonance over dijet background







Signal masses per model from mZ'=1.5 ->3.5 TeV in bins of 250 GeV, and bins of 500 GeV for higher masses

ATLAS HEPData Recommendations: Statistically Limited Searches

- **Preservation of analysis logic:**
  - cutflow table for each signal region, for several reproducible or well-defined signal samples
- **Preservation of numerical values:**
  - central values for the search regions in each bin
- **Systematic uncertainties/correlations:**
  - A simple stat/syst breakdown is sufficient if the analysis is dominated by statistical uncertainties
- **SM background predictions**
- **Limits**
  - Exclusion limits + necessary information for a theorist to emulate the analysis

Di-jet invariant mass distribution

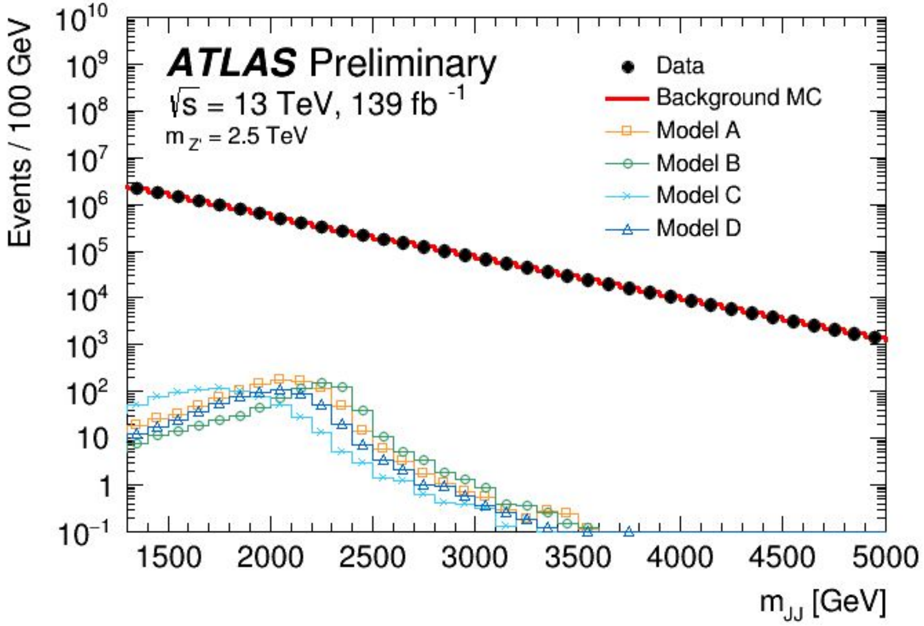
preselection

$$m_{JJ} > 1.3 \text{ TeV}$$

$$m(j_{1,2}) > 50 \text{ GeV} \quad |\eta(j_{1,2})| < 2.0$$

$$p_T(j_1) > 500 \text{ GeV}, p_T(j_2) > 400 \text{ GeV}$$

- Data, signal, background counts for mjj in preselection
- Signal mjj shape before signal region cuts for reinterpretation validation
- Normalisation:
    - simulated background is normalised to the data
    - signals are normalised to a production cross-section of 10 fb.



Distributions of the number of tracks associated with the two leading jets

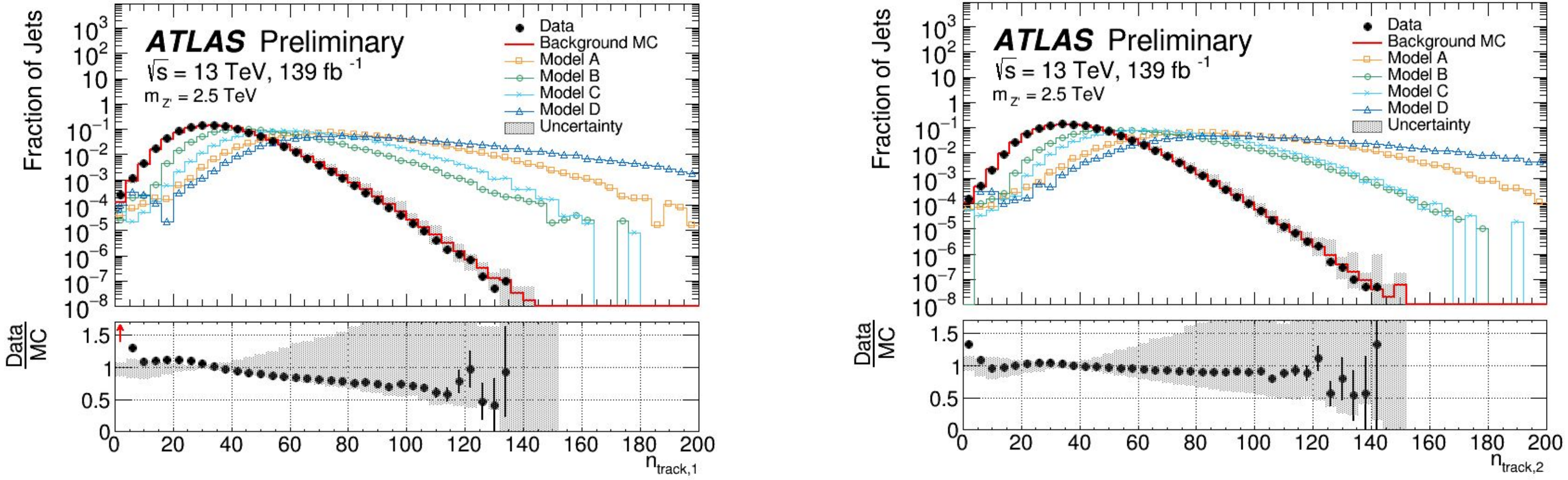
preselection

$m_{JJ} > 1.3$ TeV

$m(j_{1,2}) > 50$ GeV $\quad |\eta(j_{1,2})| < 2.0$

$p_T(j_1) > 500$ GeV , $p_T(j_2) > 400$ GeV

- The number of tracks matched to the untrimmed jets via ghost association
- Uncertainty band around background : modelling uncertainty



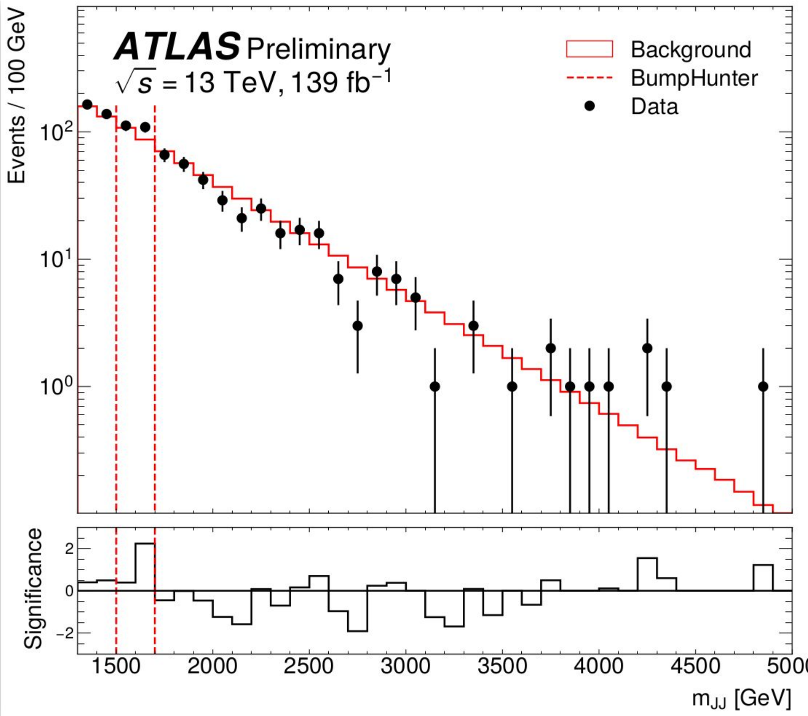
Model Independent Limits

Signal Region (SR)

$n^{\epsilon}_{\mathrm{track},1} > 0$ and $n^{\epsilon}_{\mathrm{track},2} > 0$

- Bump hunter test performed to find localised excess in model-independent manner
- Algorithm scans the distributions using a sliding window of variable width between two bins and half the total number of bins.
- Highest excess: between 1500 < mjj < 1700 GeV, with a global p-value of 0.63

- Data / background yield for mjj in the SR
  - + statisitical error on data

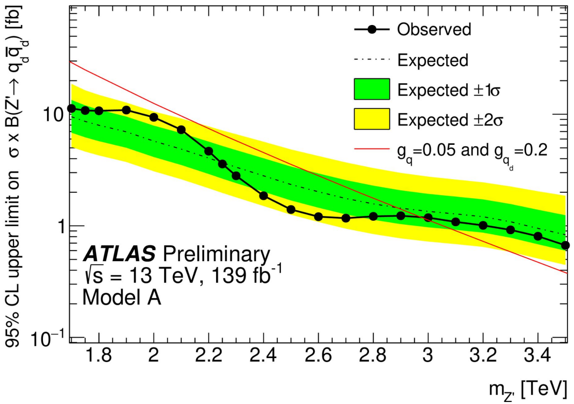


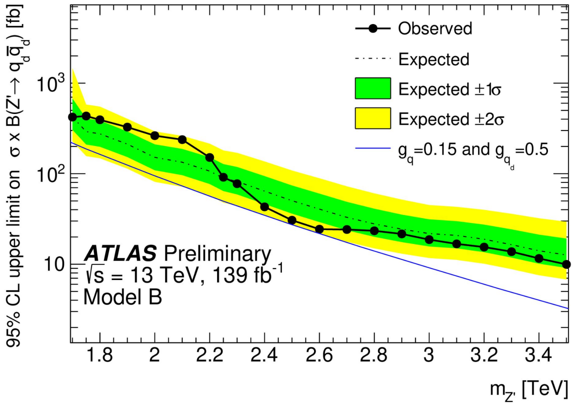
Model Dependent Limits

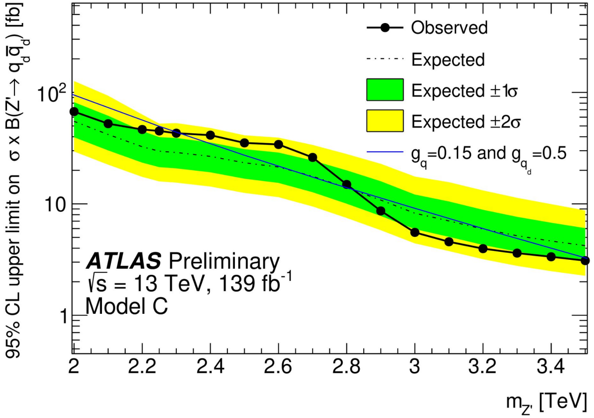
Signal Region (SR)

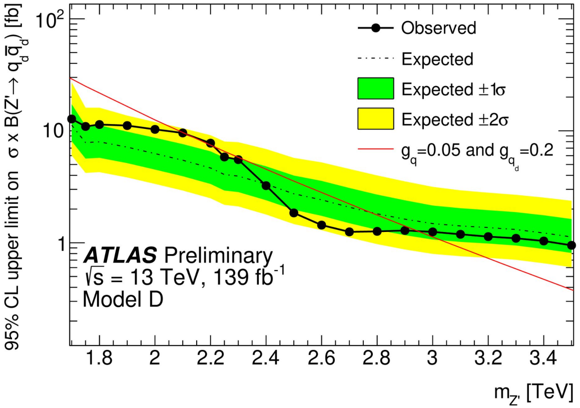
$n^{\epsilon}_{\text{track},1} > 0$ and $n^{\epsilon}_{\text{track},2} > 0$

- Expect limits
- Observed limits
- $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty ranges
- Look into storing likelihoods as **pyhf** objects
- Theory predictions typically not provided for HEPData









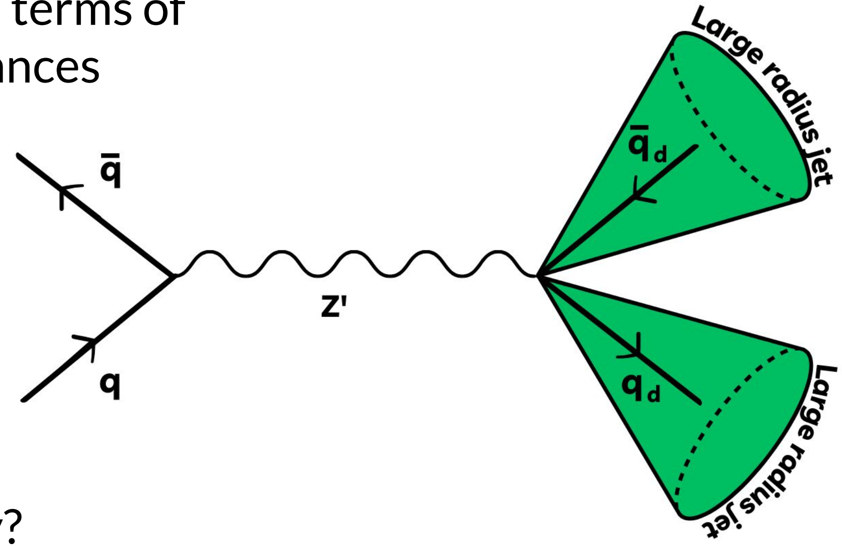
Cutflow Table : Relative Signal Efficiencies

Selection / Model	A	B	C	D
$m_{\mathrm{JJ}} > 1.3$ TeV	92.9	94.8	80.9	91.8
Jet trigger	93.0	93.2	92.5	92.3
$m_{\mathrm{J}_{1,2}} > 50$ GeV, $p_{\mathrm{T,J}_1} > 500$ GeV, $p_{\mathrm{T,J}_2} > 400$ GeV	88.5	60.0	81.3	56.1
$\lvert\eta_{\mathrm{J}_{1,2}}\rvert < 2$	99.9	99.9	100	100
$m_{\mathrm{J}_{1,2}} < 600$ GeV, $p_{\mathrm{T,J}_{1,2}} < 3000$ GeV	99.8	99.7	99.9	99.8
Signal Region ($n^{\epsilon}_{\mathrm{track},1} > 0$ and $n^{\epsilon}_{\mathrm{track},2} > 0$)	37.0	2.7	11.6	55.5

Table 3: Relative efficiency (in %) of each analysis selection with respect to the previous one for models A through D and $m'_Z = 2.5$ TeV. Approximately 90000 events were generated per signal.

Discussion Points

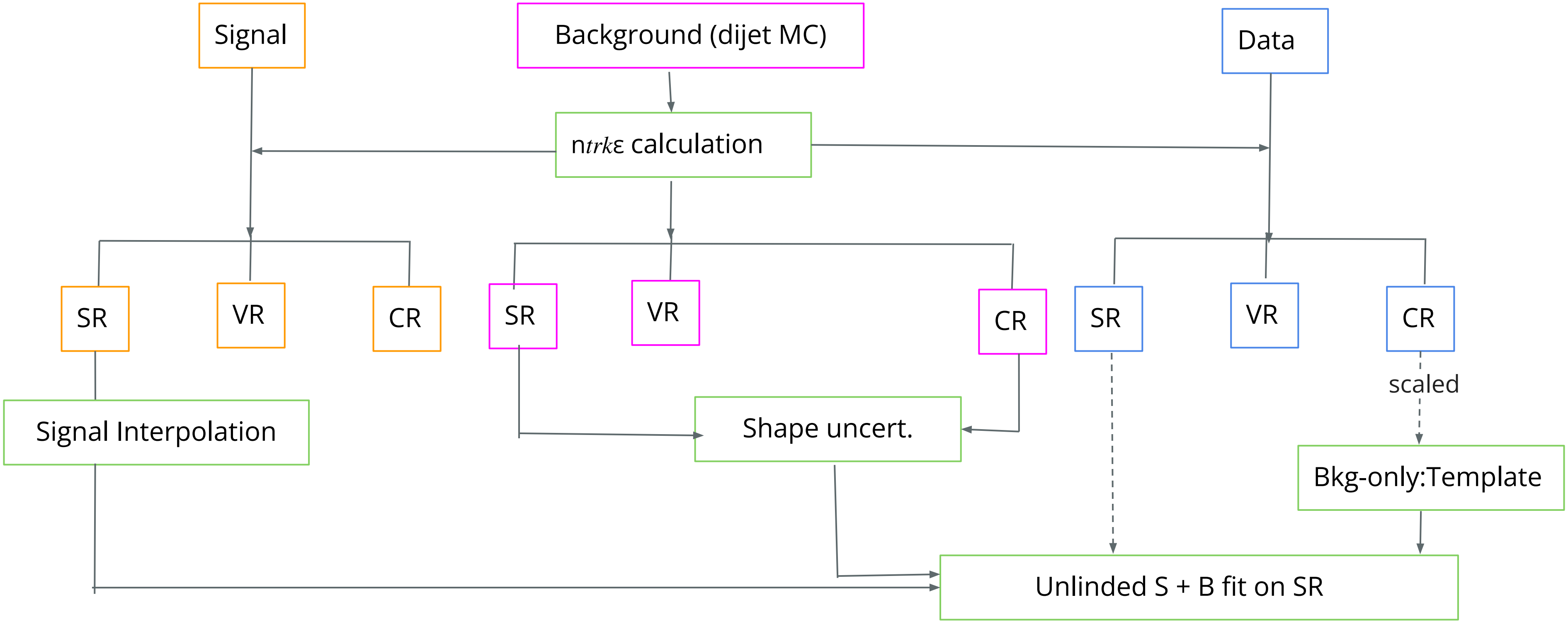
- 4 benchmark models considered: by no means an exhaustive selection of dark QCD scenarios!
    - First time interpreting ATLAS data in terms of promptly-decaying dark QCD resonances



- **Theorists:** What other information would be useful to include in our HEPData entry?
    - We would love to maximize the re-interpretability of our results.

Backup

Analysis workflow

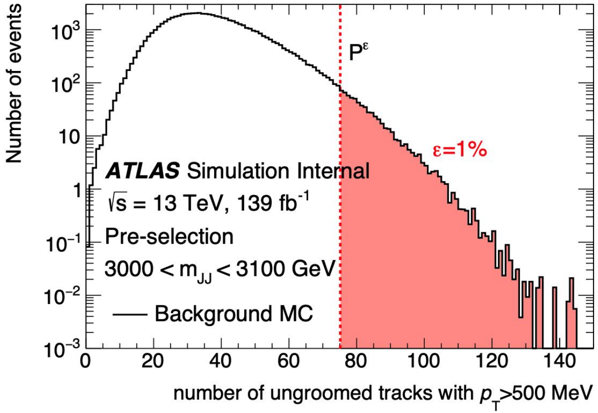


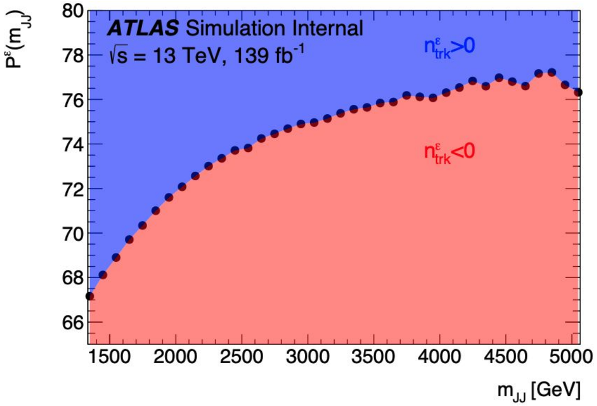
SR: n*trk*ε(j1) > 0 and n*trk*ε(j2) > 0
CR: n*trk*ε(j1) < 0 and n*trk*ε(j2) < 0
VR: n*trk*ε(j1) > 0 and n*trk*ε(j2) < 0

VR & CR scaled to expected SR when testing the fitting

Decorrelation with ntrkε

1. In bins of mjj, evaluate the cut value on **ntrk** that corresponds to a given target background efficiency (ε=1%)



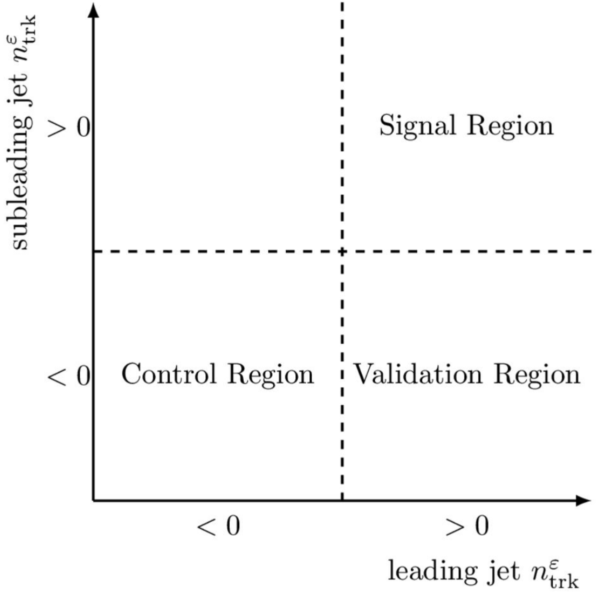


2. Fit this profile in order to get a smooth distribution of expected ntrk percentiles $\longrightarrow P^{\epsilon}$

3. Define ntrkε:

$$n^{\epsilon}_{\mathrm{trk}}(m_{\mathrm{JJ}}) = n_{\mathrm{trk}} - P^{\epsilon}(m_{\mathrm{JJ}})$$

analysis regions



Percentiles are defined from full MC * , since ..

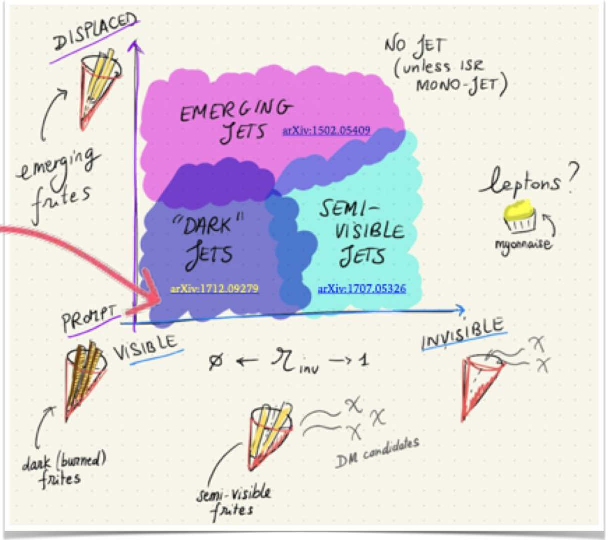
- Defining from full data => unblinding the SR
- Defining from data CR would change the shape of the mjj distribution

* *ntrk data monte carlo differences are covered with systematics (more info backup slides)*

Target Model

Unusual dijet signatures that could arise from a QCD-like dark sector.

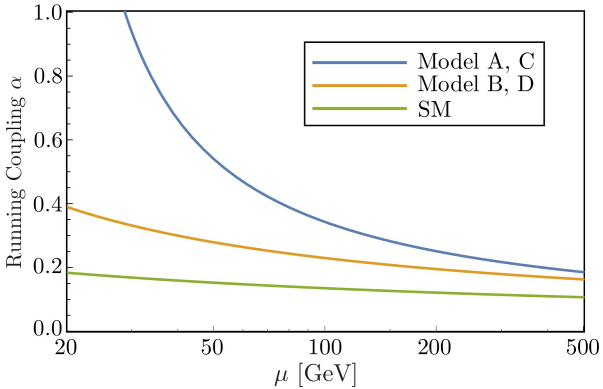
- Assume a heavy mediator Z' links dark sector and visible sector



Graphics by: C. Doglioni, K. Pedro

[arXiv:1712.09279]

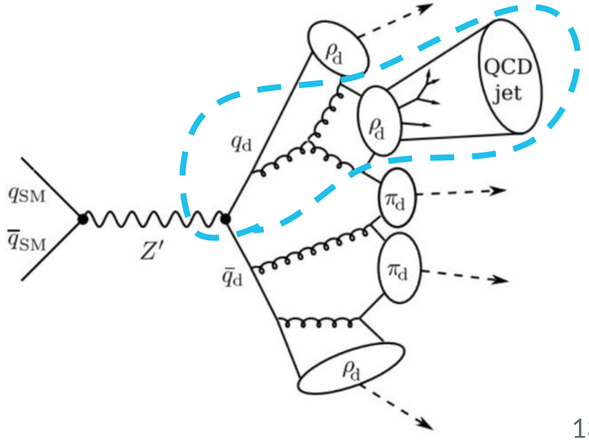
	N_d	n_f	Λ_d (GeV)	$\tilde{m}_{q'}$ (GeV)	m_{π_d} (GeV)	m_{ρ_d} (GeV)
A	3	2	15	20	10	50
B	3	6	2	2	2	4.67
C	3	2	15	20	10	50
D	3	6	2	2	2	4.67



Consider **4 benchmark models..**

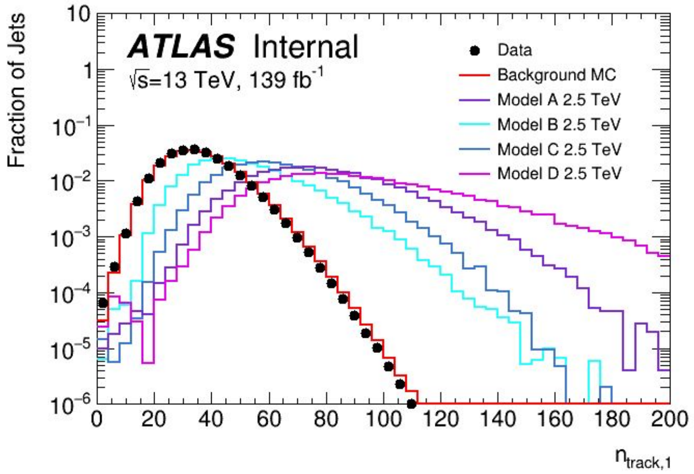
Compared to the SM, the dark jets have:

- stronger running couplings
- more soft particles (=> more tracks)
- larger jets, due to **double hadronization**

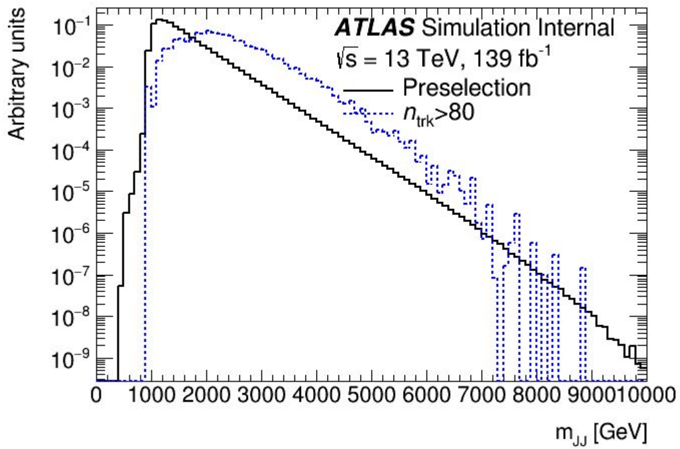


Tagging Dark Jets

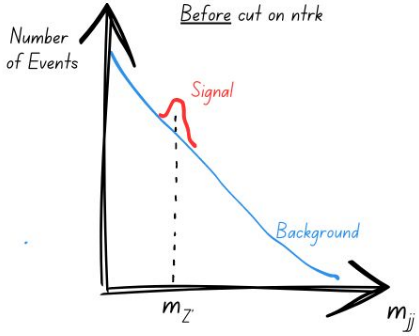
The ungroomed number of tracks inside the jet, **ntrk**, is a strong discriminating variable.

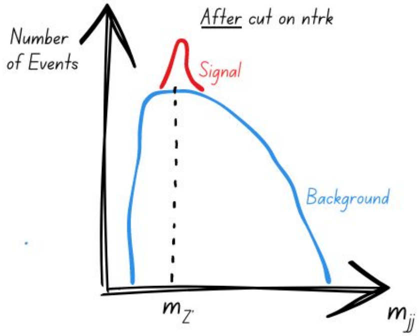


But a sharp **ntrk** cut significantly sculpts the background dijet mass spectrum



Which impacts our bump hunt!





Solution:

Define a new key discriminator n_{trk}^{ϵ} decorrelated from the dijet mass

Preselection

Triggers:
HLT_j460_a10_sub_L1J100 for 2015 and 2016
HLT_j460_a10t_lcw_jes_L1J100 for 2017 and 2018

Preselection cuts: (Twiki)
Two R = 1.0 trimmed LCTopo jets

$|\eta(j_{1,2})| < 2.0$

$p_T(j_1) > 500$ GeV , $p_T(j_2) > 400$ GeV

$m(j_{1,2}) > 50$ GeV

$p_T(j_{1,2}) < 3000$ GeV

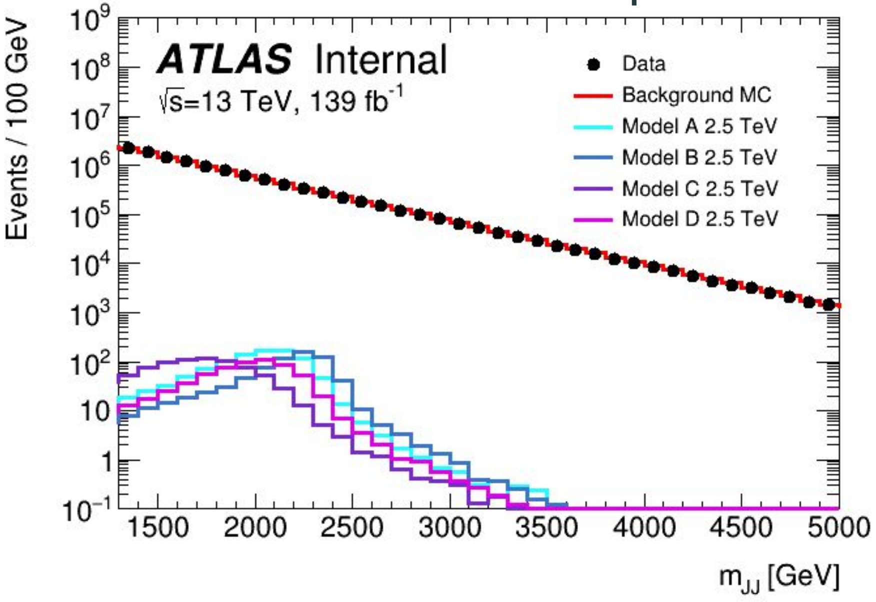
$m(j_{1,2}) < 600$ GeV

For background Monte Carlo (di-jet quality cut):

$$0.7 < \frac{p_T^{\text{leading jet}} + p_T^{\text{subleading jet}}}{2 \times p_T^{\text{leading truth jet}}} < 1.3$$

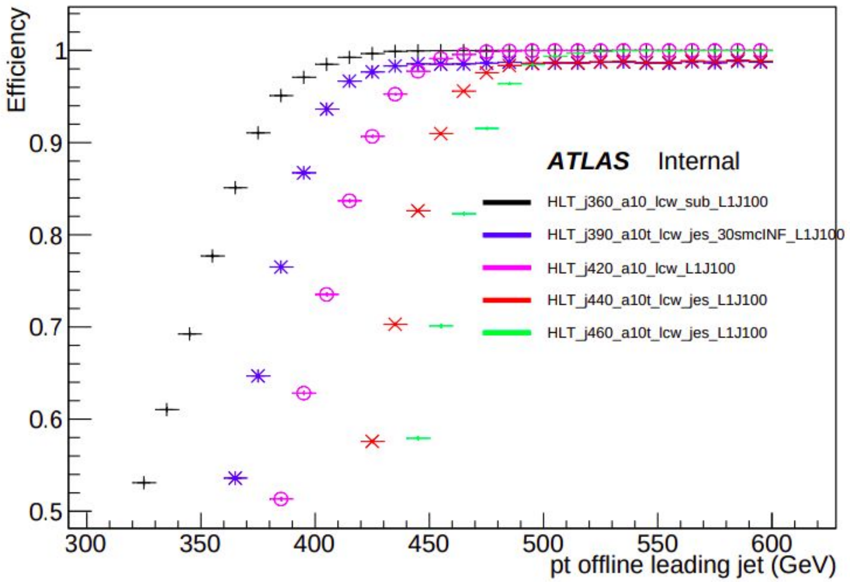
Derivation: EXOT3
Analysis framework: xAODAnaHelpers.
Useful links: Glance, Analysis code , Internal note.

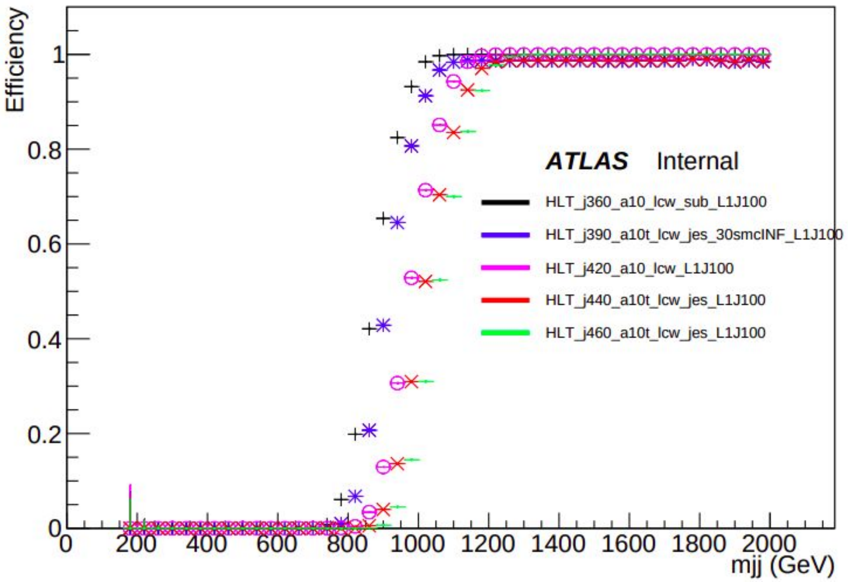
after preselection



- signal distributions are:
  - relatively broad
  - shifted to lower values than Z' mass
- no trimming => narrower signal peak BUT would increase background significantly

Trigger Turn Ons



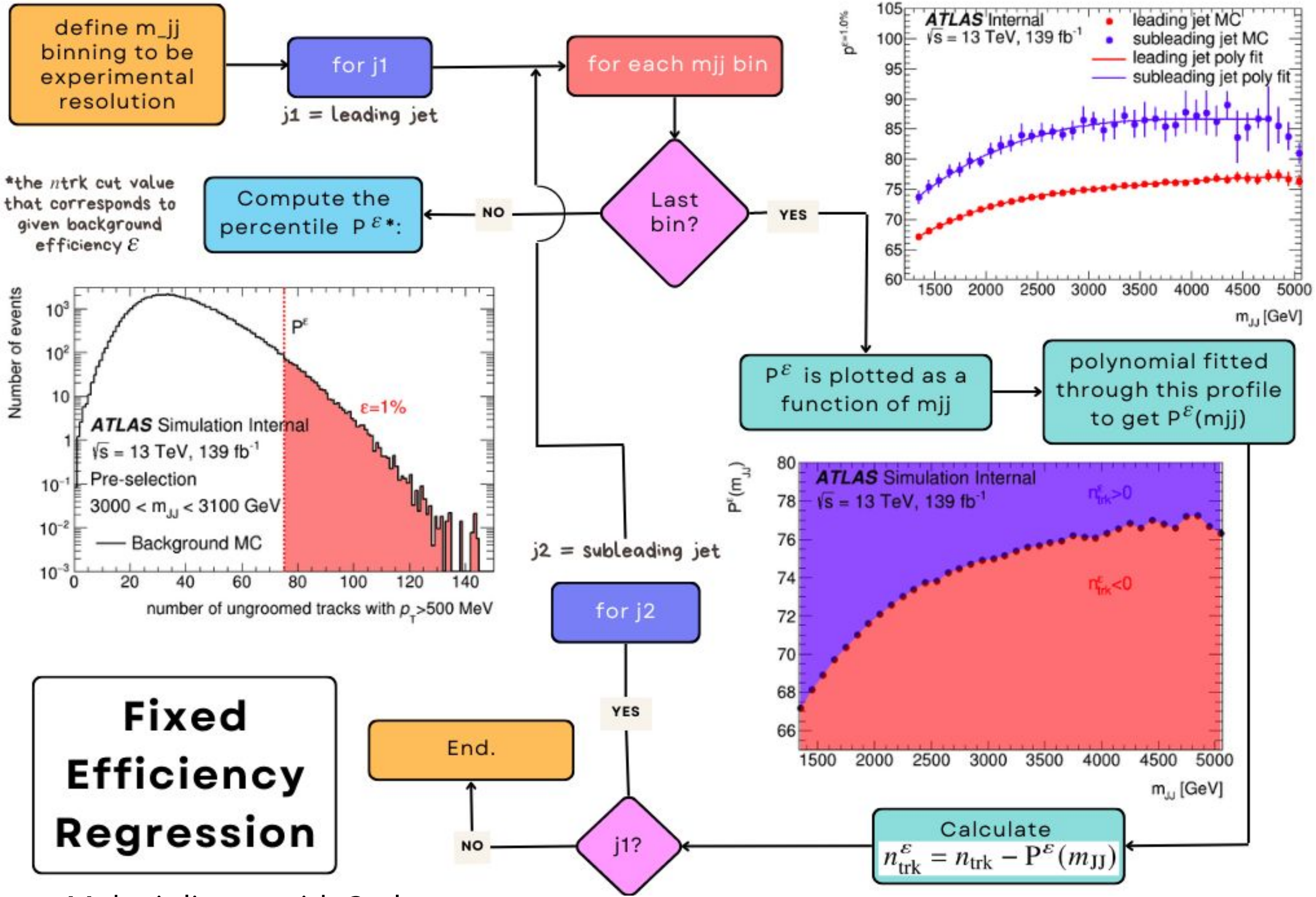


Decorrelation with ntrkε

Percentiles are defined from full MC, since ..

- Defining from full data => unblinding the SR
- Defining from data CR would change the shape of the mjj distribution

n_{trk} data monte carlo differences are covered with systematics



Make it linear with 3 plots...