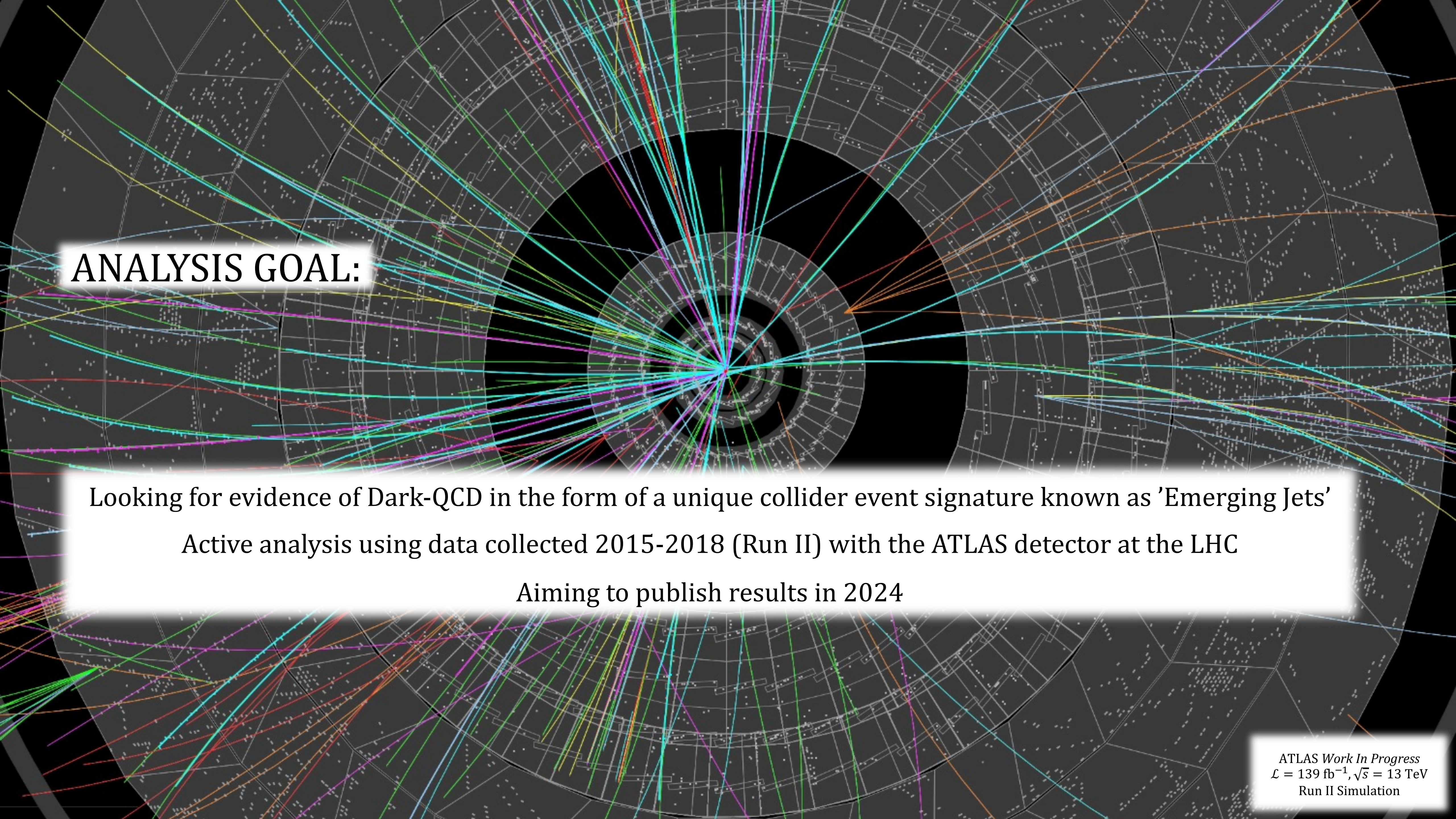
SEARCH FOR EMERGING JETS WITH THE ATLAS DETECTOR AT THE LHC

WINTER NUCLEAR & PARTICLE PHYSICS CONFERENCE

IAN RAMIREZ-BEREND

17 FEBRUARY 2024



ANALYSIS GOAL:

Looking for evidence of Dark-QCD in the form of a unique collider event signature known as ’Emerging Jets’

Active analysis using data collected 2015-2018 (Run II) with the ATLAS detector at the LHC

Aiming to publish results in 2024

THE EMERGING JET SIGNATURE

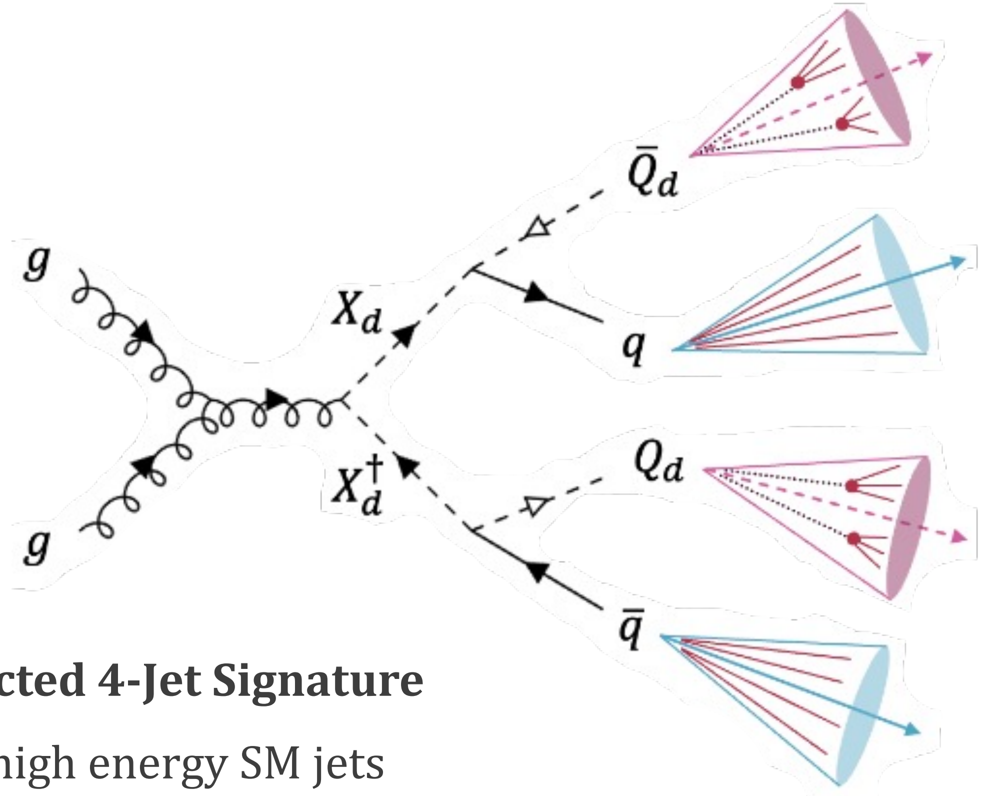
- $pp \rightarrow X_d^\dagger X_d \rightarrow 2\ q + 2\ Q_d$

  Dark Jets ($2\ Q_d$)

  SM Jets ($2\ q$)

- X_d: TeV-Scale Dark Mediator

- Q_d form GeV-Scale dark hadrons (long lived)

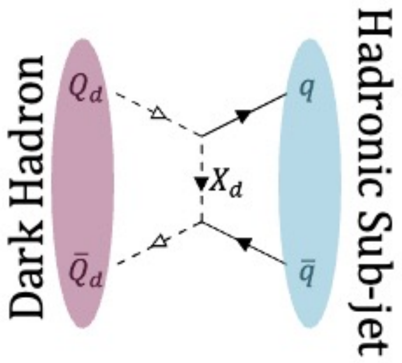


Expected 4-Jet Signature

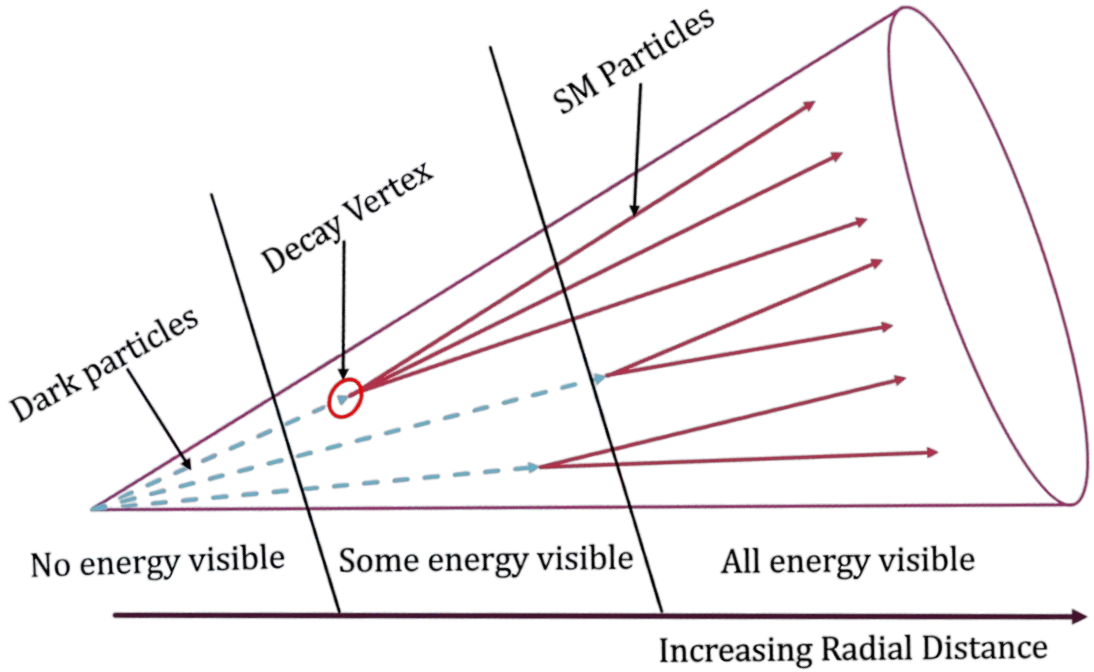
- 2 high energy SM jets
- 2 ‘Emerging” jets:
  - many displaced vertices
  - Few tracks close to the collision point

Carleton ATLAS

DARK JETS

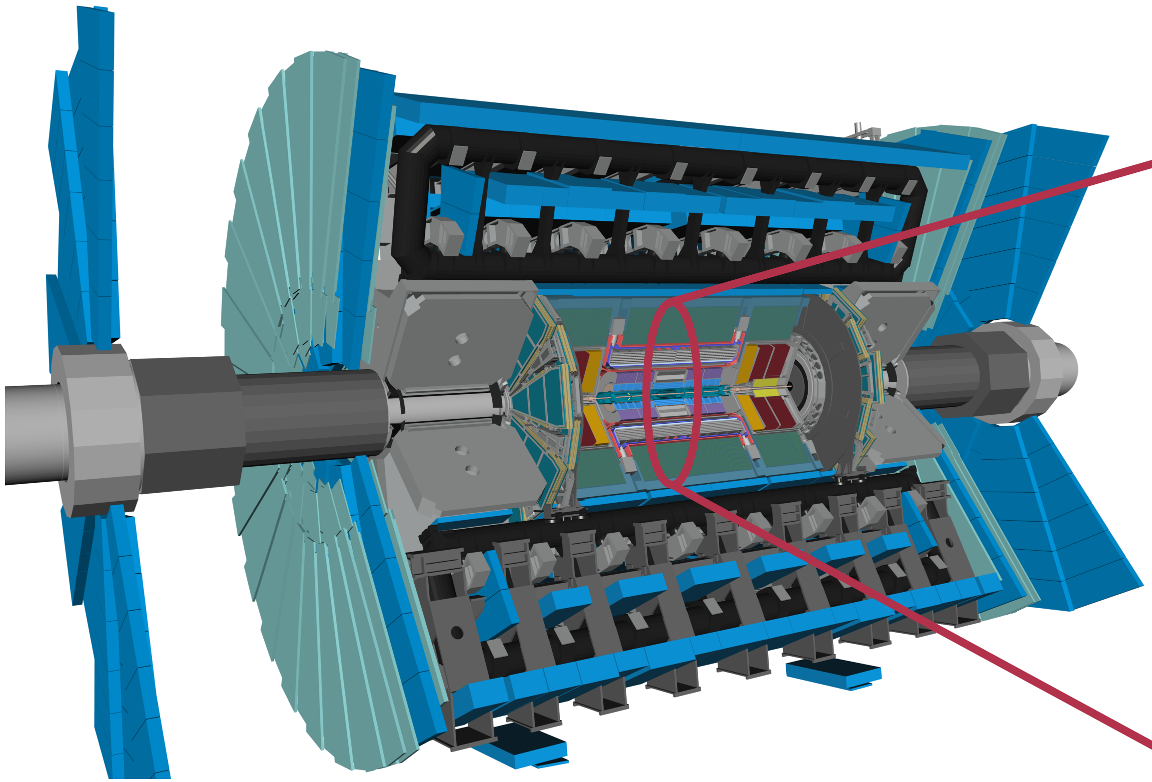


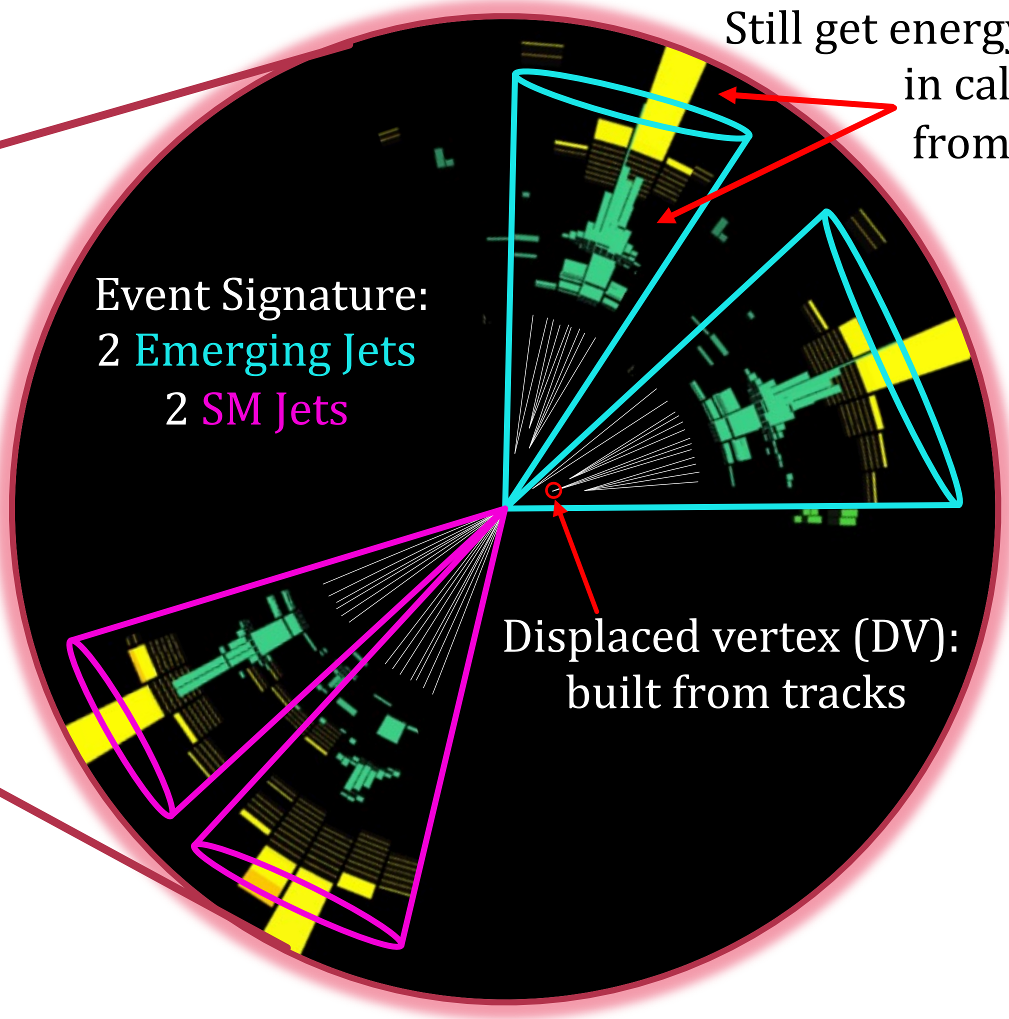
- X_d decay at interaction point
- Dark jet made of π_d (invisible)
- Each π_d decay leaves **Displaced Decay Vertex (DV)**



Emerging Jet with 3 DVs

EMERGING JETS TOPOLOGY IN ATLAS



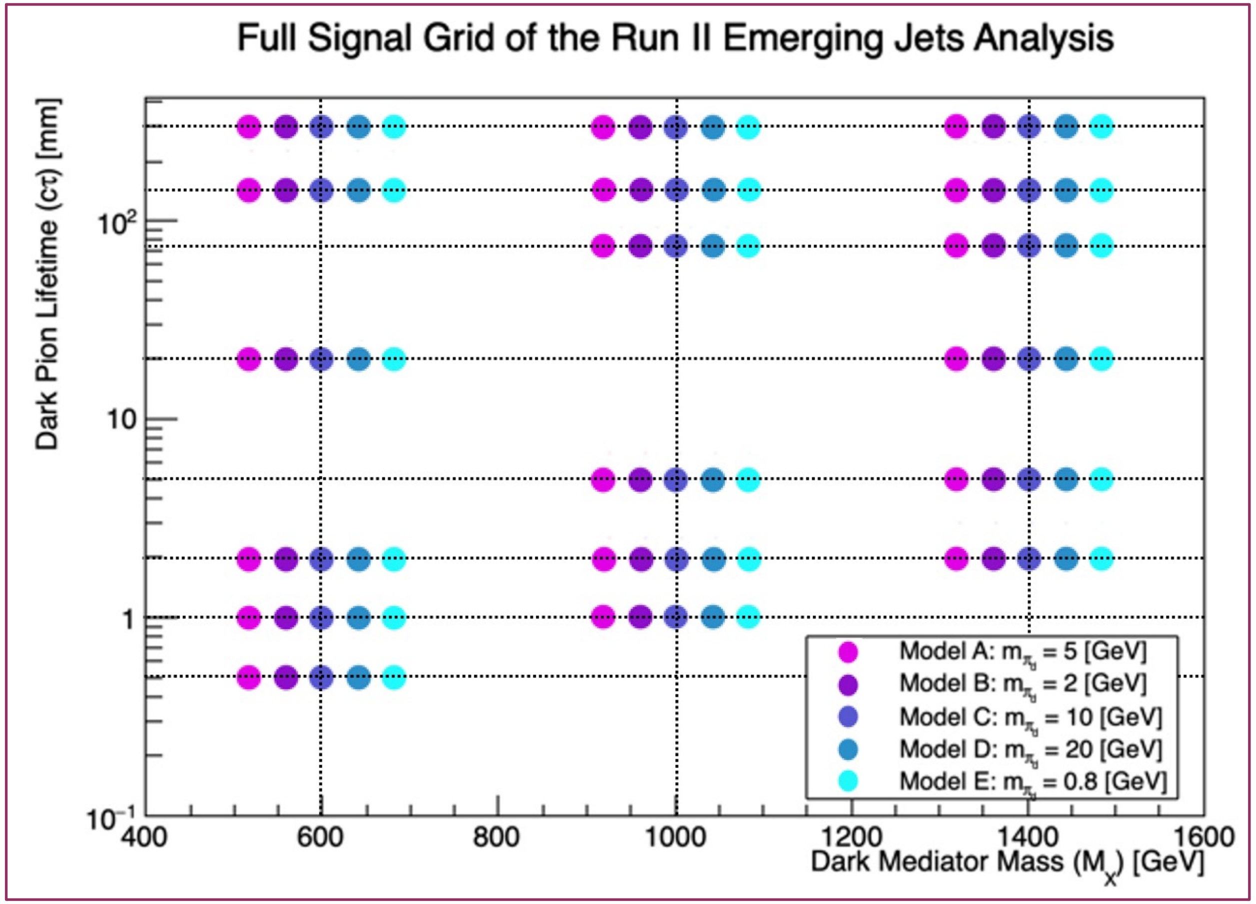


SM background events:
4-jet events with very few DVs
(QCD multi-jet processes)

Carleton ATLAS

SIMULATED SIGNAL MODELS

- Three parameters of interest change the phenomenology of emerging jets:
  - Dark Mediator Mass (M_X)
  - Dark Pion Lifetime ($c\tau$)
  - Dark Pion Mass (m_{π_d})
- Define a 90 signal point grid
  - 5 GeV-scale dark pion masses: $0.8 - 20$ GeV
  - 3 TeV-scale mediator masses: $600, \; 1000, \; 1400$ GeV
  - 6 lifetimes per M_X in the range: $0.5 - 300$ mm



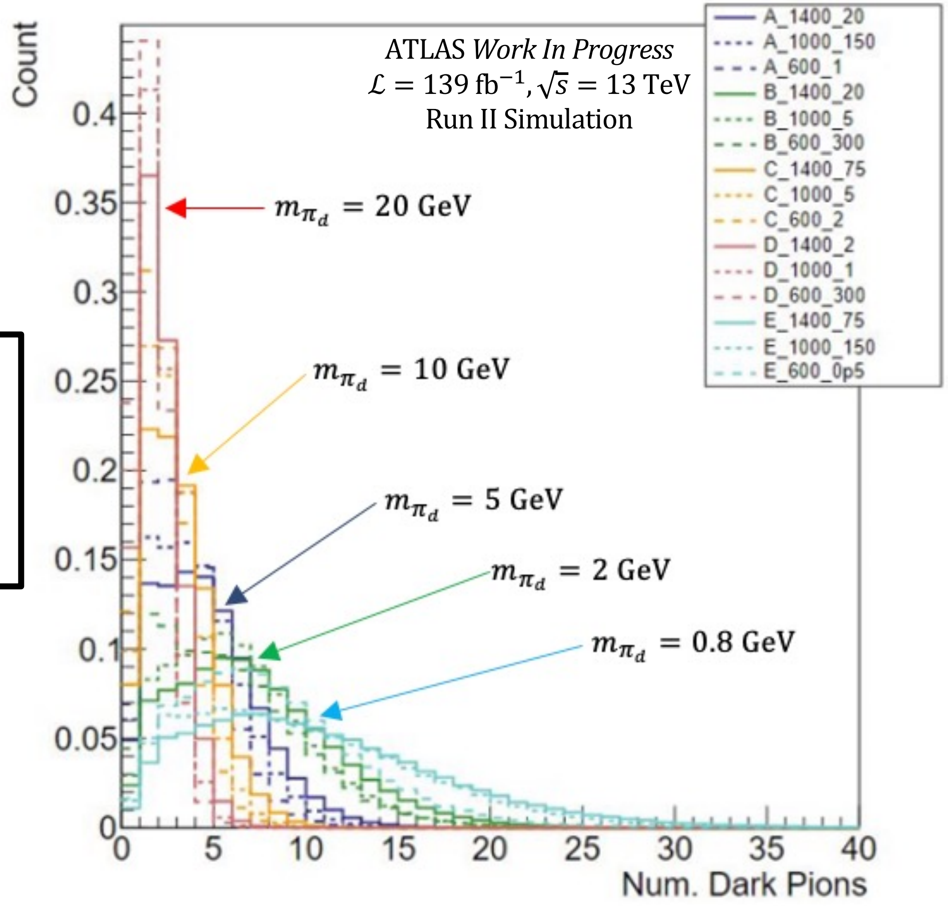
Carleton ATLAS

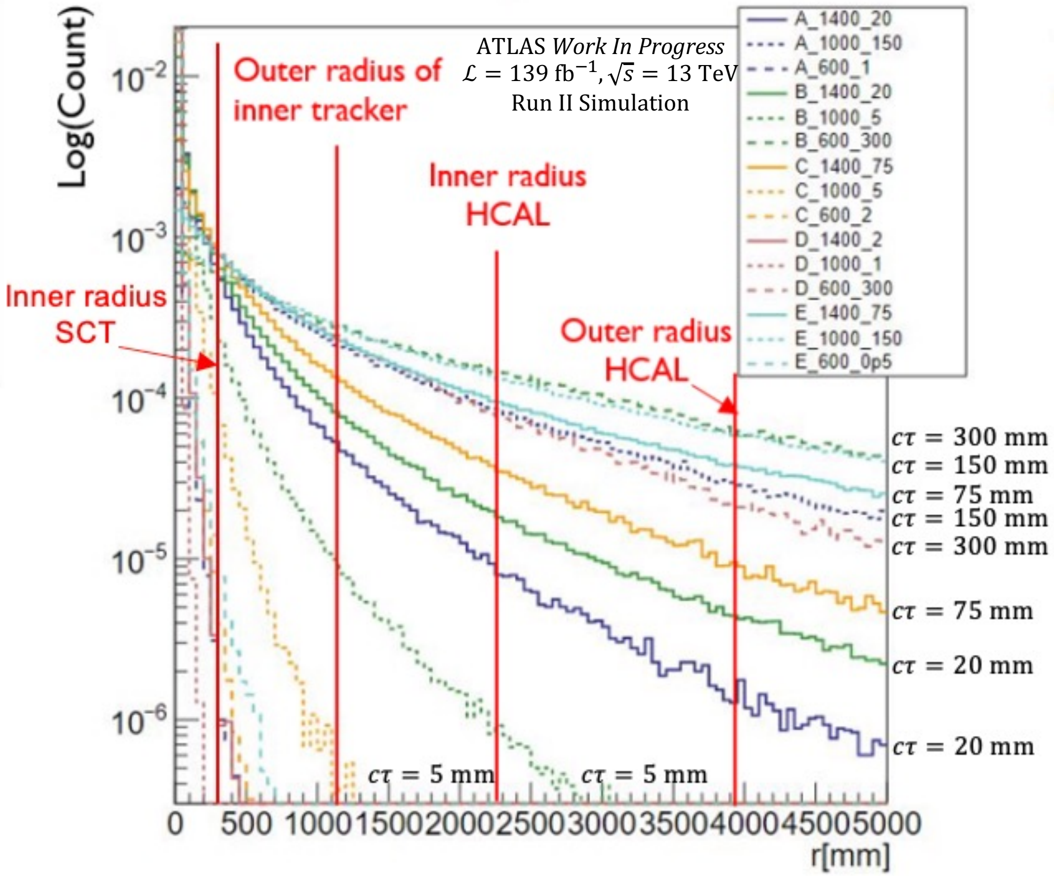
EMERGING JET KINEMATICS

$$m_{\pi d} \propto \frac{1}{N_{\pi d}} \rightarrow \frac{1}{\text{nDV}}$$

$$c\tau_{\pi d} \propto \mathrm{r}_{\pi d}^{decay} \rightarrow \mathrm{r}_{DV}$$

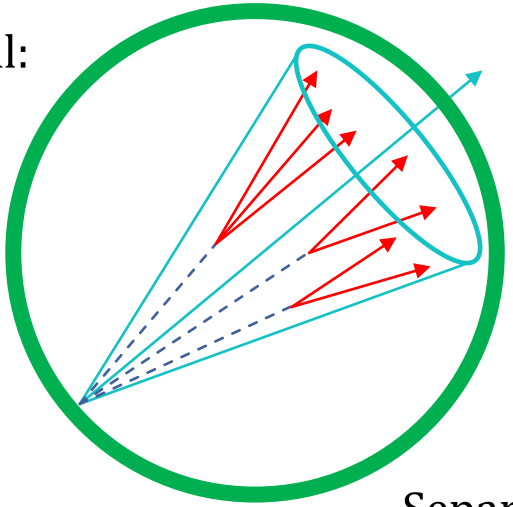
Signal events can look very different depending on the parameters of the model $(m_{\pi d}, M_X, c\tau_{\pi d})$

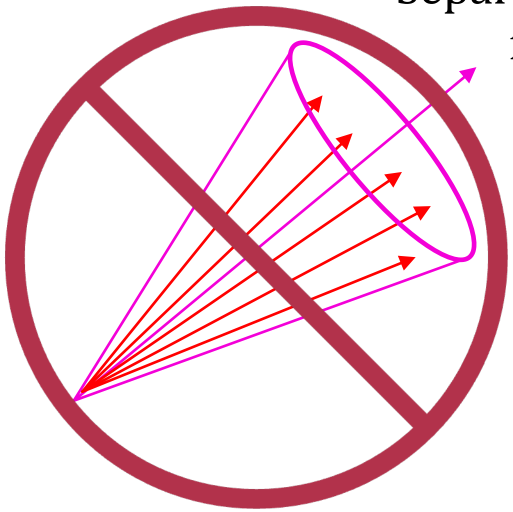




Carleton ATLAS

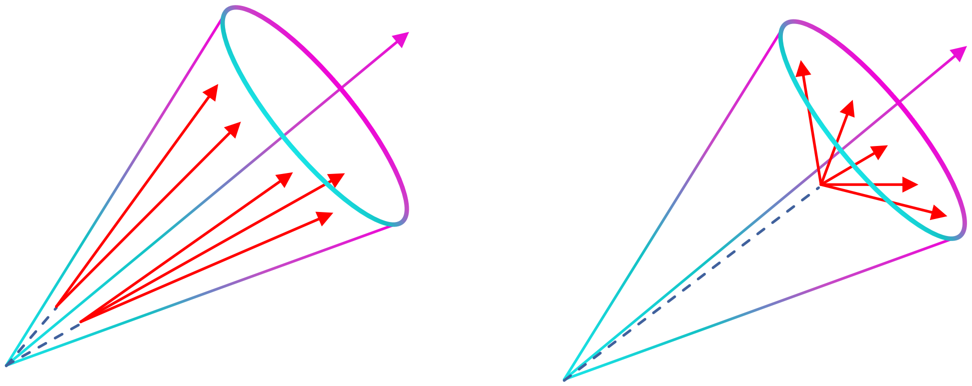
The Goal:



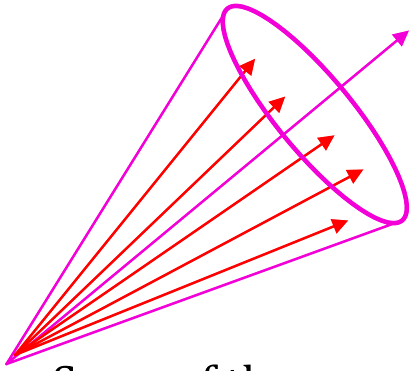


Separate Emerging Jets from SM Jets

The Problem:



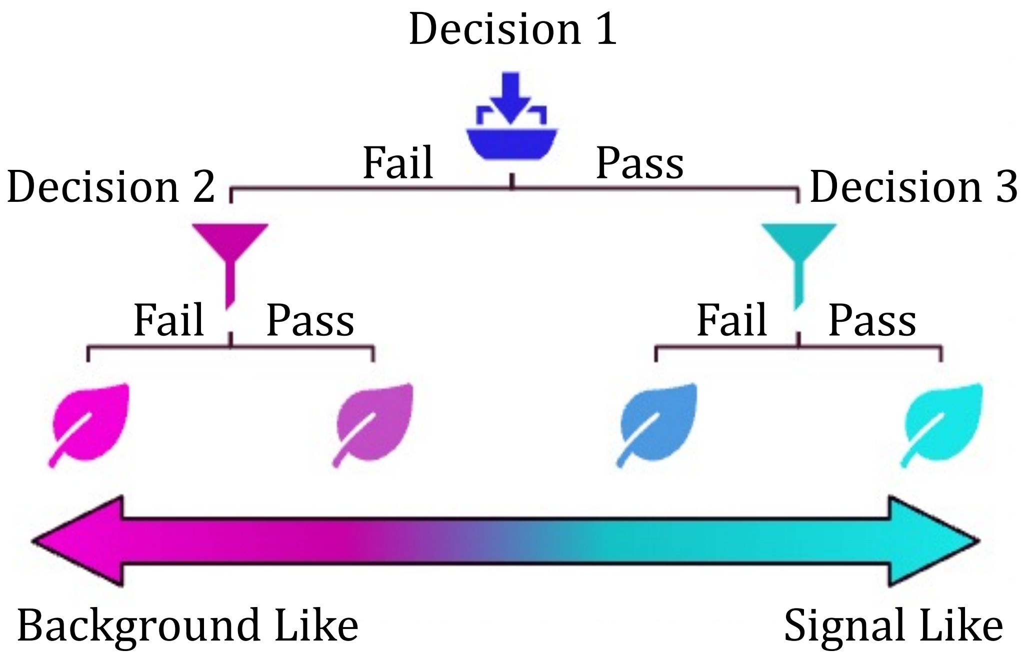
Signal Models can look very different



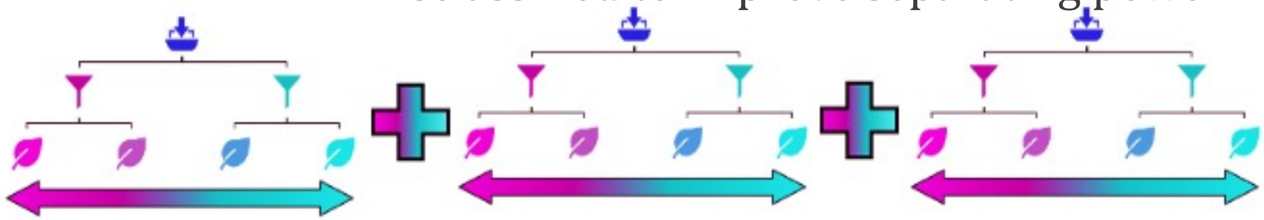
Some of them even resemble the background

The Solution: Use Machine Learning!

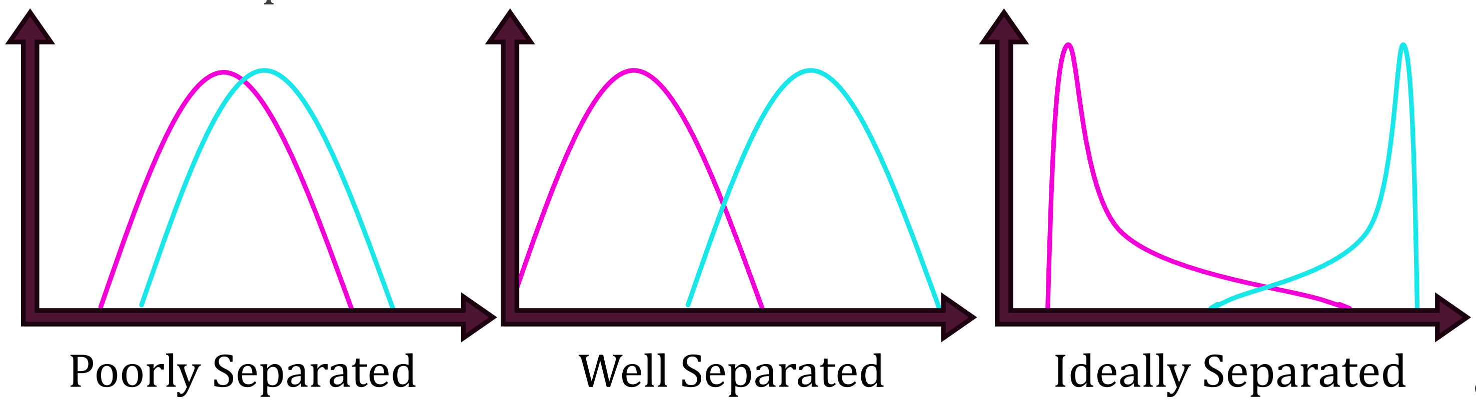
BOOSTED DECISION TREES (BDTs)



- Pass events through a series of binary decisions
- Sort events into background-like (-1) and signal-like (1)
- *Gradient Boosting*: re-train BDT several times on events which were misclassified to improve separating power

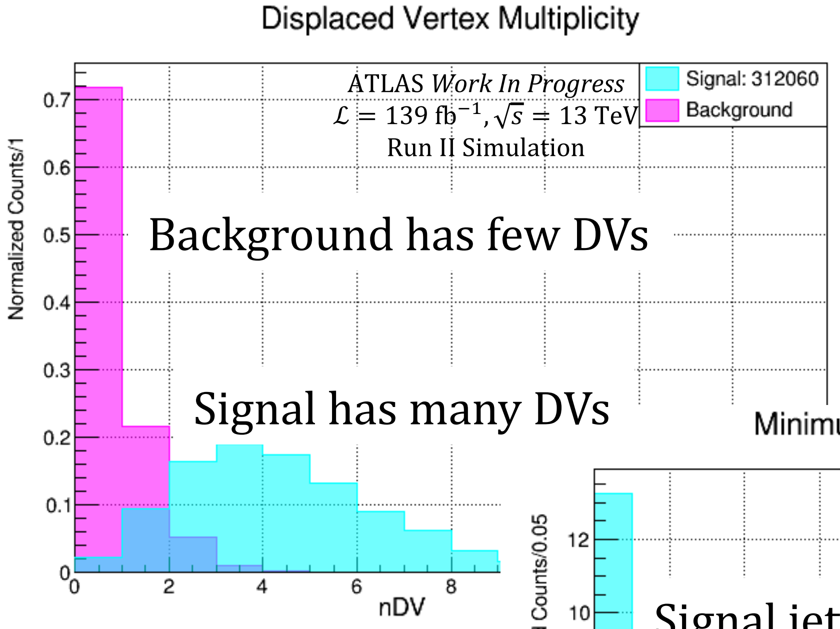


- Each event given a score from [-1, 1], aggregate of all events gives BDT Response Distribution:

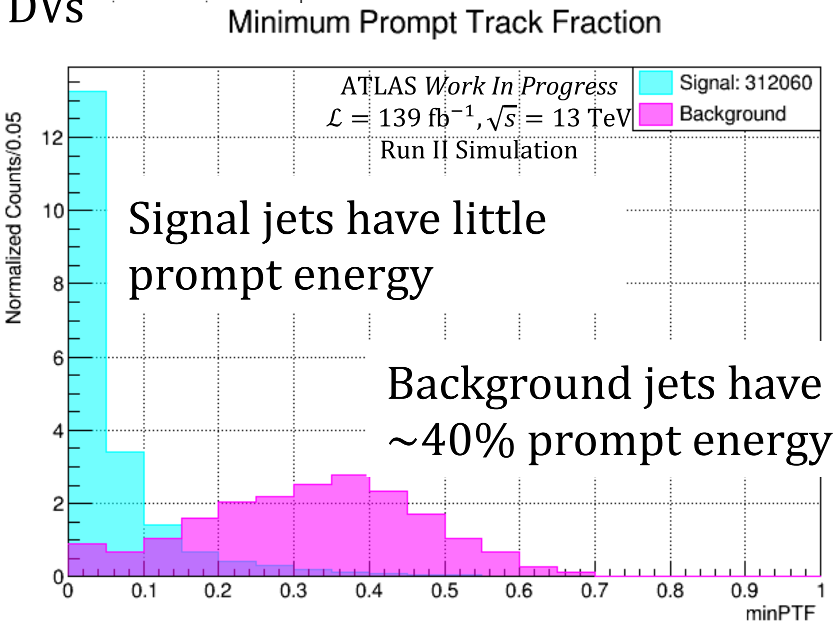


Carleton ATLAS

KEY ANALYSIS VARIABLES



- Baseline event signature: 4 high p_T jets
  - Require at least 4 jets per event, sum of 4-jet $p_T \geq 1000$ GeV
- Expect signal jets to have many DVs → use **jet-matched DV multiplicity**



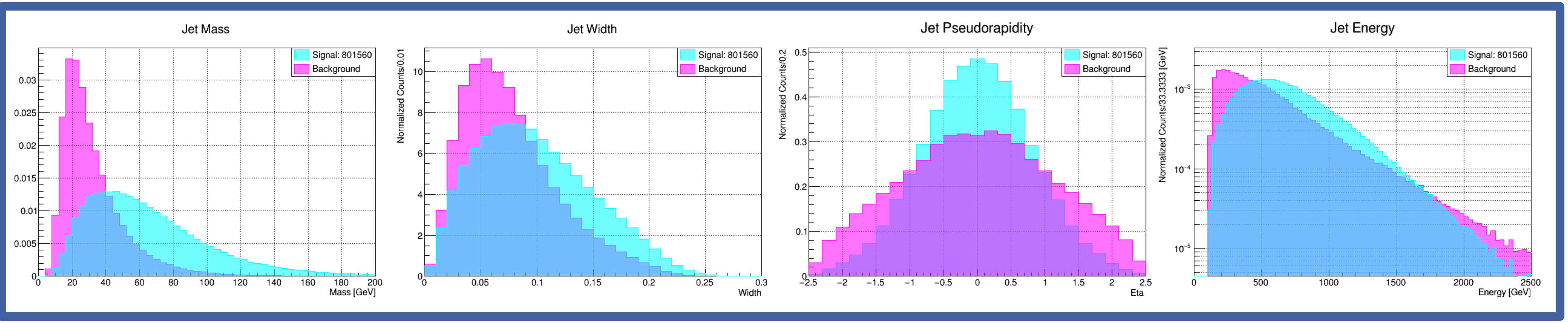
- Expect background jets to have lots of prompt tracks, signal to have few
- **Prompt Track Fraction (PTF)**: measure of jet's energy from prompt tracks
  - Use minimum of 4 leading jets to separate signal and background

EMERGING JETS BDT STRUCTURE

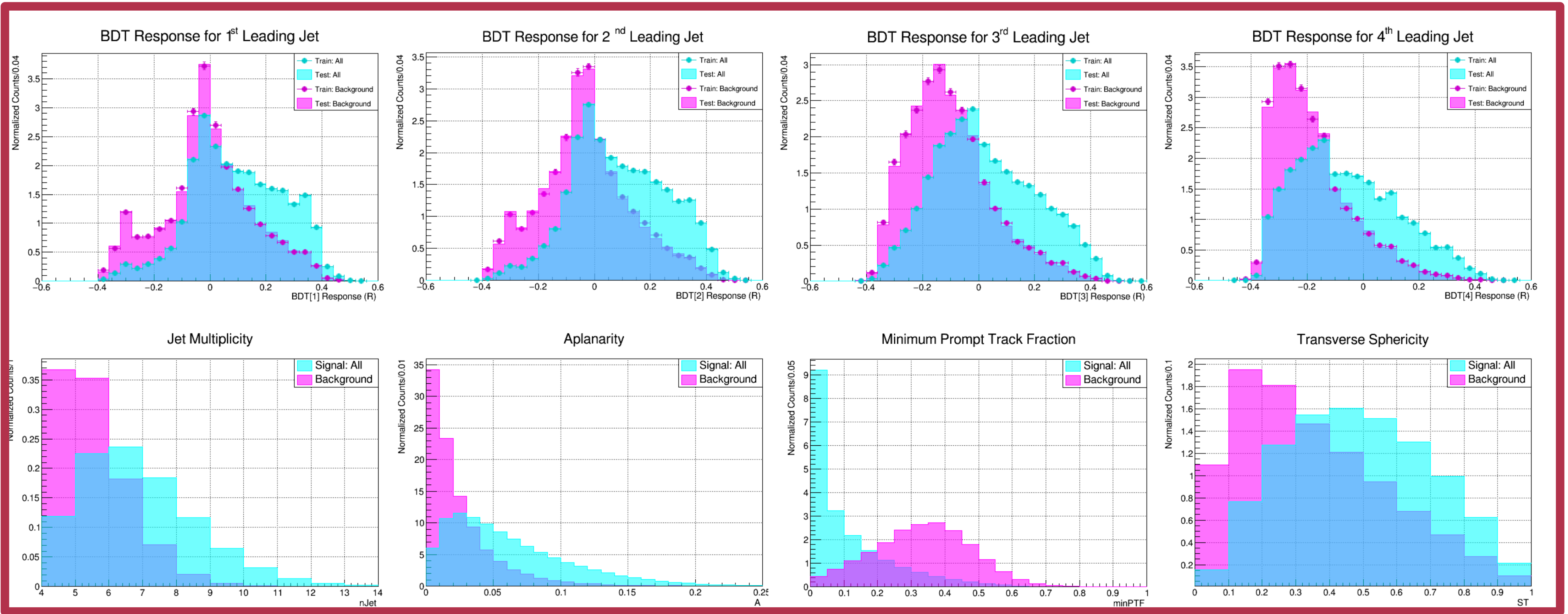
- Combine all signal models together for training
- Train BDT in two steps:

1. Train on jet-level information
   - Energy, Mass, Width, η ← Calorimeter information from the visible part of jets
2. Train on event-level information
   - 4 jet BDT scores
   - minPTF, sum of jet p_T, jet multiplicity ← Variables which characterize the base event selection
   - event-shape information ← Variables which characterize the topology of multi-jet events

Carleton ATLAS



Jet-Level Variables

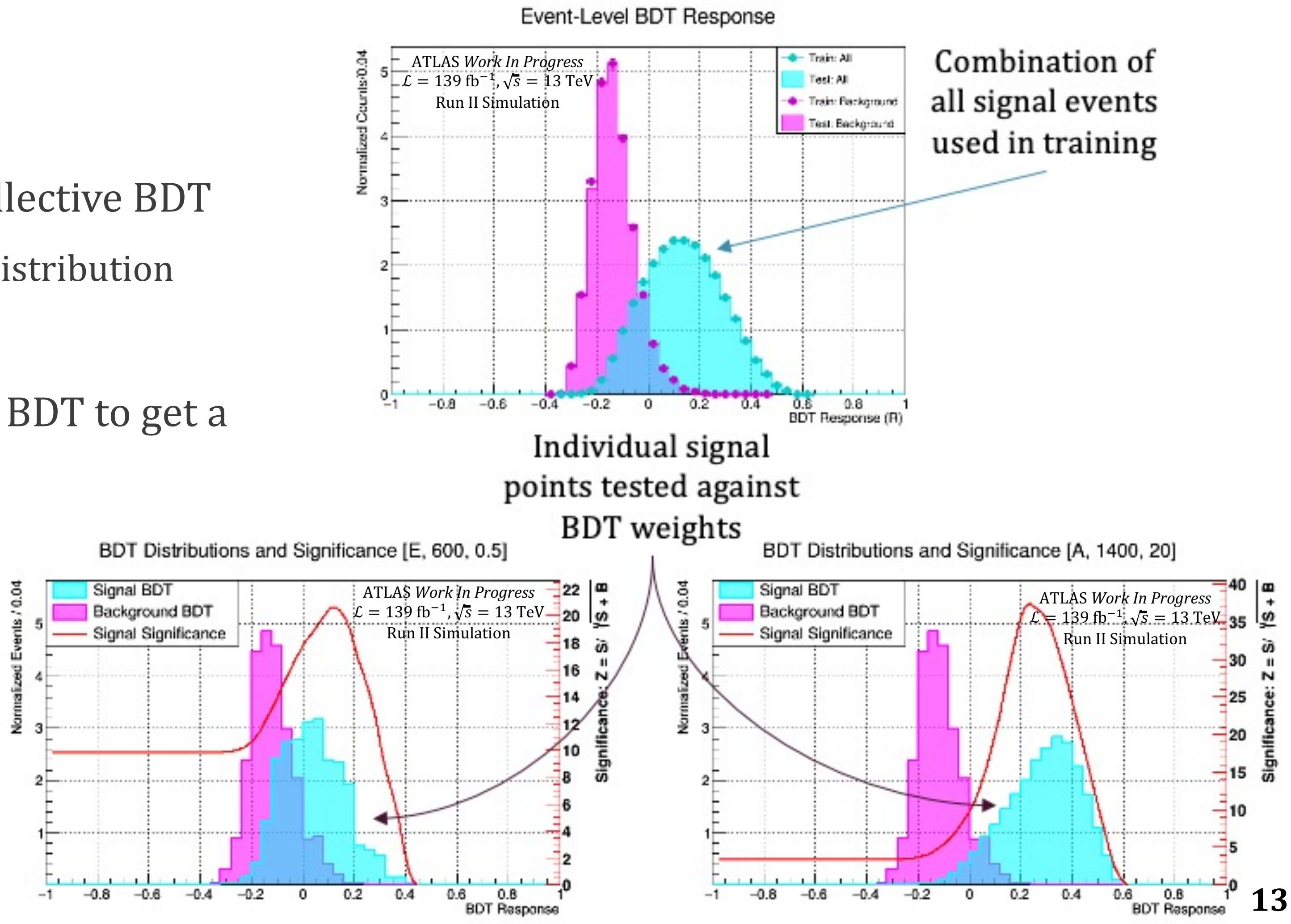


Jet-Level BDT Response

Event-Level Variables

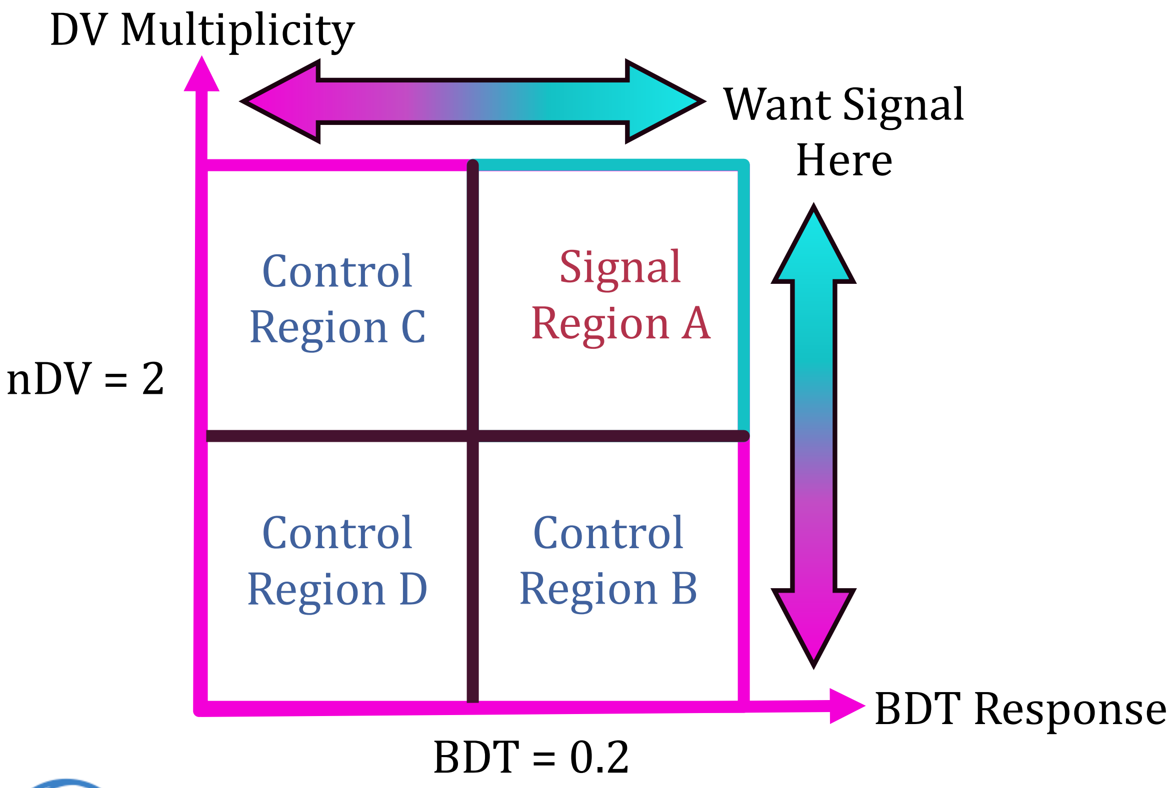
EVENT-LEVEL BDT RESPONSE

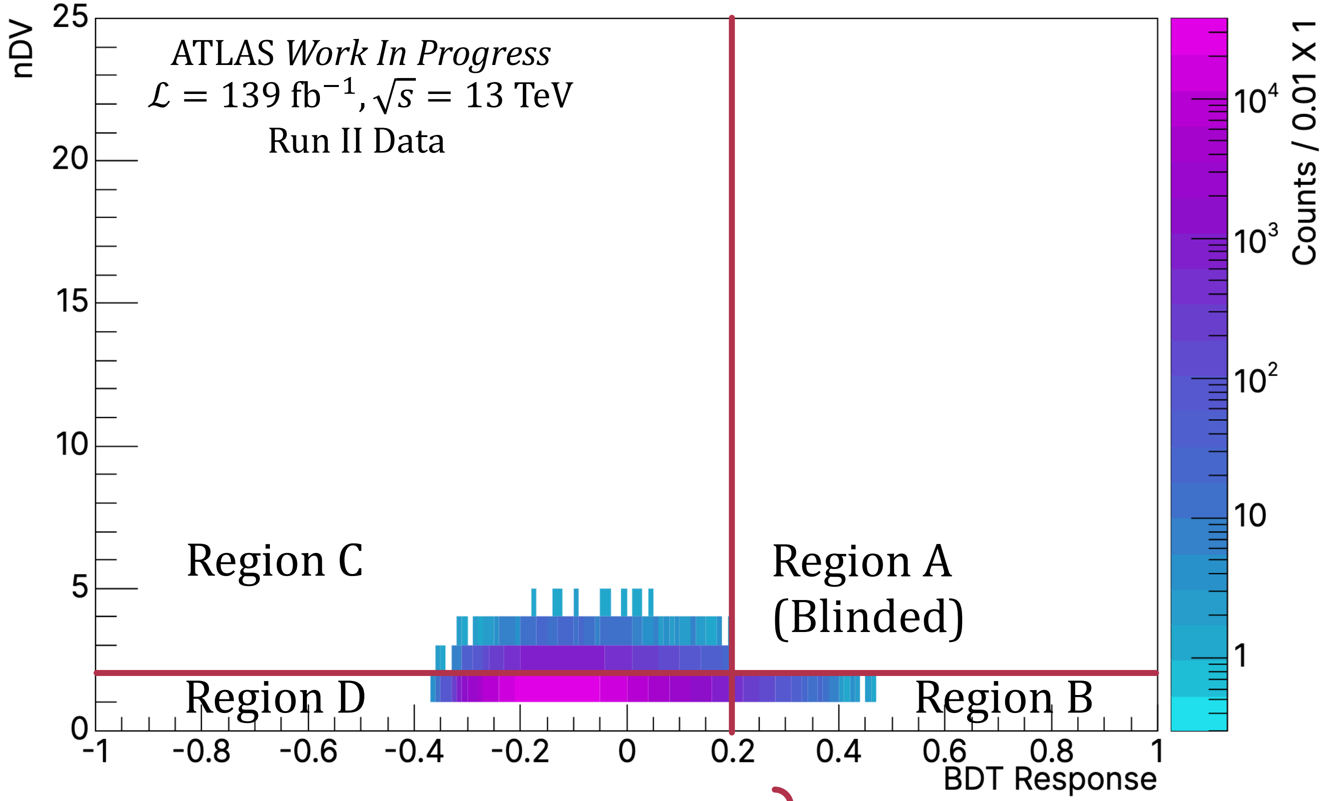
- Test individual signal models against the collective BDT
  - Each signal model produces a different BDT distribution
  - Only one Background BDT distribution
- ATLAS Run II data is then tested against the BDT to get a data BDT distribution



Carleton ATLAS

DATA-DRIVEN BACKGROUND ESTIMATION



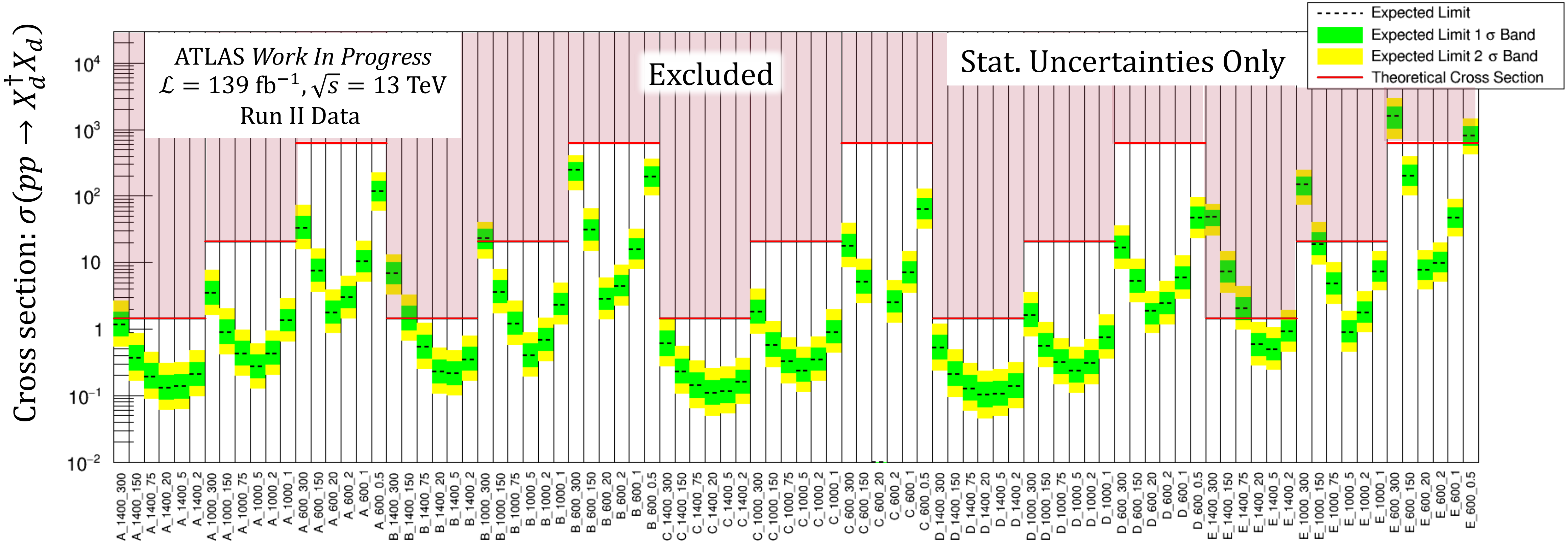


$$N_A^{est} = \frac{N_C}{N_D} \times N_B$$

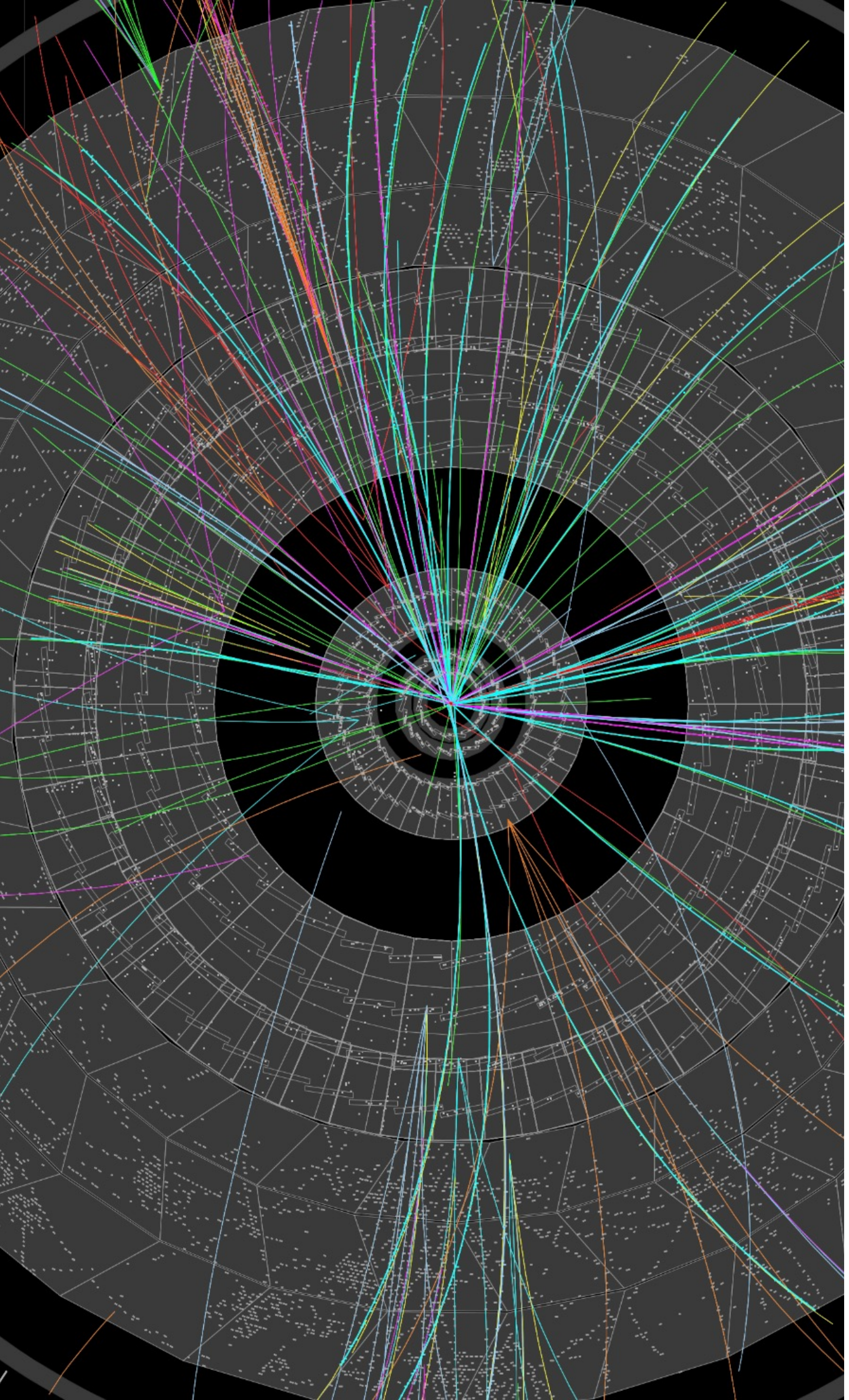
Assumption 1: Variables are Uncorrelated

Assumption 2: Signal Contained in Region A

DATA-PREDICTED RESULTS



Signal Models



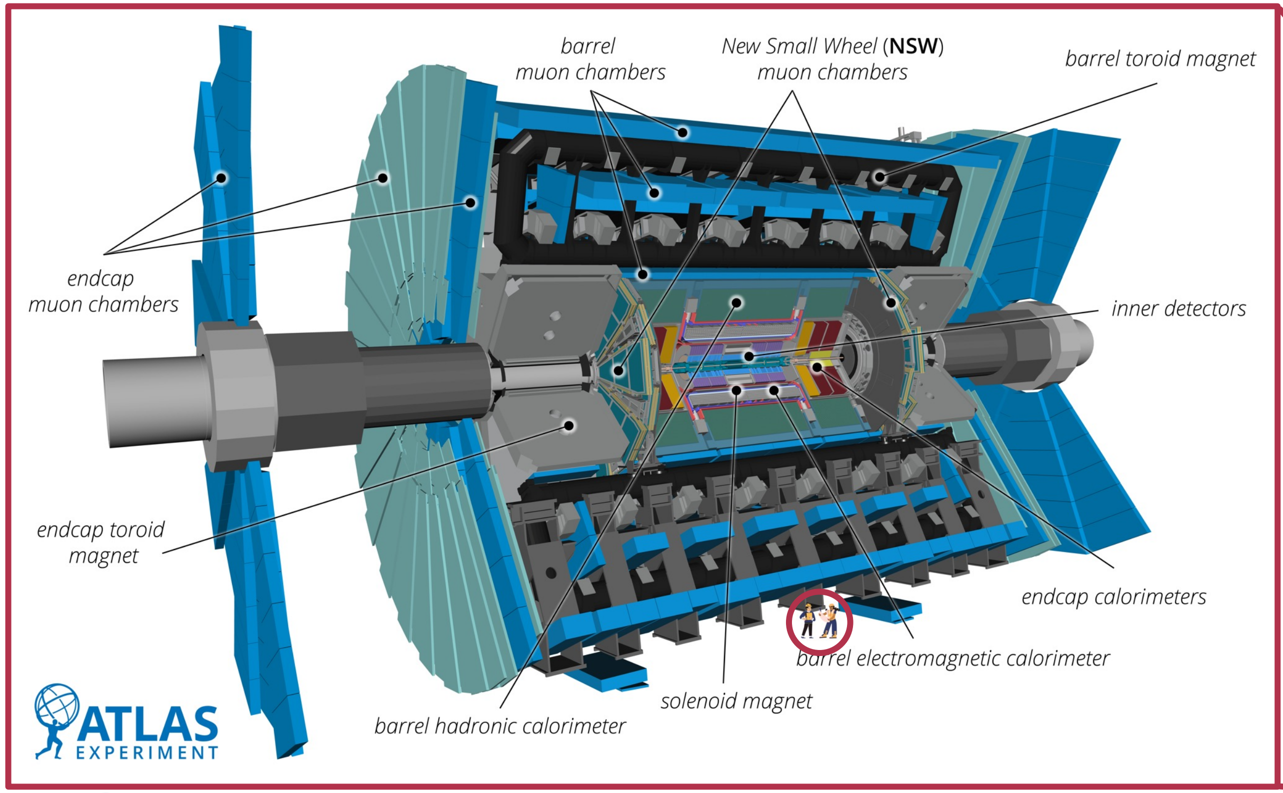
SUMMARY

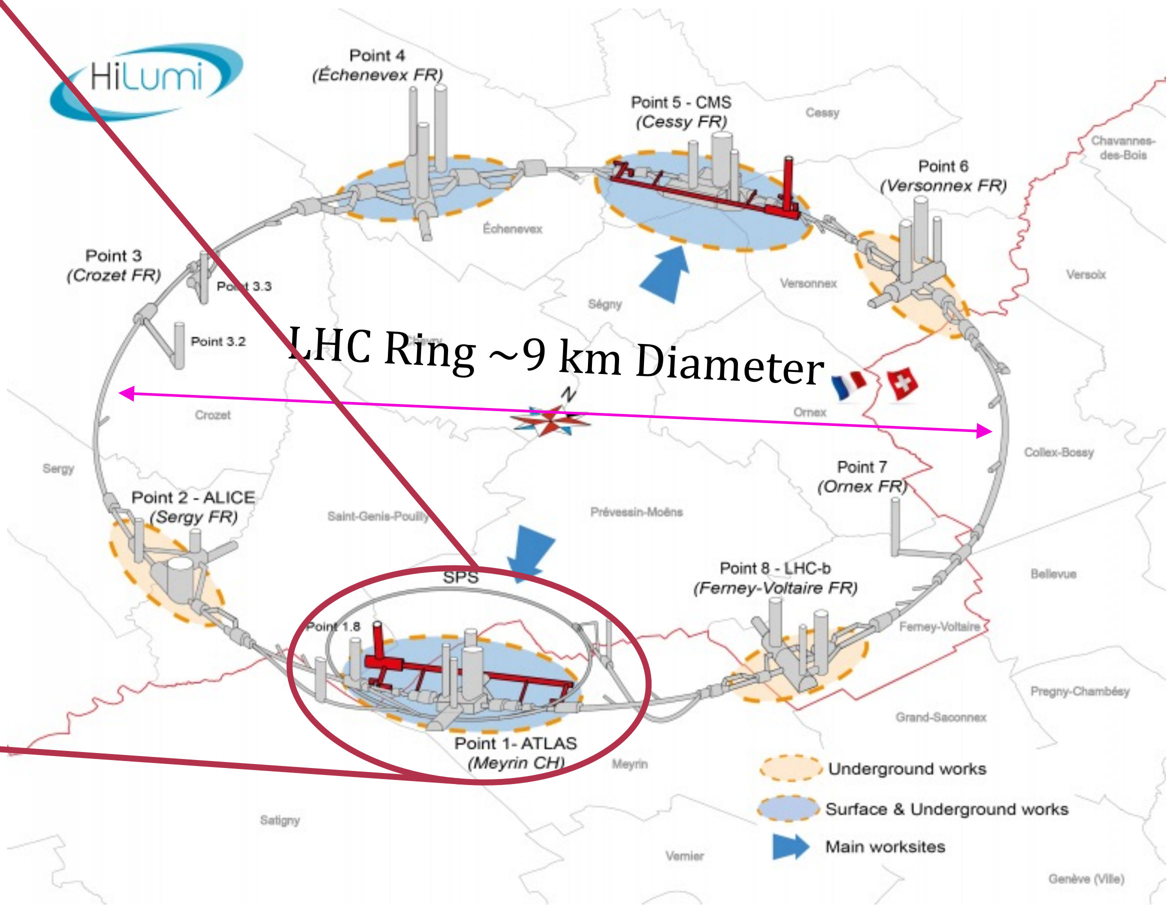
- Finalizing this ATLAS analysis, very close to publishing!
- Using Run II data, we predict sensitivity to most of our Emerging Jets models
- 1st of its kind analysis for ATLAS, expands the model space being tested for emerging-jet-like scenarios

Thank You For Listening!

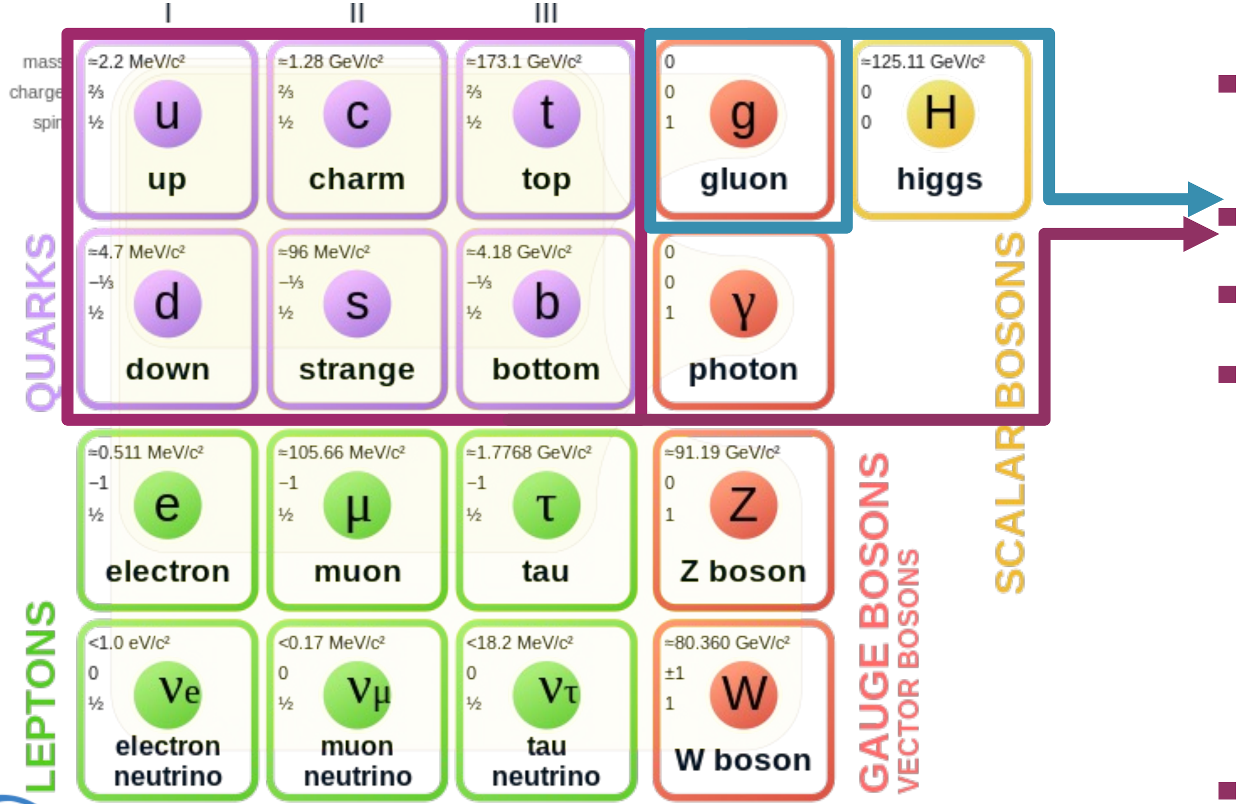
ADDITIONAL MATERIAL

ATLAS Experiment and the LHC at CERN





THE STANDARD MODEL AND QCD



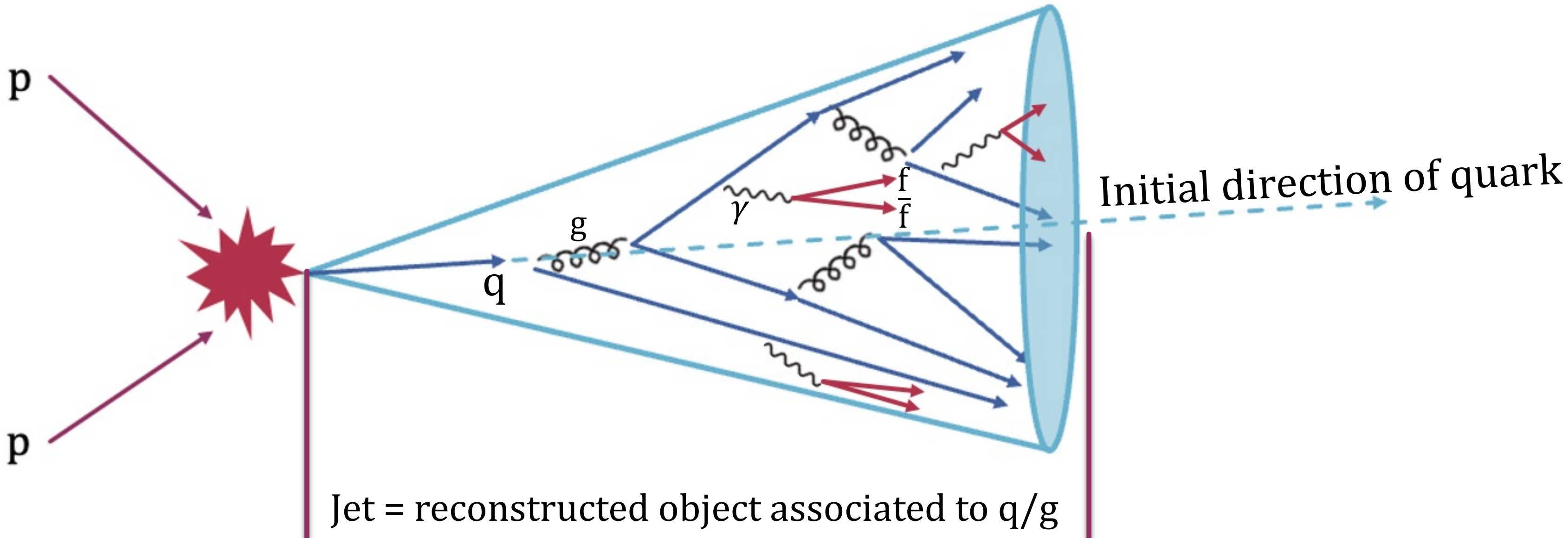
- Quantum Chromodynamics (QCD): *model* of Strong Interaction
- Describes interactions between quarks and gluons
- Introduces three colour charges (r, g, b)
- **Colour Confinement:** all physical states are colour-neutral
  - Quarks pair up into groups of 2 or 3:

$$\underbrace{g + \bar{g}}_{\text{Mesons}} \quad \text{or} \quad \underbrace{r + g + b}_{\text{Baryons}}$$

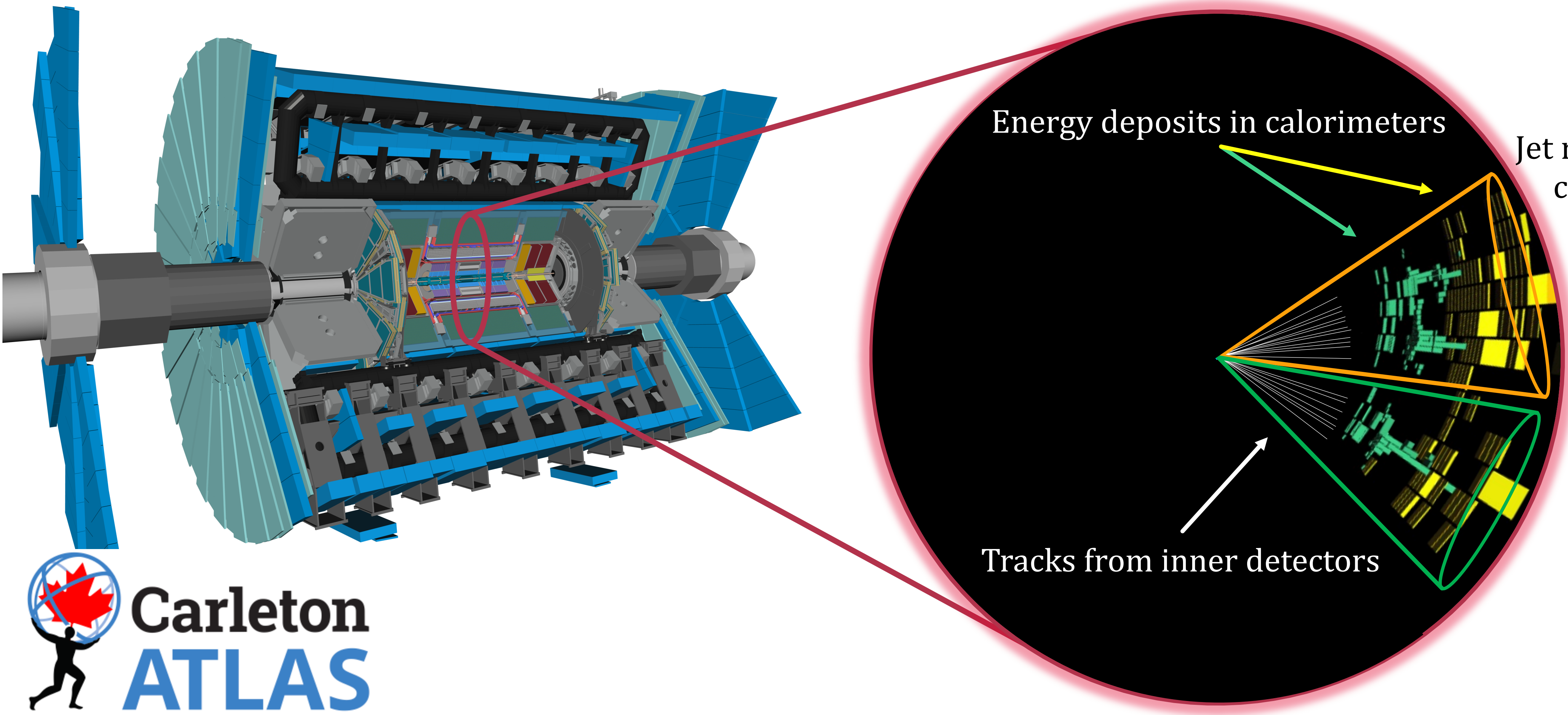
- **Asymptotic Freedom:** coupling is inversely proportional to energy transfer
  - Need high energy (collider) environments to study perturbative QCD

PARTICLE JETS

- When quarks/gluons are produced in high-energy collisions they:
  - Hadronize: quarks pull other quarks out of the vacuum to form mesons and baryons
  - Shower: hadronization creates a collimated spray of particles
- What can we measure?
  - Tracks: charged particles produced in the shower leave tracks in our detectors
  - Energy Deposits (calo clusters): particles interact (either via EM or QCD) to leave energy deposits in calorimeters



JETS IN ATLAS



Jet reconstruction clusters energy deposits into cones which point back to the initial interaction

Carleton ATLAS

VARIABLE DEFINITIONS

$$H_T = \sum_{i=1}^{4} p_{T,i}^{Jet}$$

$$\mathrm{min}PTF = \frac{1}{p_T^{Jet}} \sum_i p_{T,i}^{Track} \left(d_{0,i}^{Track} < 3\,\sigma_{d_0,i}\right) \;\forall\, i = Tracks \in Jet$$

$$E_{Jet} = \sum_i E_i \;\forall\, i = CaloCells \in Jet$$

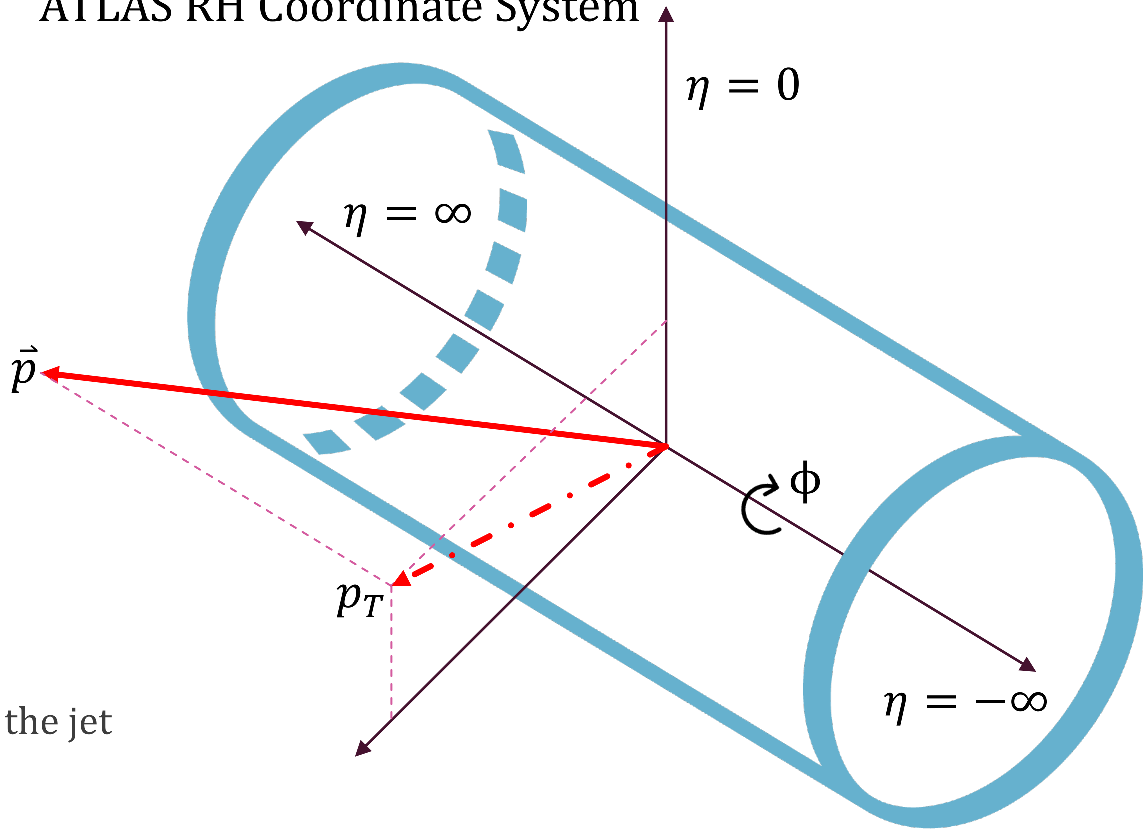
$$M_{Jet} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p_i}\right)^2} \;\forall\, i = CaloCells \in Jet$$

η_{Jet}: Measured at the central axis of the jet

$$W_{Jet} = \frac{1}{p_T^{Jet}} \sum_i p_{T,i}\,\Delta R_i \forall\, i = CaloCells \in Jet \quad \Delta R \text{ measured w.r.t. central axis of the jet}$$

$$\Delta \mathrm{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

ATLAS RH Coordinate System



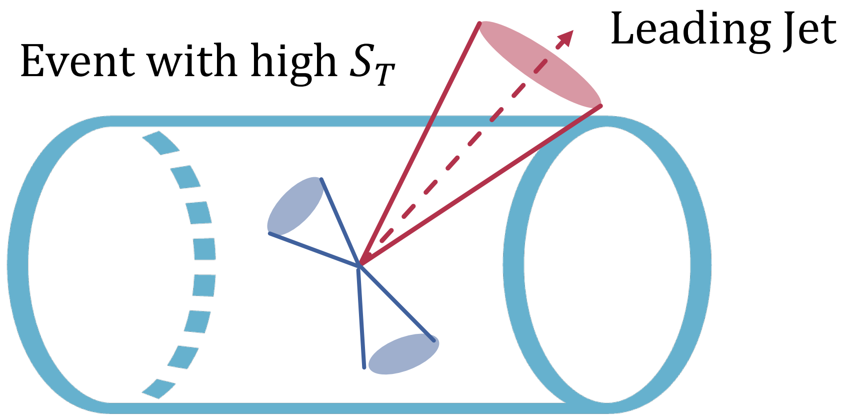
Carleton ATLAS

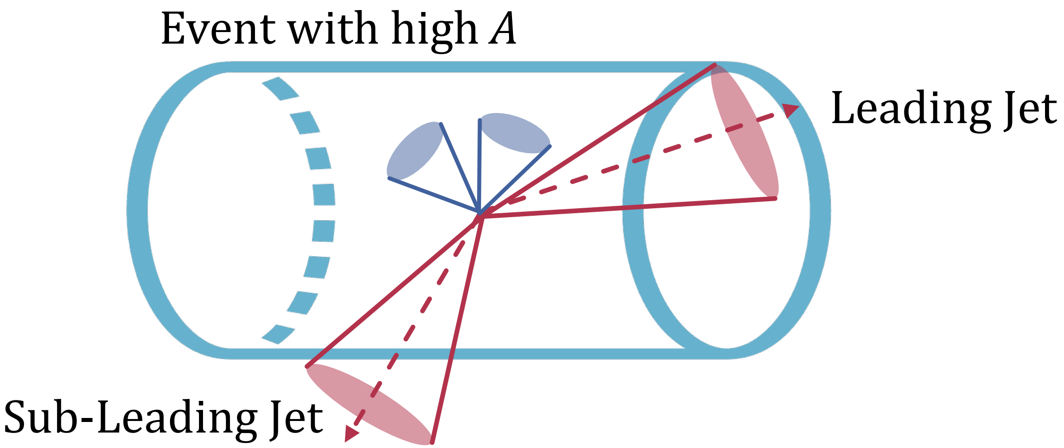
VARIABLE DEFINITIONS [CONT.]

- Event Shape Variables: try to characterize the topology of multi-jet events
- Based on Jet-Momentum Tensor:

$$\mathbb{M}_{ij} = \sum_{k=0}^{3} \frac{(p_i p_j)_k}{(p_i p^i)_k}$$

- Produces Eigenvalues $(\lambda_0, \lambda_1, \lambda_2)$ which define the variables
  - S_T: Define a plane perpendicular to the leading jet, measure of how much energy is along that plane
  - A: Define a plane through two leading jets, measure energy perpendicular to that plane





Carleton ATLAS

RECONSTRUCTED ANALYSIS OBJECTS

Calorimeter Jets (r21)

- EM-topo clusters
- anti- $k_T = 0.4$
- $p_T^{jet} \geq 50$ GeV

- Event Preselection:
  - 4-jet trigger: HLT_4j90-150
  - $nJet \geq 4$
  - $p_T^{4\ leading\ jets} \geq 120$ GeV
  - $|\eta|^{4\ leading\ jets} < 2.4$

Tracks

- Combination of standard tracks and large radius tracks (LRT)
- Standard Tracks:
  - $p_T^{track} \geq 0.5$ GeV
  - $|\eta|^{tracks} < 2.7$
- LRT:
  - $p_T^{track} \geq 0.9$ GeV
  - $|\eta|^{tracks} < 5$
- Event Preselection :
  - $p_T^{track} \geq 1$ GeV

Displaced Vertices

- Built with VSI vertexing
- Tight Working Point:
  - $r\ ,\ |z| < 300$ mm
  - $|d_0| < 10$mm, $|z_0| < 100$ mm
  - $p_T^{DV} \geq 2.5$ GeV
  - $m^{DV} > 0.7$ GeV
  - Pass Material Map Veto
- Event Preselection :
  - Must be jet-matched
  - $nDV \geq 1$

VARIABLE RANKS FROM SINGLE BDT TRAINING

Jet BDT Ranks		
Variable	Importance	Separation
Mass	1	1
Energy	2	4
Width	3	2
η	4	3

Event BDT Ranks		
Variable	Importance	Separation
minPTF	1	1
nJet	4	2
A	2	3
S_T	7	7
H_T	9	9
BDT[0]	8	8
BDT[1]	6	6
BDT[2]	5	5
BDT[3]	3	4

Carleton ATLAS

SIGNAL SYSTEMATICS

CP Jet Systematics:

- JES: Strong Reduction Configuration
- JER: Simple JER Configuration
- JMS: Frozen Configuration

- Up and down shifts from all NPs are symmetrized and combined in quadrature to give single values for each source
- Each jet systematic is then combined to give one overall systematic uncertainty

CP Pileup Systematic:

- Up and down pileup re-weightings are symmetrized

Tracking and Vertexting Systematic:

- Assume 2% for standard tracks, for LRT:
- Compare K-short vertices between data and MC
- Create a per-track uncertainty based on radial DV position
- Randomly remove tracks based on their per-track uncertainty
- Difference between modified and original vertex selection is taken as systematic uncertainty

PromptTrackFrac Systematic:

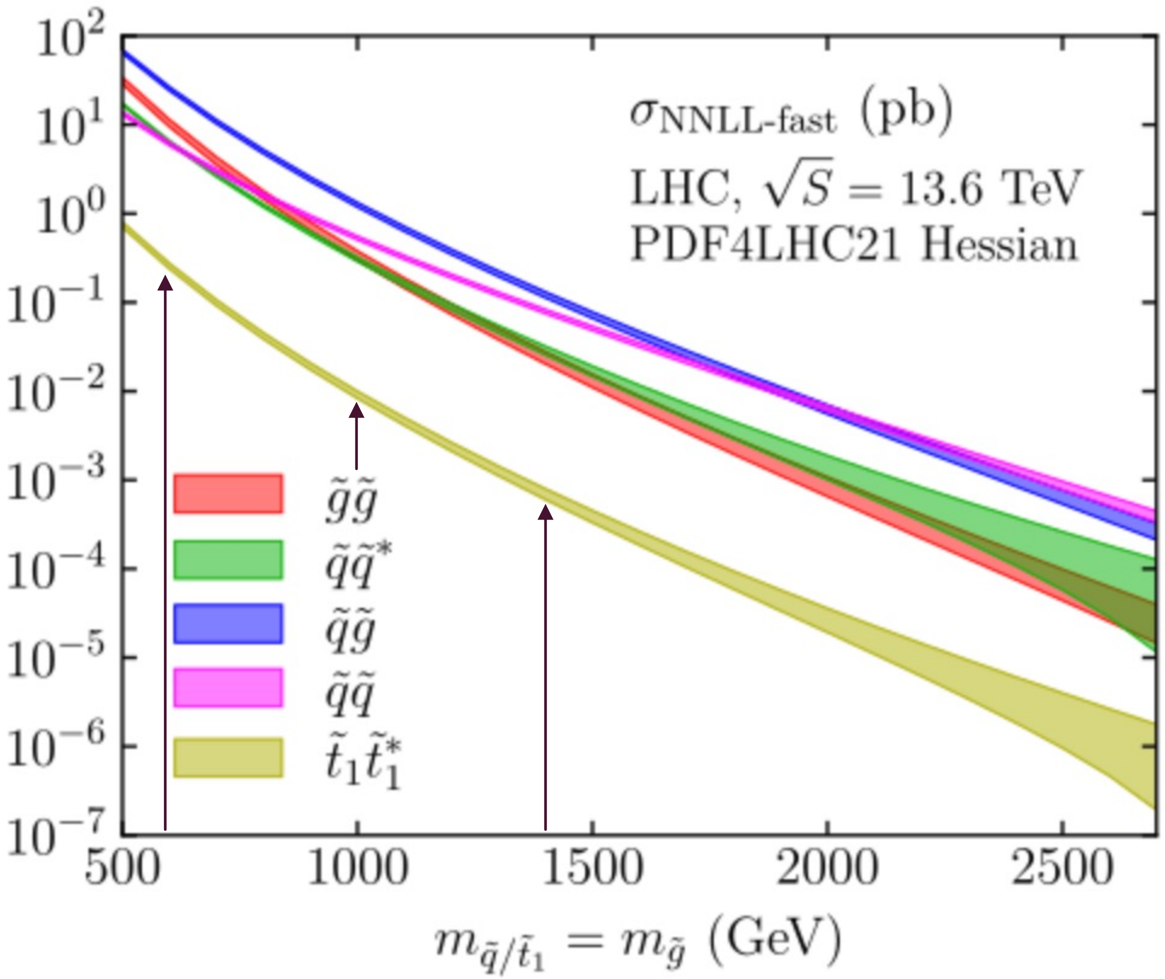
- Compare minPTF distributions between data and MC
- Ratios give a per-event weight used to scale the search region distribution
- Differences in signal yield gives systematic uncertainty

Carleton ATLAS

THEORETICAL CROSS-SECTION UNCERTAINTY

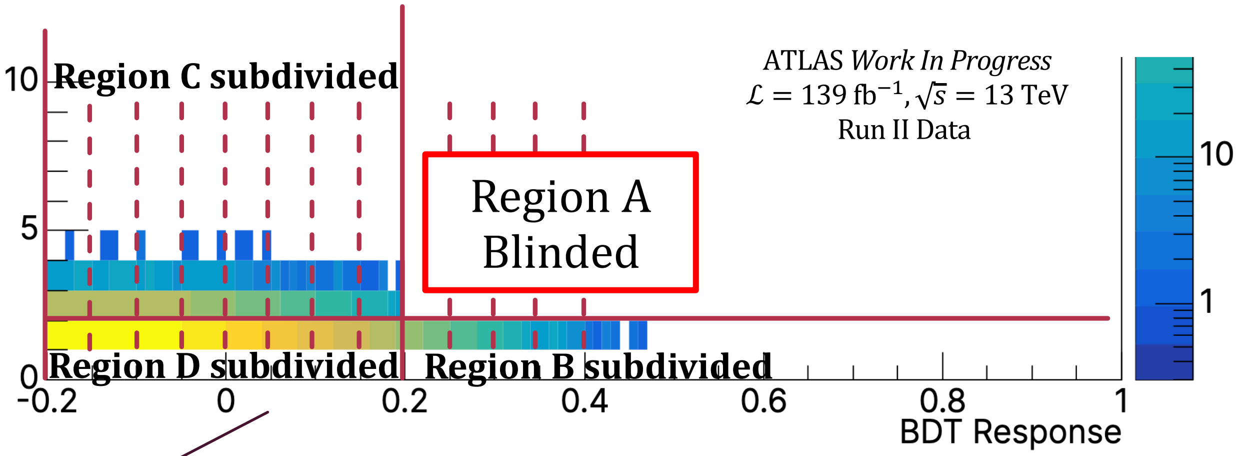
- Determined using stop pair production (in the limit where other squarks and gluinos have decoupled)
- Values taken from SUSY cross-section Twiki
- Since we use 3-color model, cross-sections are multiplied by a factor of 3

	M_X = 600 GeV	M_X = 1000 GeV	M_X = 1400 GeV
Old Values	430 fb	15.2 fb	1.08 fb
Updated (with uncertainty)	(650 ± 50) fb	(20.5 ± 2.3) fb	(1.42 ± 0.22) fb



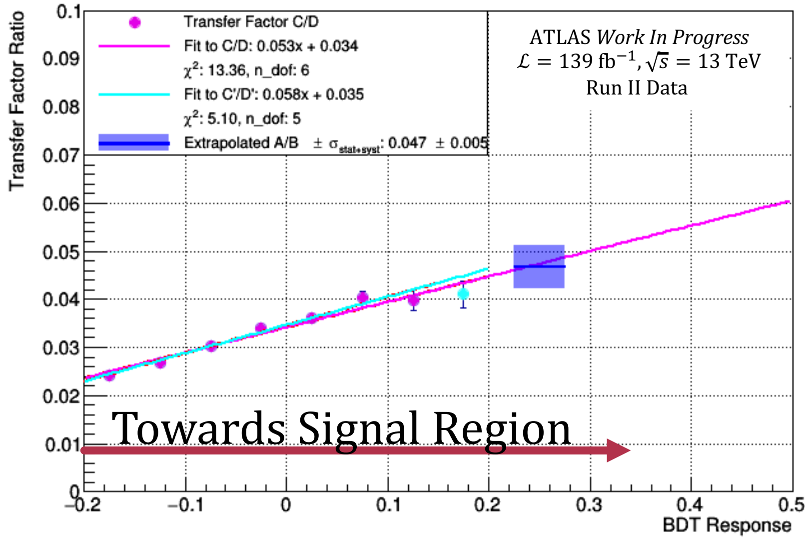
Carleton ATLAS

LINEAR FIT FOR ABCD METHOD



Take ratio of $\frac{N_{C,s}}{N_{D,s}}$

Fit linear function



- Instead of using just 4 regions, can break up the plane into many smaller regions
- Approximate the shape of the ratio $\frac{N_{C,s}}{N_{D,s}}$ as scanned over BDT
- Can fit a linear function to regions C and D, and extrapolate trend into A to estimate background

$$N_A^{est} = \sum_{s \in B} \underbrace{(p_1 \cdot x_s + p_0)}_{\text{Linear Function}} \cdot \underbrace{N_{B,s}}_{\text{Counts in each sub-region of B}}$$

LIKELIHOOD FIT AND CONFIDENCE LIMITS

- Simultaneous fit to entire ABCD plane, subdivided into sub-regions to fit linear function

$$\mathcal{L}(n_{i,s}|p_0, p_1, \mu, N_{i,s}) = \prod_{i=A,B,C,D} \prod_{s} \frac{e^{-N_{i,s}} N_{i,s}^{n_{i,s}}}{n_{i,s}!} \prod_{j=1}^{4} \frac{e^{-\frac{1}{2}(\frac{1-\alpha_j}{\sigma_j})^2}}{\sigma_j \sqrt{2\pi}}$$

control + signal regions

Sub-divisions

Poisson yields

Gaussian nuisance parameters

- Fit takes:
  1. Data ABCD plane
  2. Signal ABCD plane for signal subtraction
  3. 4 Gaussian nuisance parameters for systematic uncertainties

- Background prediction given by:

$$N_A^{est} = \sum_{s \in B} (p_1 \cdot x_s + p_0) \cdot N_{B,s} + \mu \cdot N_{A,s}^{Sig}$$

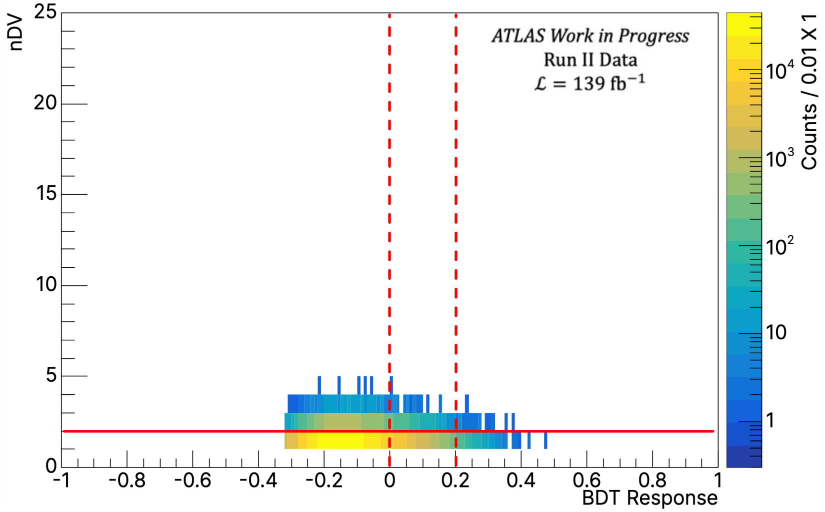
- Can then perform hypothesis test to find μ at 95% confidence limit
- In most cases use asymptotic formula to extract limit
  - In cases where asymptotic limits do not converge, can manually run toys to extract a limit

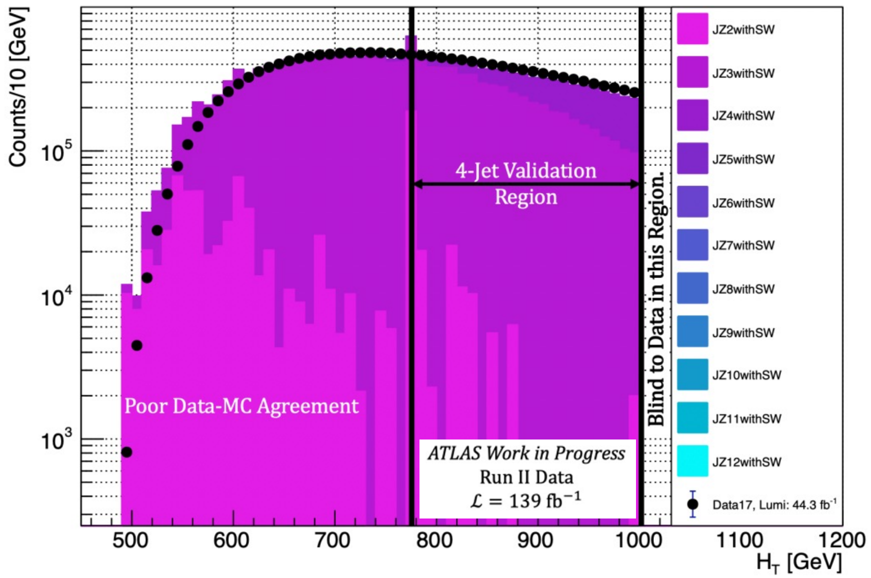
Carleton ATLAS

VALIDATION REGION DEFINITIONS

Use orthogonal H_T selection to define Low$-H_T$ validation region, restrict lower edge of H_T for better Data-MC agreement

$$775\ \text{GeV} \leq H_T < 1000\ \text{GeV}$$





Test two different BDT cuts in Low$-H_T$:
MC background BDT distribution runs out of stats at 0.2