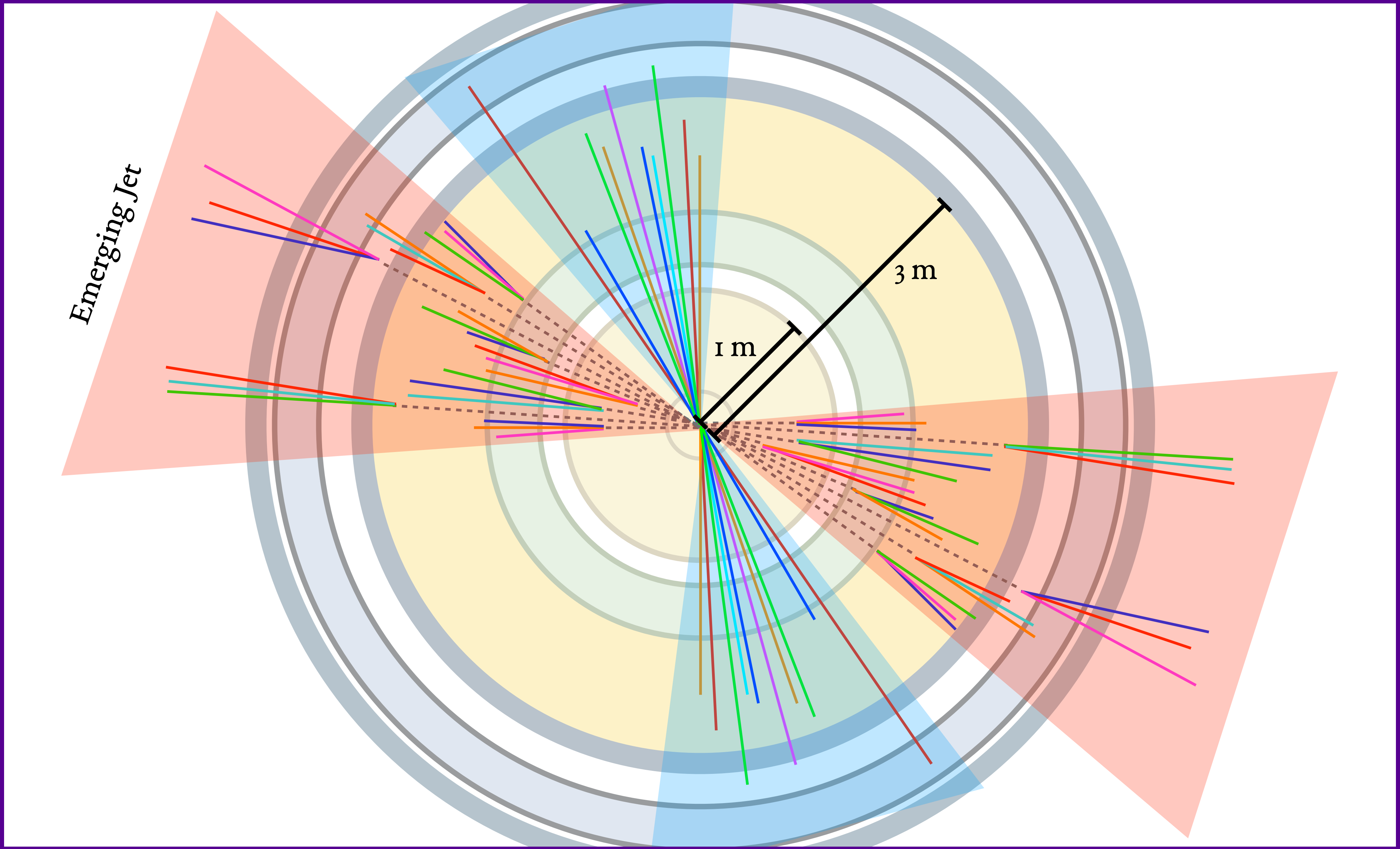
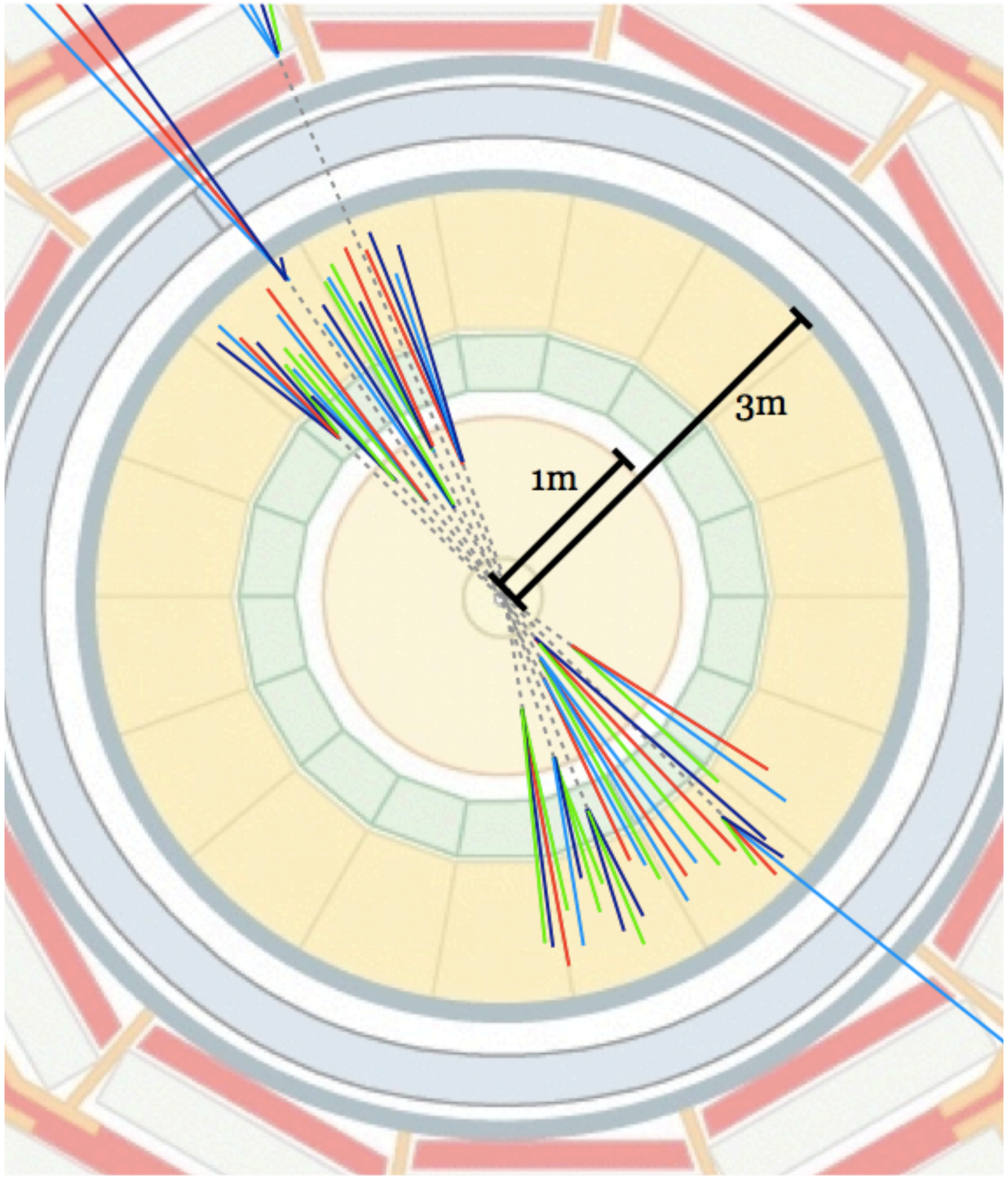


A Search for Emerging Jets at 13 TeV at ATLAS Run 2

Colleen Treado
Advisor: Andy Haas
New York University Thesis Defense
30 October 2020

Outline

- **Introduction**
  - The Standard Model (SM) and beyond (BSM)
  - Emerging jets theory: dark matter + dark QCD and EJs collider phenomenology
  - The Large Hadron Collider (LHC)
  - The ATLAS detector
- **Search for emerging jets**
  - Emerging jets at a glance
  - Signal modeling
  - RPV/LL reconstruction + LLPs at the LHC
  - Signal object and event selections
  - Background estimation
- **Conclusion and outlook**
  - Summary of analysis + plans for the future



Introduction

The Standard Model

- Gauge theory of **strong** and **electroweak** interactions
- **Fermions** interact through forces mediated by exchange of **gauge bosons**

SM GAUGE SYMMETRY GROUP:
$SU(3) \times SU(2)_L \times U(1)_Y$

STRONG (QCD)

- Eight massless, bicolored gluons
- Color-charged interactions
  - 3 colors (rgb)
  - Gluon self-interactions

ELECTROWEAK

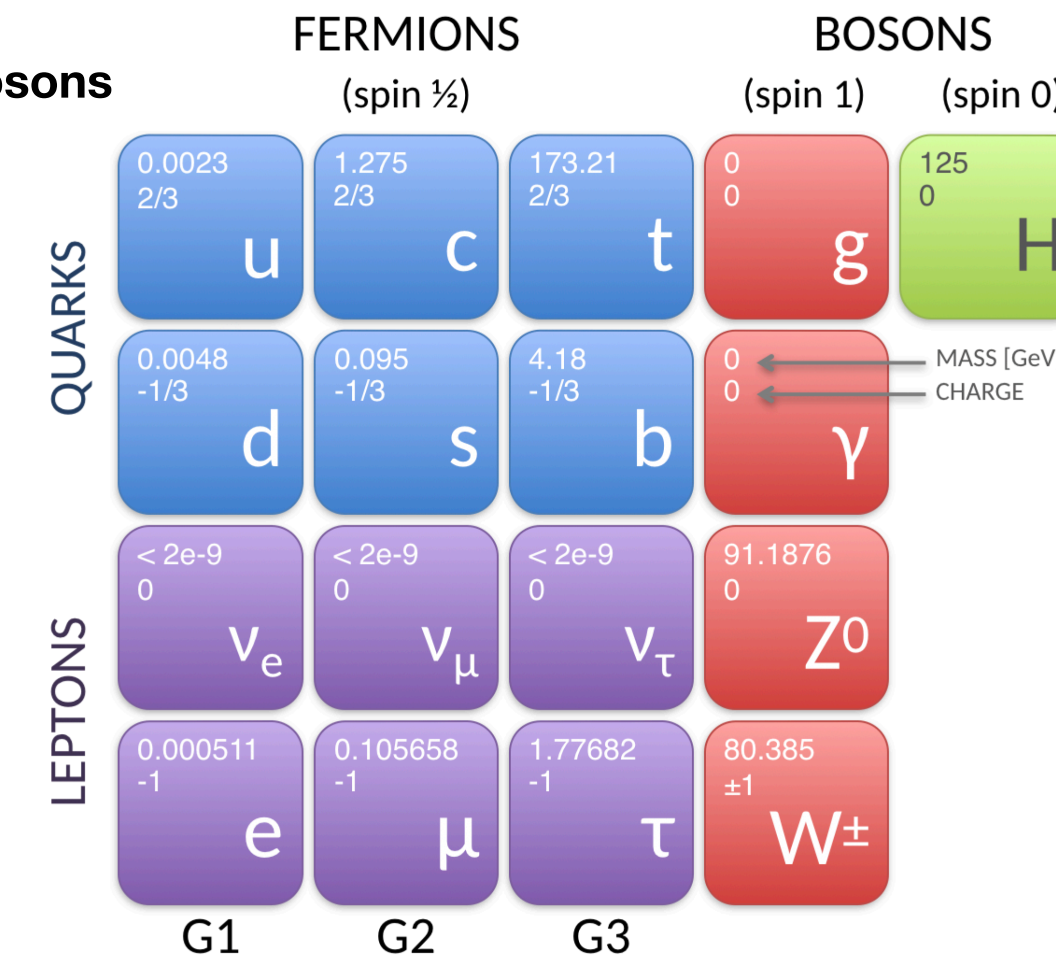
- Four massless gauge bosons: W^a of SU(2) and B of U(1)
- EW symmetry breaking down to $U(1)_{EM}$

WEAK

- Massive $W^{\pm}$ and Z
- Fermion interactions

EM (QED)

- Massless photon
- Electrically charged interactions

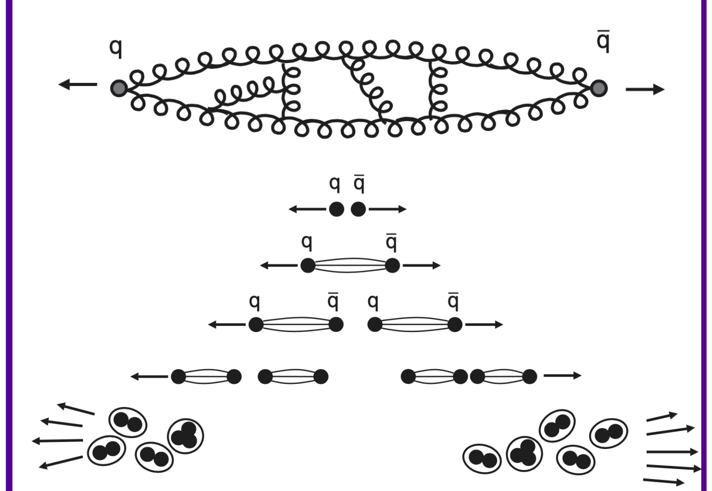


Asymptotic freedom

- Strong running coupling large at low energy and small at high momentum → QCD perturbative well above Λ_{QCD}

Color confinement, hadronization, and jets

- Colored particles confined to colorless hadrons
- High energy partons radiate gluons and quark-antiquark pairs in cascade, producing narrow collimated jets of hadrons



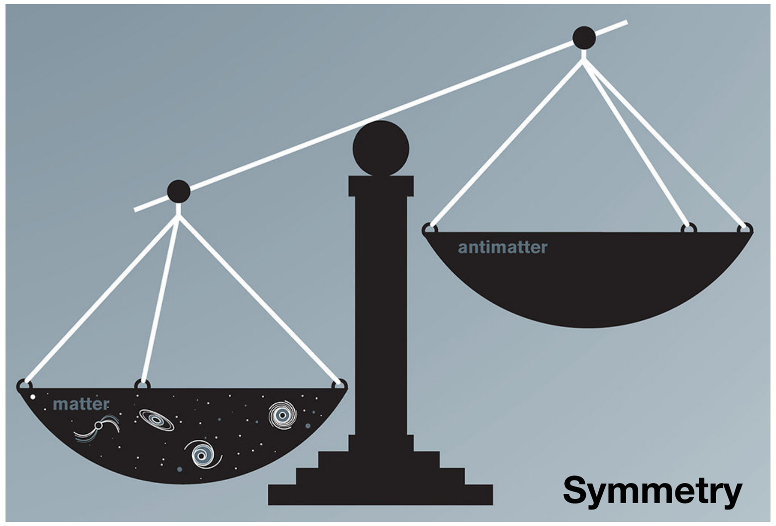
M. Thomson

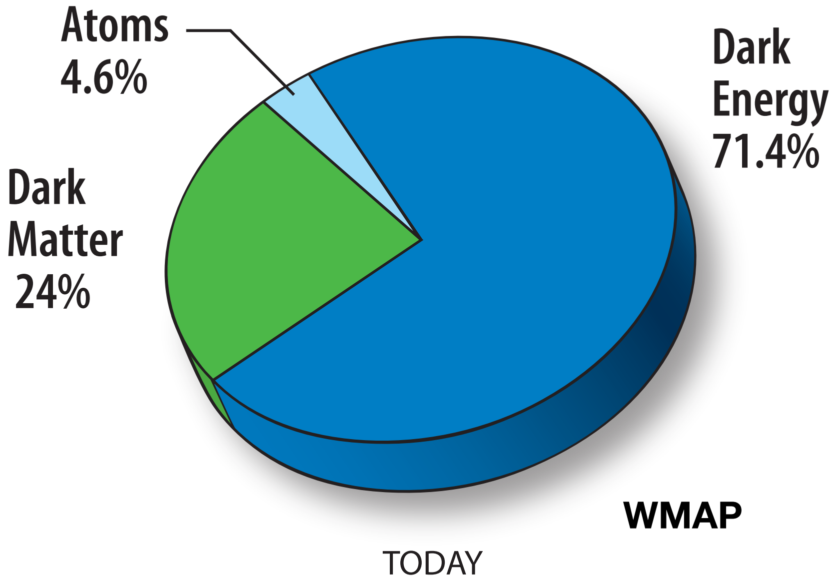
EWSB + Higgs mechanism

- Higgs field has non-zero vev → spontaneously breaks EW symmetry → massive W/Z + massless photon

Beyond the Standard Model

- SM hugely successful…
  - Higgs discovery and precision measurements of W/Z, etc.
- …but incomplete picture of known universe
  - Gravity, hierarchy problem, grand unification, baryon asymmetry, dark matter, etc.
- BSM extensions of SM gauge group may explain one or more of these issues
- → new physics accessible at the LHC, with new particles to discover and new realms to explore





BARYON ASYMMETRY

- Observed abundance of baryonic matter over antimatter of unknown origin → some mechanism of baryogenesis in early universe required to generate asymmetry
- Sakharov conditions:

$$\eta = \frac{n_b - n_{\bar{b}}}{n_\gamma} \approx \frac{n_b}{n_\gamma} \approx 5 \times 10^{-10}$$

  1. baryon number B violation
  2. C- and CP-symmetry violation
  3. departure from thermal equilibrium
  - Baryogenesis in SM could not produce large enough asymmetry to reproduce observed baryon abundance → look to BSM for effective mechanism

DARK MATTER

- DM = invisible, non-baryonic matter that only interacts gravitationally and weakly → ~25% critical density and ~80% matter density of universe
- Evidence from astrophysical observations and cosmological measurements
  - Velocity studies of motion of galaxies and gravitational lensing
  - BBN predictions and CMB measurements of relic abundances
- No SM candidate → most popular = WIMP (weakly interacting massive particle)
  - Neutral, non-colored, and stable, with relic density indicative of weak-scale particle
  - → lack of detection motivating alternative candidates

Dark Matter + Dark QCD

- Baryon asymmetry and dark matter problems can be resolved in models of **asymmetric dark matter** with QCD-like confining hidden sectors at the GeV scale, known as **dark QCD**
  - Standard CDM / WIMP paradigm under tension due to astrophysical anomalies and null experimental results → motivation to consider more complicated dark sectors
  - Asymmetric DM explains baryon asymmetry and coincidence between baryonic and dark matter densities

$\Omega_{DM} / \Omega_B \approx 5 \ (\Omega_i = n_i \times M_i)$ → **motivation**

→ present-day DM abundance originates from same particle-antiparticle asymmetry responsible for baryon abundance

Baryonic matter density: $\eta = n_B / n_\gamma \approx 5 \times 10^{-10}$

Dark matter density: $n_{DM} \sim n_B$ → **assumptions**

Baryon mass: $M_p = 938$ MeV

Dark matter mass: $M_{DM} \approx 5$ GeV

→ suggests dark matter at GeV scale, comparable to proton mass

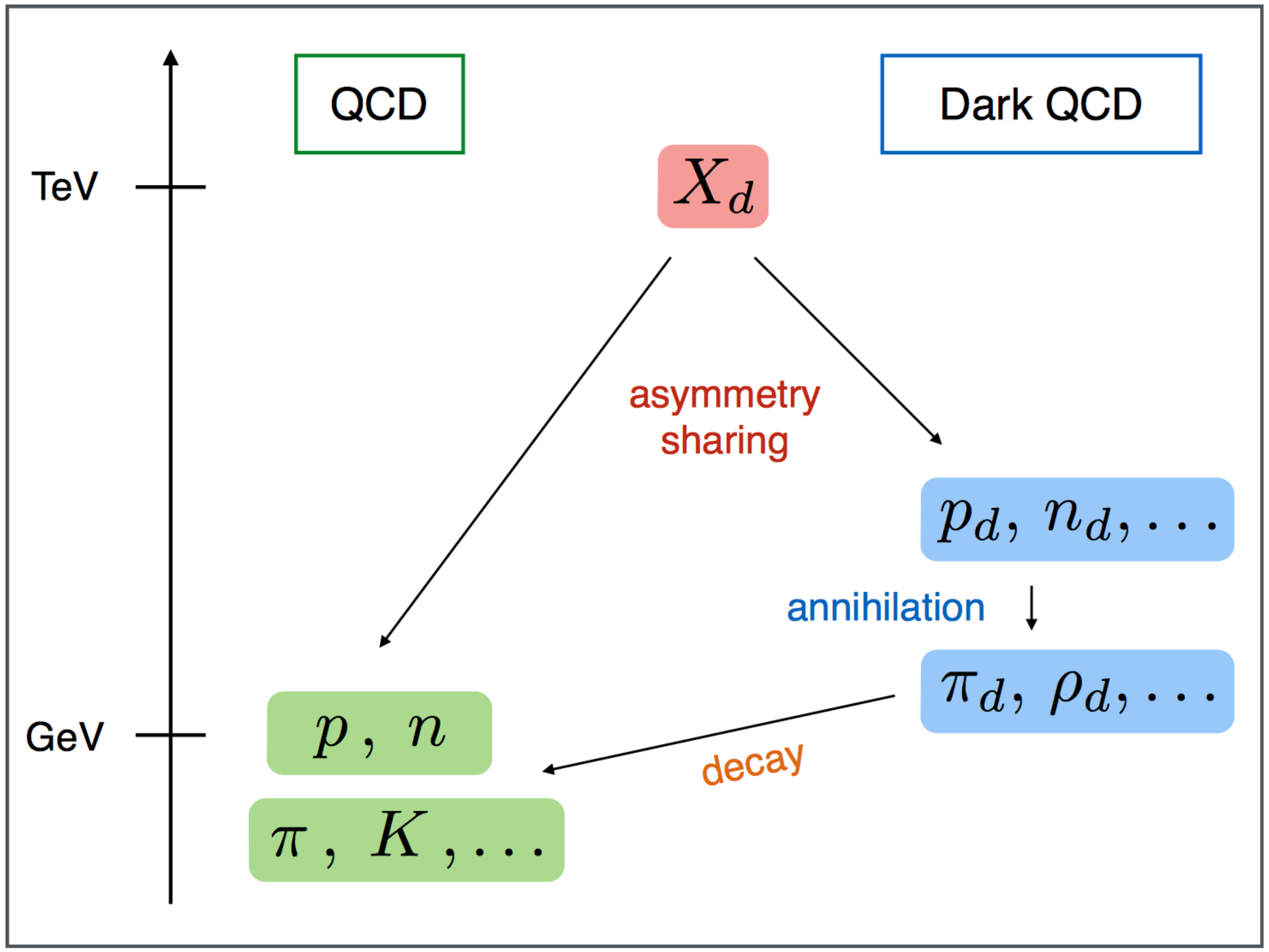
$M_{DM} \approx 5M_p$ → **unexplained**

- Dark QCD introduces new **TeV-scale mediator**, charged under QCD and dark QCD, that generates shared asymmetry and acts as portal between dark and visible sectors → TeV-scale mediator yields dark confinement scale close to QCD at GeV scale
  - Lightest dark baryon = stable **dark proton** = dark matter candidate

Dark QCD Model

Dark QCD = SM extension: $\mathbf{G_{SM} \times SU(N_d)}$ → proposed by EJ theorists: Schwaller, Stolarski, and Weiler

[arXiv:1502.05409]

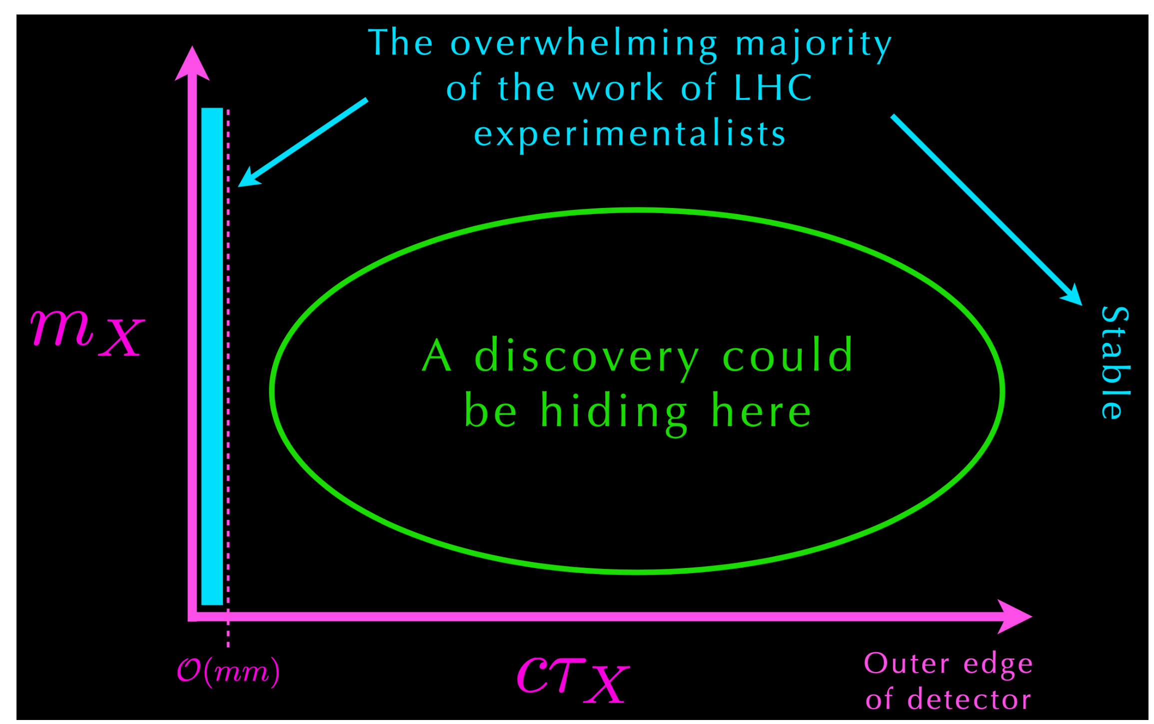


Λ_d (dark confinement scale) + $m_{\pi d}$ (dark pion mass) of order 1-10 GeV

Field	$SU(3) \times SU(2) \times U(1)$	$SU(3)_d$	Mass	Spin
Q_d	(1,1,0)	(3)	$\mathcal{O}$(GeV)	1/2
X_d	$(3,1,\frac{1}{3})$	(3)	$\mathcal{O}$(TeV)	0

- Potentially discoverable phenomenology of dark QCD
  - **GeV-scale dark matter** + **TeV-scale mediator** generating asymmetry and connecting hidden and visible sectors
  - Zoo of unstable dark hadrons, similar to QCD, with lightest (**dark pions**) decaying to SM
  - GeV-to-TeV hierarchy yields slow decays, resulting in **long-lived dark particles** with macroscopic decay lengths of order centimeters to meters

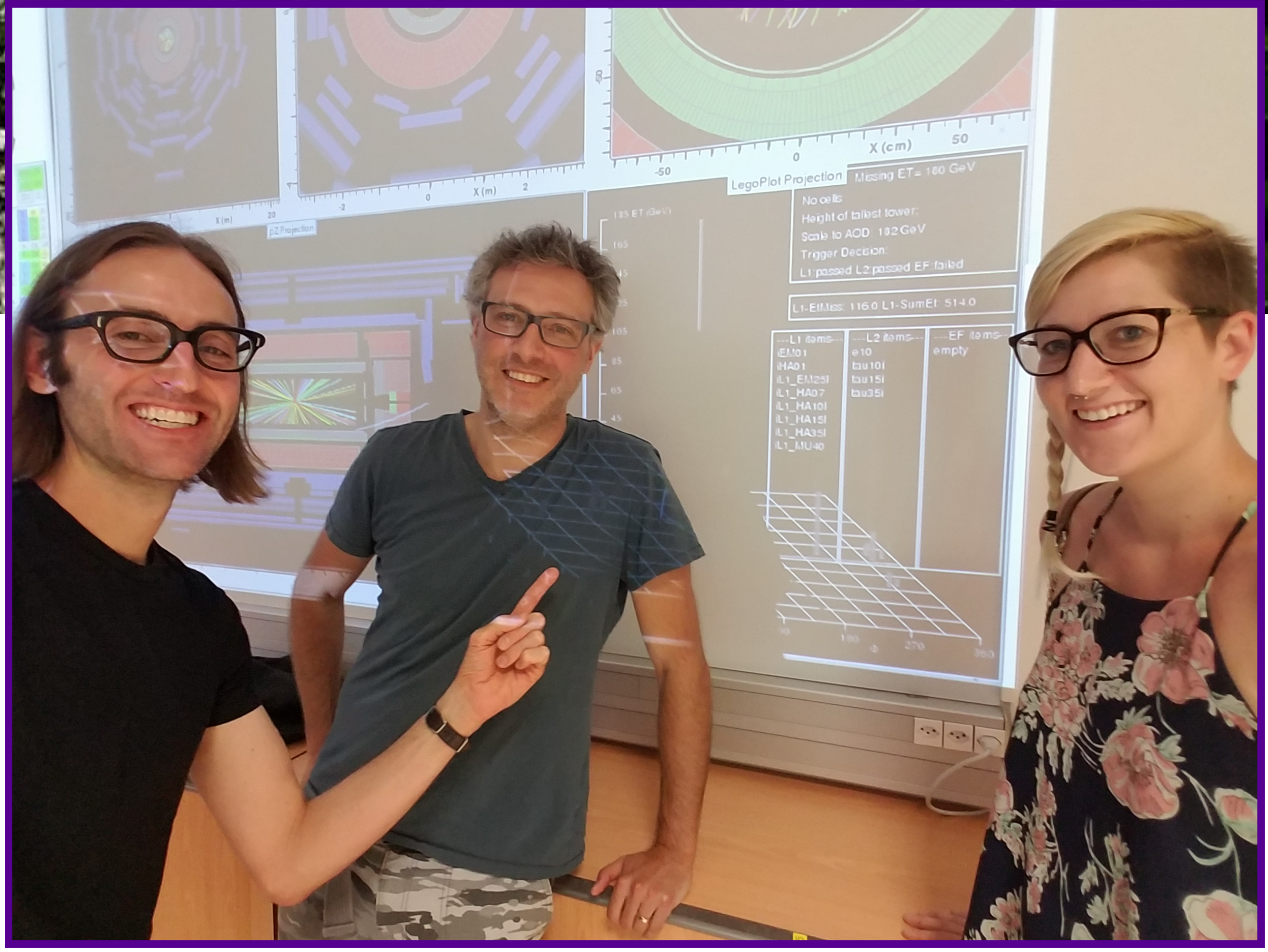
$$c\tau_0 = \frac{c\hbar}{\Gamma} \approx 80\,\mathrm{mm} \times \frac{1}{\kappa^4} \times \left(\frac{2\,\mathrm{GeV}}{f_{\pi_d}}\right)^2 \left(\frac{100\,\mathrm{MeV}}{m_{\mathrm{down}}}\right)^2 \left(\frac{2\,\mathrm{GeV}}{m_{\pi_d}}\right) \left(\frac{M_{X_d}}{1\,\mathrm{TeV}}\right)^4$$



J. Beacham

The Large Hadron Collider (LHC)

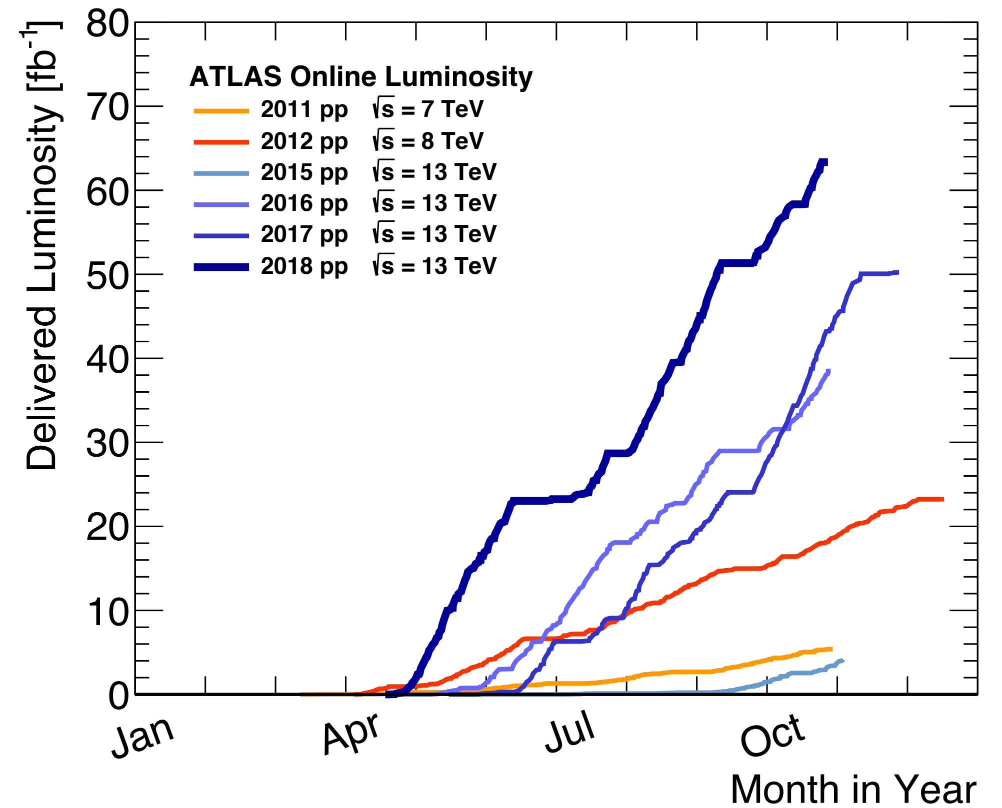
- High energy **circular proton-proton collider** at CERN (Geneva, Switzerland) — largest, highest-energy, and highest-luminosity particle accelerator to date
  - 26.7 km in circumference
  - **14 TeV** design energy + **1e34 cm^{-2}s^{-1}** design luminosity
  - → designed to test SM and search for new BSM physics
- Proton beams accelerated to ultra-relativistic speeds and circulated through two superconducting rings with collisions at four interaction sites within detectors along ring
- General purpose detectors **ATLAS** + CMS at **P1** + P5

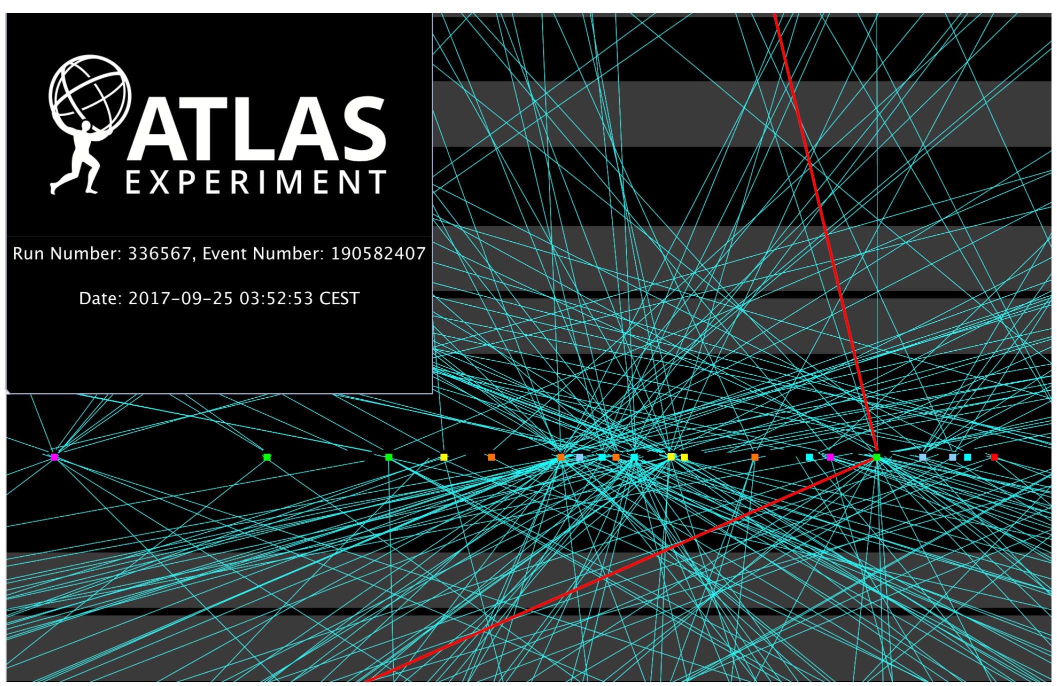


The EJs team in the ATLAS control room

LHC Overview

- Run 1: 2010-2012 at 7-8 TeV → 20 fb^{-1} of total data + Higgs discovered in July 2012 with 5 fb^{-1} of data
- **Run 2: 2015-2018 at 13 TeV** → 139 fb^{-1} of data = 10 trillion collisions
- → data taking periods between April and November + year-end shutdowns for maintenance and repairs
- → data-taking conditions improved between Run 1 + Run 2 and year to year, increasing luminosity and ability to produce rare events
- → increased luminosity means increased pileup = multiple overlapping collisions in single event
- Run 3 begins in 2021 at 14 TeV





$$N = \sigma \times \int L\, dt$$

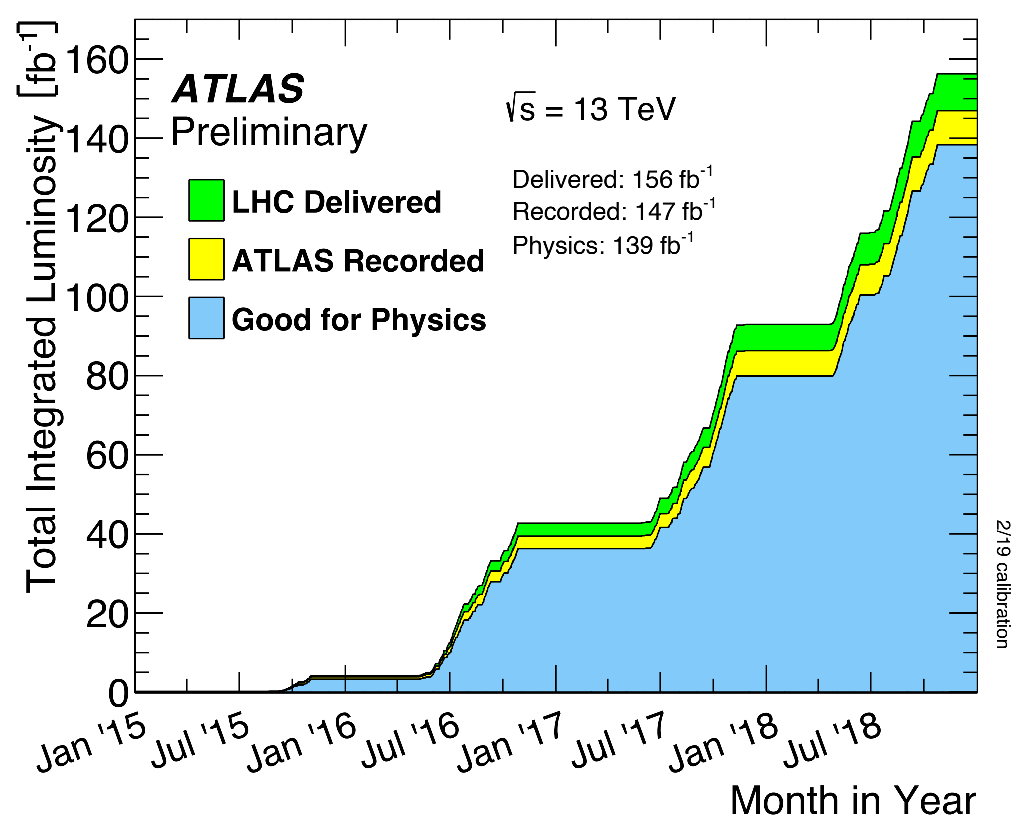
$$= \sigma \times L_{int}$$

→ L depends on beam parameters

→ L_{int} = amount of data

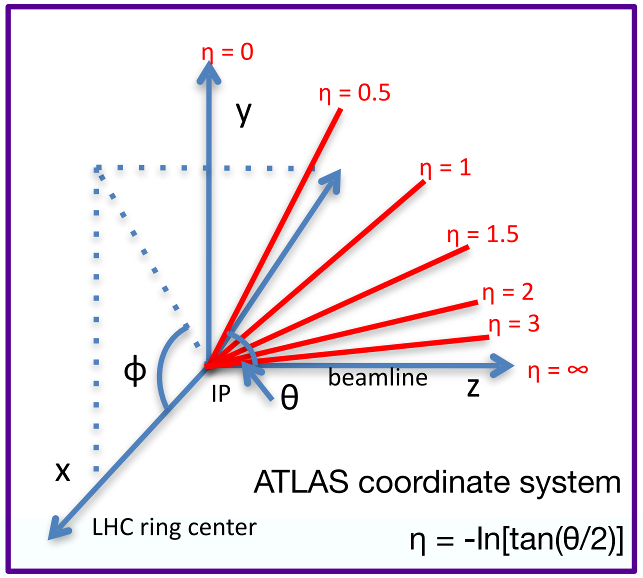
$$\langle \mu \rangle = N_{PV}$$

→ average number of inelastic interactions per bunch crossing



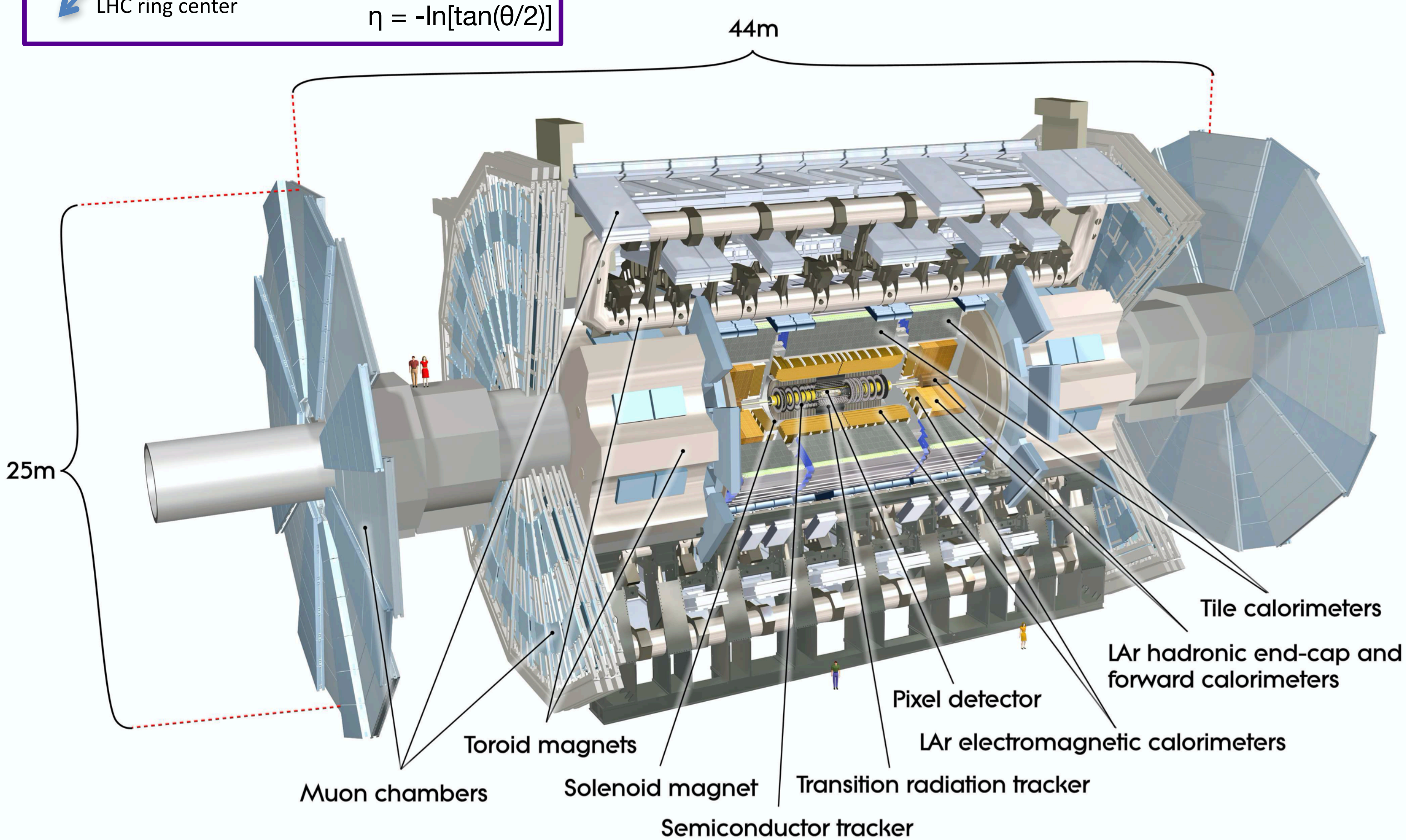
The ATLAS Detector

A **T**oroidal **L**HC **A**pparatu**S** = general purpose detector with sensitivity to wide range of signatures



LAYOUT + DESIGN

- Speed, efficiency, precision, and accuracy of selection, measurement, and reconstruction of physics processes of interest
- → concentric layers around collision to identify and measure different particle types and properties



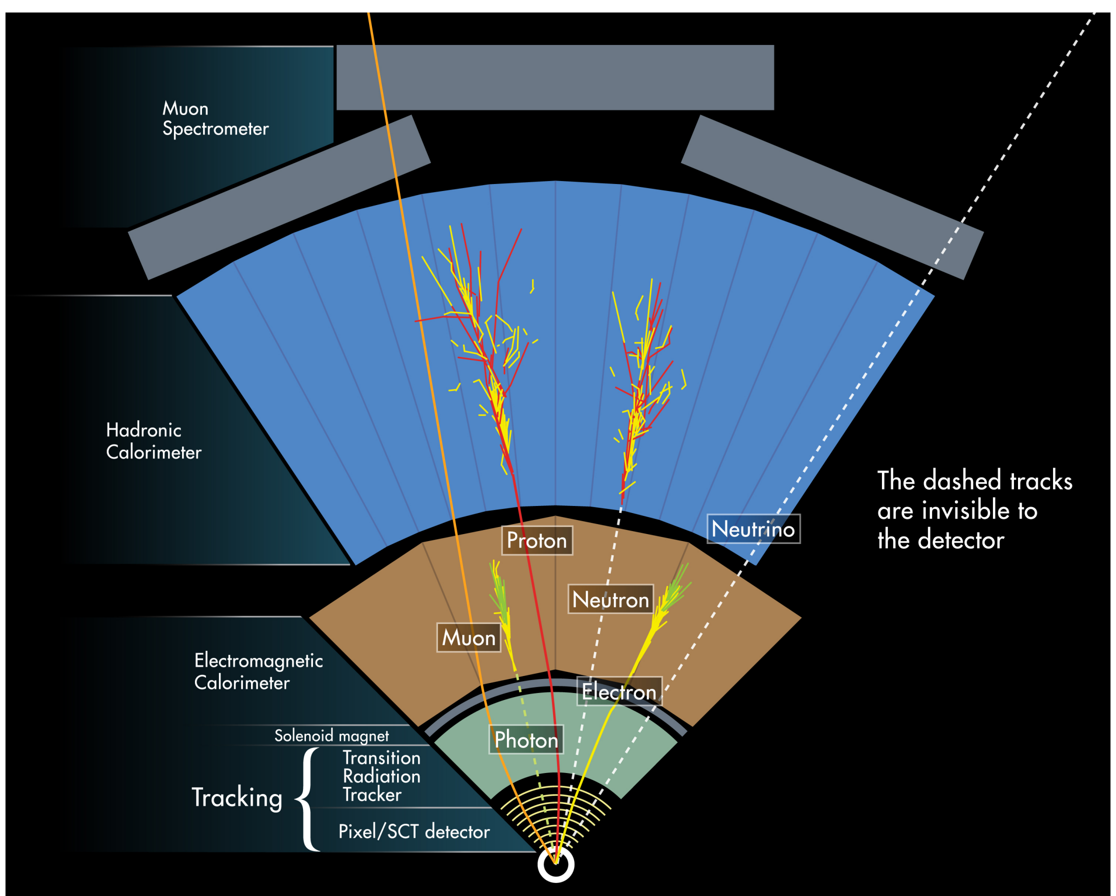
DETECTOR COMPONENTS

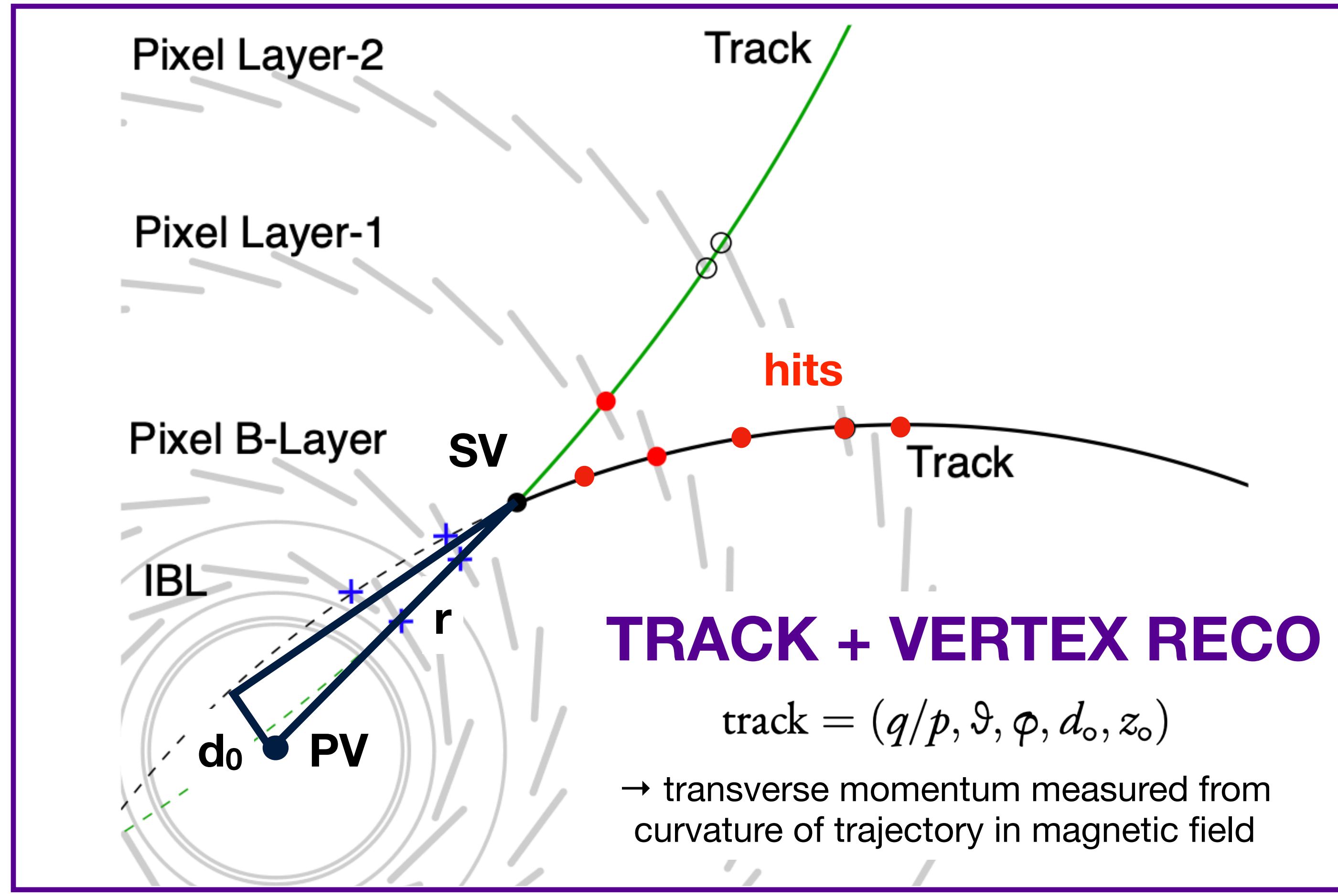
- **Inner Detector**: reconstruction of charged particle tracks + vertices and precise momentum measurements
  - → **Pixel** + **SCT** (Si detector) perform precision tracking + vertex finding
  - → **TRT** extends tracking + identifies electrons
- **Calorimeters**: energy measurements of EM + strongly interacting particles → absorbers induce showers + samplers measure energy deposits
  - → **EM** (**LAr**)
  - → **Hadronic** (**Tile** + LAr HEC + LAr FCal)
- **Muon Spectrometer**: muon tracking system
- **Trigger** system: two-level trigger (**L1** + **HLT**) to reduce rate from **40 MHz to 1 kHz** → selects subset of events containing majority of interesting physics
- Superconducting **magnets**: **solenoid** around ID + **toroids** around MS to bend charged particle trajectories for charge and momentum measurements

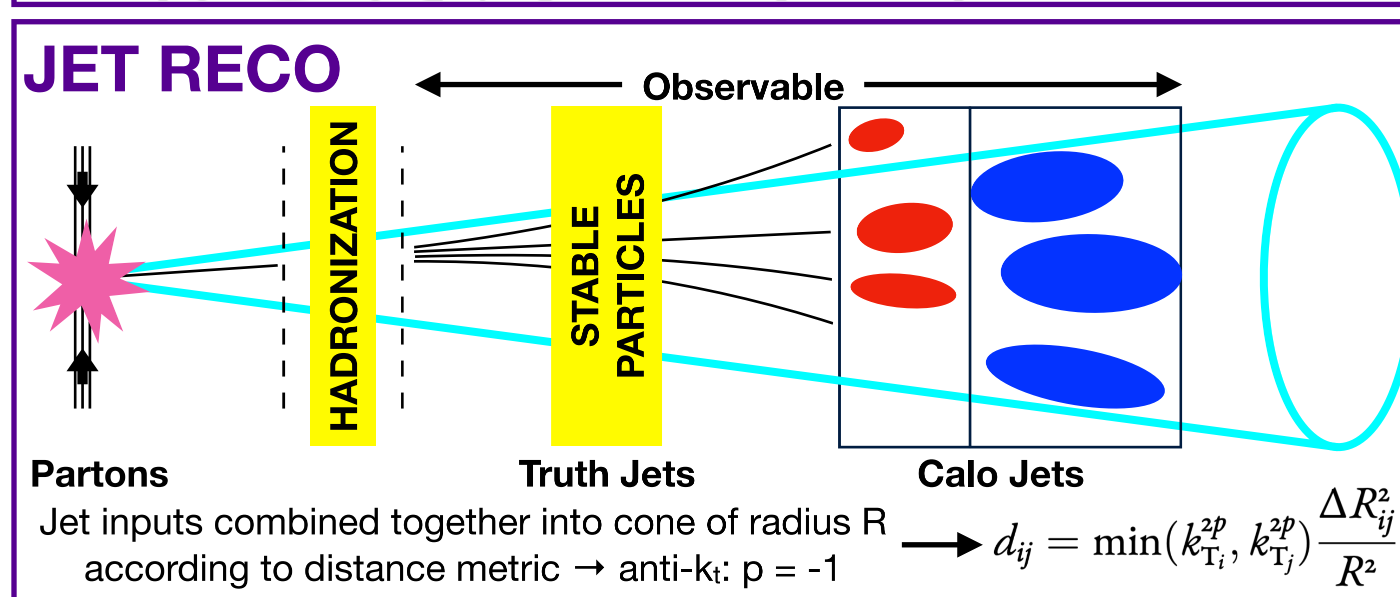
Reconstruction

Reconstruction combines and converts digitized detector readout signals from various detector elements into identifiable physics objects

1. **Identification of hits** + **formation of intermediate objects**, i.e. tracks, vertices, clusters
2. **Measurement of physical properties**, i.e. charge, momentum, energy, from particle-like signatures
   - **Tracking**: identification of charged particle tracks and measurement of momenta through pattern recognition and fitting algorithms
   - **Vertexing**: reconstruction of vertices where tracks meet
   - **Calo-clustering**: combination of energy deposits in calorimeter cells into clusters using topological clustering algorithm for use in jet reconstruction
3. **Particle identification** through dedicated algorithms combining signatures from different parts of detector



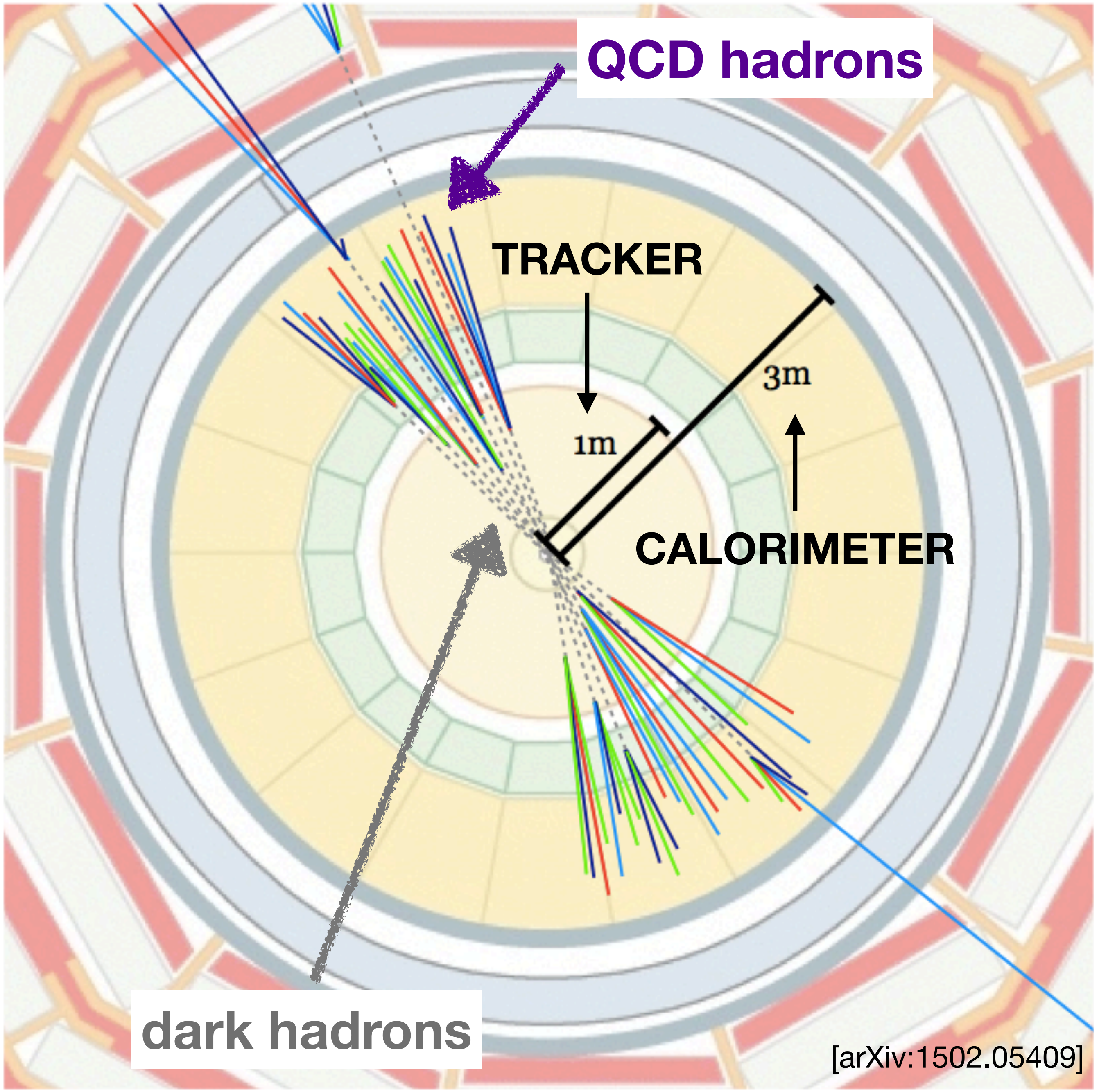




Search

Emerging Jets at a Glance

- Novel LHC signature arising from **multiple displaced decays** of dark or hidden sector hadrons to the visible sector within a **single jet** object at **various macroscopic distances** from the interaction point
    - Decay = spontaneous process where particle transforms into multiple other particles → "displaced" means this occurs far from the initial collision point
    - Vertex = point in space where tracks from particles produced in decay meet
- Jet is neither prompt nor displaced with a single vertex, but ***emerging*** into the detector with **many** small, nearby **displaced vertices (DVs)**
- **Emerging Jet (EJ)** object characterized by mostly **displaced tracks** with large impact parameters and many distinct **displaced vertices** within the individual jet cone = new and unique **exotic BSM** detector object
- → EJs analysis = dedicated signature-based **long-lived particle (LLP)** search within exotics working group → first of its kind at ATLAS



$$p\,p \rightarrow X_d X_d^\dagger \rightarrow \bar{q}\,Q_d\,\bar{Q}_d\,q$$

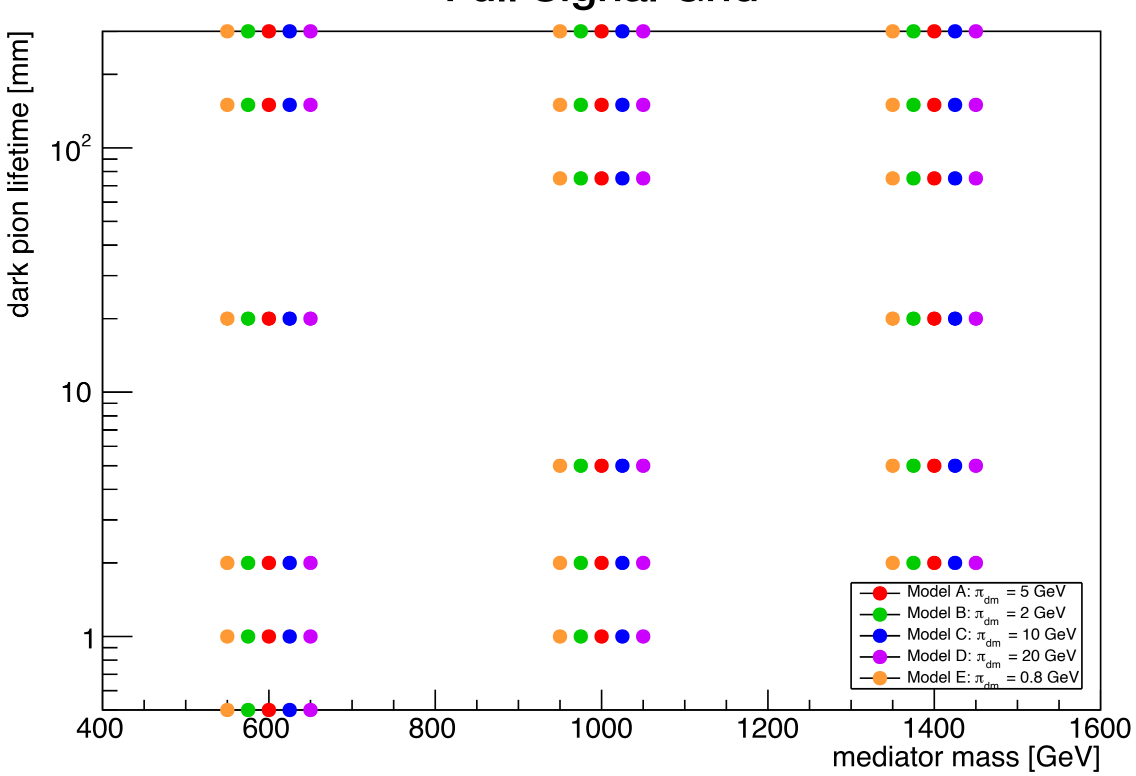
FINAL STATE: 2 QCD JETS + 2 EMERGING JETS

Signal Modeling

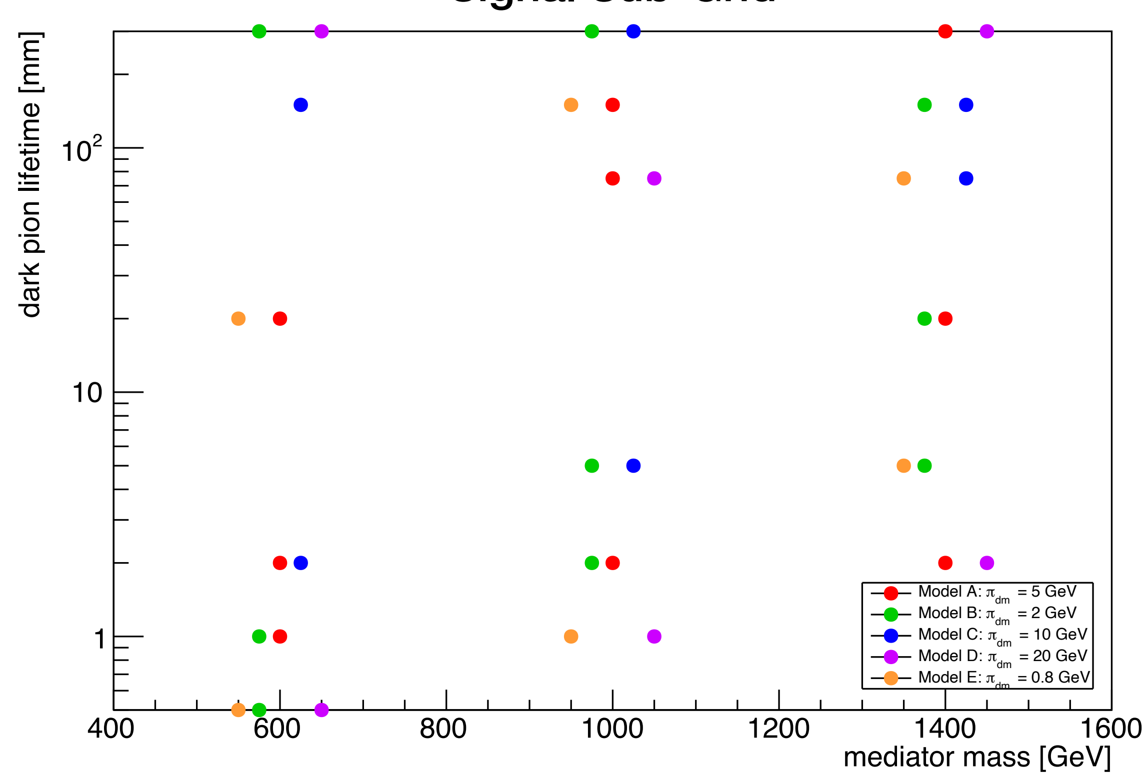
- **Five benchmark signal models** = simplified models of dark QCD in context of HV and defined by mass of dark sector, with **mediator mass** and **dark pion lifetime** left as free parameters
- **90 signal points**: **5 models** with **3 mediator masses** and **6 dark pion lifetimes** (intended to span parameter space entirely) → **36** out of 90 points currently available

	Model A	Model B	Model C	Model D	Model E
Dark QCD mass	10 GeV	4 GeV	20 GeV	40 GeV	1.6 GeV
Dark Rho Mass	20 GeV	8 GeV	40 GeV	80 GeV	3.2 GeV
Dark Pion Mass	5 GeV	2 GeV	10 GeV	20 GeV	0.8 GeV
Mediator Mass	600; 1000; 1400 GeV				
Dark Pion Lifetime	0.5, 1, 2, 20, 150, 300; 1, 2, 5, 75, 150, 300; 2, 5, 20, 75, 150 300 mm				

Full Signal Grid

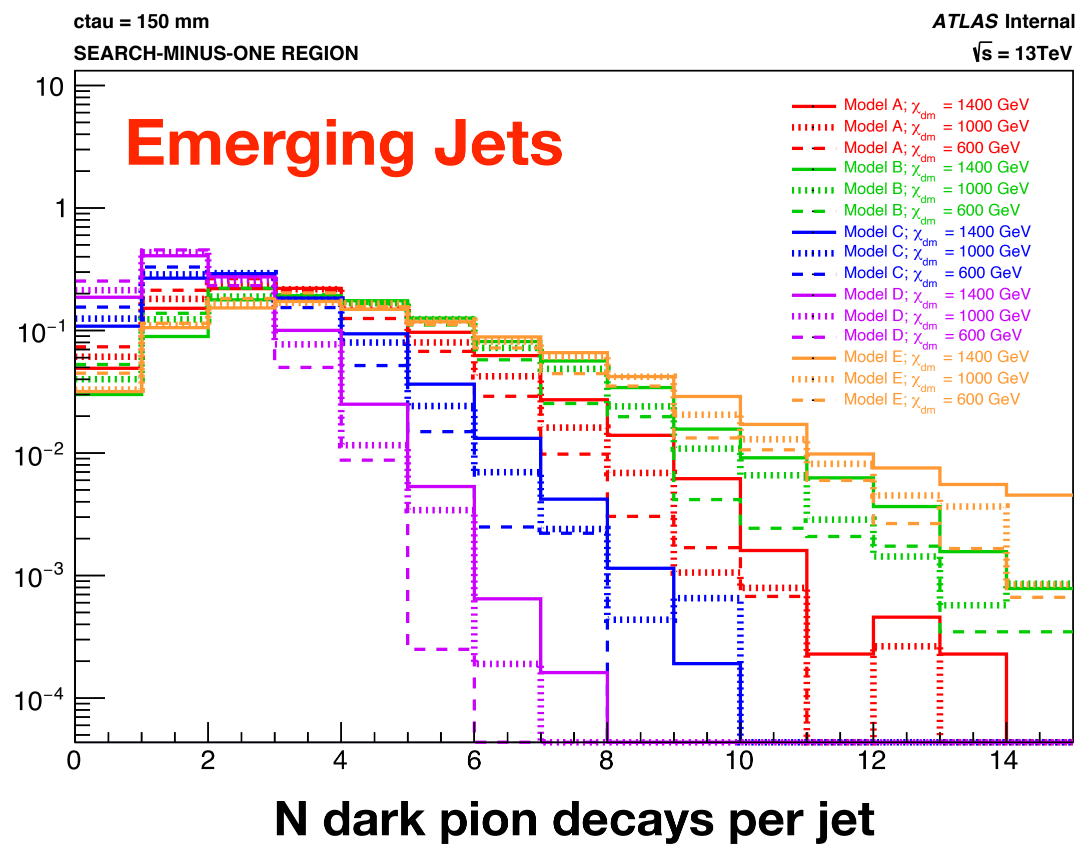


Signal Sub-Grid



Signal Event Generation

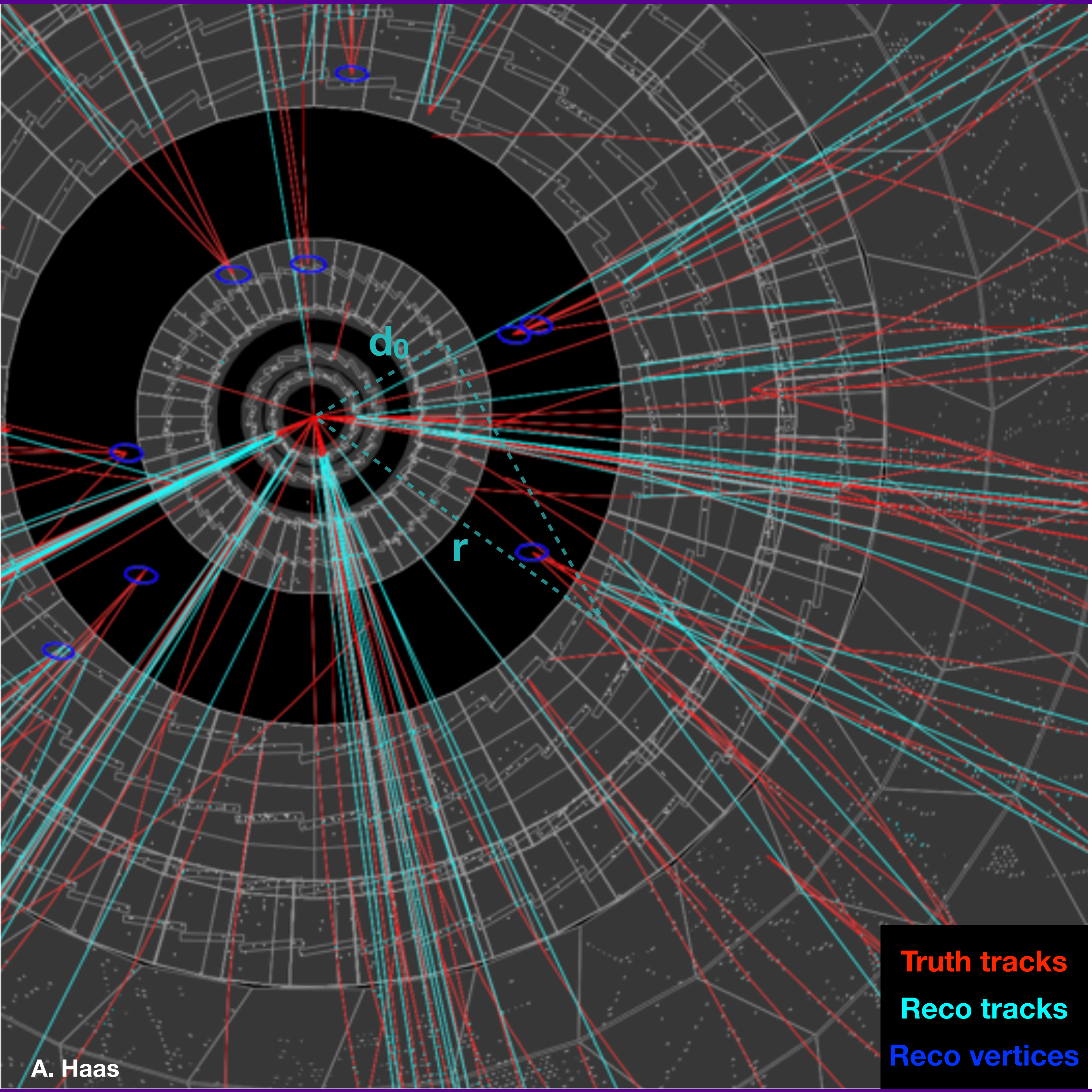
- Simulated Monte Carlo events generated with **Pythia 8.230 Hidden Valley** scenario with **running dark gauge coupling** and truth-level **four-jet generator filter**
  - Running coupling needed to produce realistic events with QCD-like dark sector → more accurately reproduces QCD-like parton showering, yielding narrower, more conical dark jet objects
  - Dark quark masses set by dark confinement scale
  - Generator filter applied to increase statistics for signal points with otherwise low efficiency: four truth jets with $p_T > 100$ GeV and $|\eta| < 2.7$ → looser versions of offline reco-level signal jet cuts



Generated truth-level events show multiple dark pion decays per EJ, as expected

- After event generation, MC signal samples run through full detector simulation and reconstruction → LLP searches require dedicated RPVLL reconstruction with specialized algorithms…

RPV/LL Reconstruction

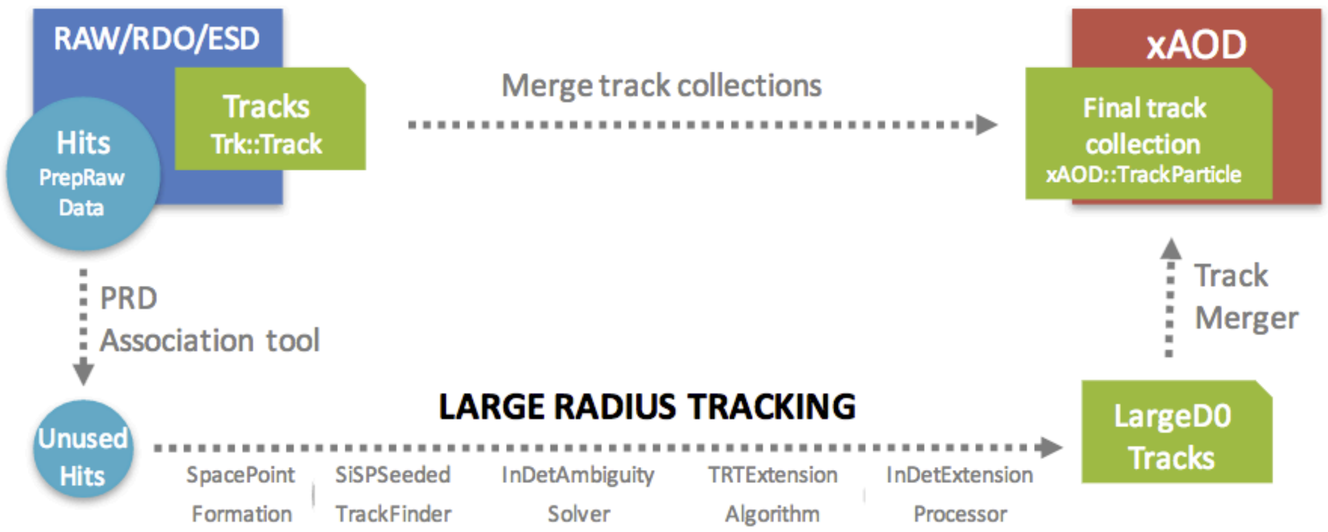


- LLP searches look for evidence of neutral long-lived particles decaying to charged tracks at significant distances from PV
- → LLP reconstruction objects = displaced secondary tracks and vertices with large impact parameters and large radial positions
- Standard ATLAS reconstruction inefficient for LLP searches
    - Standard reconstruction optimized for prompt tracks and vertices with $|d_0| < 10$ mm and $r < 20$ mm to conserve resources
- …can recover **displaced tracks and vertices** with special **LLP-optimized reconstruction** in **dedicated stream = RPVLL reconstruction**
    - **Large-d0 tracking (LRT)** + **secondary vertexing (VSI)** algorithms run on top of standard track and vertex reconstruction
    - → RPVLL setup applied to **filtered subset of data** in dedicated RPVLL stream during central reprocessing of data

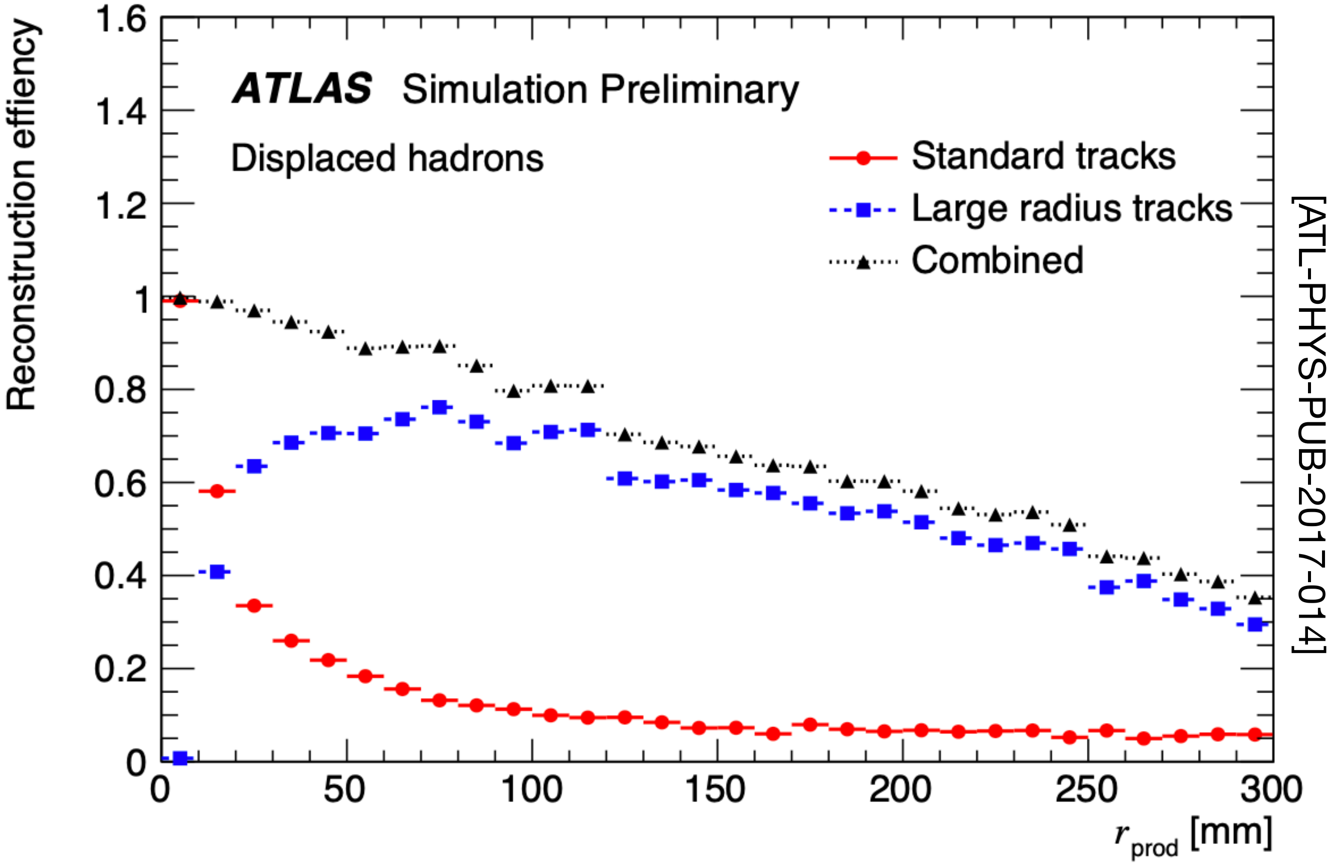
Large-Radius Tracking (LRT)

- Additional "third-pass" tracking → uses leftover hits and applies relaxed selection criteria on track impact parameters and hit multiplicities
  - Analogous to first-pass "inside-out" tracking of standard algorithm but for loosened requirements

J. Duarte-Campderros



Tracking algorithm	Standard inside-out	Large radius
Maximum d_0 [mm]	10	300
Maximum z_0 [mm]	250	1500
Maximum $\|\eta\|$	2.7	5
Minimum p_T [MeV]	500	900
Maximum shared silicon modules	1	2
Minimum unshared silicon hits	6	5
Minimum silicon hits	7	7
Seed extension	Combinatorial	Sequential

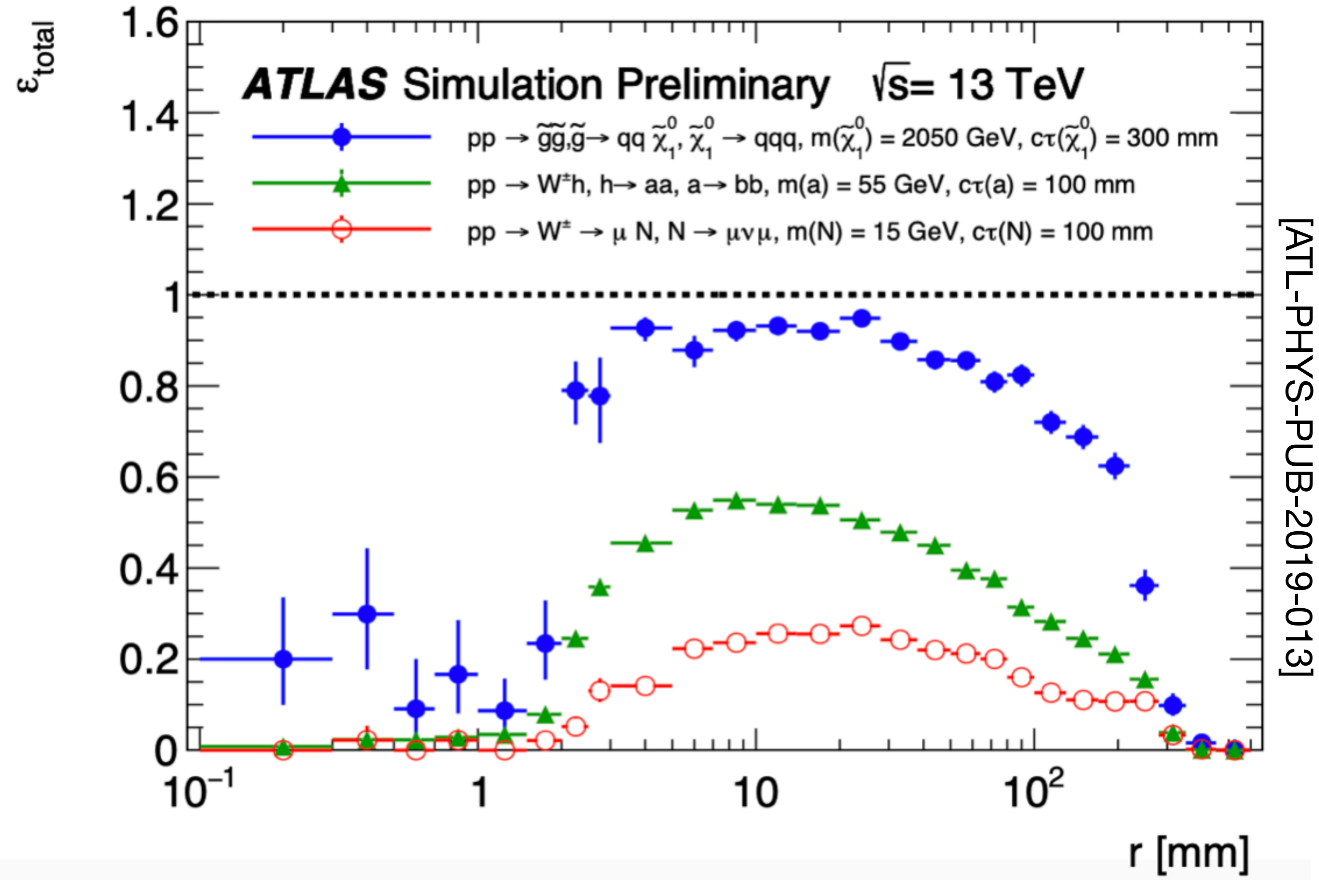


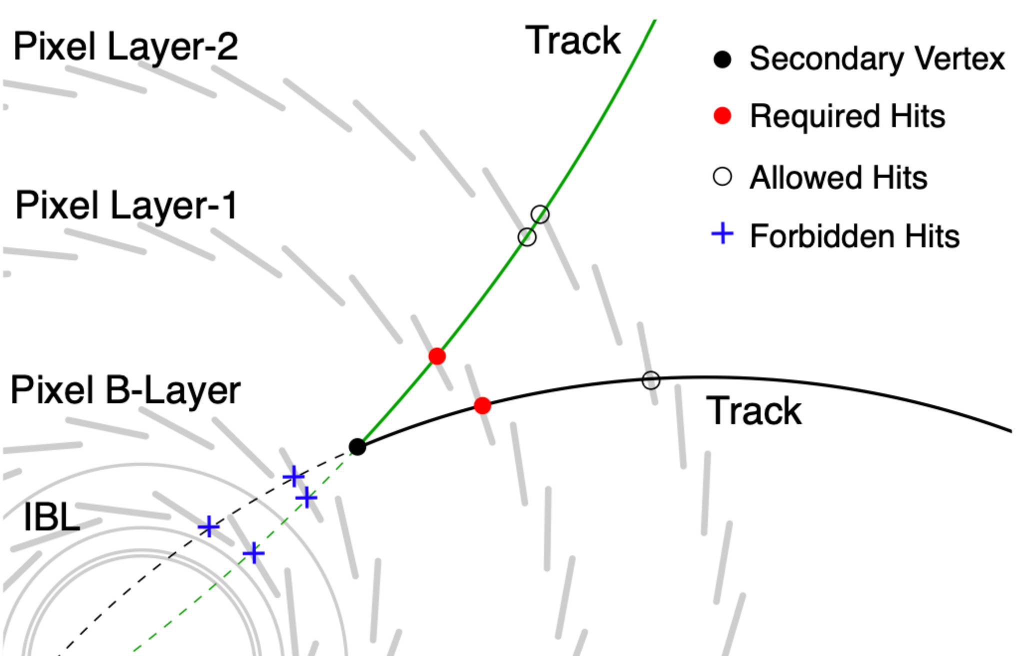
- Significant increase in combined tracking efficiency, particularly for r > 50 mm, but drops off above r ≈ 300 mm → r = radial position of decay vertex producing track
- Large fake rate ~80% presents challenge for DV reco

Secondary Vertexing (VSI)

- Dedicated secondary vertexing algorithm designed for LLP searches → uses combined collection of standard and LRT tracks to reconstruct displaced decay vertices produced far from IP

1. **Seed track selection** based on quality criteria to reject tracks incompatible with LLP decay → "selected" tracks
2. **Two-track seed finding** of set of pairs of selected tracks loosely compatible with LLP decay, based on initial position estimate, precision fit, and hit pattern
3. **Multi-track vertex forming** from two-track seeds with incompatibility graph method and **track rearrangement** to resolve ambiguities
4. **Vertex merging** of split and nearby vertices
5. **Track attachment** of tracks initially unselected for vertex seeding but compatible with reconstructed vertices → "attached" or "associated" tracks





- Efficiency decreases with increasing radial position and decreasing track multiplicity
- → LRT hit requirements fail for r ≳ 300 mm and fewer outgoing particles means lower probability of sufficient number of reconstructed tracks for vertex formation
- VSI algorithm suffers large fake rates from random track crossings and material interactions in detector

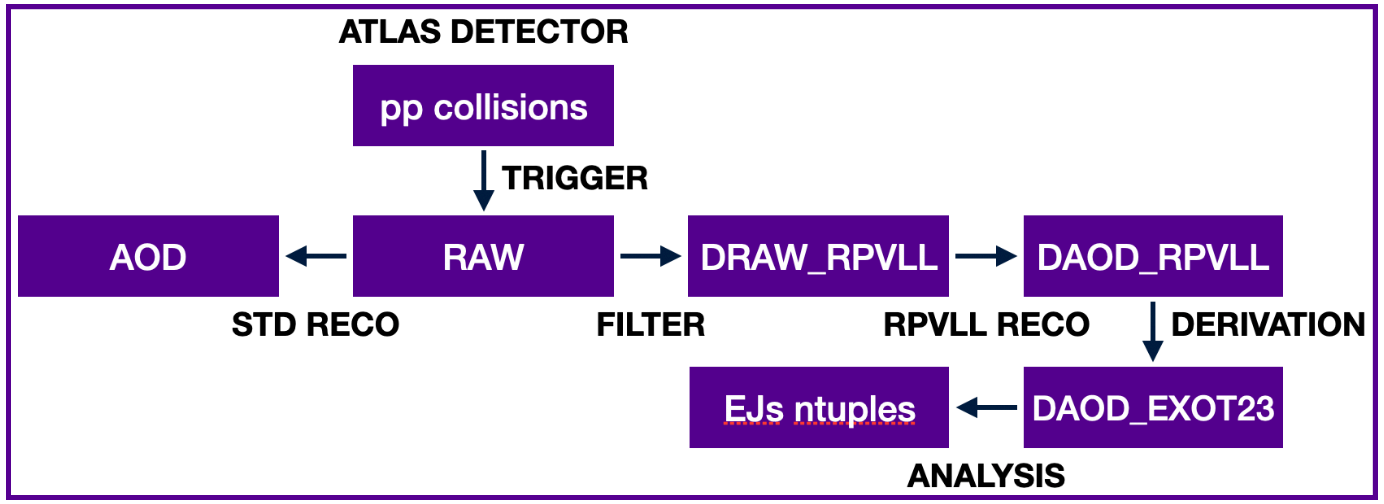
DRAW_RPVLL Filter

- RPVLL reconstruction expensive → RAW events containing LLP signatures filtered online into dedicated **DRAW_RPVLL stream** and later reconstructed with RPVLL re-tracking and re-vertexing

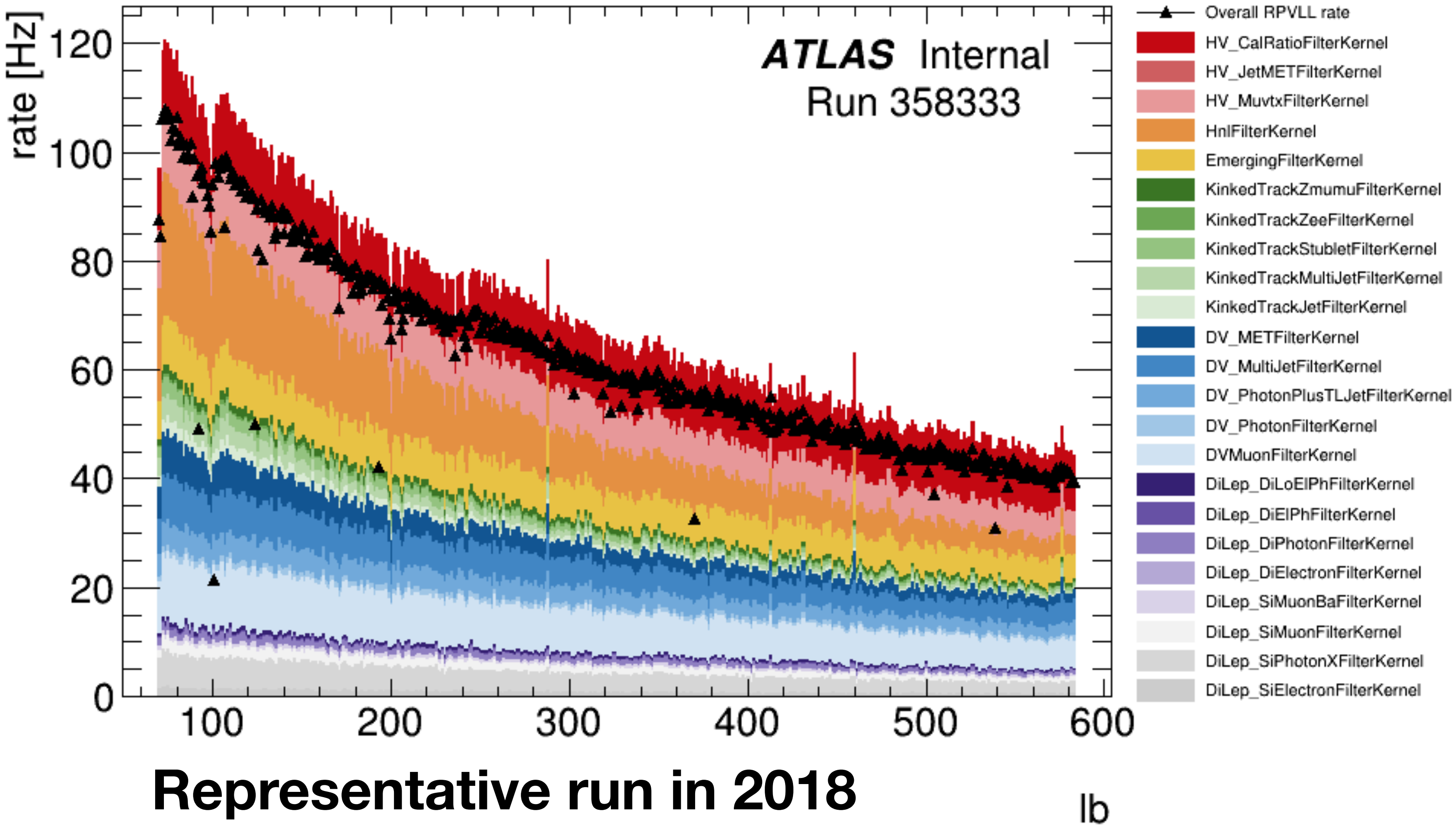
- **DRAW_RPVLL filters** select events relevant to LLP searches → filters defined by selections on triggers and loose cuts on standard reconstruction objects

  - Events selected by filters collected into DRAW_RPVLL stream, reducing size of RPVLL data to ~5-10% of full dataset → events in DRAW_RPVLL stream reconstructed with LRT + VSI

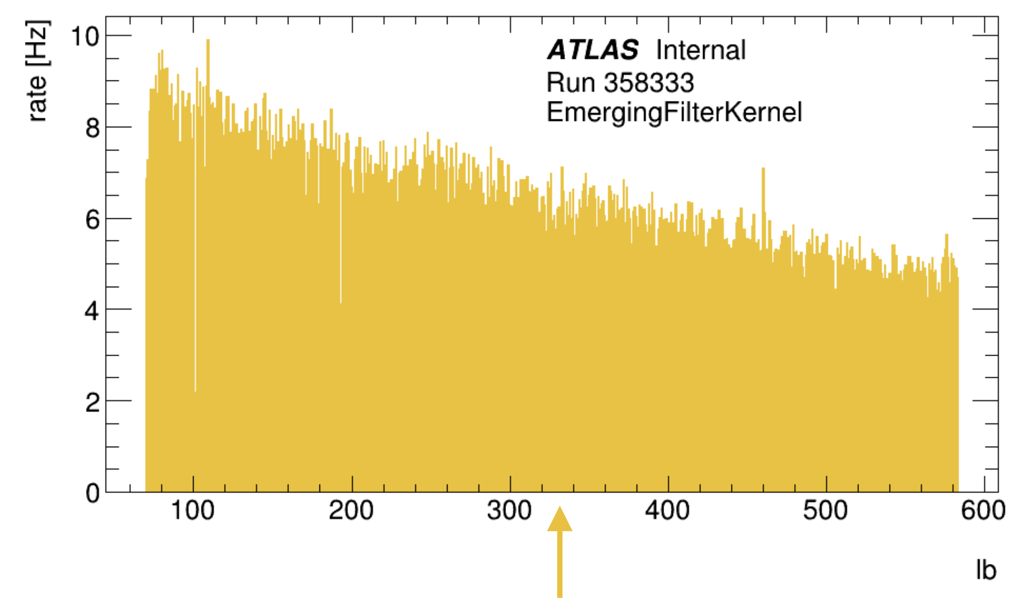
  - Average overall filter rate required low (~30 Hz) without sacrificing signal efficiency → DRAW_RPVLL monitor keeps rates in check…



Standard reco rate ~10 s/event → 100x increase for RPVLL reco with rate ~1000 s / event → 250 billion events to reconstruct → 100,000 cores to reprocess full dataset in one month



Representative run in 2018



EMERGING FILTER

- **Four-jet signal** events: lowest unprescaled **four-jet trigger** + ≥ **4 jets** with **p_T > 120 GeV** and **$|\eta|$ < 2.5**
- **Dijet background** events: heavily prescaled **single jet trigger** + ≥ **2 jets** with p_T > 120 GeV and $|\eta|$ < 2.5

Signal Selection Overview

Aim of analysis to select signal events containing EJs signature characteristic of simplified dark QCD models → selection criteria define signal region

Event Preselection

- DRAW_RPVLL + DAOD_EXOT skimming
- Event cleaning + quality cuts
- HS: ≥1 PV with ≥ 2 tracks
- ≥ 2 baseline jets

Search Region Selection

- **Four-jet trigger(s)**:
  - HLT_4j100 or HLT_4j120
- **Four jets** with
  - **p_T > 120 GeV**
  - **$|\eta|$ < 2.5**
- Leading 4-jet **H_T > 1000 GeV**

Validation Region Selection

- **Single-jet trigger(s) + two or three jets**

Signal Selection

- **N_{DV}** ≥ 2, 3, or 4
- **N_{EJ}** ≥ 0, 1, 2
- **X^{NJet}** > Y

→ signal DVs + EJs subject to quality criteria
→ X^{NJet} = leading four-jet or dijet observable

Loose
Material Map veto
Track cleaning
Fiducial volume: $r < 300$ mm and $\lvert z \rvert < 300$ mm
Quality of fit: $\chi^2/N_{\mathrm{DoF}} < 5$
K-short mass cut: $m > 0.7$ GeV
Medium
Loose selections
$p_{\mathrm{T}} > 2.5$ GeV
Tight
Medium selections
minimum track $\lvert d_0 \rvert < 10$ mm
minimum track $\lvert z_0 \rvert < 100$ mm
minimum square-root track d_0-error < 0.5
minimum square-root track z_0-error < 1.5

DV SELECTIONS

→ reject fake / SM background DVs from hadronic interactions and random track crossings

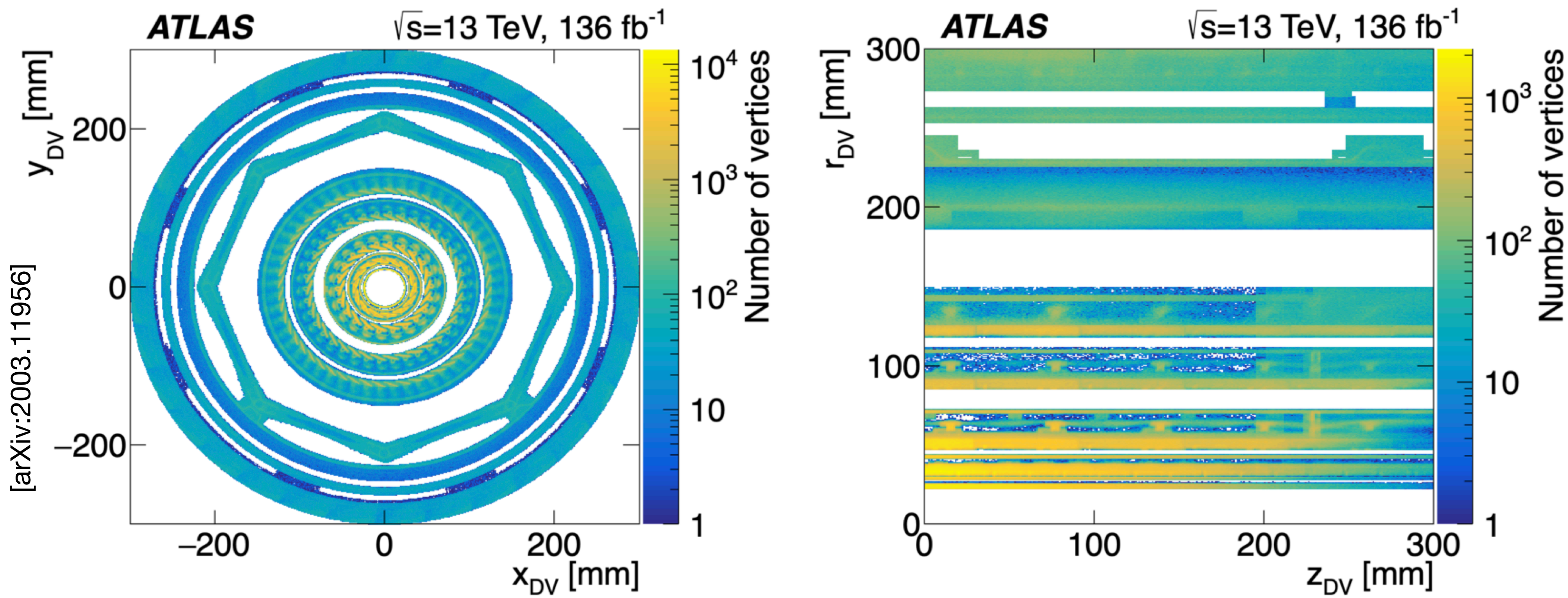
→ signal DVs w/in dR < 0.6 of leading jet

EJ SELECTIONS

→ discriminate between signal jets originating from dark pion decays within dark jets and QCD jets from SM quarks

Loose
$p_{\mathrm{T}} > 200$ GeV
$\lvert\eta\rvert < 2.0$
$m > 25$ GeV
$p_{\mathrm{T}}^{N_{SV}} > 5$ GeV
$N_{\mathrm{jet-trk}}^{N_{SV}} > 2$
Tight
Loose selections
$p_{\mathrm{T}}^{N_{SV}} > 7.5$ GeV
$N_{\mathrm{jet-trk}}^{N_{SV}} > 4$

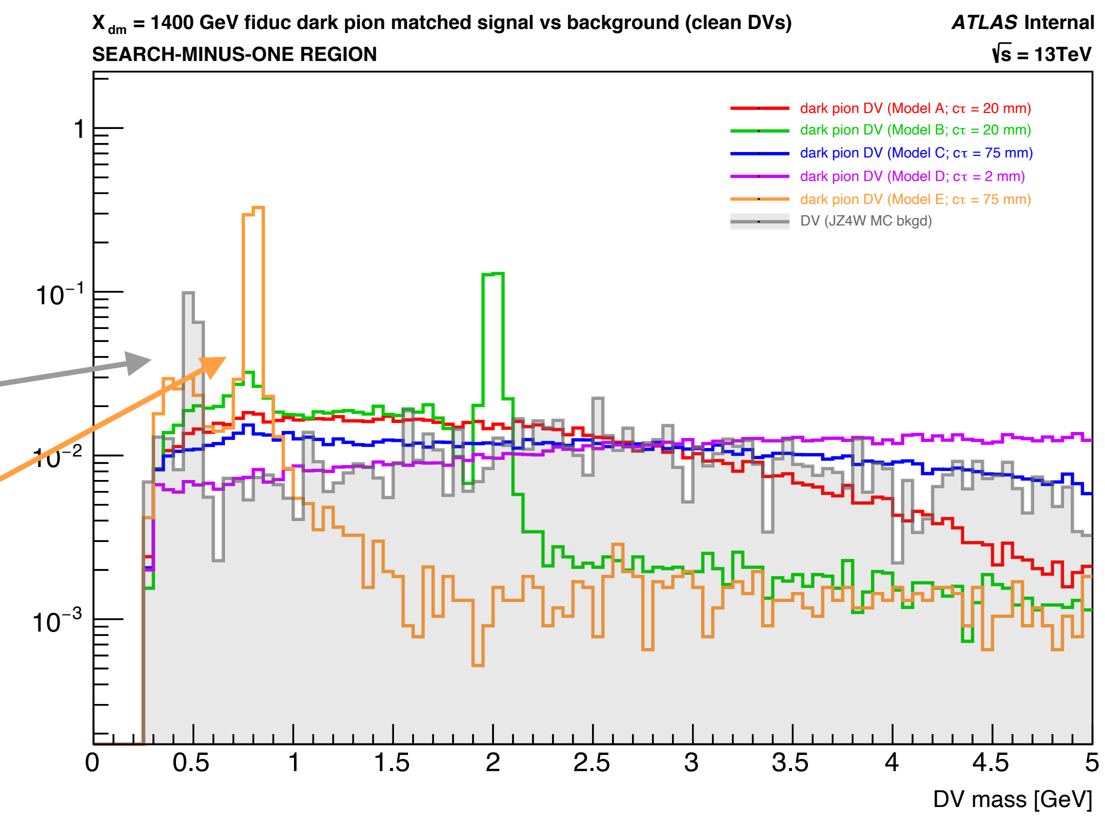
Material Map Veto



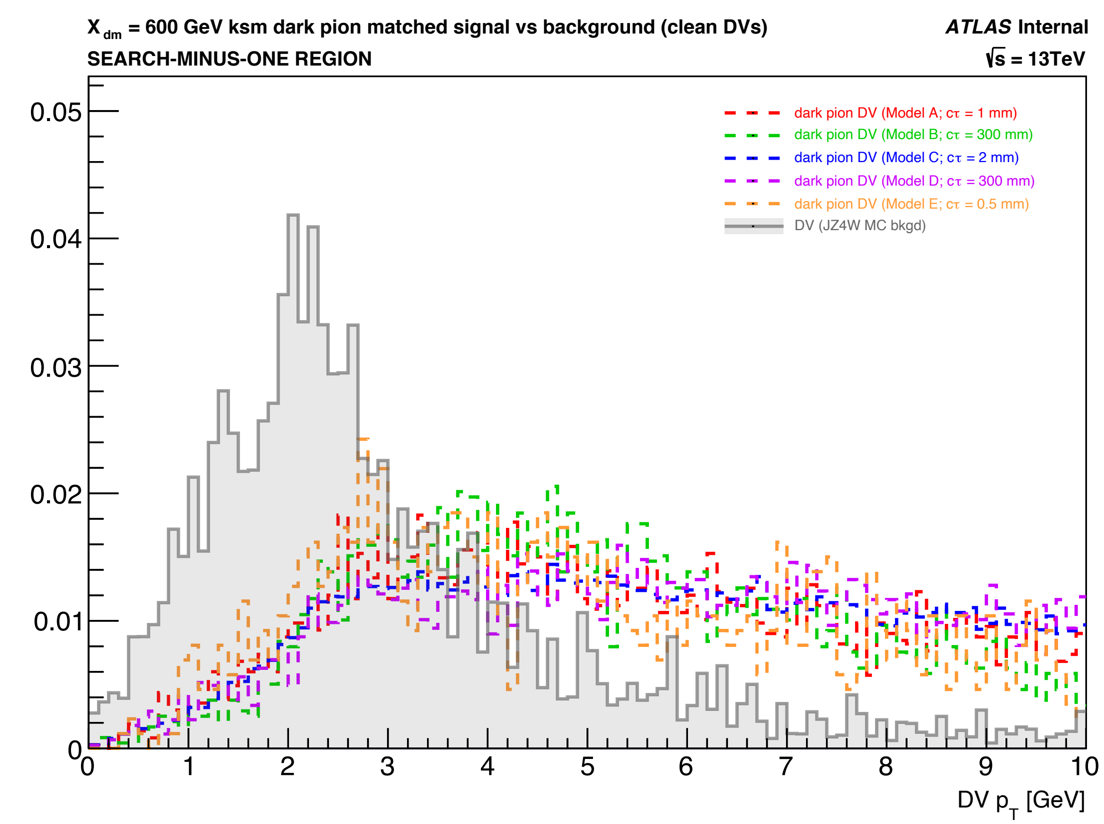
- DVs rejected if reconstructed inside material-rich regions of detector, as determined by 3D material map
- Hadronic interactions with detector material involve SM particles produced in pp collision hitting detector and creating showers of secondary particles that look like DVs → always occur in regions material → remove all DVs found in detector material
- Material map = simplified map of detector material, covering $r < 300$ mm, $|z| < 300$ mm, and $|\phi| < \pi$, indicating presence or absence of material at given point in detector
  - Derived from measurements of DVs in data → high density of DVs in region indicates material
- → material map veto applied by comparing locations of material in map to locations of DVs in samples

Kinematic Cuts

- K-shorts decay to charged pions with macroscopic lifetimes: $c\tau_{K_S} \sim 27$ mm, yielding SM background DVs → remove DVs around K-short mass window
  - K-short mass: $m_{K_S} \sim 0.5$ GeV
  - Lower limit of EJs signal: $m_{\pi_d} = 0.8$ GeV
  - → select DVs with **m > 0.7 GeV**



- After Loose selections, DVs dominated by fakes produced from random track crossings → fakes tend to be low mass and low multiplicity
  - EJs signal DVs also low mass and multiplicity, so cuts on these variables risk loss of signal efficiency → cut on p_T instead (more freedom in selection)
  - → DVs required to have **p_T > 2.5 GeV**

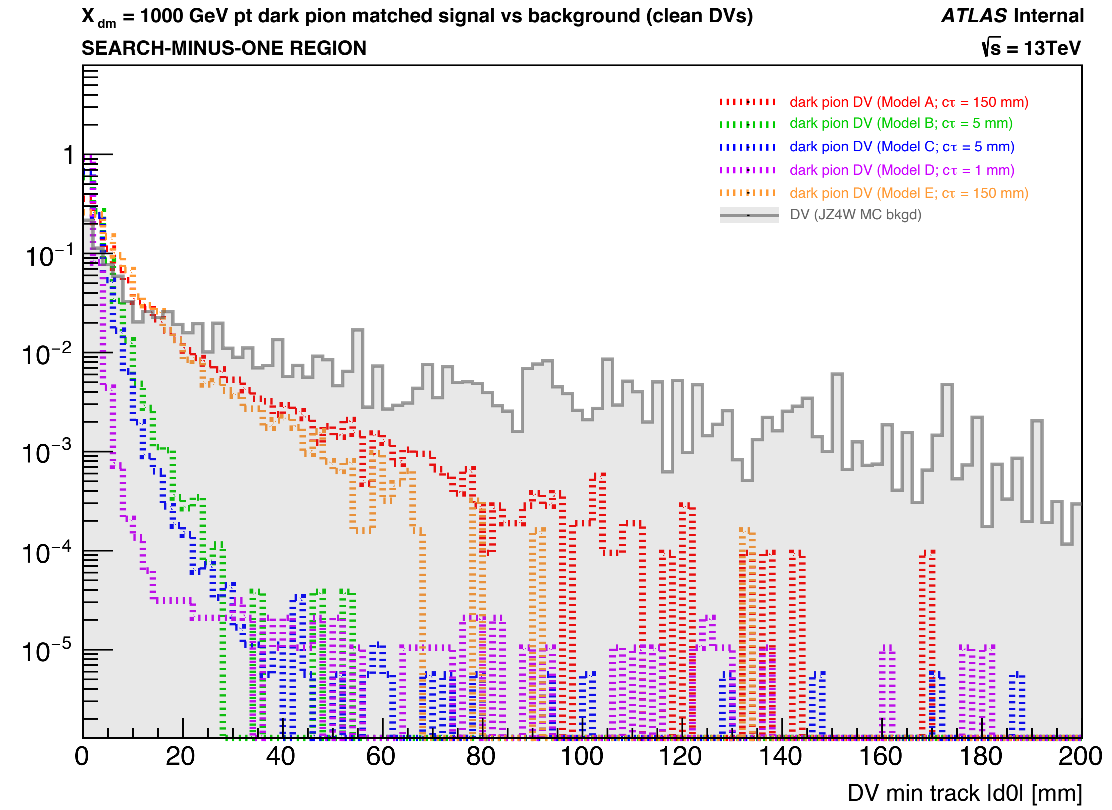


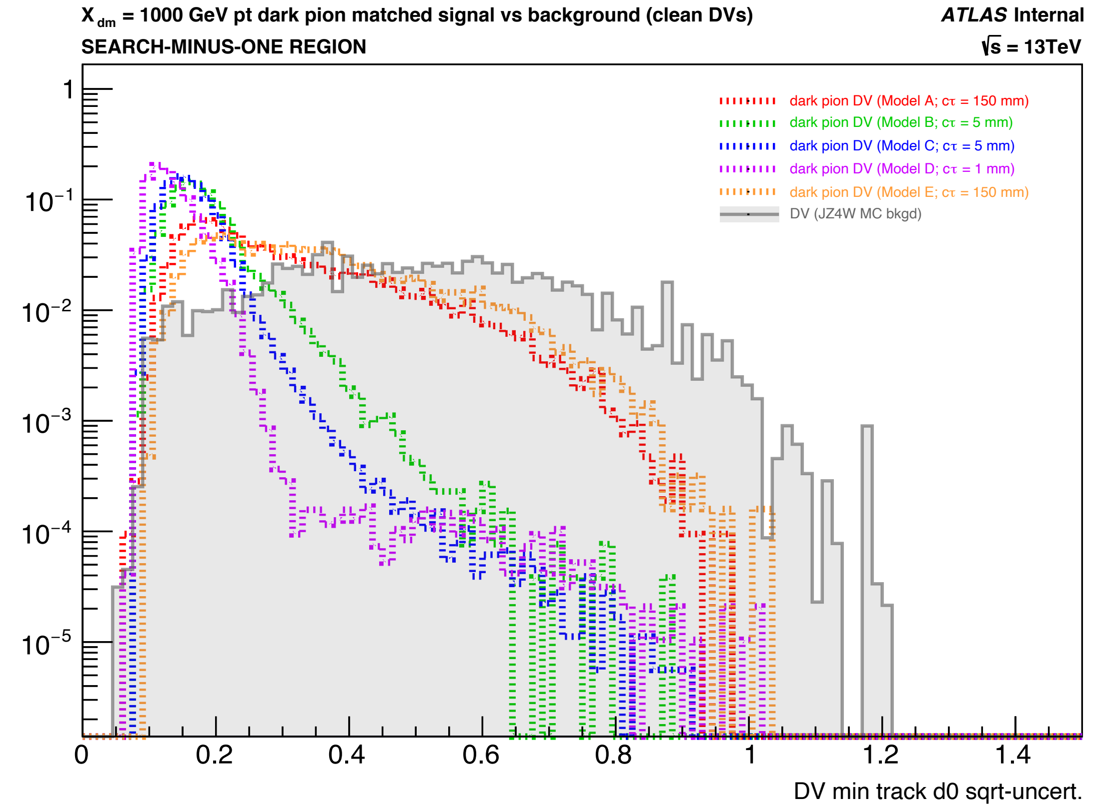
Impact Parameter Cuts

- Loose selections of LRT yield large rates of fake tracks, which randomly cross to produce fake reconstructed DVs → background DVs more likely to contain tracks with large impact parameters and large related uncertainties, indicative of fake tracks

- Cuts applied to minimum of track impact parameters and impact parameter uncertainties

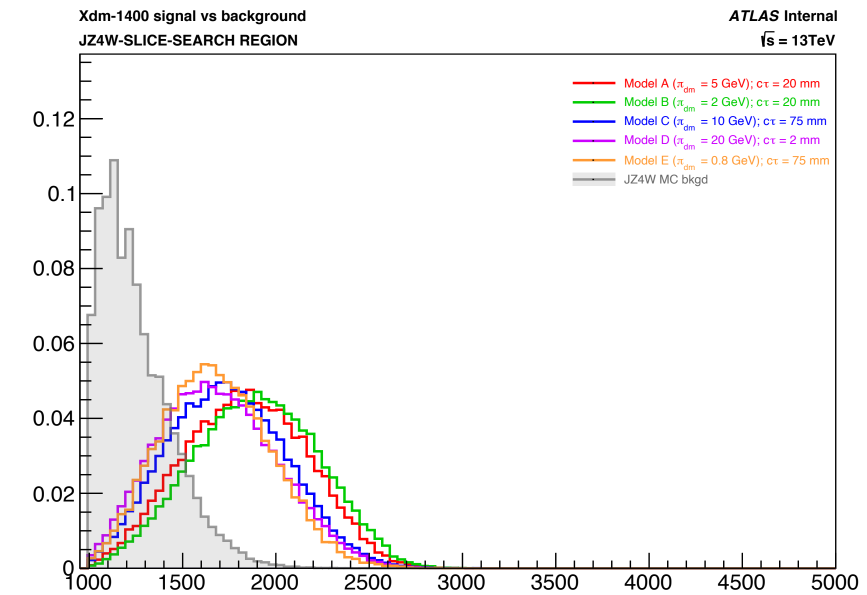
- DVs required to have:
  - **min-|d_0| < 10 mm + min-|z_0| < 100 mm**
  - **min-d_0-sqrt-error < 0.5 +min-z_0-sqrt-error < 1.5**





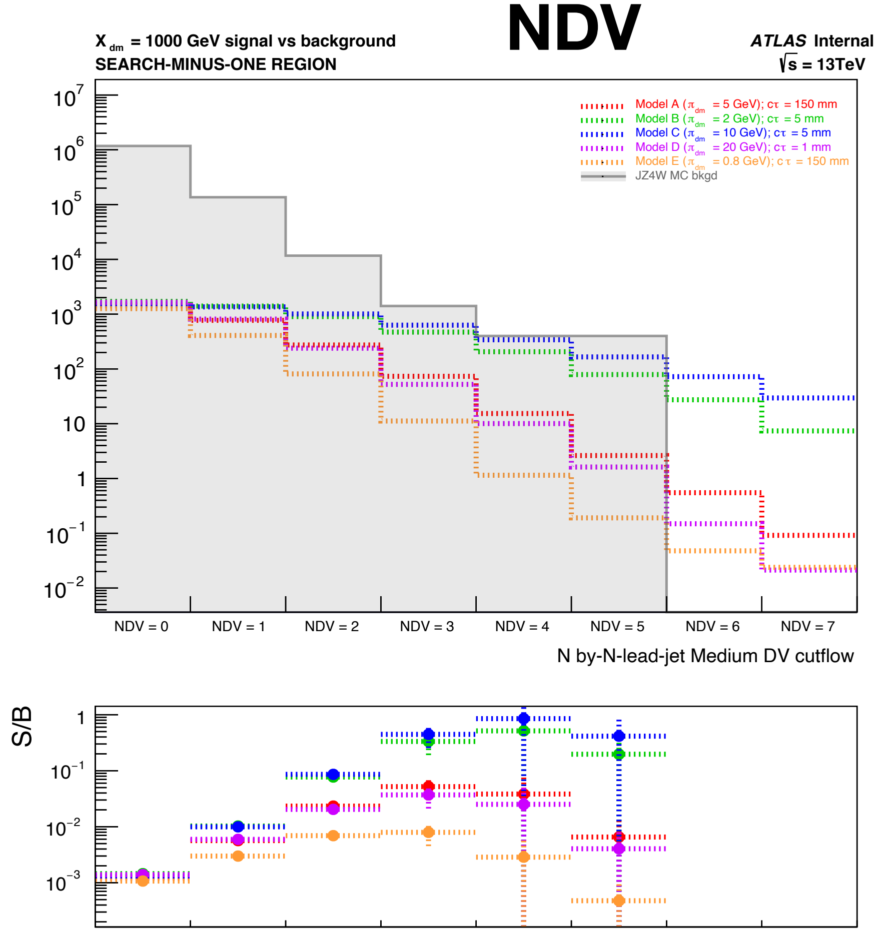
NDV, NEJ, NJetX Cutflows

- Individual signal selections effective at reducing background while maintaining signal
- Selections still to be optimized across all models to maximize sensitivity → want to keep S/B as high as possible for all points

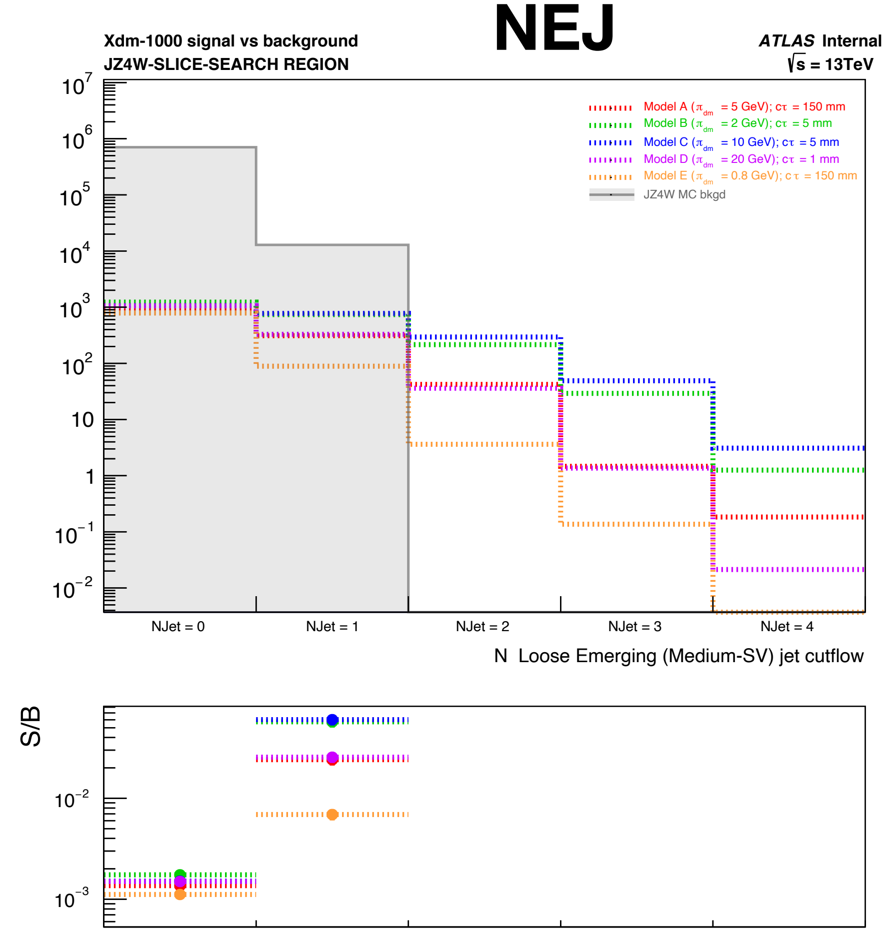


Leading four-jet H_T

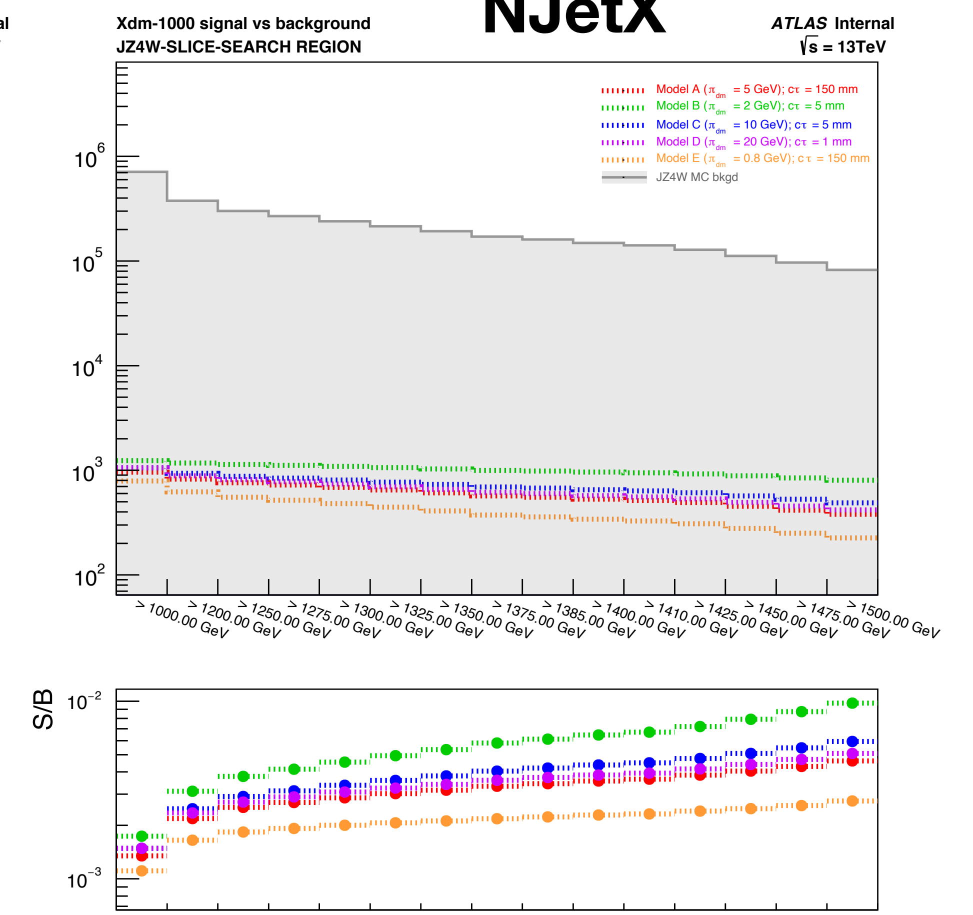
NDV



NEJ



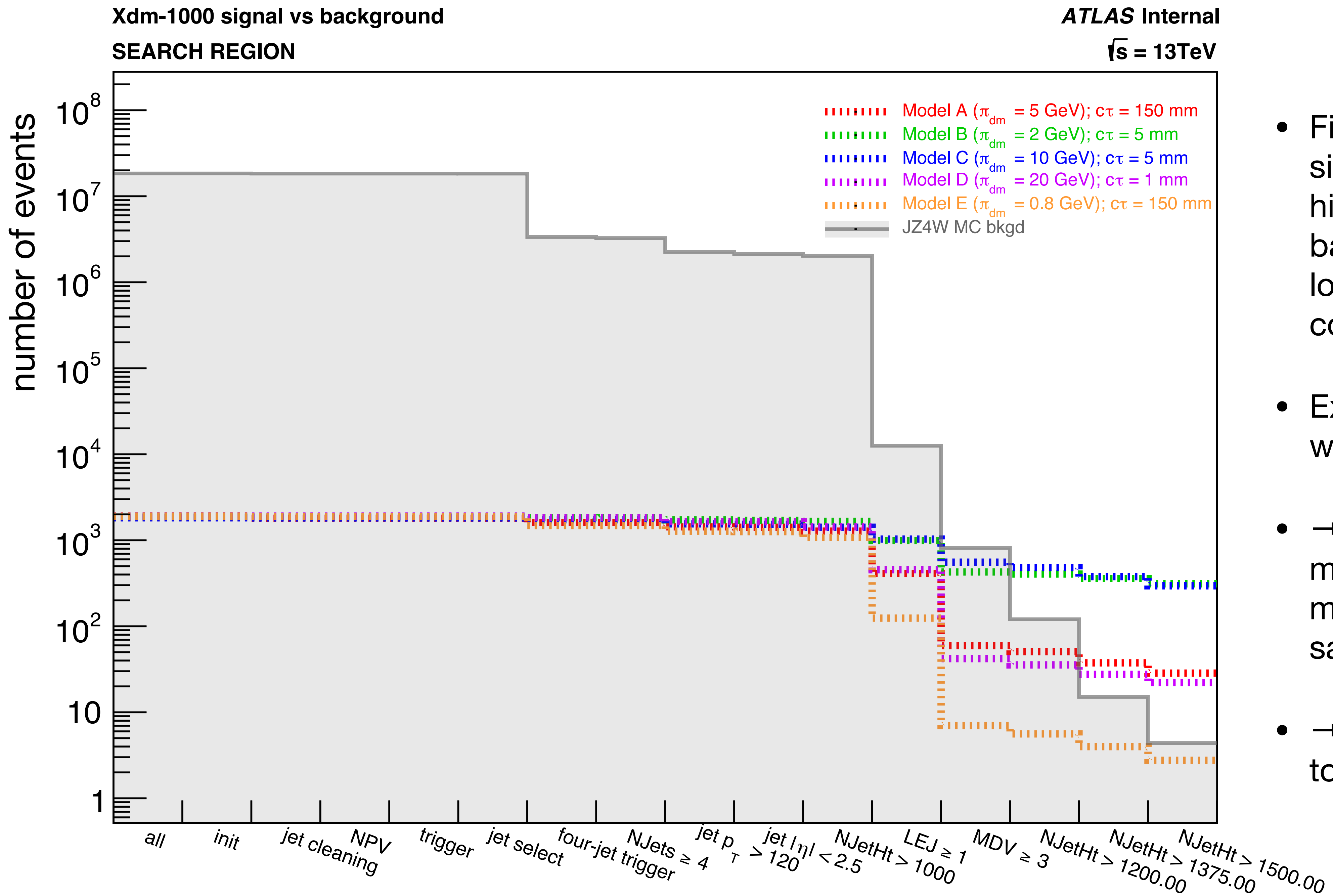
NJetX



Full Signal Region Cutflow

Final criteria defining potential signal regions = **N_{EJ} + N_{DV} + X^{NJet}** →
→ still to be tuned for maximal sensitivity across full signal grid

$N_{LEJ} \geq 1$ + $N_{MDV} \geq 3$ + $H_T^{NJet} > 1200\text{-}1500$



- Final N_{DV} + N_{EJ} + X^{NJet} signal region selections highly effective at reducing background without total loss of signal for certain combinations of cuts
- Example selection performs well for most signal points
- → low sensitivity for low mediator mass, low model mass, and extreme lifetime samples
- → optimal signal region still to be determined…

Background Estimation: ABCD Method

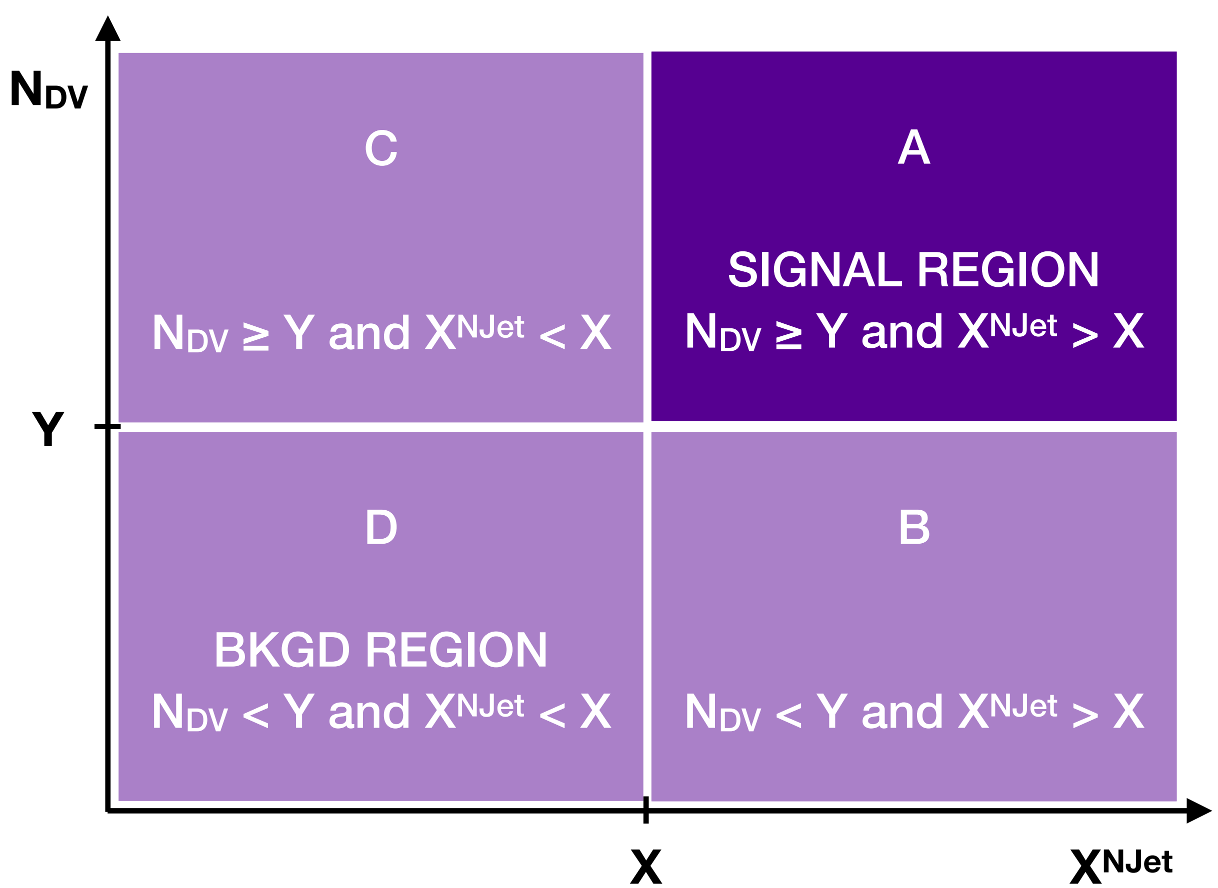
- Background = **SM QCD multijets** → estimation of background yield in signal region performed using **data-driven ABCD method**
- 2D ABCD plane where signal and background separated into four regions: A, B, C, and D → defined by two independent selections on **uncorrelated variables** forming part of final signal region: **DV multiplicity (N_{DV})** + **N_{jet} leading jets (X^{NJet}) = DV × NJetX** ABCD plane
- MC imperfect approximation of SM background → ABCD method allows us to measure background in data without unblinding analysis
- Measure **rate** of **background** events **faking signal** in one variable in **region free from signal** using **data** → if ABCD observables sufficiently **uncorrelated**, ratio holds across other variable
- Initial studies and validation tests performed with MC background, but actual background estimates will be done with full Run 2 data

ABCD method:

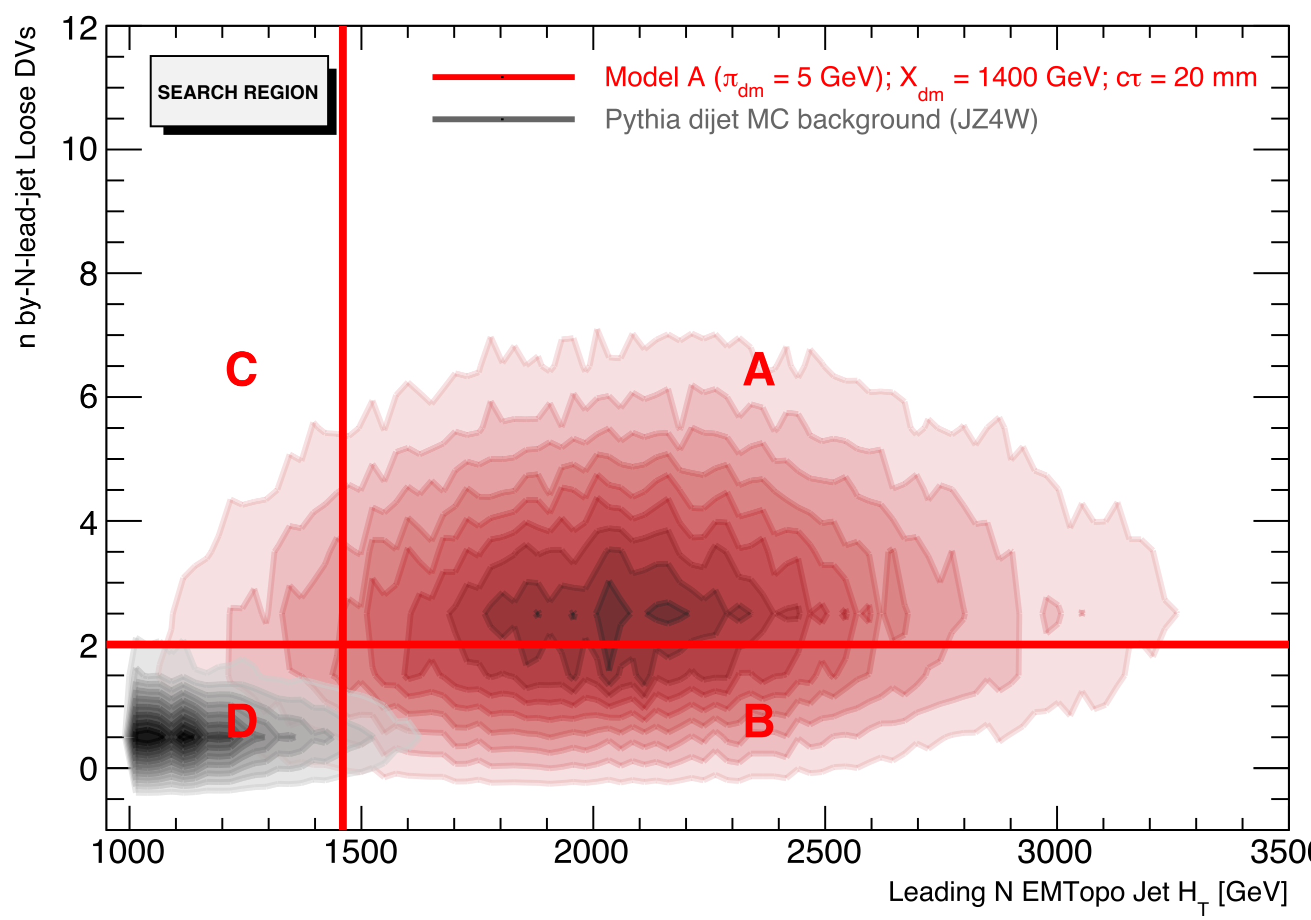
$$N_A^{\text{bkgd}} = N_B^{\text{bkgd}} \times N_C^{\text{bkgd}} / N_D^{\text{bkgd}}$$

Assumption:

$$N_C^{\text{bkgd}} / N_D^{\text{bkgd}} = N_A^{\text{bkgd}} / N_B^{\text{bkgd}}$$



ABCD Plane



Conclusion

Summary + Outlook

- Ongoing EJs search for evidence of new physics producing energetic four-jet + multi-DVs signal at ATLAS Run 2 using 139 fb^{-1} of data at sqrt{s} = 13 TeV → extensive work done designing analysis from ground up and putting all the pieces in place…
  - Overall design of analysis developed and implemented to target EJs signal amidst large SM QCD multijet background
  - Dedicated Emerging DRAW_RPVLL filter + DAOD_EXOT23 derivation developed and applied to full Run 2 dataset to process events of interest
  - Custom analysis code framework - EmergingJetsAnalysis - built to perform studies and produce ntuples, histograms, and plots specific to EJs search → configurable for use by other RPVLL analyses
  - Full signal grid designed and 36 of 90 points produced for initial use
  - Extensive truth and Pythia validation studies performed in early stages of analysis development to optimally design signal grid such that it fully covers parameter space where analysis sensitive
  - Careful design and investigation of signal event and object selections, including DV + EJ WPs, in development of potential signal regions
  - Initial background estimation and validation performed
- Run 2 analysis nearly finished, with just a few steps remaining…
  - Optimize signal region, including DV + EJ WPs, to maximize sensitivity across signal grid
  - Calculate signal yields and efficiencies and estimate background yield in signal region using full Run 2 data
  - Fully validate background estimation method
  - Unblind analysis and discover emerging jets!

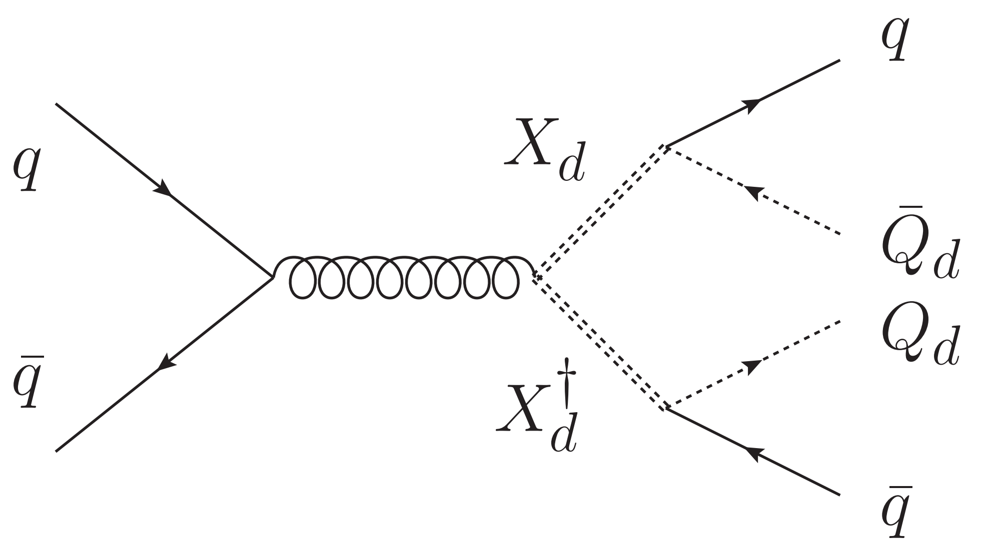
THANK YOU!

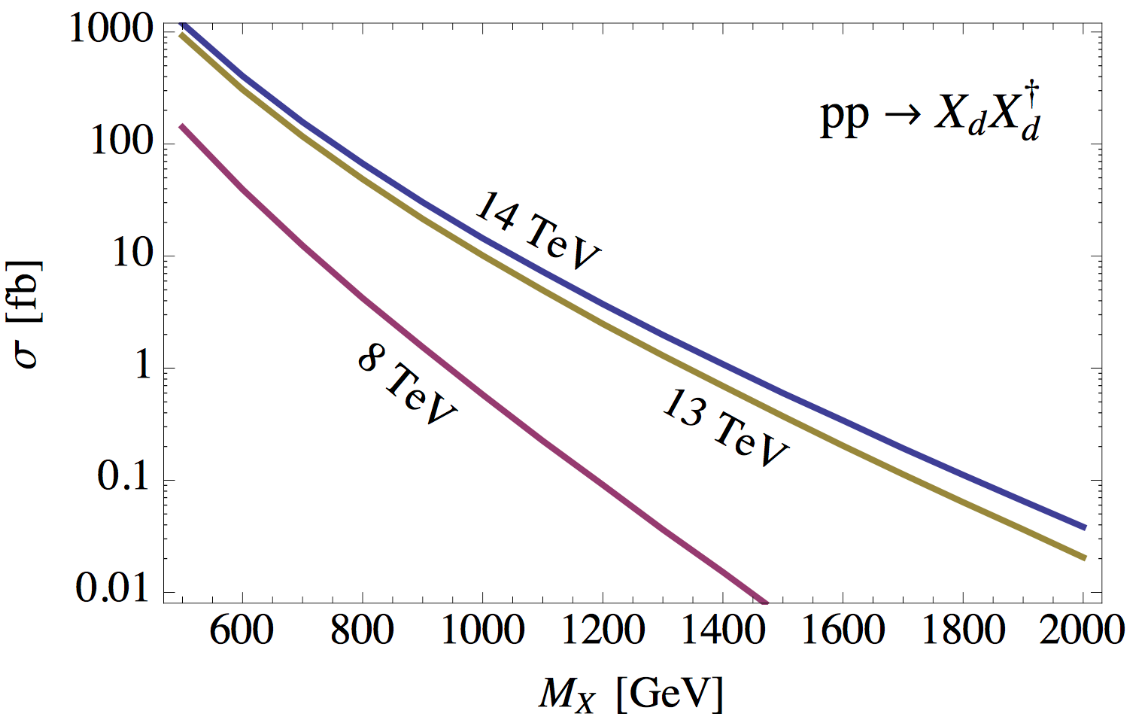
BACKUP

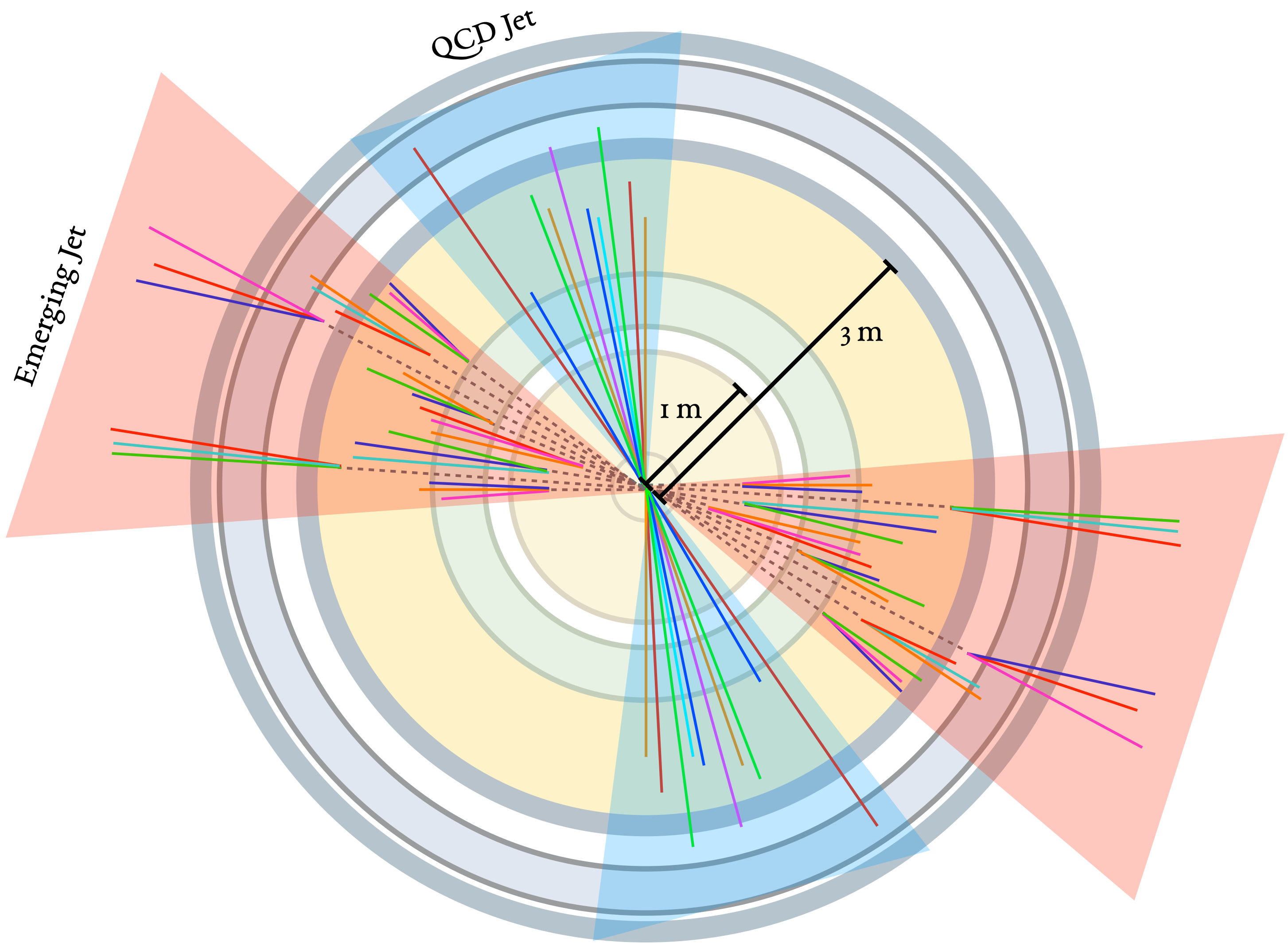
EJs Collider Phenomenology

HARD PROCESS: scalar mediator pair-production + subsequent decay

$$\boxed{p\,p \to X_d X_d^\dagger \to \bar{q}\,Q_d\,\bar{Q}_d\,q}$$





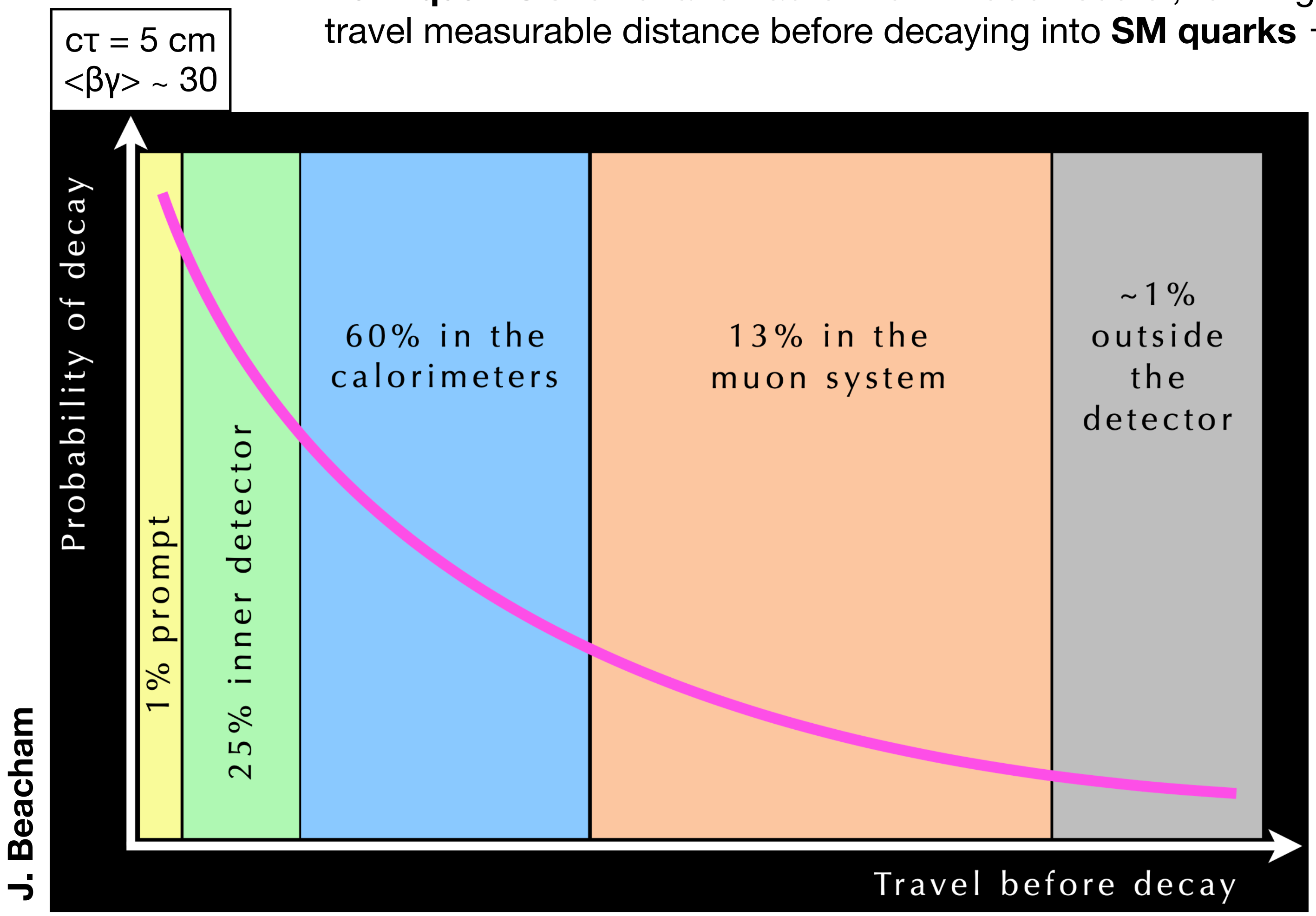


- SM quarks → standard QCD jets
- Dark quarks → dark jets of dark pions → emerging jets

FINAL STATE: 2 QCD JETS + 2 EMERGING JETS

Dark / Emerging Jet Production

- Dark quark undergoes parton showering and fragmentation in dark sector
  - **Dark parton shower** = radiation of dark gluons off dark partons and splitting of dark gluons into dark quark pairs
    - Depends on running of dark gauge coupling constant, i.e. numbers of dark colors and flavors, and produces QCD-like dark jets
  - **Dark fragmentation** = conversion of dark partons into dark hadrons (non-perturbative)
    - Dark baryon production suppressed → dark mesons = lightest dark pions π_d + heavier dark rhos ρ_d, with heavier states promptly decaying to lightest
    - **Dark quarks** shower and hadronize in hidden sector, forming prompt **dark jets** comprised mostly of dark pions → **dark pions** travel measurable distance before decaying into **SM quarks** → dark jets gradually turn visible, **emerging** into detector



- Observed lifetime of LLP in lab frame distributed exponentially according to **characteristic decay length βγcτ**
  - βγ = p/m = boost factor; cτ = proper lifetime
  - → harder particles travel further on average
- Dark pion decays produce **sub-jets of small numbers of SM hadrons from common DVs**, with more visible energy farther from IP
  - → jet ***emerges*** into detector with **multiple nearby displaced vertices** over wide range of displacements
  - → new detector object + distinct signature

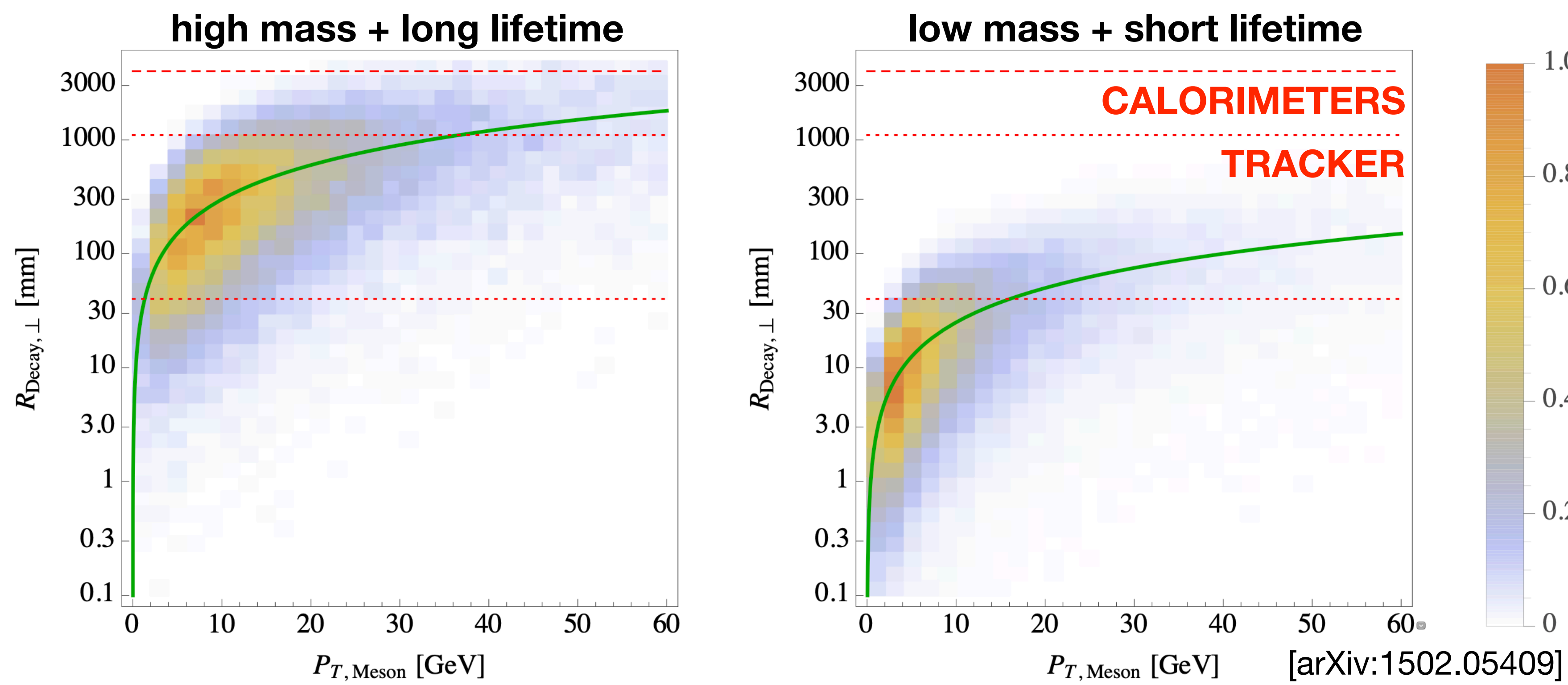
EJs Event Topology

- 2 SM quarks + 2 dark quarks → 4-jet final state: 2 QCD jets + 2 emerging jets
  - TeV-scale mediator + GeV-scale Q_d yields high-p_T jets and substantial scalar sum-p_T (H_T) per event

- Dark pion lifetime spans wide range and sampled from exponential
  - Most decays occur far from beam but well within tracker → clustered around average transverse decay length $\beta_T\gamma_T c\tau_{\pi d} = (p_T/m_{\pi d})c\tau_{\pi d}$

- → EJs contain high multiplicity of highly-displaced vertices within standard jet cone

- → **EJs event topology** = **four-jet** final state with **high-p_T jets, large H_T**, and **multiple DVs**



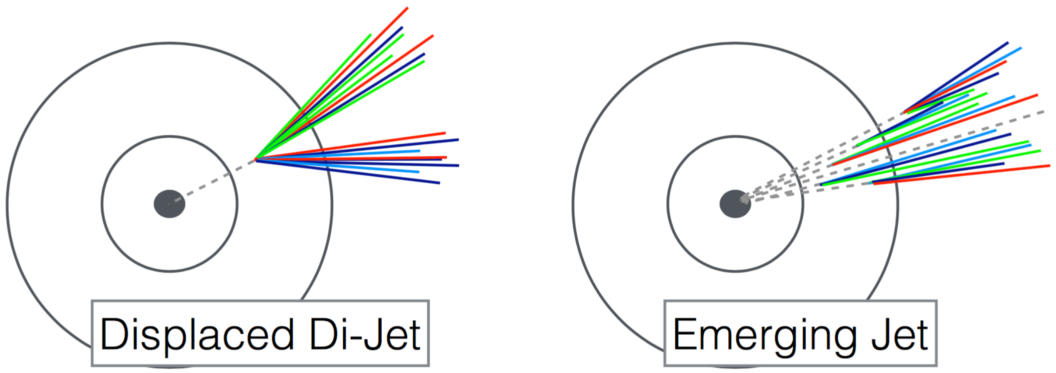
[arXiv:1502.05409]

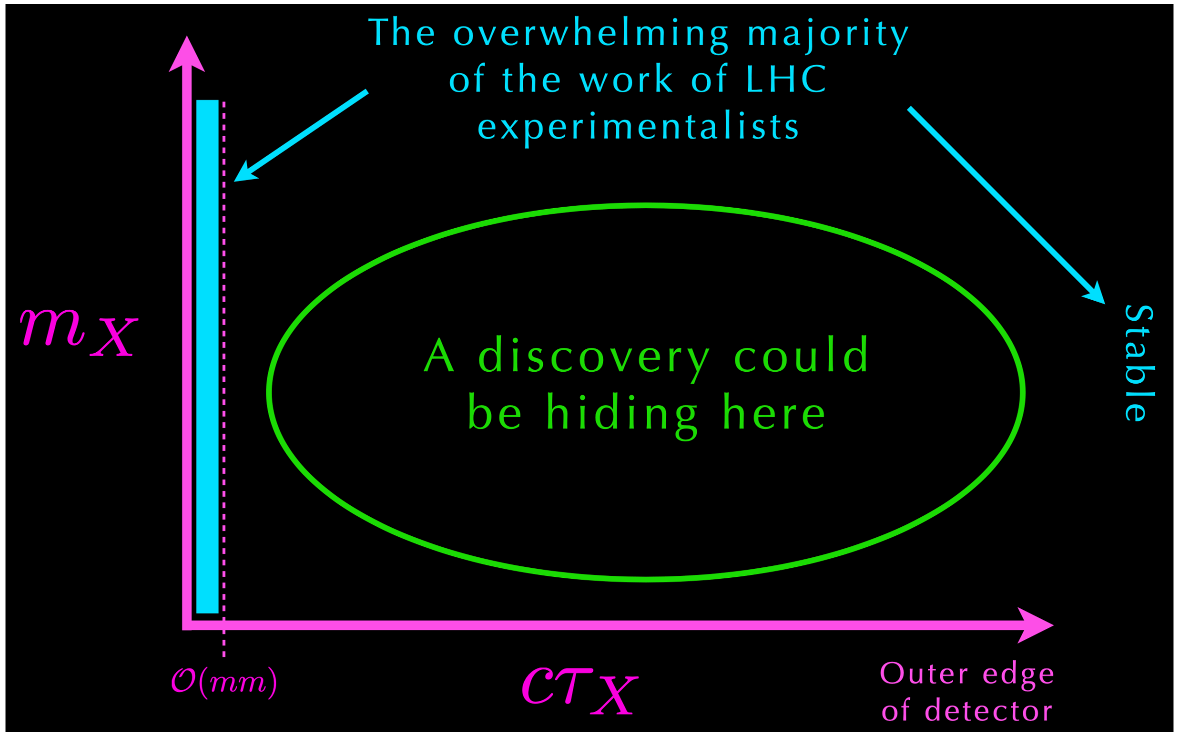
COLLIDER PARAMETERS

- Λ_d — **dark confinement scale** = approximate mass of dark hadrons, with 1-10 GeV range motivated by DM
- $m_{\rho d}$ — **mass of dark vector meson**, which decays to dark pion
- $m_{\pi d}$ — **mass of dark pseudoscalar meson**, which decays to SM with non-negligible lifetime
- $c\tau_{\pi d}$ — **lifetime of dark pion**
- m_{Xd} — **mass of bifundamental scalar mediator**, which decays to SM quark + dark quark
- → define dark sector and determine EJs event topology, i.e. where jets will emerge within detector and with what energy and DV multiplicity…

Analysis Overview + Search Strategy

- Ongoing search for EJs at ATLAS with full Run 2 dataset (139 fb^{-1}) → dedicated LLP search with unique signature
  - EJs = new detector objects = distinct signature → EJs events have DVs with lower masses and track multiplicities versus displaced multi-jets, along with multiple DVs in single jet cone
  - New signature-based search for **long-lived dark particles** with **displaced decays** within **calorimeter jets** → energetic **four-jet** final state with some number of **emerging-like jets** and large **displaced vertex multiplicity**
  - Search strategy hinges upon identifying secondary vertices from signal LLP decays
- LLPs open up new, largely unexplored territory at LHC…
  - …but also bring about own set of challenges at detector → require specialized **RPVLL techniques** for analysis
  - Common goal of LLP analyses to design signature-based but **model-independent** searches to target events with topologies consistent with signal while being sensitive to other related models involving similar scenarios, i.e. dark hadronization, of which details unknown





J. Beacham

- **Aim of analysis** to select pure enough sample of signal events, i.e. to **isolate EJs signal** from large SM QCD background
  - Carefully design selections to maximize EJs signal events and minimize SM background events in region of parameter space expected to be rich in signal → background reduction primary obstacle to search for rare and small signal process
  - Benchmark signal models in expansive signal grid used to examine EJs phenomenology across (m_{X_d}, $c\tau_{\pi_d}$) parameter space in order to study sensitivity of detector to EJs signal and to design analysis selections

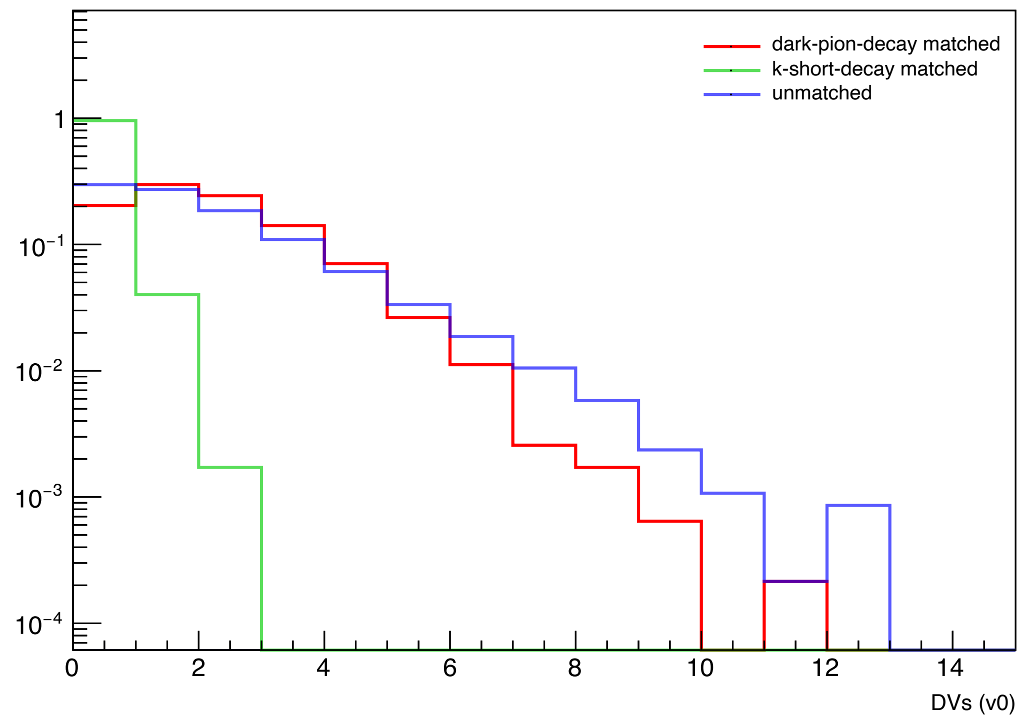
EJs Signal Grid Motivation

- **Benchmark models** span range of **dark QCD sector scale**
  - **Models A, B, and C = intermediate mass regime**, where search expected to be most sensitive
  - **Model D = high mass regime**, representing upper limit of search sensitivity → EJs topology breaks down and signal looks more like displaced dijets, as most jet energy contained in single massive dark pion
  - **Model E = low mass regime**, representing lower limit of search sensitivity → LLP reconstruction inefficient, as lighter dark pions yield softer and fewer tracks, making DVs harder to reconstruct
  - → heavier dark sectors produce larger DV track multiplicities and shorter average transverse decay lengths → lighter states are more boosted, so more likely to decay outside of the calorimeter and not be clustered into jets
  - → dark pion multiplicity inversely relate to model mass
- **Mediator masses** and **dark pion lifetimes** span theoretically-motivated **parameter space**
  - Lower mediator masses produce softer jets → if too low, no background discrimination
  - Average dark pion decay distance proportional to lifetime → longer lifetimes yield more decays in or outside of calorimeters, resulting in jets with lower momentum and different shapes
    - → particles from longest-lived and most energetic dark pions not clustered into reconstructed jets and thus not counted towards jet energies
  - RPVLL reconstruction algorithms highly inefficient for very short and very long lifetimes

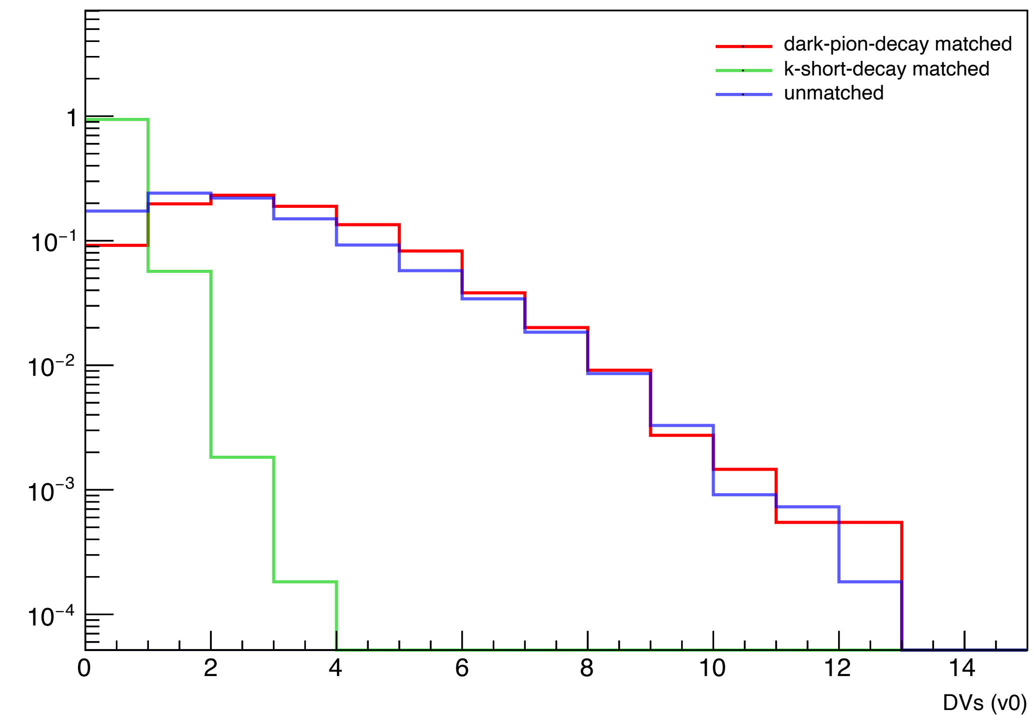
EJs Signal Grid: N Events

- Tens to hundreds of thousand events generated per point, depending on estimated expected signal efficiencies
  - Require large statistics to understand signal behavior and trends in distributions
  - Aiming for ~500 signal events per point in final signal region
- Based on estimated signal efficiencies averaged over ABCD test planes and vertex cleaning loss factor → N = ~**500** / (average signal efficiency * estimated good vertex efficiency factor) for lowest efficiency point
  - → good vertex efficiency factor estimated between 0.1 and 0.5
  - → about half of DVs matched to dark pions for intermediate lifetimes → can expect less for more extreme lifetimes and lower dark pion masses, as vertex reconstruction efficiency decreases

ModA_5_1000_150 (search-1 region)



ModB_2_1000_5 (search-1 region)

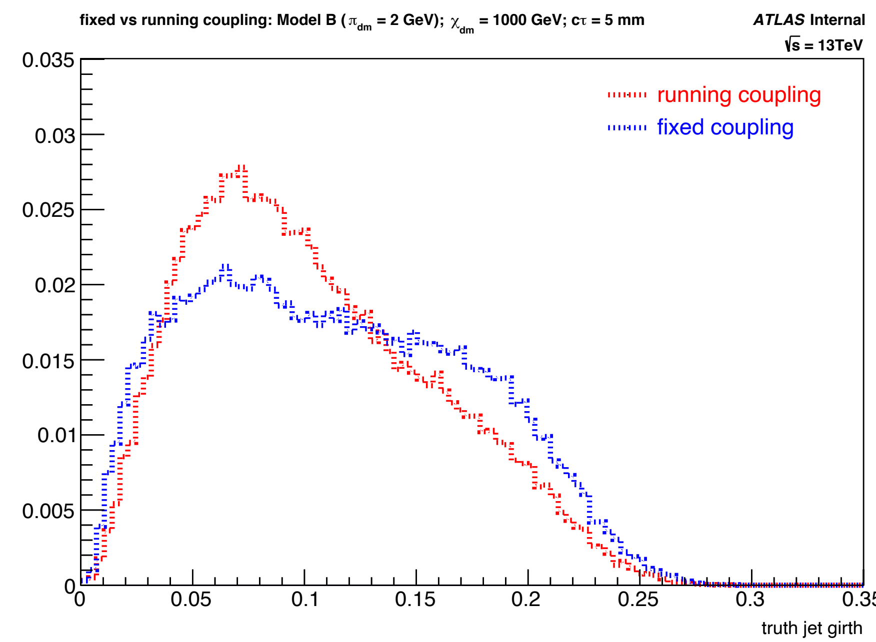


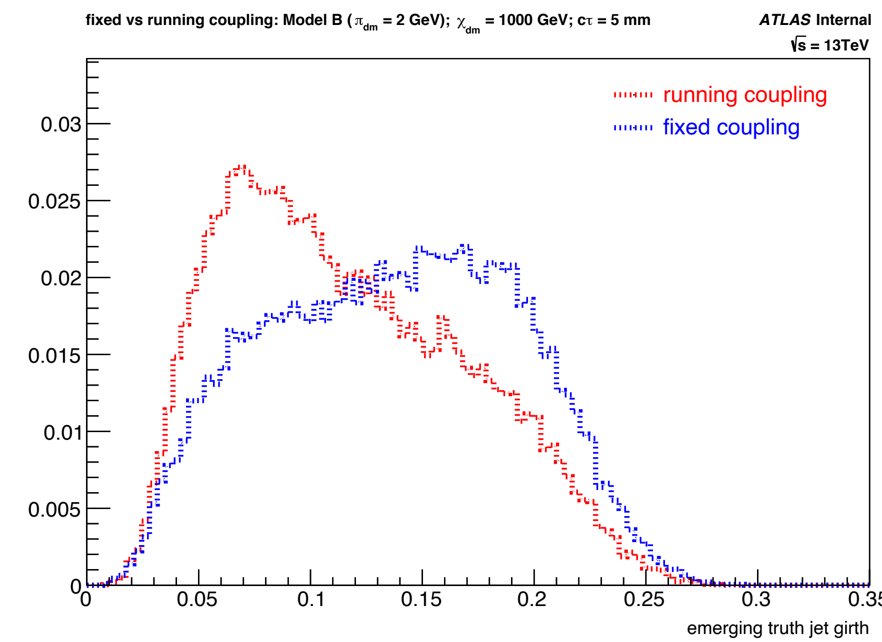
EJs Signal Sub-Grid

Signal Point				Number of Events			Metadata	
DSID	Model	m_{χ_d}	$c\tau_{\pi_d}$	MC16a	MC16d	MC16e	Cross-Section [nb]	Gen. Filter Efficiency
312001	A	1400	300	99,000	118,000	160,000	1.0785×10^{-6}	0.93601
312004	A	1400	20	60,000	80,000	100,000	1.0786×10^{-6}	0.93528
312006	A	1400	2	50,000	60,000	80,000	1.0791×10^{-6}	0.93595
312008	A	1000	150	80,000	100,000	130,000	1.5195×10^{-5}	0.86813
312009	A	1000	75	60,000	70,000	100,000	1.5191×10^{-5}	0.86835
312011	A	1000	2	60,000	70,000	100,000	1.5180×10^{-5}	0.86679
312015	A	600	20	99,000	120,000	160,000	4.2962×10^{-4}	0.60561
312016	A	600	2	100,000	120,000	159,000	4.3019×10^{-4}	0.60450
312017	A	600	1	220,000	280,000	360,000	4.3009×10^{-4}	0.60564
312020	B	1400	150	100,000	120,000	160,000	1.0777×10^{-6}	0.94283
312022	B	1400	20	60,000	80,000	99,000	1.0783×10^{-6}	0.94249
312023	B	1400	5	50,000	60,000	79,000	1.0783×10^{-6}	0.94287
312025	B	1000	300	120,000	146,000	199,000	1.5190×10^{-5}	0.88997
312028	B	1000	5	70,000	90,000	119,000	1.5216×10^{-5}	0.89052
312029	B	1000	2	60,000	70,000	100,000	1.5208×10^{-5}	0.88949
312031	B	600	300	150,000	190,000	242,000	4.3017×10^{-4}	0.67869
312035	B	600	1	200,000	250,000	300,000	4.3013×10^{-4}	0.67772
312036	B	600	0.5	200,000	250,000	300,000	4.3042×10^{-4}	0.67886
312038	C	1400	150	100,000	120,000	159,000	1.0784×10^{-6}	0.94193
312039	C	1400	75	78,000	100,000	129,000	1.0779×10^{-6}	0.94363
312043	C	1000	300	119,000	150,000	196,000	1.5197×10^{-5}	0.86652
312046	C	1000	5	99,000	130,000	170,000	1.5183×10^{-5}	0.86615
312050	C	600	150	200,000	250,000	300,000	4.3074×10^{-4}	0.57394
312052	C	600	2	320,000	400,000	530,000	4.2297×10^{-4}	0.57309
312055	D	1400	300	98,000	120,000	160,000	1.0791×10^{-6}	0.96925
312060	D	1400	2	90,000	110,000	148,000	1.0775×10^{-6}	0.96897
312063	D	1000	75	60,000	70,000	100,000	1.5168×10^{-5}	0.90922
312066	D	1000	1	130,000	160,000	220,000	1.5188×10^{-5}	0.90867
312067	D	600	300	410,000	500,000	680,000	4.3001×10^{-4}	0.63178
312072	D	600	0.5	200,000	249,000	299,000	4.3014×10^{-4}	0.63165
312075	E	1400	75	50,000	60,000	80,000	1.0773×10^{-6}	0.94790
312077	E	1400	5	50,000	60,000	80,000	1.0780×10^{-6}	0.94723
312080	E	1000	150	60,000	80,000	100,000	1.5184×10^{-5}	0.90574
312084	E	1000	1	120,000	150,000	200,000	1.5197×10^{-5}	0.90591
312087	E	600	20	100,000	119,000	160,000	4.3053×10^{-4}	0.72697
312090	E	600	0.5	110,000	140,000	180,000	4.3026×10^{-4}	0.72834

HV Running Coupling

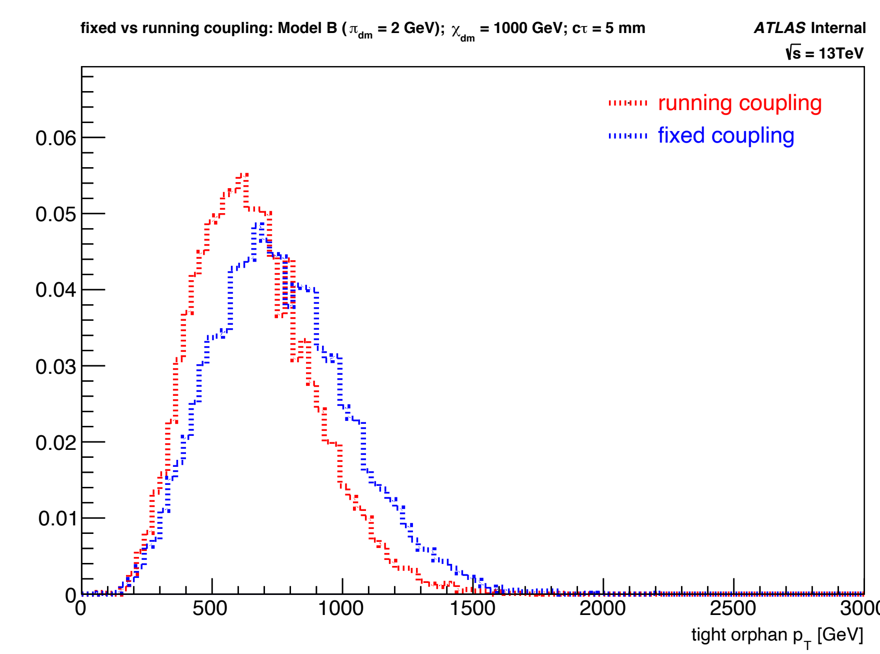
$$\text{girth} = \frac{1}{p_{\mathrm{T}}^{\mathrm{jet}}} \sum_i p_{\mathrm{T}}^i \, \Delta R_i$$

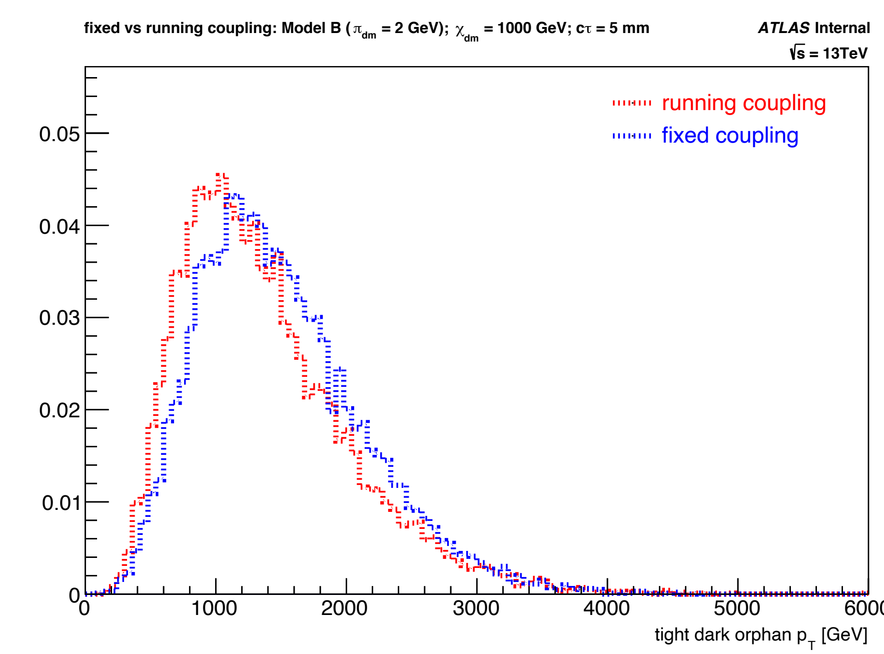




- HV running coupling required to produce QCD-like behavior in the hidden sector
- → fixed coupling produces more spherical, rather than jet-like, events

$$\text{orphan } p_{\mathrm{T}} = \sum_i p_{\mathrm{T}}^i - \sum_i \left(p_{\mathrm{T}}^i \,|\, \text{constituent of hard jet}\right)$$

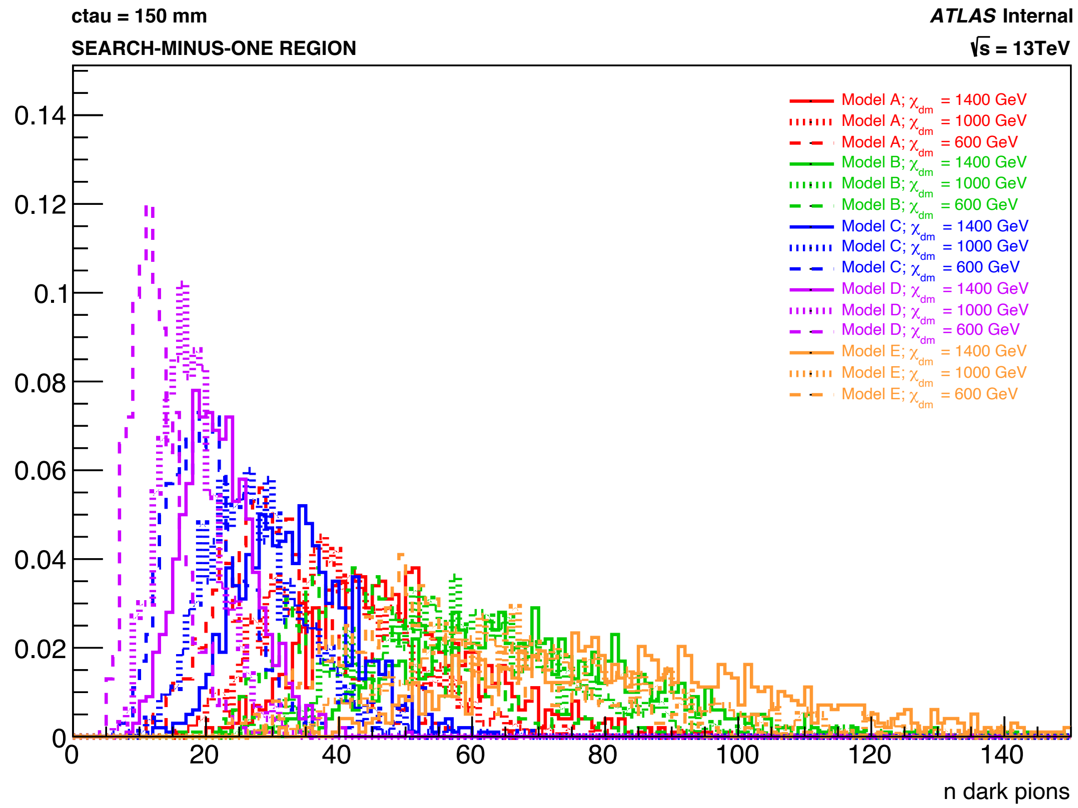




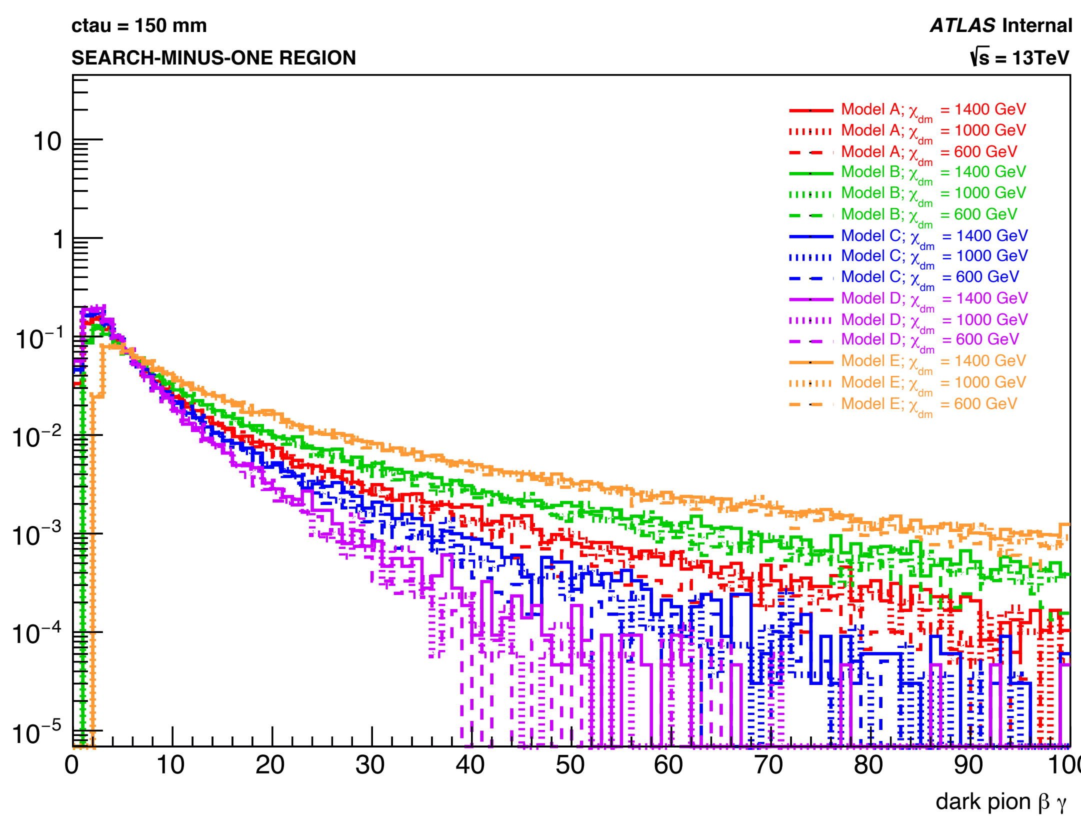
Signal Event Generation

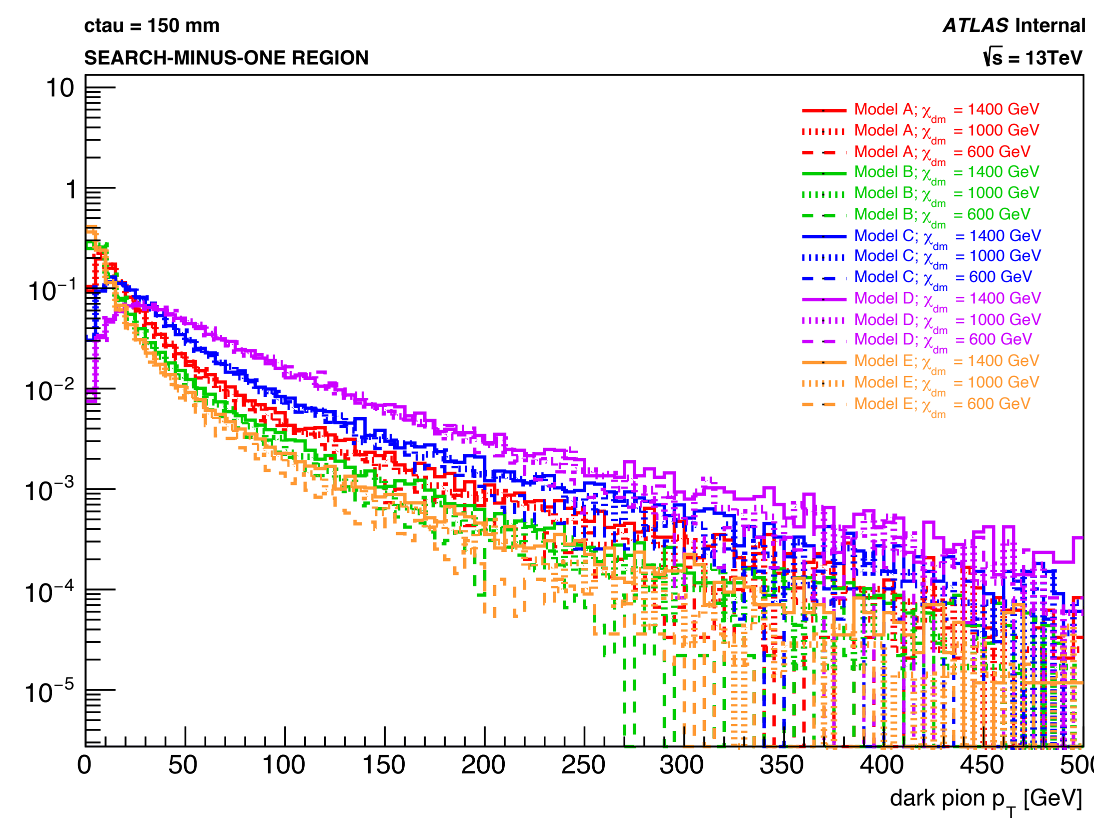
- Events generated with **Pythia 8.230 Hidden Valley** scenario with running dark gauge coupling and truth-level four-jet generator filter
  - Hidden Valley (HV) = light confining hidden sector ("valley"), containing invisible light particles and communicator between hidden and visible sectors
  - Event processes: X_d pair production through initial quark or gluon states → each mediator subsequently decays to down quark + dark quark
  - Dark sector gauge coupling allowed to run to first order, with β-function based on numbers of dark colors and dark flavors → produces QCD-like dark sector
    - Running coupling more accurately reproduces QCD-like parton showering in dark sector, yielding narrower, more conical dark jet objects, while fixed coupling produces more spherical events, with energy distributed more widely throughout detector
    - Running coupling option newly added to recent Pythia releases, motivated partially by EJs theorists and validated extensively by EJs analysis team
  - Simplified fragmentation, producing only dark pion and dark rho mesons (no dark baryons) → dark rhos promptly decay to dark pions (BR = 0.999) or down quarks (BR = 0.001), and dark pions decay to down quarks (BR = 1.0) with displacements set by signal point lifetime
  - Dark quark masses set by dark confinement scale
  - Generator filter applied to increase statistics for signal points with otherwise low efficiency:
    four truth jets with $p_T > 100$ GeV and $|\eta| < 2.7$ → looser versions of offline reco-level signal jet cuts
- After event generation, MC signal samples run through full detector simulation and reconstruction → LLP searches require dedicated RPVLL reconstruction with specialized algorithms…

Dark Pions

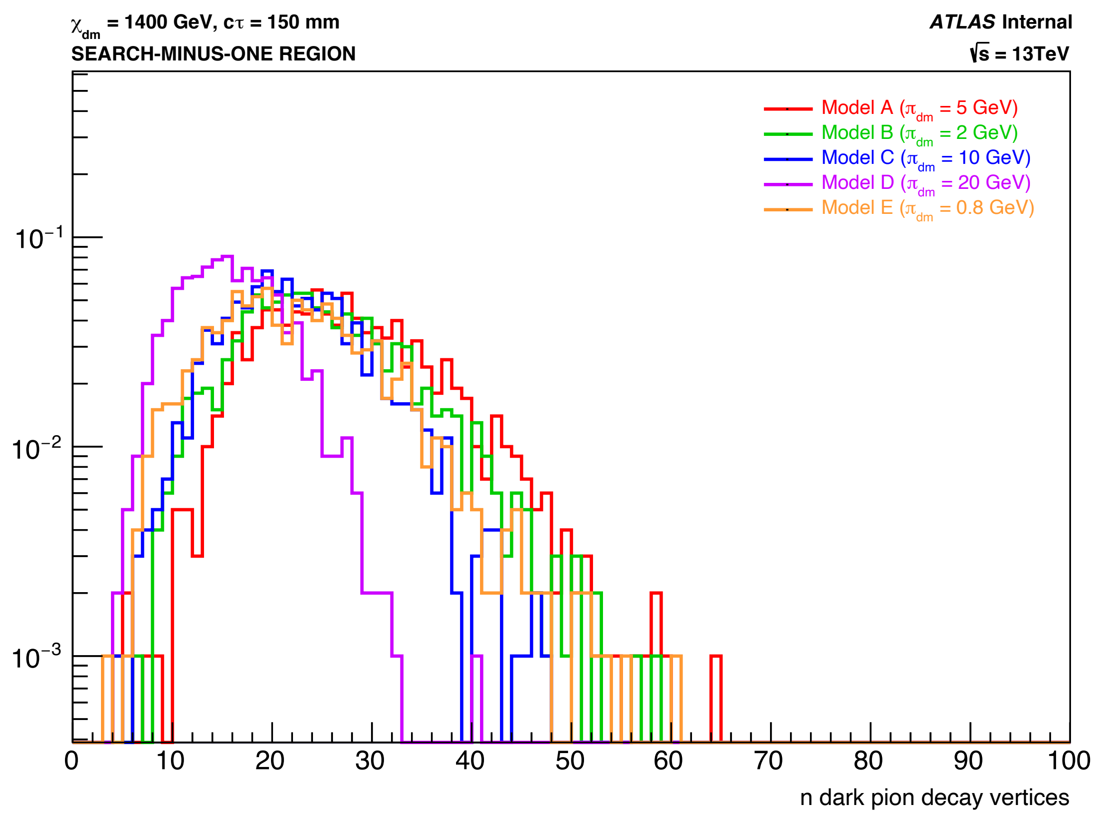


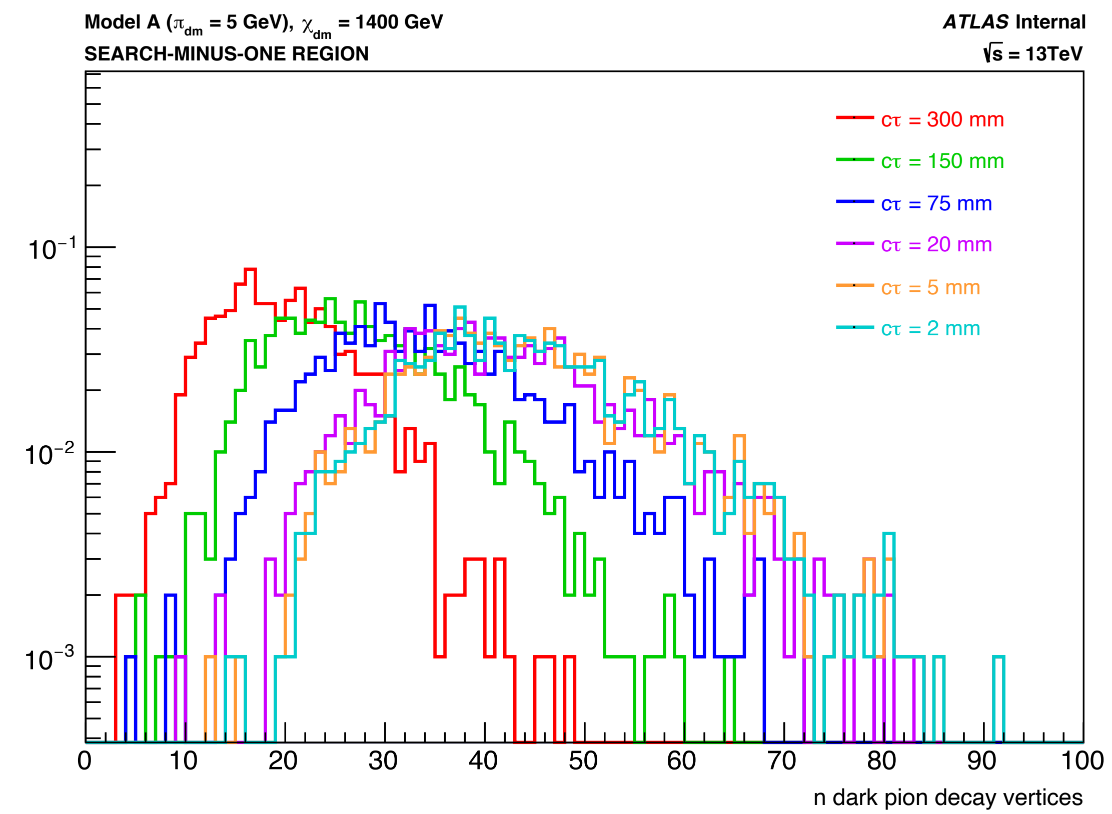
- Dark pion multiplicity inversely proportional to model mass scale and directly proportional to mediator mass
- Lighter dark pions are more boosted, and heavier dark pions have larger momenta





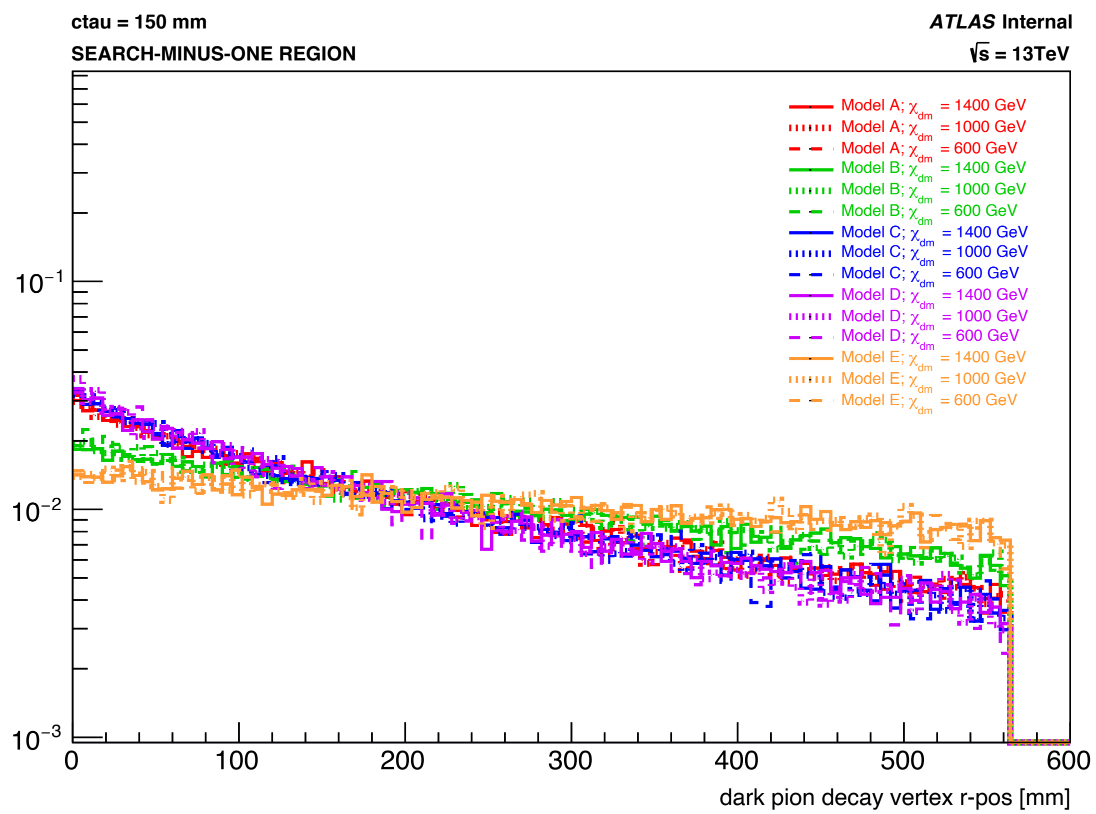
Dark Pion Decay Multiplicity

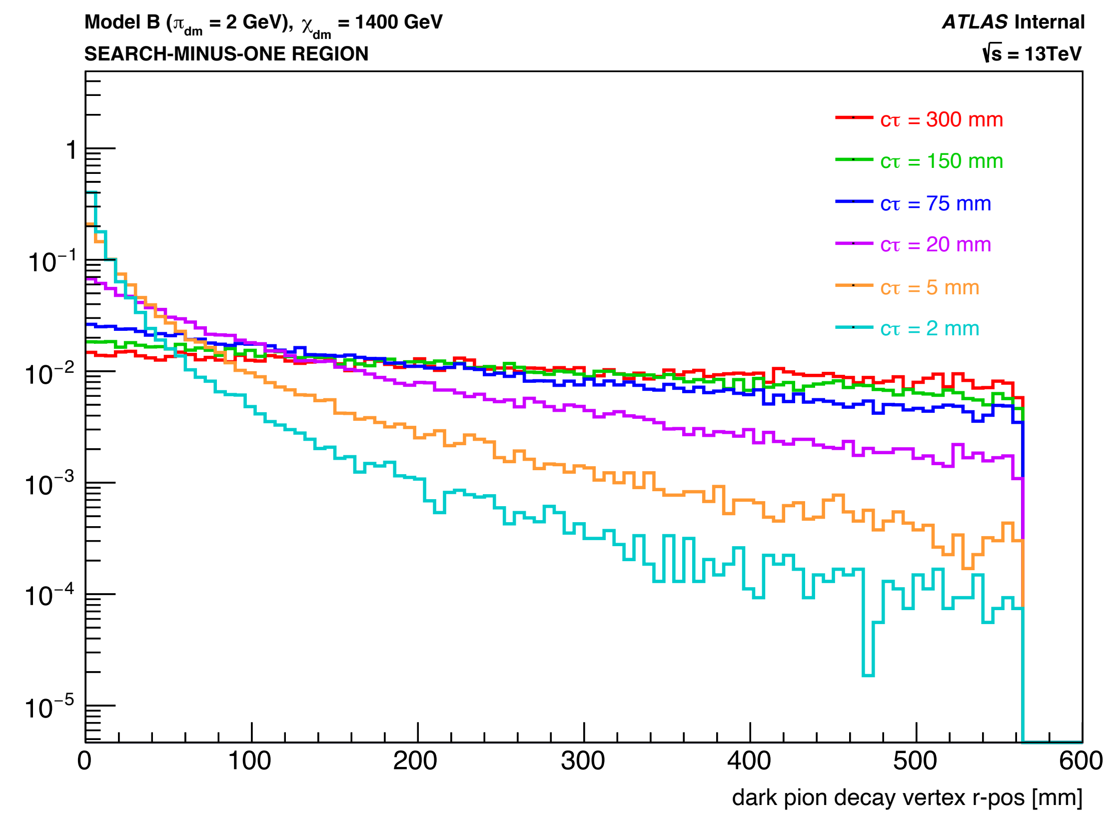




- Observed dependency of dark pion decay vertex multiplicity on dark pion lifetime due to fiducial volume cuts requiring decay within SCT: r < 563 mm and |z| < 2720
  - → dark pions with longer lifetimes more likely to decay outside silicon detector and thus less likely to be reconstructed as DVs

Dark Pion Decay Position





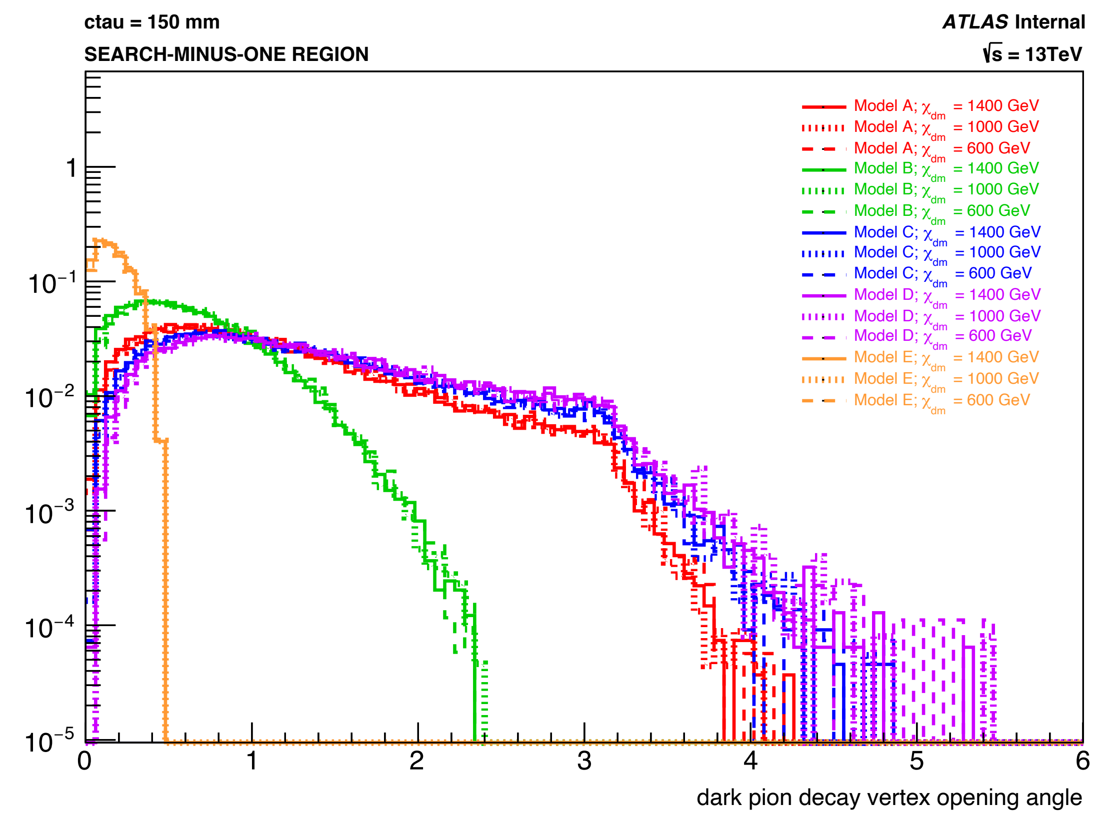
- Position of dark pion decay vertex follows exponentially falling distribution with mean decay length $\langle \Delta L \rangle = \beta\gamma c\tau$ →
- → dark pion decays slightly further displaced in lighter models, as lighter dark pions more boosted, so travel further before decaying
- → displacement of decay directly dependent on dark pion lifetime

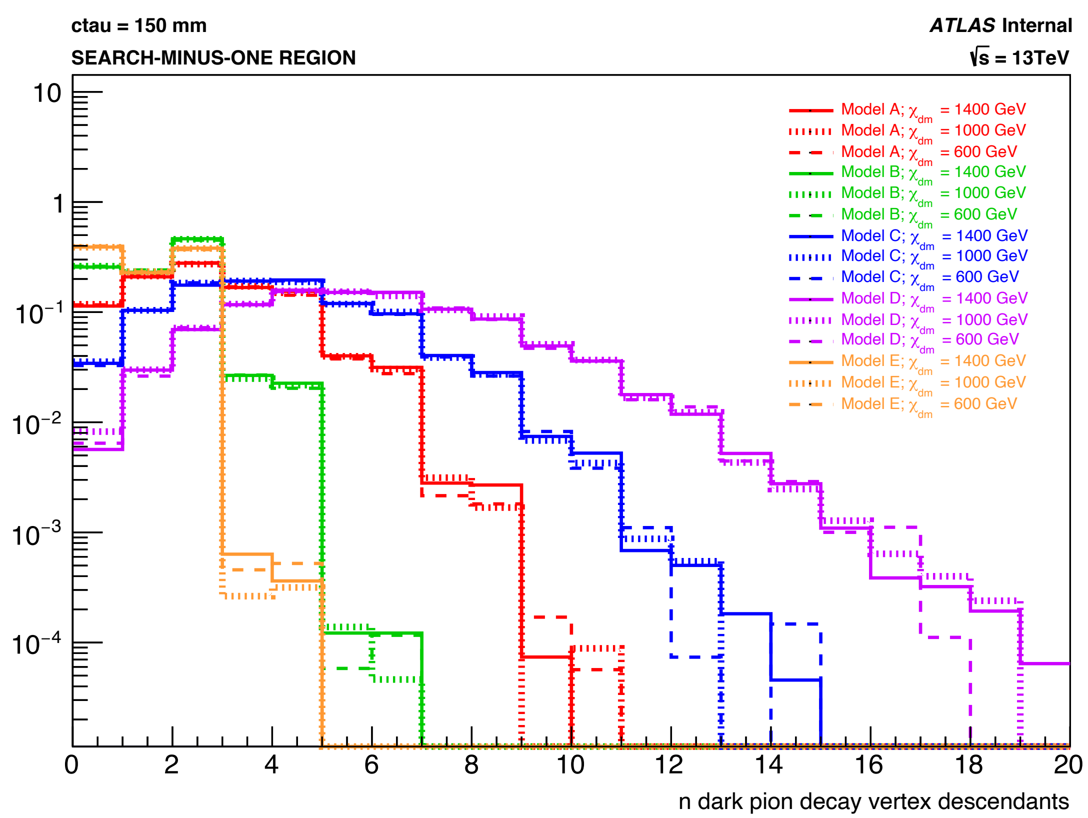
Dark Pion Decays

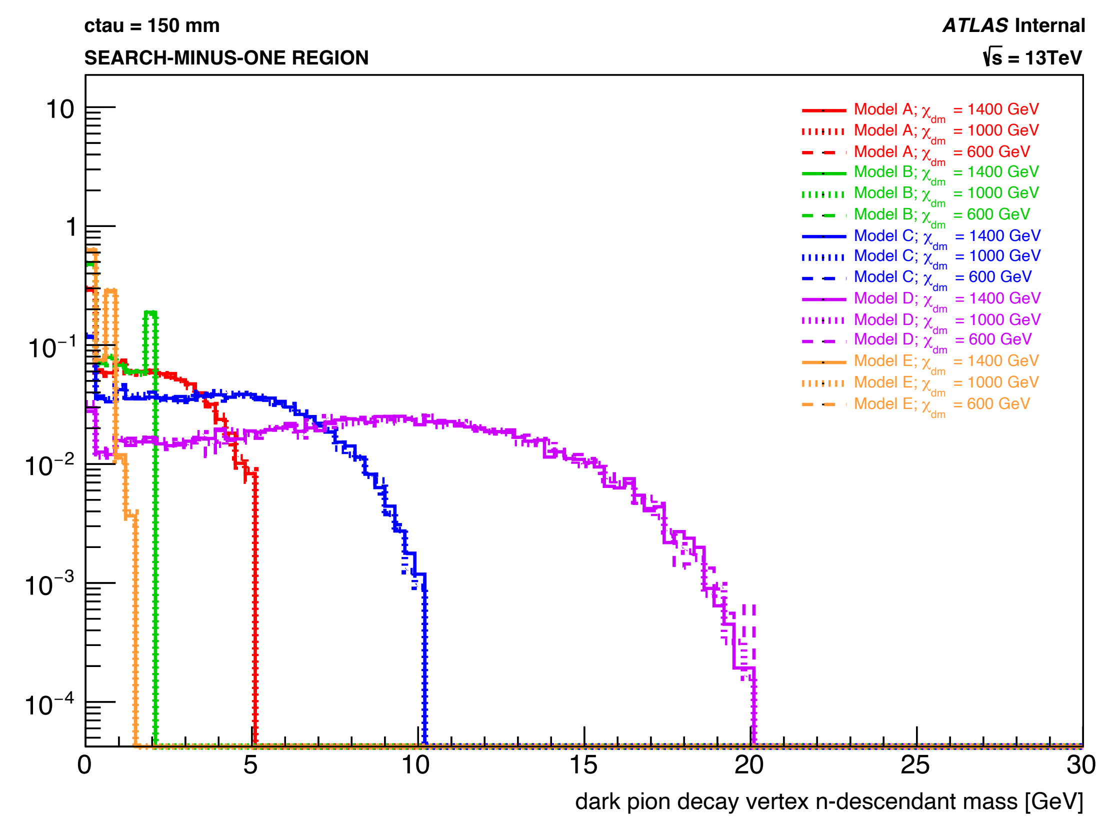
- Opening angle = maximum angle distance between outgoing particles → related to mass of model, with heavier models producing wider dark pion decays and thus wider jets

- Descendant = charged particle with $p_T > 1$ GeV in decay chain → descendant multiplicity + mass directly dependent on model mass

  - Descendant mass = invariant mass of decay vertex, calculated from four-vector sum of descendants → distribution centered around dark pion mass

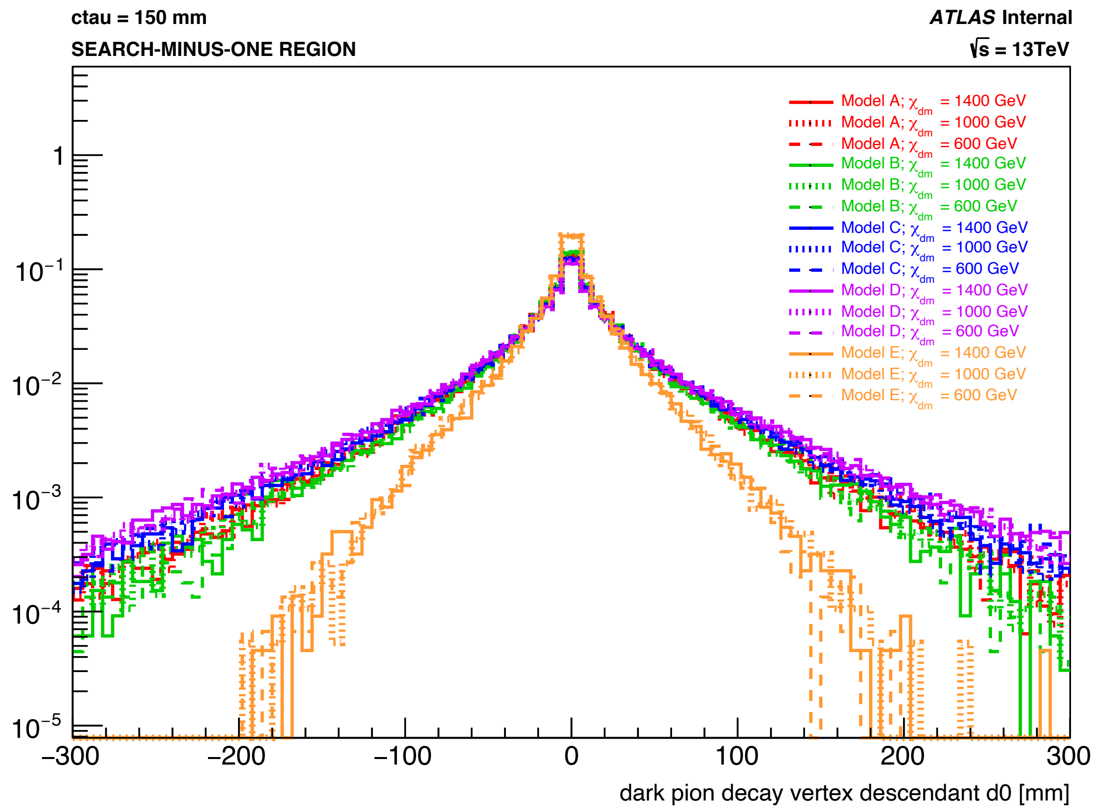


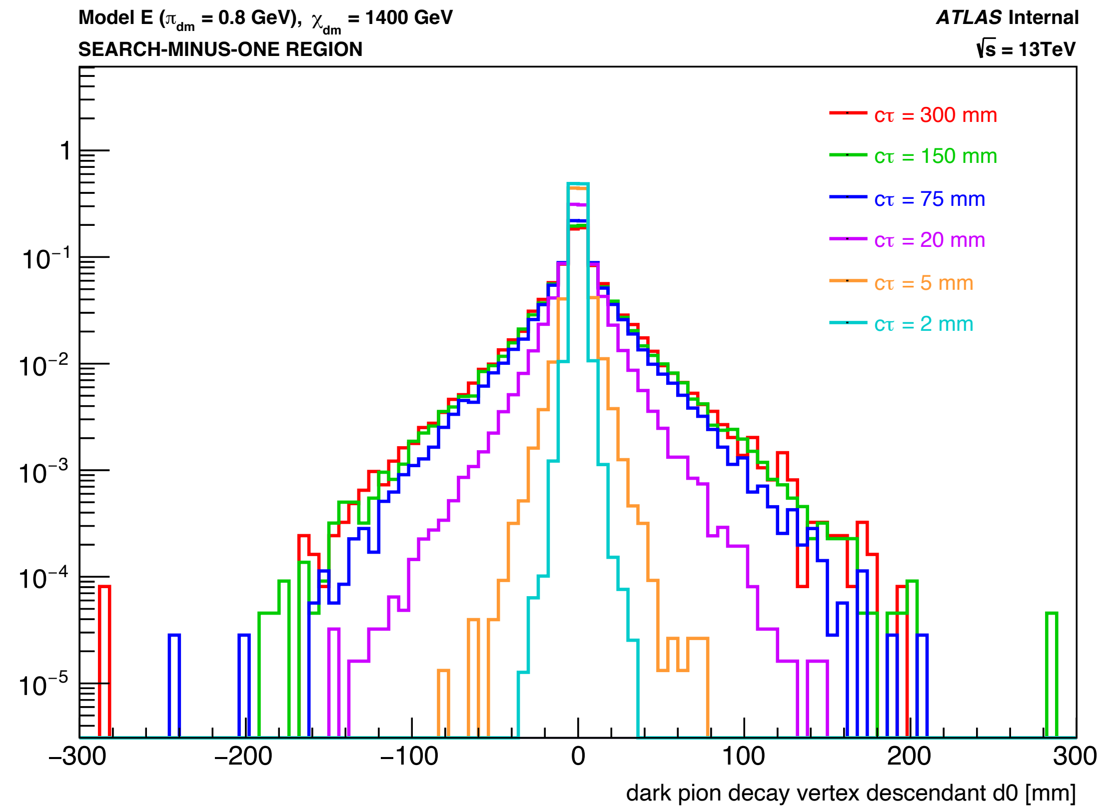




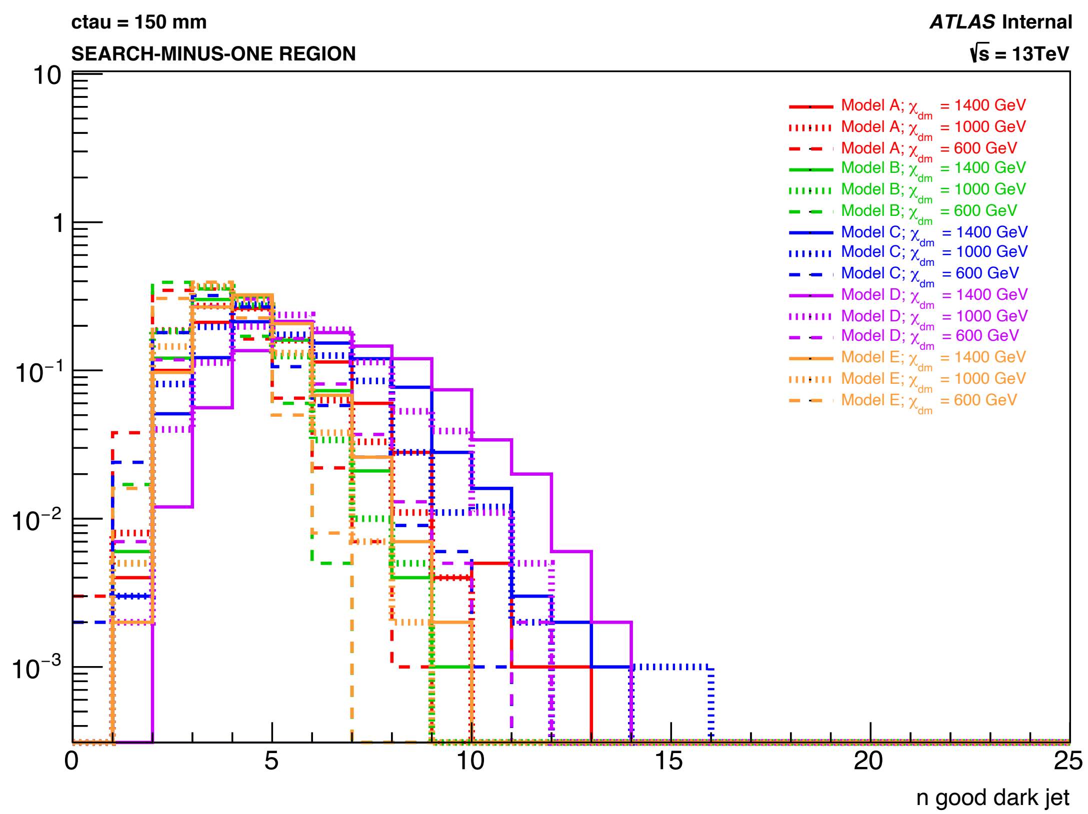
Dark Pion Descendants

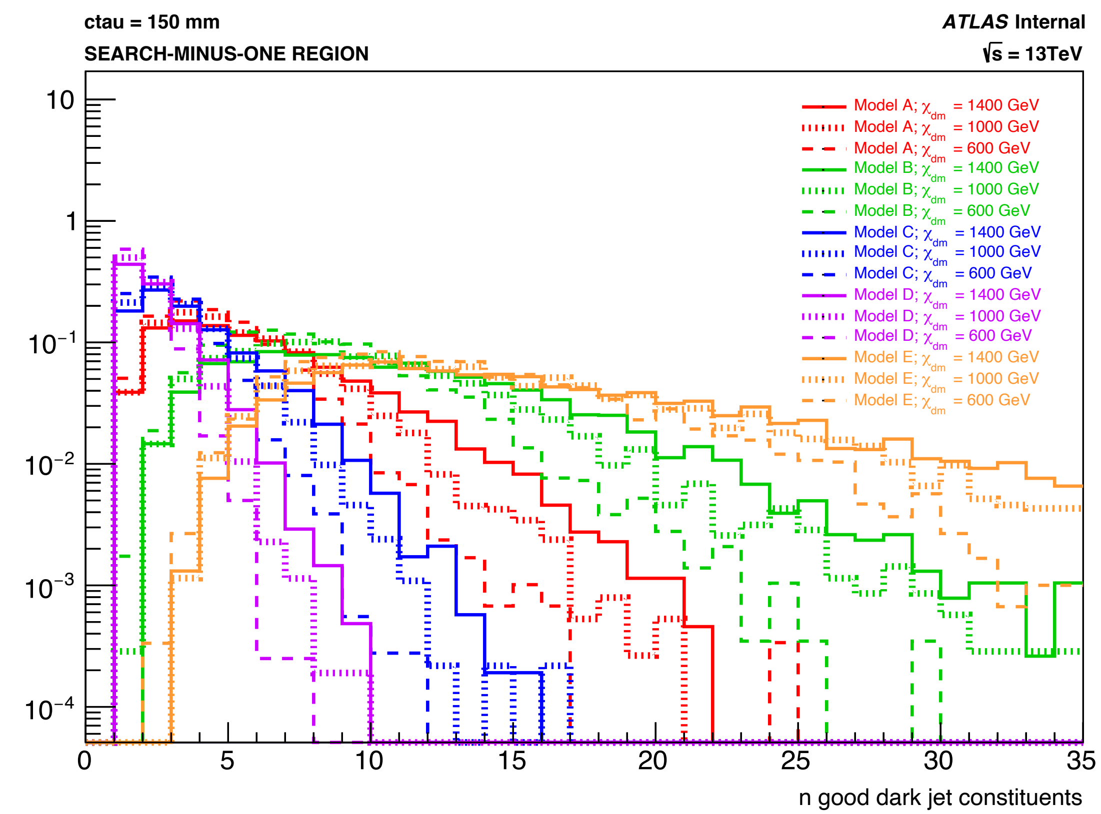
- Descendants represent displaced truth tracks capable of being reconstructed in inner detector
- Transverse impact parameter of truth tracks approximated from decay vertex position and angle from descendant: $d_0 \approx r \times \sin\Delta\phi$
  - → directly dependent on mass scale, as consequence of decay vertex opening angle → lighter models produce descendants more parallel to direction of vertex pointing vector
  - → directly dependent on dark pion lifetime





Truth Dark Jets





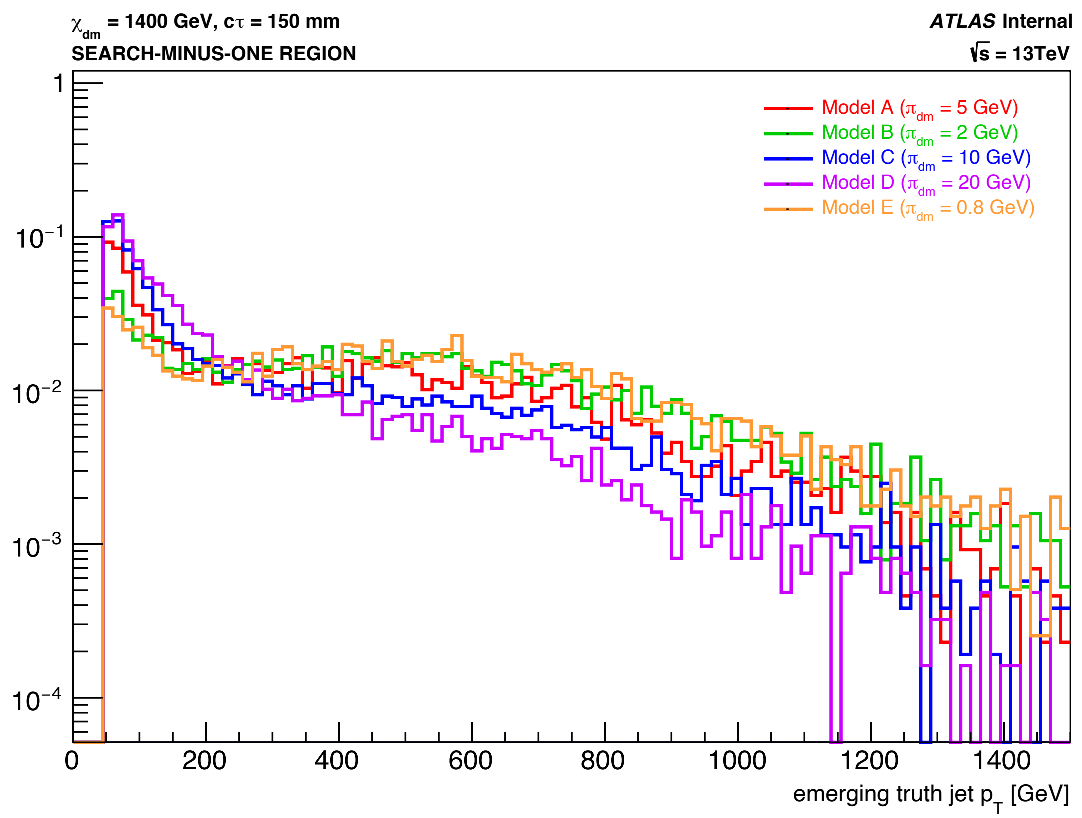
- Truth dark jet = truth jet clustered only from dark particles, representative of initial dark quark from hard process → “good” dark jet requires pT > 30 GeV (excludes poorly clustered dark jets consisting of single constituents)

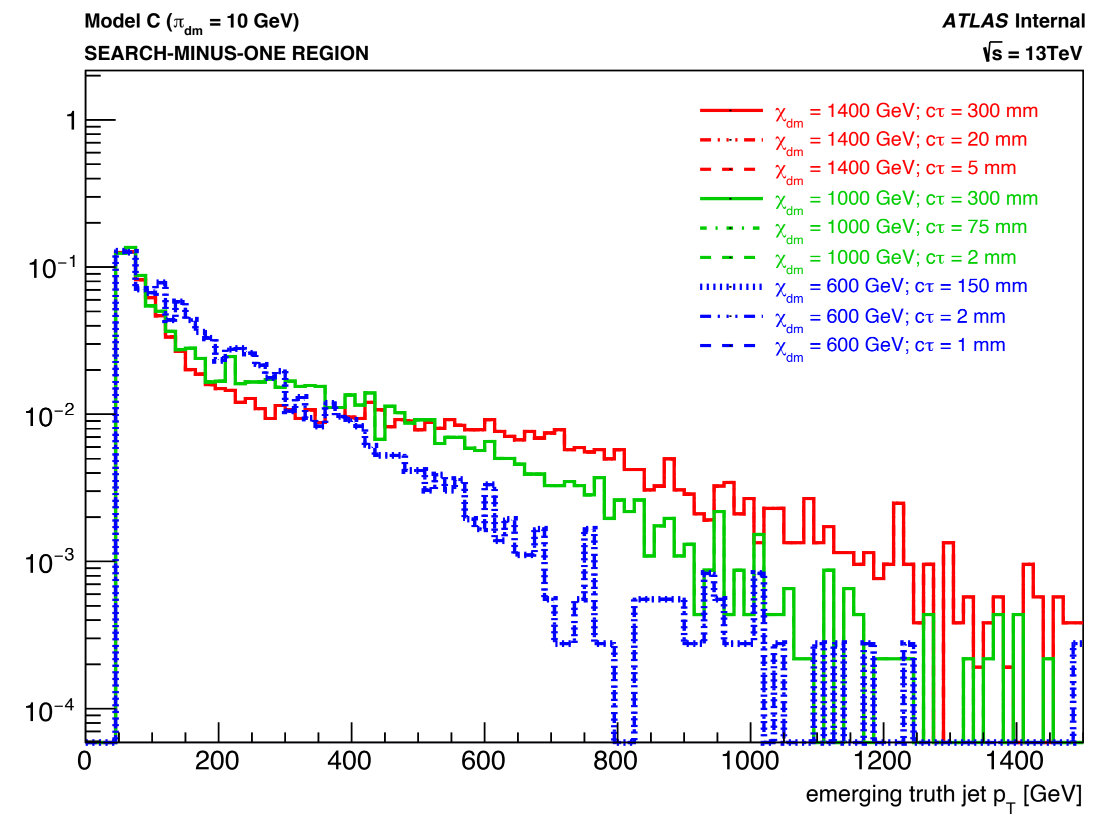
- Truth jet multiplicity increases with increasing model and mediator mass → potentially consequence of poor dark jet clustering, with heavier models producing dark sectors with more widely spread out dark hadrons, then accidentally reconstructed as separate dark jets

- Dark jet constituent multiplicity increases with mediator mass but decreases with model mass, as dark pion multiplicity is directly related to mediator mass but inversely related to scale of dark QCD

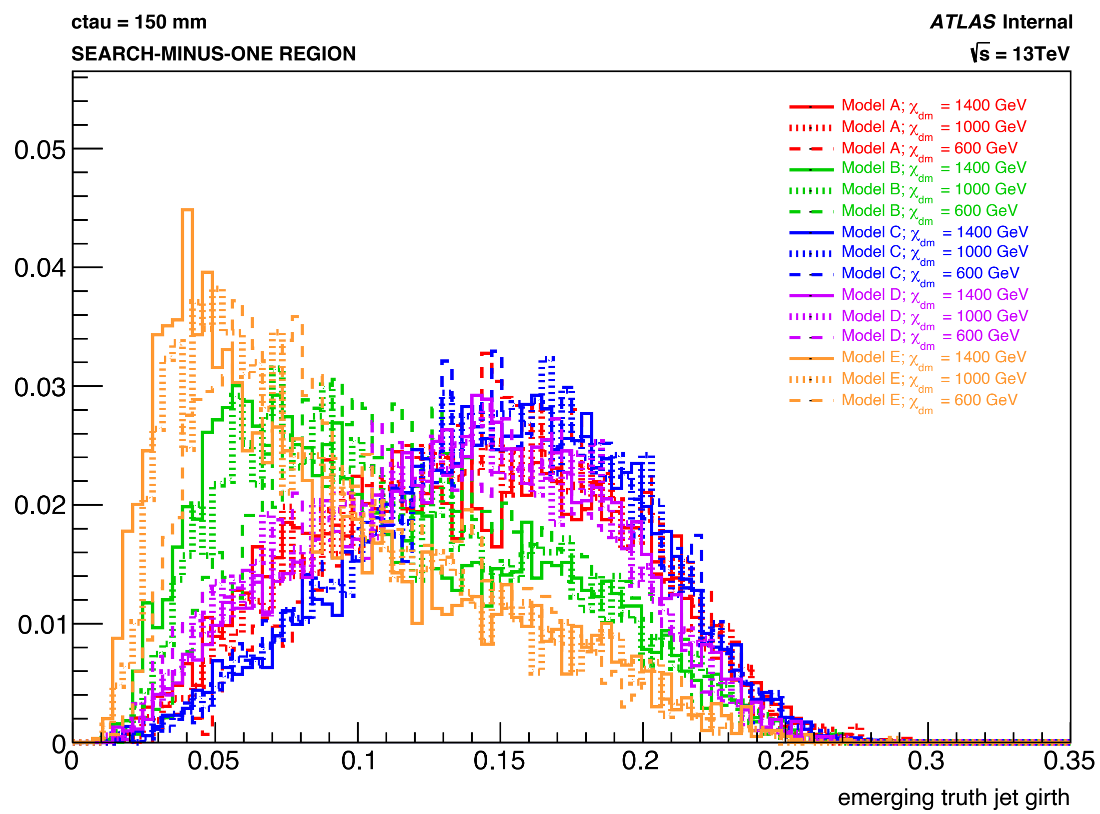
Truth Emerging Jets

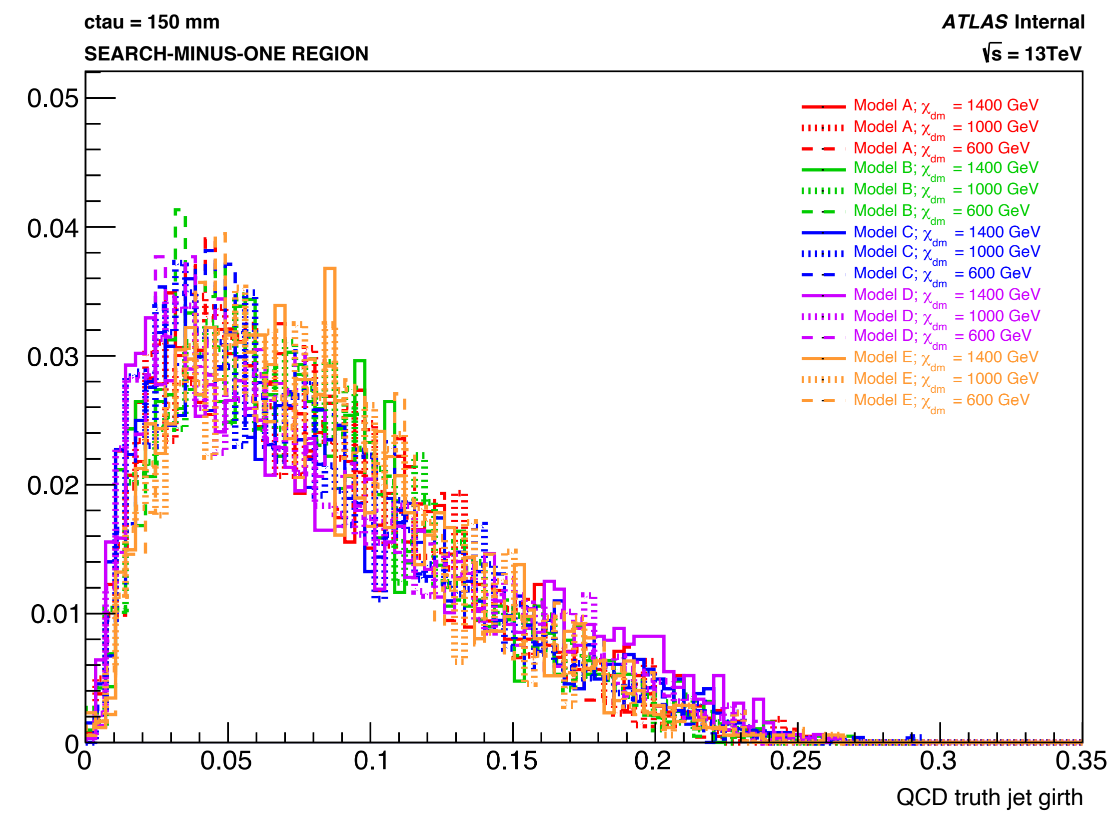
- Truth emerging jet = truth jet matched to good dark jet, where match if within $\Delta R < 0.3$ of dark jet
- Transverse momentum inversely related to model mass scale and directly related to mediator mass → heavier mediators produce higher-p_T objects, and lower-mass models produce more dark pions, which contribute to emerging jet





Truth Emerging vs QCD Jets: Girth





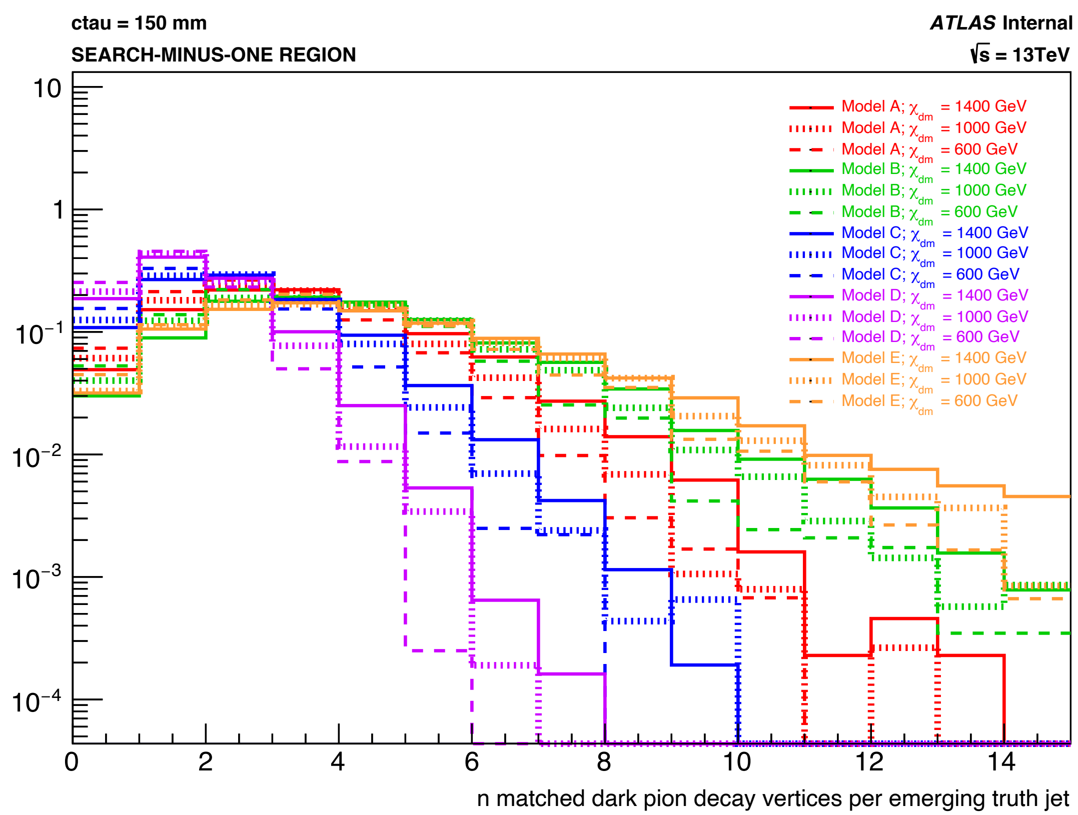
- Truth QCD jet = visible truth jet not matched to truth dark jet
- Truth QCD jet girth independent of signal point parameters
- Truth emerging jet girth directly dependent on model mass scale, with heavier models producing wider emerging jets, as consequence of dark pion decay vertex opening angle → also inversely related to mediator mass, with heavier mediators yielding slightly narrower jets

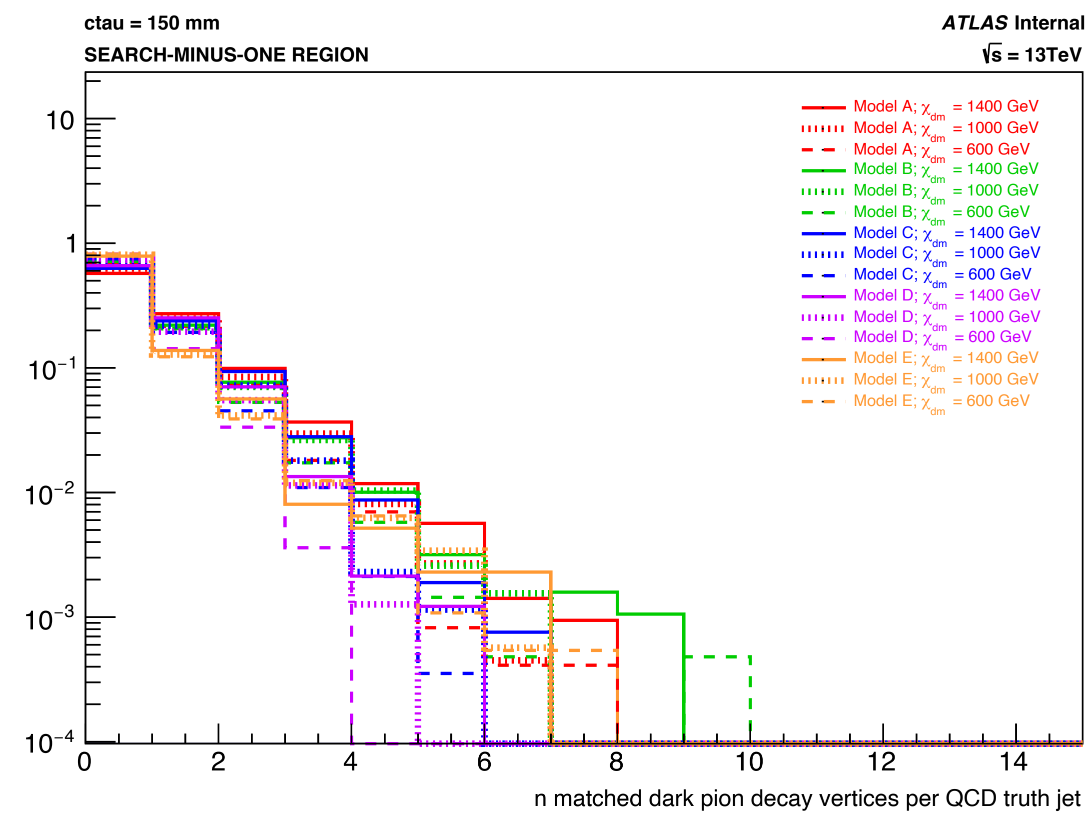
Truth Emerging vs QCD Jets: Associated Vertices

- Dark pion decay vertex associated to truth jet if within ΔR < 0.6 → more dark pion decays found around truth emerging jets than QCD

- Lower model mass scales produce larger numbers of jet-associated vertices and smaller numbers of jet-associated descendants, consistent with overall model dependency of event-level vertex and descendant multiplicity

- Jet-associated dark pion decay vertex and descendant multiplicity inversely related to dark pion lifetime, as longer lifetime dark pions more likely to decay outside ID or calorimeter and thus not be associated to jets



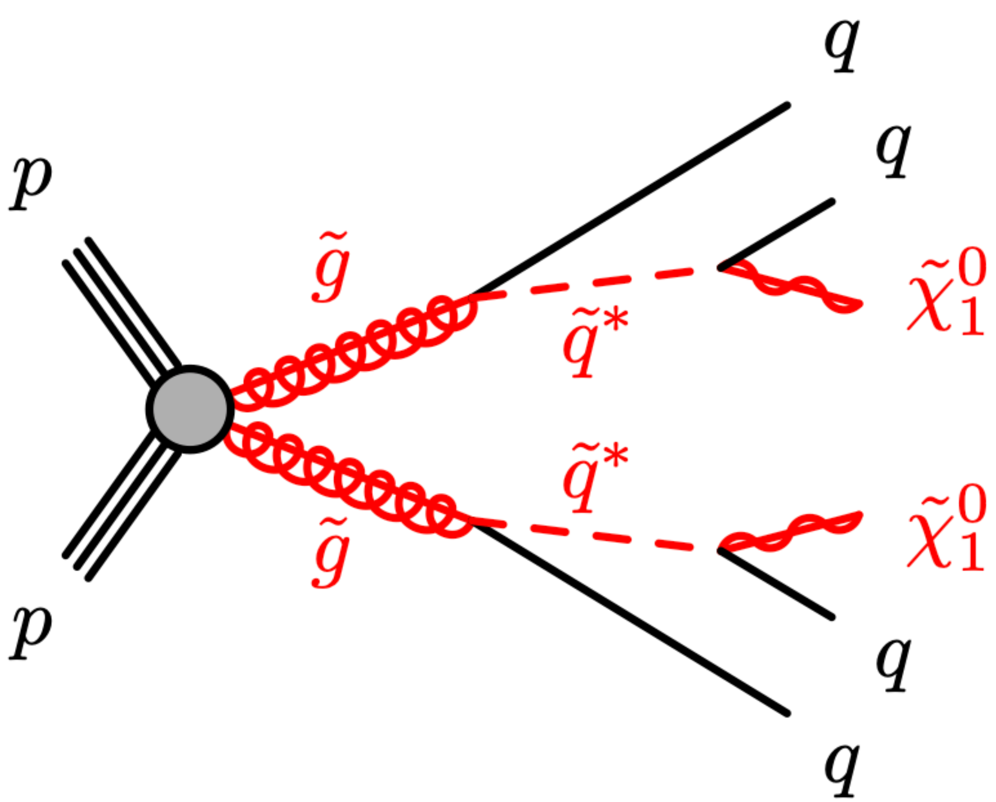


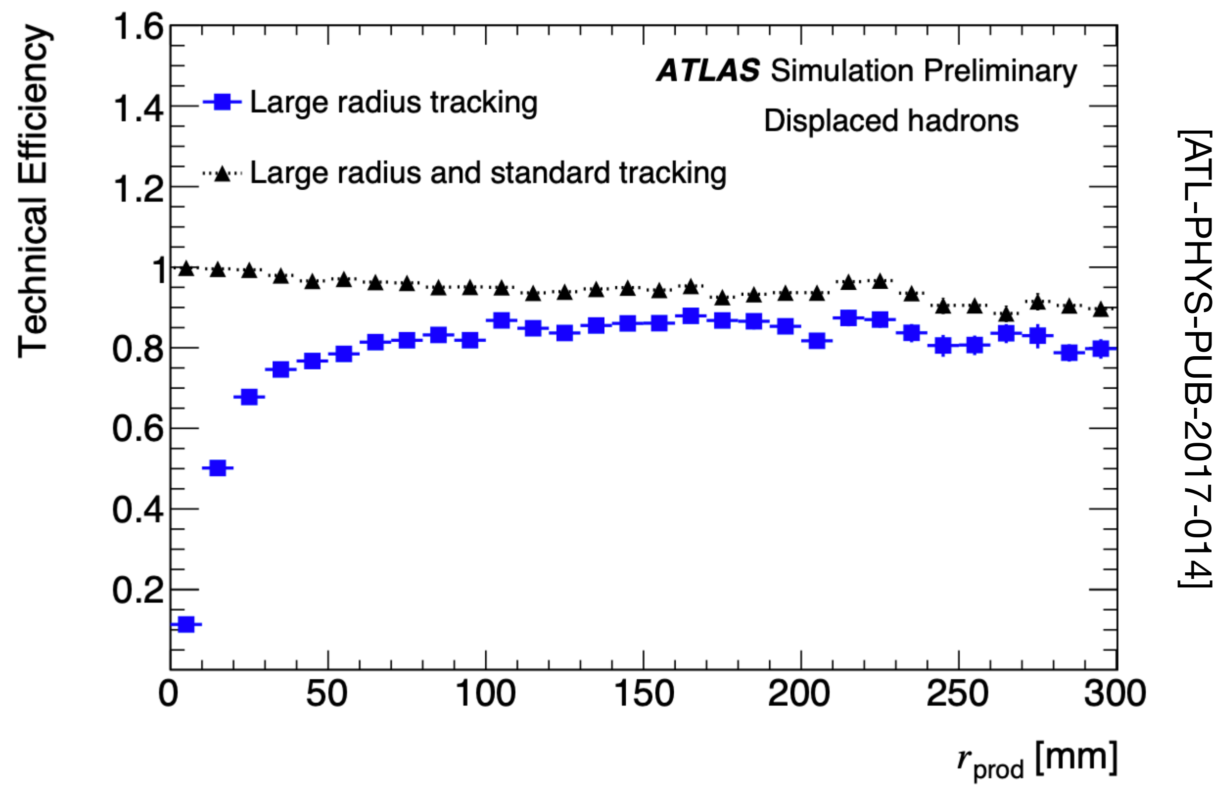
LRT Performance

- LRT algorithm performance studied with simplified split SUSY model yielding displaced hadrons: pair-produced gluinos form R-hadrons going to quarks and neutralinos → small mass difference between gluino and neutralino results in soft p_T spectrum for hadronic decay of ~few GeV

- Performance studies use hit-based truth matching scheme, where each track assigned match score with respect to most hit-associated track
  - → match score = weighted fraction of track hits originating from truth particle = probability track corresponds to real charged particle trajectory
  - → match score > 0.5 required for truth match

- Technical efficiency defined to isolate algorithmic efficiency of track reconstruction from effect of detector geometry → only considers truth particles that leave behind enough hits in silicon detector to be reconstructible via silicon-seeded tracking methods
  - Technical efficiency between 80% and 100% for hadrons with r_{prod} < 300 mm





Secondary Vertexing (VSI)

- Dedicated secondary vertexing algorithm designed for LLP searches → uses combined collection of standard to LRT tracks to reconstruct displaced decay vertices produced far from IP with sequence of steps:

   1. Seed track selection: DV reconstruction seeded by track pairs approximately compatible with LLP decay → seed tracks preselected based on quality criteria = “selected” tracks

   2. Two-track seed finding: set of all pairs of selected tracks loosely compatible with LLP decay used to form set of two-track seed vertices → compatibility based on initial position estimate, precision fit, and hit pattern

   3. Multi-track vertex forming and track rearrangement: N-track vertices formed by combining two-track seeds using incompatibility graph method → tracks rearranged to resolve ambiguities

   4. Vertex merging: split and nearby vertices merged using one of set of merging algorithms → “split” vertex = single LLP decay accidentally reconstructed as multiple distinct vertices

   5. Track attachment: tracks initially unselected for vertex seeding but compatible with reconstructed vertices attached to existing DVs = “attached” or “associated” tracks → can help to improve vertex kinematics, i.e. track multiplicity and invariant mass

VSI Seed Track Selection

- VSI seeded by tracks reconstructed from both standard tracking and LRT algorithms, as DVs can be expected to contain tracks with a range of impact parameters

All track criteria
$p_T > 1$ GeV
$\chi^2/N_{\mathrm{DoF}} < 50$
if $N_{\mathrm{Pixel}} = 0$, $N_{\mathrm{SCT}} \geq 6$
if $N_{\mathrm{Pixel}} < 2$, $N_{\mathrm{TRT}} \geq 1$
$2\,\mathrm{mm} < \lvert d_0 \rvert < 300\,\mathrm{mm}$
$\lvert z_0 \rvert < 1500\,\mathrm{mm}$
not already associated to any HS or PU PV
Low-p_T track criteria
if $p_T < 25$ GeV, $N_{\mathrm{SCT}} \geq 7$
if $p_T < 25$ GeV and $\lvert \eta \rvert < 1.7$, $N_{\mathrm{TRT}} \geq 20$

- Selection criteria designed to be loose enough to effectively reconstruct DVs across various physics models and tight enough to reduce inclusion of fake / SM tracks in DV seeds

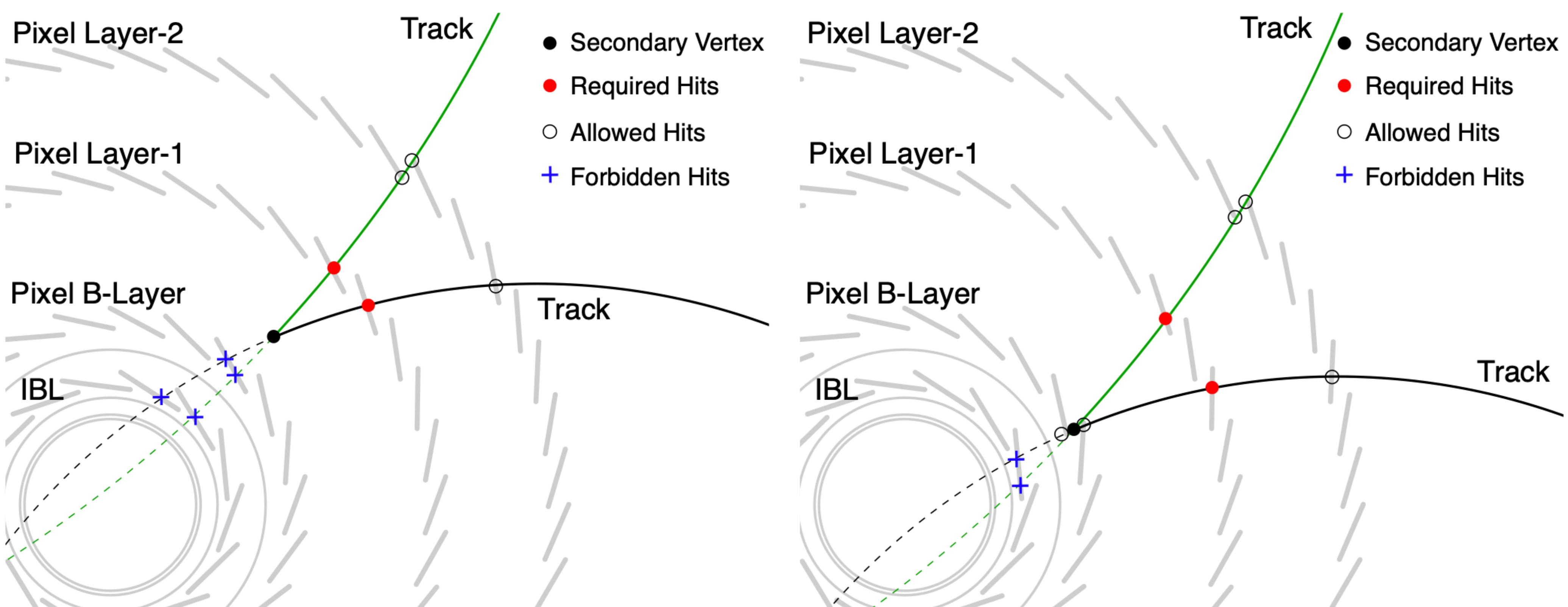
VSI Two-Track Seed Finding

initial estimate seed vertex criteria
$r_{\mathrm{vtx}}^{\mathrm{init}} < 563\,\mathrm{mm}$
$\lvert d_0^{\mathrm{wrt\ init\ SV}} \rvert < 100\,\mathrm{mm}$
$\lvert z_0^{\mathrm{wrt\ init\ SV}} \rvert < 50\,\mathrm{mm}$
precision fit seed vertex criteria
$r_{\mathrm{vtx}}^{\mathrm{fit}} < 563\,\mathrm{mm}$
$\chi^2/N_{\mathrm{DoF}} > 5$
LLP decay compatibility criteria
hits required outside and forbidden inside seed position
$\cos(\varphi^{(1)} - \varphi_{\vec{r}}) > -0.8$ and $\cos(\varphi^{(2)} - \varphi_{\vec{r}}) > -0.8$
$\cos(\varphi^{\mathrm{sum}} - \varphi_{v}) > -0.8$

- Initial vertex seed position estimated from fast algorithm using track parameters measured at perigee
- Precision fitting with full track extrapolation on remaining track pairs → minimize χ² with Kalman filter
- Hit pattern consistency check and additional kinematic cuts ensure vertex seeds and tracks compatible with LLP decay

[ATL-PHYS-PUB-2019-013]

- Hit pattern consistency check based on expectation that tracks travel outward (not inward) from vertex
- Kinematic cuts remove vertex seeds and tracks pointing backwards to PV

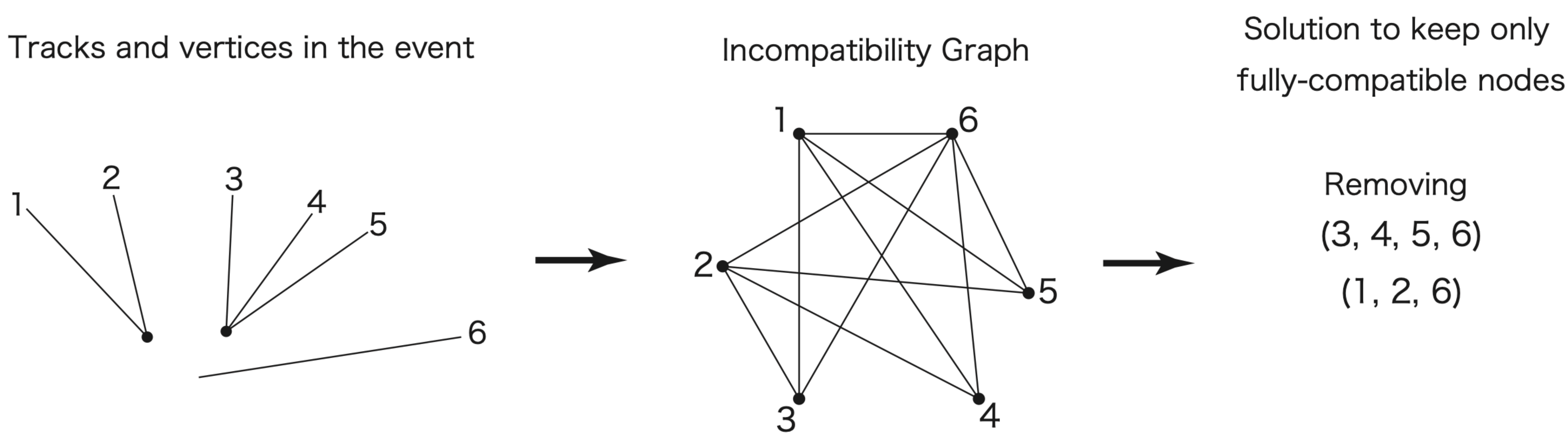


VSI Multi-Track Vertex Forming

- Two-track vertices sharing common tracks iteratively combined into all possible multi-track vertices based on incompatibility graph

- Incompatible track pairs represented by incompatibility graph → nodes represent tracks, and edges connect tracks incompatible with each other for vertex seeding

  - Graph solved and compatible tracks extracted by removing minimum number of fully incompatible nodes (and associated edges) such that only fully-isolated nodes remain → remaining tracks compatible and fitted as N-track vertex candidate



[ATL-PHYS-PUB-2019-013]

VSI Track Rearrangement

- Ambiguities = tracks associated to multiple vertex candidates → resolved by iteratively rearranging tracks to optimize goodness of fit to vertices until all tracks maximally associated to single DV

1. ID track with most number of associated vertices and calculate χ^2 of track with each associated vertex
2. ID associated vertex to which track has worst χ^2 → if larger than some threshold, remove track from vertex and repeat
3. Otherwise, calculate distance significance and number of shared tracks between all vertices sharing track of interest → if shared vertex position compatible within 10σ, or if shared vertices share at least two tracks, merge vertices

   → σ = distance uncertainty between vertices = quadrature sum of covariance matrices of vertex fits

4. Otherwise, remove track from worst-χ^2 vertex
5. Repeat procedure until all tracks associated to maximum of one vertex

VSI Vertex Merging

- Vertices recombined by sequentially passing through set of merging algorithms

  1. Merge by shuffling: smaller vertex seeded by position of larger and refit → if updated position compatible with position of larger vertex, vertices merged

  2. Magnetic merging: tracks from larger vertex iteratively associated to smaller, and smaller vertex refit, seeded by position of larger → if compatible with larger, vertices merged

  3. Wild merging: all tracks from both larger and smaller vertices combined into single vertex, refit with position of original larger vertex → if refit vertex compatible with larger vertex position, vertices merged

  4. Final / proximity merging: remaining vertices within 1 mm of each other forced to merge, regardless of compatibility

- → start by attempting to merge lowest track-multiplicity DV (most likely to be incorrectly reconstructed) with higher-multiplicity DV (from highest to lowest) and increase multiplicity of DV of interest at each iteration

- → compatibility = positions within 10σ

- → algorithms tested in order, and testing stops once any single algorithm succeeds

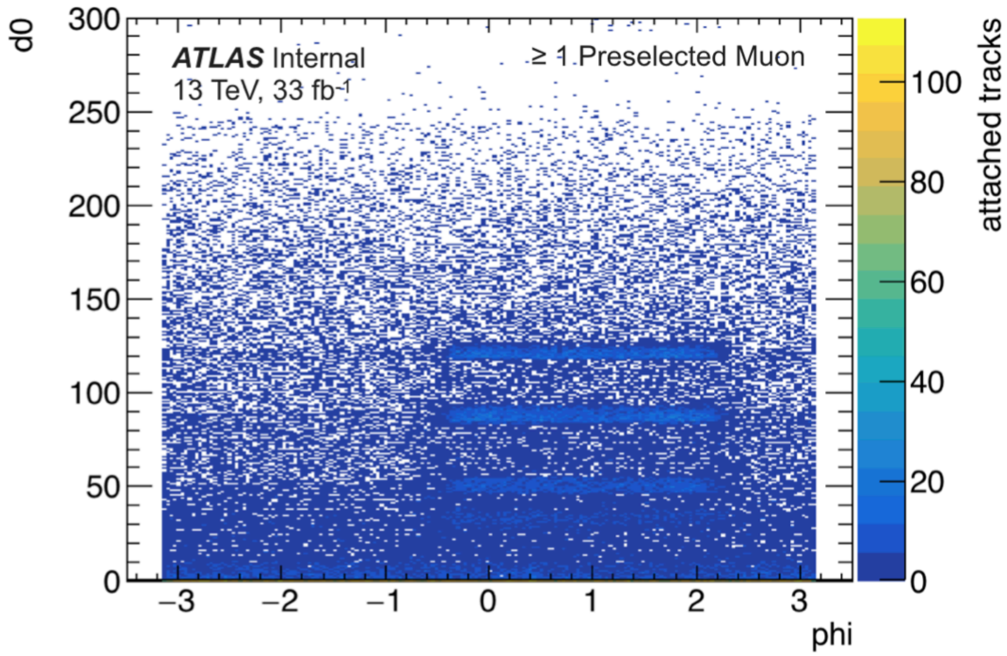
VSI Track Attachment

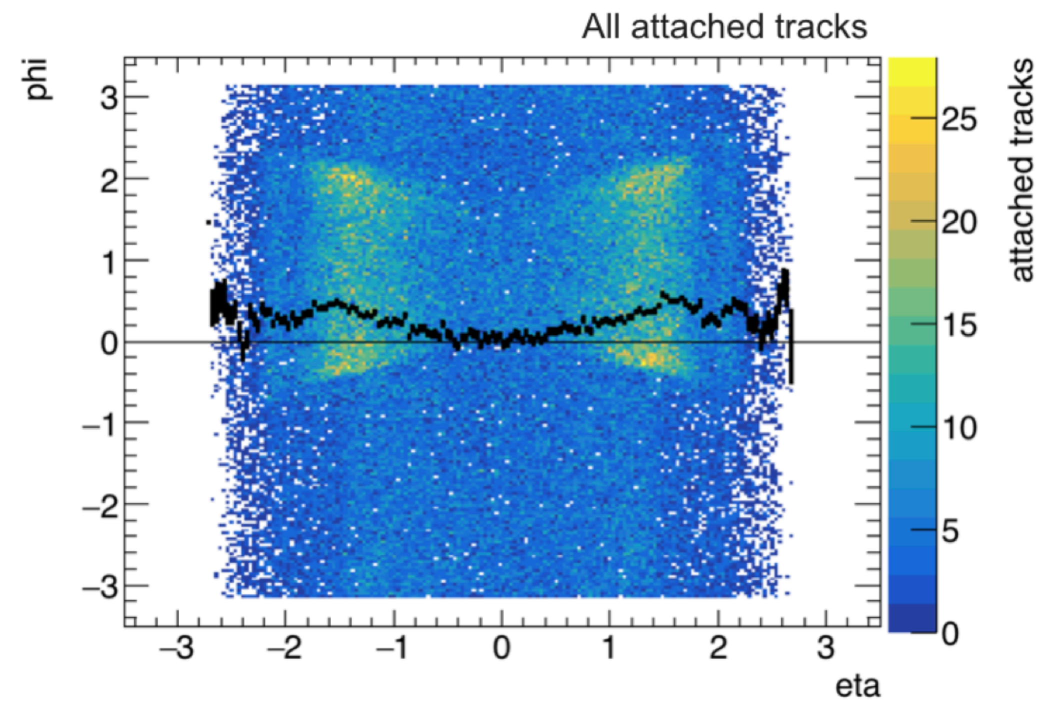
$p_T > 1\,\text{GeV}$
$\chi^2/N_{\text{DoF}} < 5$
$\lvert d_0^{\text{wrt SV}}\rvert$-significance < 5
$\lvert z_0^{\text{wrt SV}}\rvert$-significance < 5
relaxed hit pattern consistency requirement

- Loosened criteria versus selected tracks
- Attachment procedure begins with highest-multiplicity vertex and stops as soon as track attached to single vertex
- Vertex refit after attachment → if reduced goodness of fit $\chi^2/N_{\text{DoF}} < 20$, track attached

VSI Track Cleaning

- Anomalous phi asymmetry in distribution of DV associated tracks with d_0 values roughly corresponding to radial positions of Pixel layers → bug in track-attachment step of VSI algorithm related to simplified extrapolation method

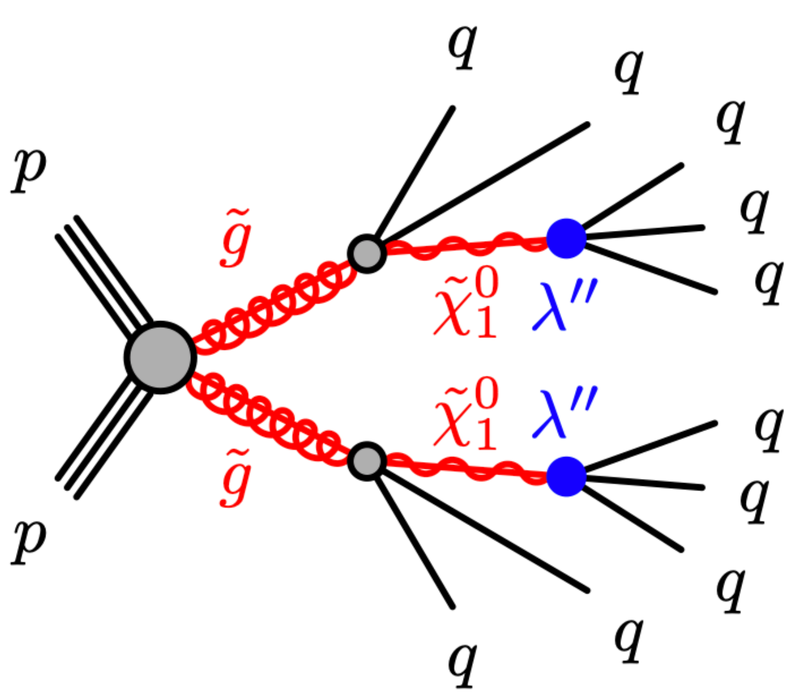


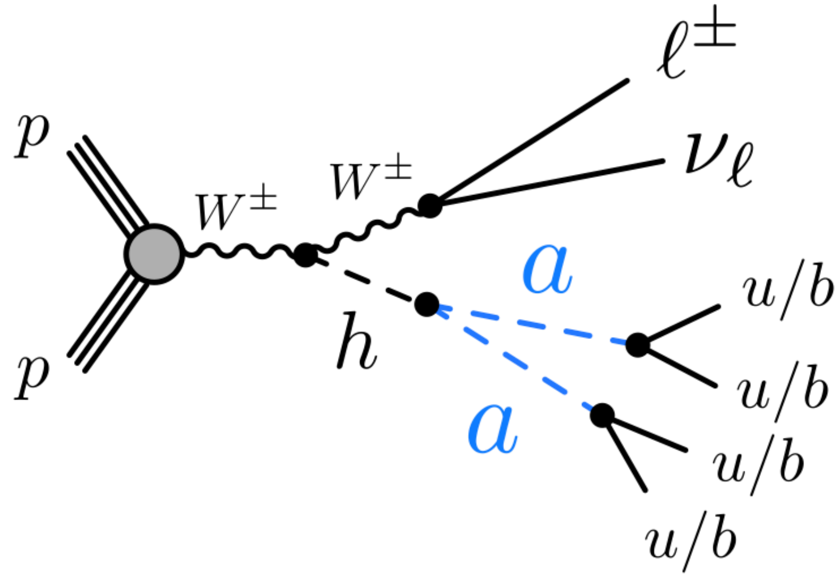


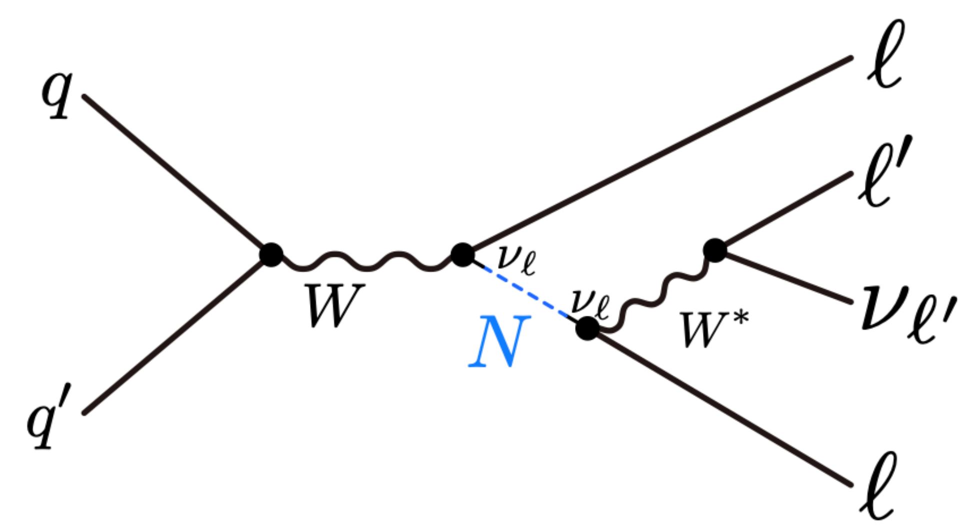
- Offline cleaning procedure to remove problem tracks from DVs and recompute DV kinematics using remaining “clean” tracks:
  - $|d_0| < 2$ mm
  - $26.0 < |d_0| < 39.0$ mm
  - $46.0 < |d_0| < 56.5$ mm
  - $83.5 < |d_0| < 93.5$ mm
  - $117.5 < |d_0| < 128.0$ mm

VSI Performance: Simplified SUSY Models

- Benchmark LLP models used to study VSI algorithm reconstruction performance:
    - Long-lived neutralinos: simplified SUSY model yielding high-track-multiplicity hadronic DVs
    - Higgs portal: hidden-sector model yielding multiple nearby DVs (reconstructed due to B-hadron lifetime)
    - Heavy neutral leptons: BSM model where HNL has small mixing with light neutralinos and thus long lifetimes and yields low-track-multiplicity DVs + lepton pairs

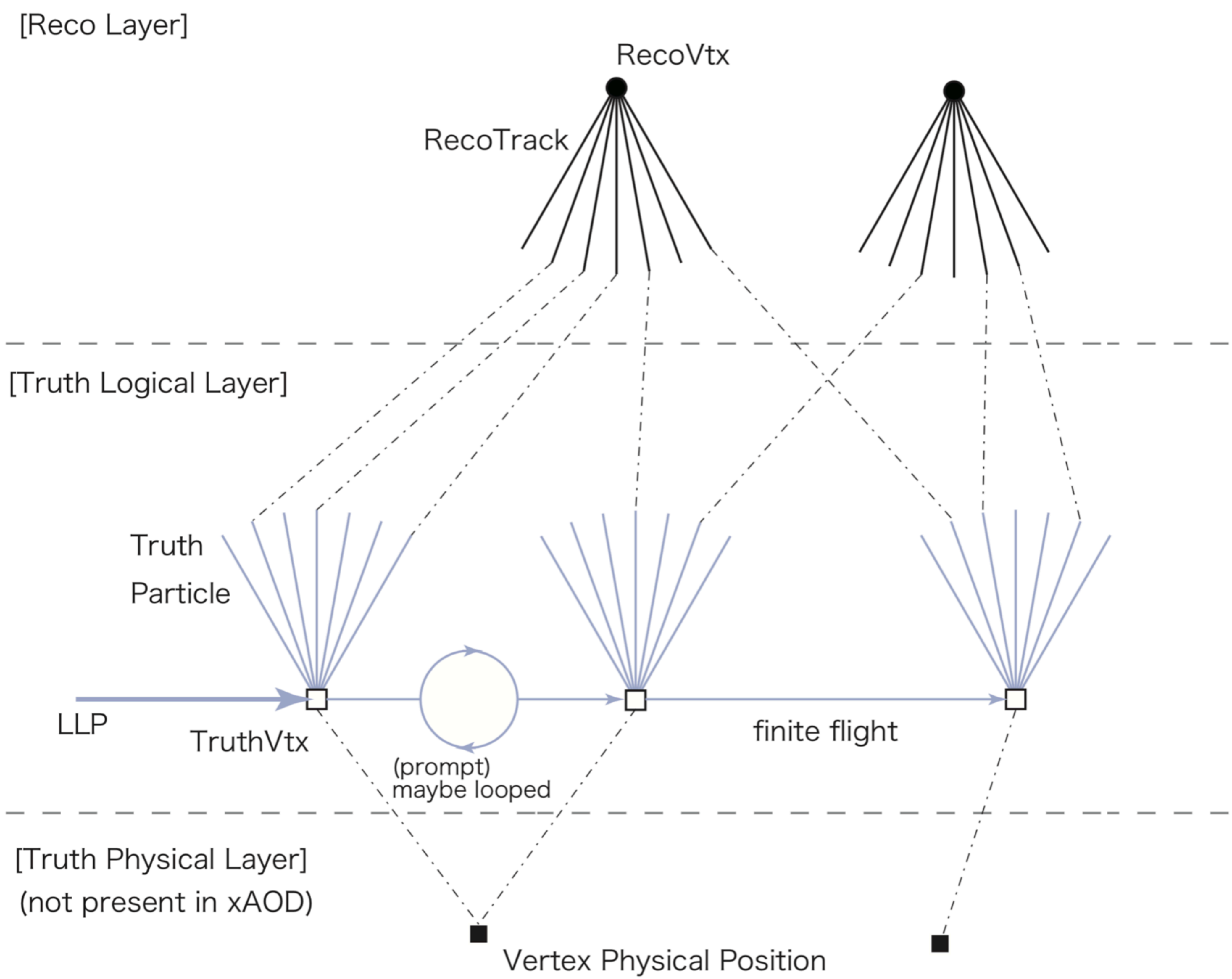






VSI Performance: Truth Matching

- VSI reconstruction efficiency evaluated using truth matching of reconstructed DVs to truth signal LLP decays
- Distance-based matching insufficient, as LLP decays can contain short-lived particles that displace reco DV from LLP position
- → instead look at composition of reco DV in terms of constituent tracks corresponding to LLP decay descendants using p_T-weighted match score
  - → truth match to descendants, not daughters, since LLP decays can involve many intermediate virtual decays as result of hadronization



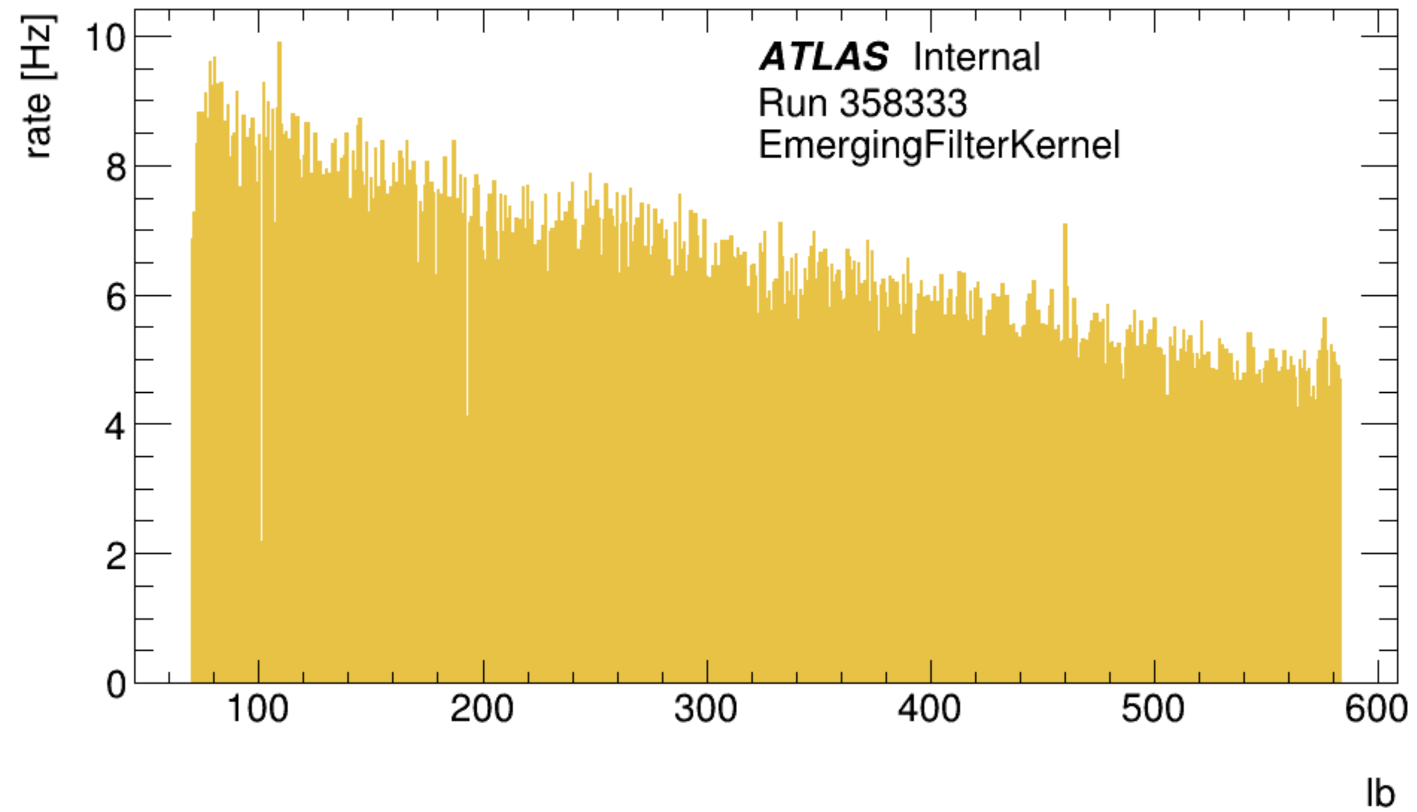
DRAW_RPVLL Monitor

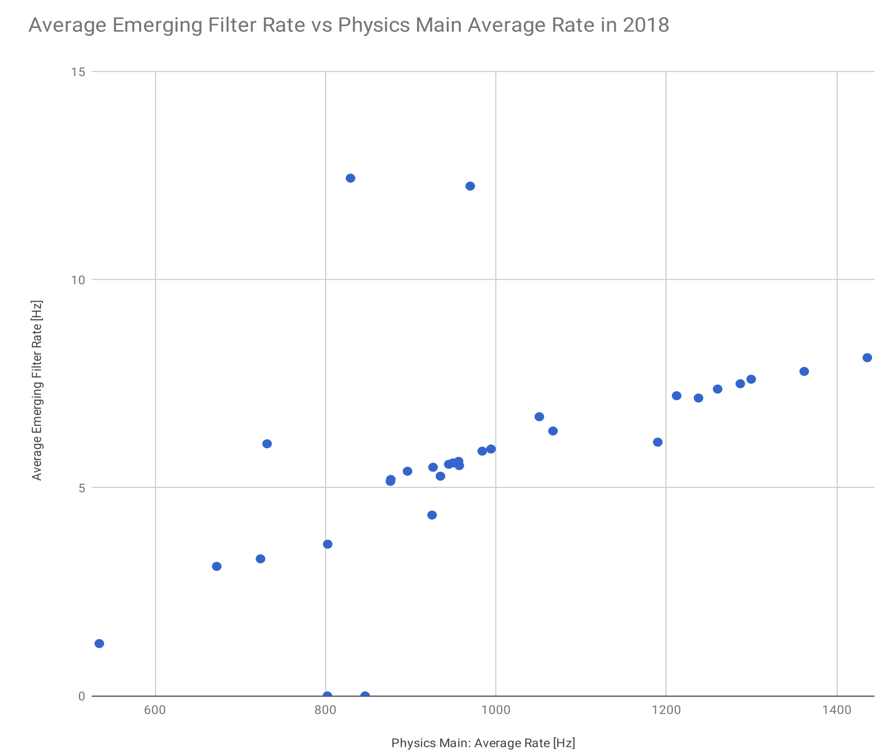
$$\text{filter rate} = \frac{\text{number of events selected by filter}}{\text{number of events in } \texttt{physics\_Main}} \times \texttt{physics\_Main} \ \text{ average rate}$$

- Role of DRAW_RPVLL stream manager:
  - Monitor individual and overall filter rates during LHC Run → track rate of events recorded per run, and watch for significant rate changes
  - Analyze rate breakdown of individual filters, and inform analyses of spikes → aid in making filter changes to reduce rate
  - Act as contact and coordinator for analyses using filters → aid in developing and updating filters and testing rates, advise on reducing rates, and inform analyses of changes affecting filters
  - Maintain filter package, including rate compilation scripts
- TriggerAPI implementation: supplement to hard-coded trigger lists with backups pulled directly from trigger menu so filters updated automatically to any changes → currently turned off

Emerging Filter

- Emerging filter developed specifically for EJs analysis and defining two sets of selections to target signal-like and background-like events
  - **Four-jet signal** events:
    - Lowest unprescaled **four-jet trigger**: HLT_4j90, **HLT_4j100**, HLT_4j110, **HLT_4j120**, HLT_4j130, HLT_4j140, HLT_4j150
    - ≥ **4 jets** with $\mathbf{p_T > 120}$ **GeV** and $\mathbf{|\eta| < 2.5}$
  - **Dijet background** events:
    - Heavily prescaled **single jet trigger**: HLT_j110
    - ≥ **2 jets** with $p_T > 120$ GeV and $|\eta| < 2.5$
  - → jet kinematic cuts designed to correspond to offline analysis cuts and to keep rate relatively low





Data and MC Samples

- Analysis uses **full Run 2 dataset**, **reprocessed** with nonstandard RPVLL reconstruction
  - **139 fb^{-1}** of **sqrt{s} = 13 TeV** pp collision data collected by ATLAS during data taking periods in 2015 - 2018
  - After filtering, data further processed into **derivations** and **ntuples** to reduce data size for storage and analysis (PB → TB → GB)
  - Specialized data flow: ***RAW*** → ***DRAW_RPVLL*** → ***DAOD_RPVLL*** → ***DAOD_EXOT23*** → ***EJs ntuples***
- MC signal and background samples produced to design analysis → **36 out of 90 signal points** + **JZ4W QCD dijets** in Pythia
  - → MC simulated with full detector simulation, reconstructed with RPVLL reconstruction, and processed into derivations and ntuples as for data

DAOD_EXOT23

- Derivation developed specifically for EJs analysis → removes unnecessary events and objects to reduce size of dataset, and adds in necessary truth jet containers
- Selects data passing on **four-jet** and **single-jet triggers** for signal and background events
- Adds in custom-built **truth dark jets** = reconstructed truth objects not available by default at ATLAS
  - → represent dark jets formed from dark sector showering + hadronization of initial dark quarks produced in hard process
  - Reconstructed by clustering “stable” dark truth particles into jets of radius R = 0.4 with anti-kt algorithm → “stable” dark particle = one decaying to SM, or not at all
  - Used to distinguish between truth jets produced from dark vs SM partons → method of tagging EJs at generator level

EmergingJetsAnalysis code framework for ntuple production and analysis

Signal Selection Overview

- Aim of analysis to select signal events containing EJs signature characteristic of simplified dark QCD models → selection criteria define signal region
- **Preselection** → process events from filtered derivations into analysis-specific ntuples containing orthogonal regions isolating signal and background
  - Loose event selection based on RPVLL / EJs analysis skimming + ATLAS-common quality criteria → EJs filter and derivation criteria, requirements on data quality, and cuts on loose jet multiplicities
  - Separate search and validation region selections based on jet triggers, multiplicities, and basic kinematics define mutually exclusive analysis regions loosely isolating signal and background events
- **Signal selection** → process events in ntuples to remove remaining background from search region and further isolate signal events
  - Additional offline selections on number of displaced vertices, number of emerging jet objects, and some event-level leading four-jet observable on top of search region selections:
NDV + NEJ + NJetX
  - Quality criteria applied to displaced vertices and emerging jet objects to isolate signal and reduce fake / background objects

Event Preselection	• DRAW_RPVLL + DAOD_EXOT skimming • Event cleaning + quality cuts • HS: ≥1 PV with ≥ 2 tracks • ≥ 2 baseline jets

Search Region Selection	• **Four-jet trigger(s)**:
	  • HLT_4j100 or HLT_4j120
	• **Four jets** with
	  • **p_T > 120 GeV**
	  • **$\vert\eta\vert$ < 2.5**
	• Leading 4-jet **H_T > 1000 GeV**

Validation Region Selection	• **Single-jet trigger(s) + two or three jets**
Signal Selection	• **N_{DV}** ≥ 2, 3, or 4 • **N_{EJ}** ≥ 0, 1, 2 • **X^{NJet}** > Y → signal DVs + EJs subject to quality criteria → X^{NJet} = leading four-jet or dijet observable

Event Preselection

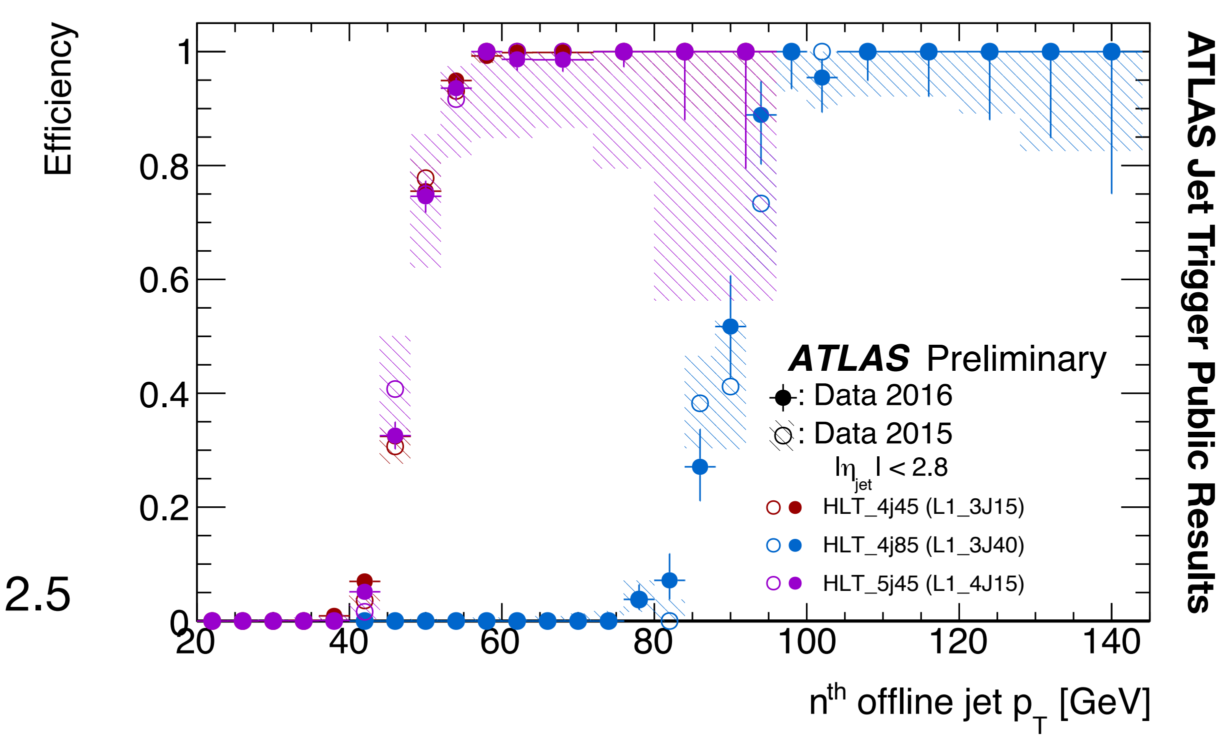
- RPVLL / EJs analysis skimming + ATLAS-common quality criteria
  - Initial **skimming** defined by DRAW_RPVLL Emerging filter and DAOD_EXOT23 derivation selections
  - **Event cleaning** procedure removes poor quality, corrupted, or incomplete data → rejects events potentially plagued by detector problems
  - **Good Runs List** = list of good lumiblocks within given runs → lumi-blocks affected by problematic detector conditions excluded , and any event in bad lumi-block rejected
  - **Jet cleaning** removes events with any unclean jets → reduces contributions / fakes from non-collision backgrounds
    - Unclean jet = $p_T > 20$ GeV and $|\eta| < 2.8$ failing Loose WP
  - **Hard scatter** criteria requires at least one primary vertex with at least two outgoing tracks → primary vertex with largest $\sum p_T^2$ taken as PV (rest pileup)
  - Baseline object selection:
    - At least **two baseline jets** per event → baseline jet = EMTopo anti-kt R=0.4 jets with $p_T > 50$ GeV and $|\eta| < 2.7$
    - All tracks in event required to pass baseline quality criteria in line with selections of LRT + VSI algs: $p_T > 1$ GeV and $\chi^2/N_{DoF} < 50$

Analysis Region Selection

- Independent analysis regions — search and validation — defined to separately study signal-like and background-like events and to investigate and validate further selections specifying ultimate signal region

- **Search Region** (signal-rich) = precursor to signal region → loosely isolates expected signal before tight selections to further remove background
  - Lowest unprescaled four-jet trigger: **HLT_4j100** in 2015 + 2016 and **HLT_4j120** in 2017 + 2018
  - At least **4 jets** with **p_T > 120 GeV** and **$|\eta|$ < 2.5** → expected to be ~ within trigger plateau
  - Leading 4-jet **H_T > 1000 GeV** → reduces QCD background

- **Validation Region** (background-rich) → isolates orthogonal region where no signal expected for purposes of studying backgrounds and validating analysis techniques
  - Single jet trigger
  - Exactly 2 or 3 jets with p_T > 120 GeV and $|\eta|$ < 2.5

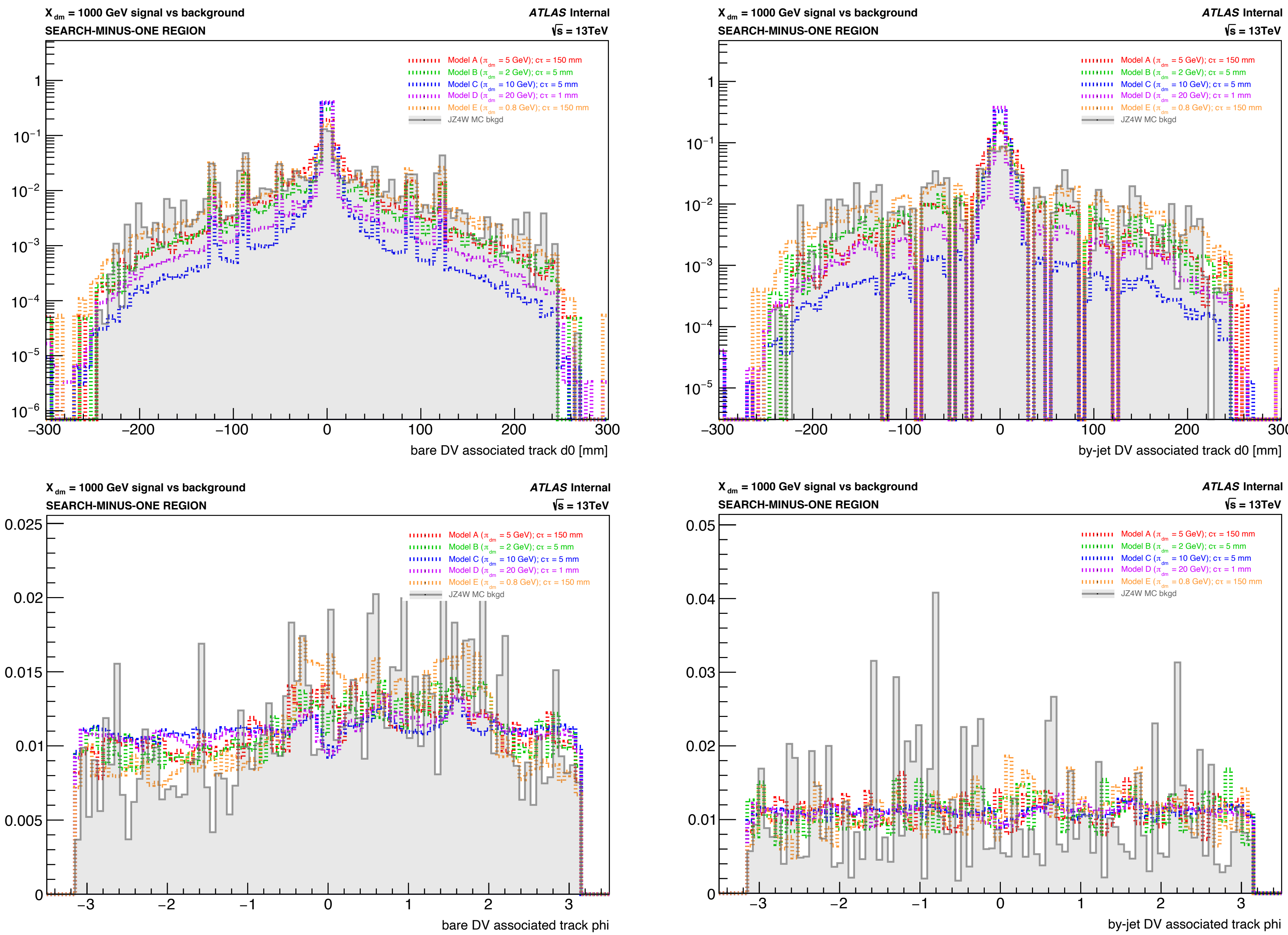


DV Selection

- Signal events required to contain some number of displaced vertices passing set of quality criteria / signal selections designed to reduce contributions from fakes and SM backgrounds: **N_{DV} ≥ 2, 3, or 4**
  - Signal DV is one passing defined working point(s) and associated to one of N_{jet} leading jets in event
- Three tentative WPs defined to optimally select signal DVs and reject background DVs:
  - **Loose** WP based on standard selections common to RPVLL DV analyses → removes known background sources, i.e. material interactions with the detector, bad tracks, k-short decays, and poorly reconstructed vertices
  - **Medium** + **Tight** WPs apply additional selections based on comparisons between signal vertices matched to LLP decays and unmatched background vertices
    - In contrast to other DV searches, EJs signal features low-mass and low-track-multiplicity DVs that are difficult to differentiate from fakes → seek alternative cuts
- DV additionally required to be associated within **$\Delta R < 0.6$** to one of **N_{jet} leading jets** in event → selection exploits EJs topology of signal, where multiple DVs expected inside single jet cone

Loose
Material Map veto
Track cleaning
Fiducial volume: $r < 300$ mm and $\lvert z \rvert < 300$ mm
Quality of fit: $\chi^2/N_{\mathrm{DoF}} < 5$
K-short mass cut: $m > 0.7$ GeV
Medium
Loose selections
$p_{\mathrm{T}} > 2.5$ GeV
Tight
Medium selections
minimum track $\lvert d_0 \rvert < 10$ mm
minimum track $\lvert z_0 \rvert < 100$ mm
minimum square-root track d_0-error < 0.5
minimum square-root track z_0-error < 1.5

DV Track Cleaning

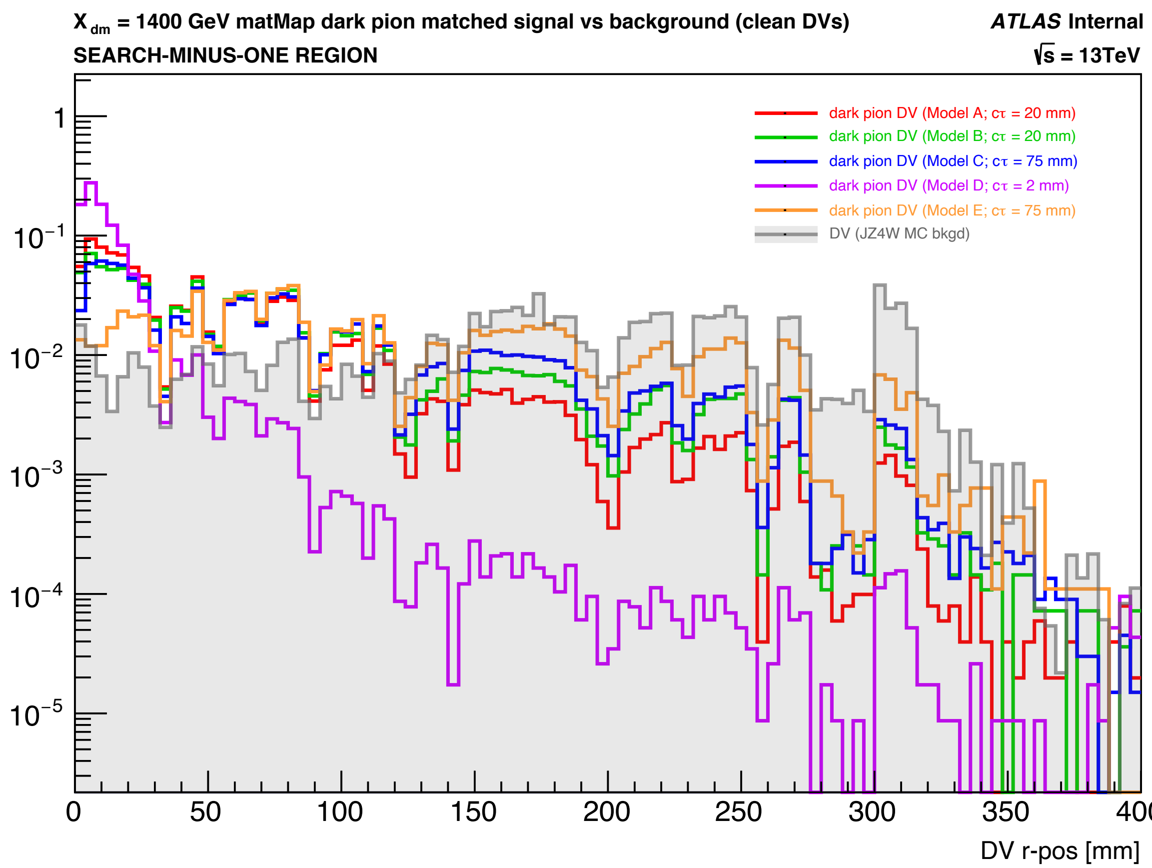


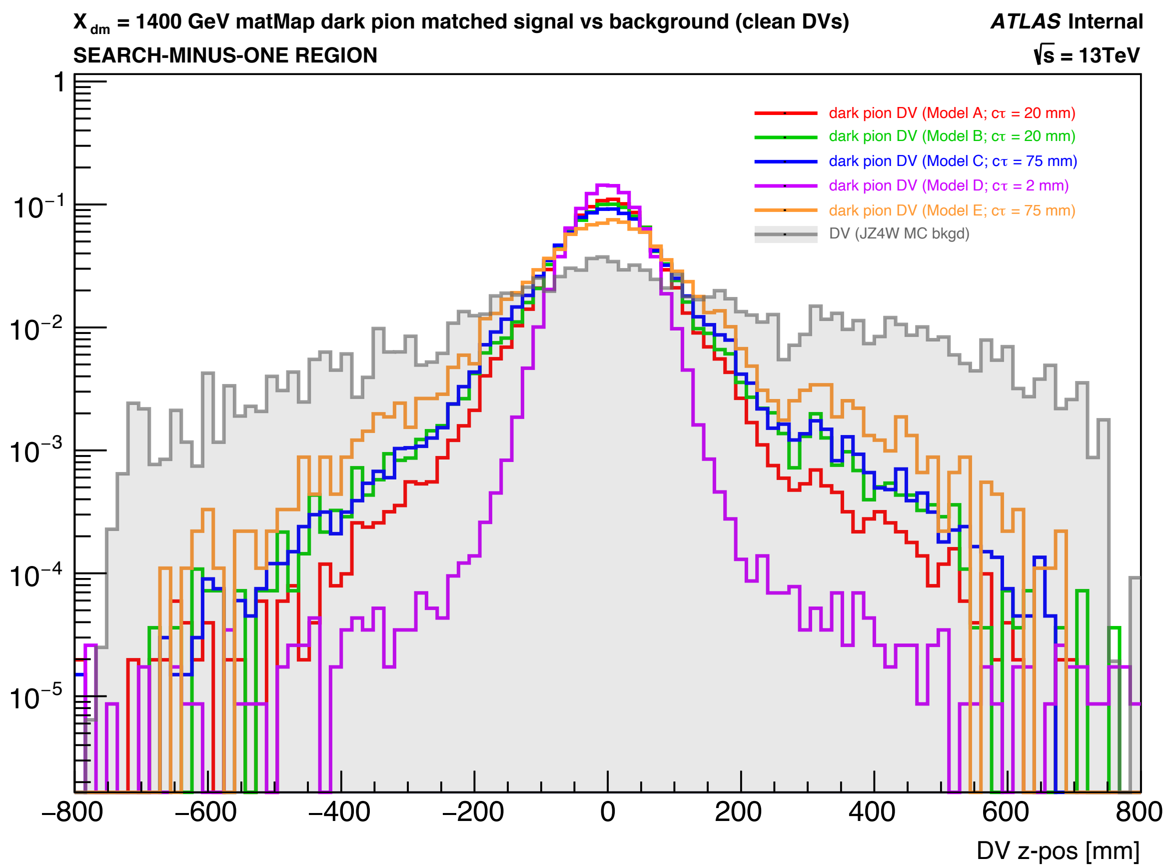
Before Cleaning **After Cleaning**

- Problematic tracks removed from vertices, and kinematics recomputed using remaining clean tracks
  - Cleaning = removal of associated tracks with d_0-values roughly corresponding to radial positions of Pixel layers → mitigates anomalous phi asymmetry due to VSI bug
  - Additional track trimming (further filtering of poor-quality or poorly-associated tracks) still to be investigated…

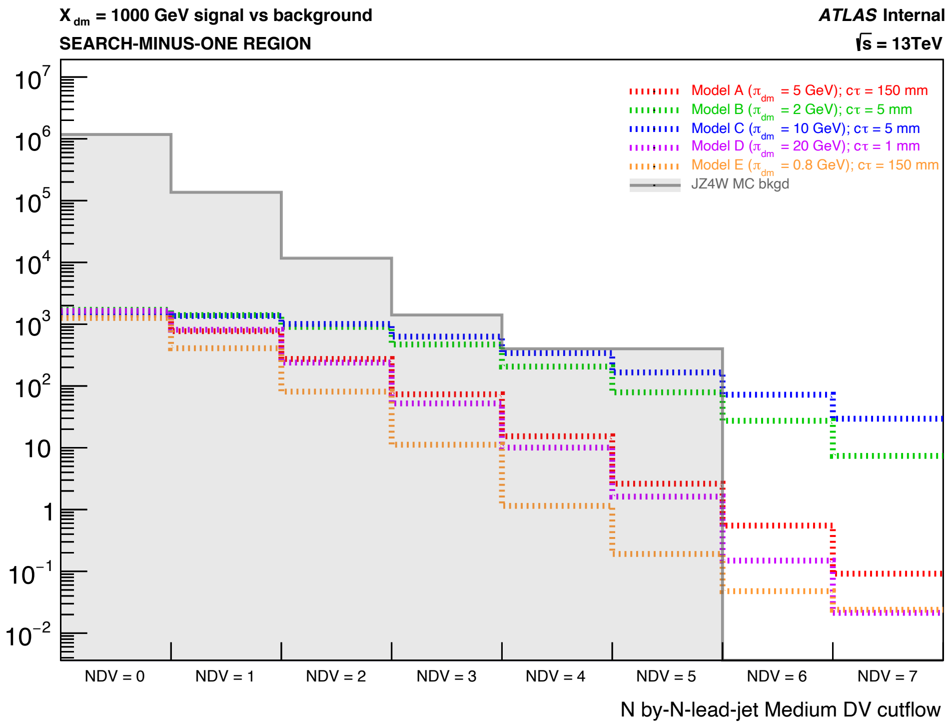
Fiducial Volume Cut

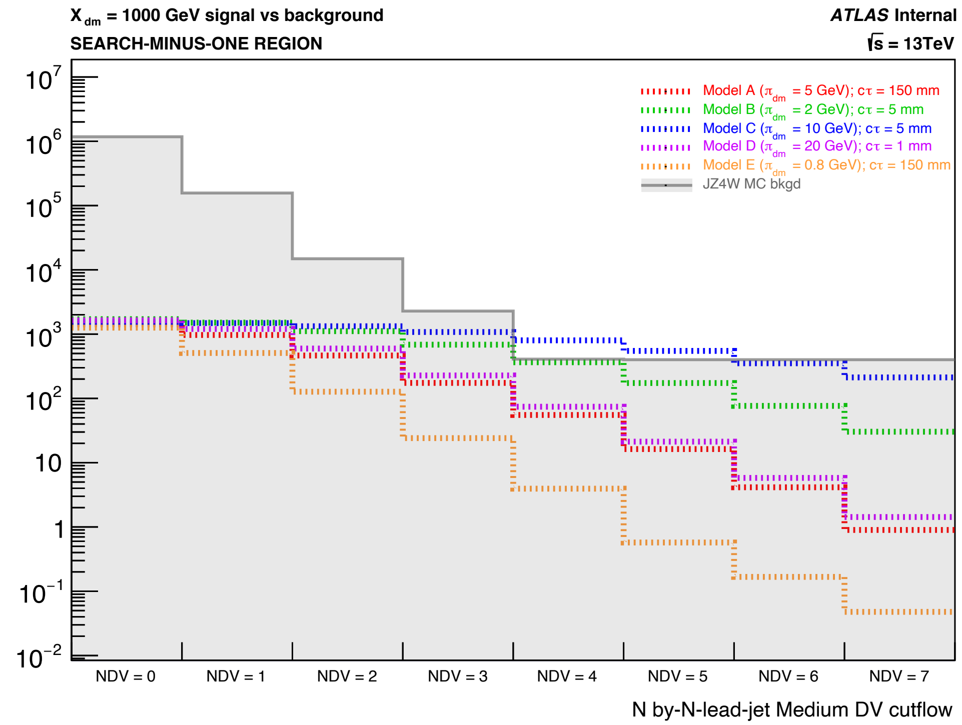
- Can only reasonably expect to efficiently reconstruct good-quality DVs within certain fiducial volume: **r < 300 mm** and **|z| < 300 mm** → DVs outside this region rejected
  - Radial and longitudinal limits set by capabilities of tracking and vertexing algorithms → reco efficiencies drop above thresholds
  - Material map only defined within fiducial volume
- Background relatively flat across all r and z → indicative of fakes reconstructed from random track crossings throughout detector

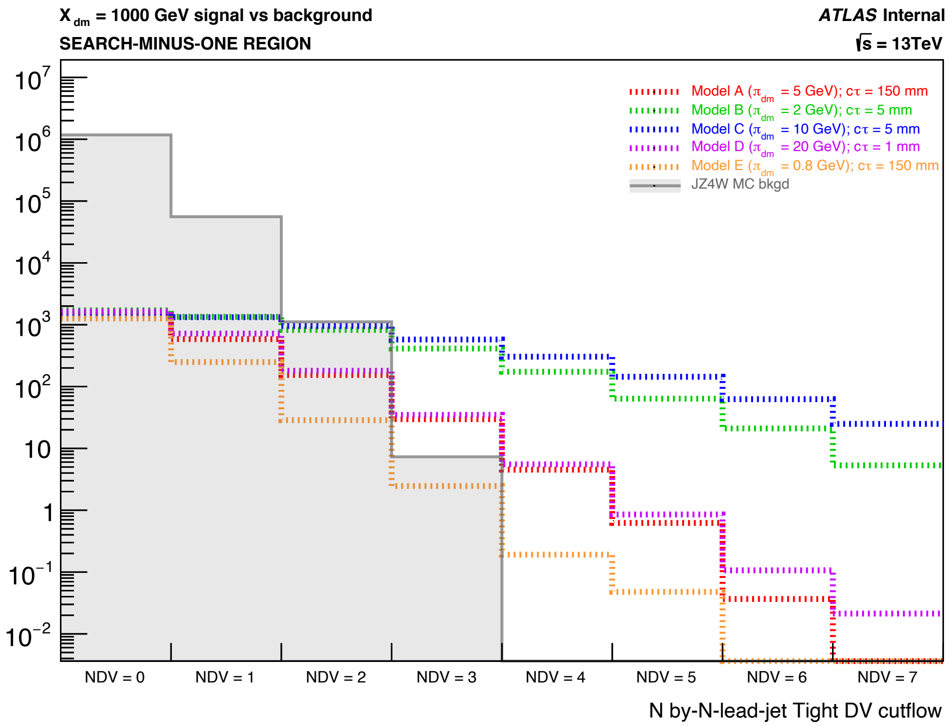


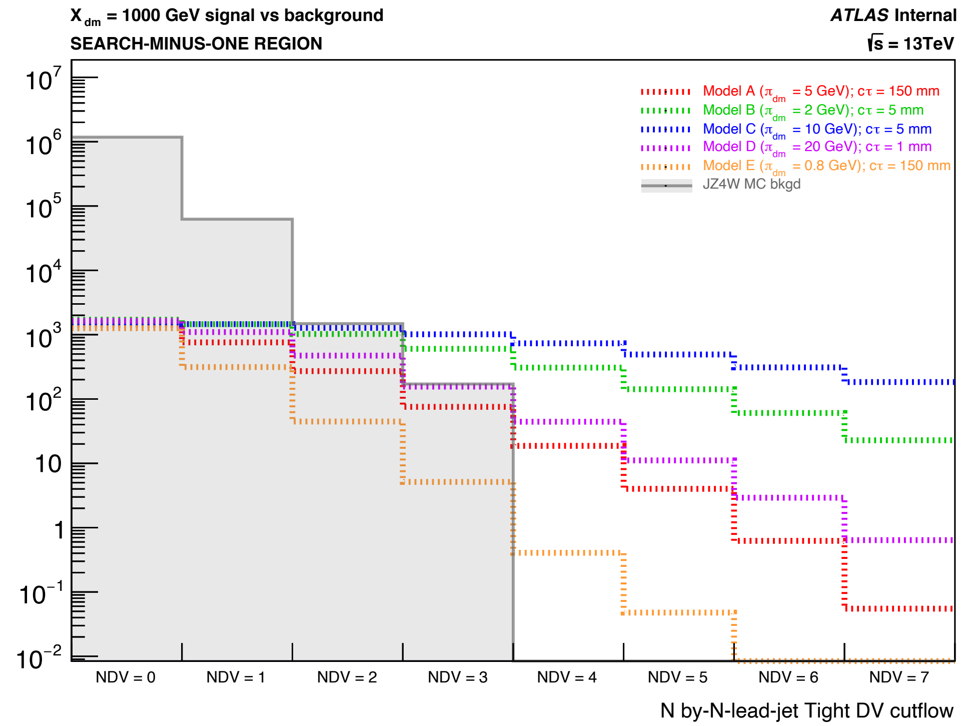


By-Jet vs By-N-Jet DVs









- Looser jet association criteria, requiring DV by any jet in event rather than leading jet, increases signal yield somewhat for all DV WPs
- → significantly increases background for Loose and Medium DV WPs, but leaves background yield relatively unaffected for Tight DVs
- → could safely loosen jet association requirement in conjunction with Tight DV selections to boost signal without risk of increased background

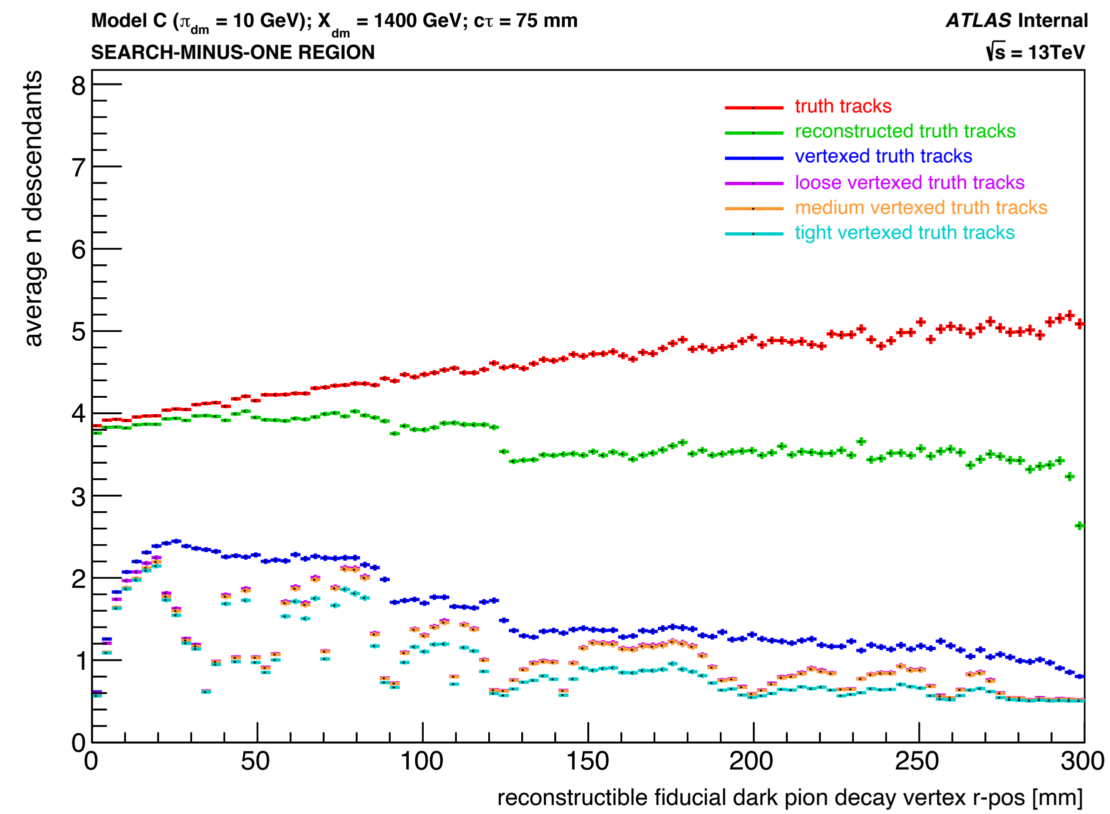
DV Reconstruction Efficiency

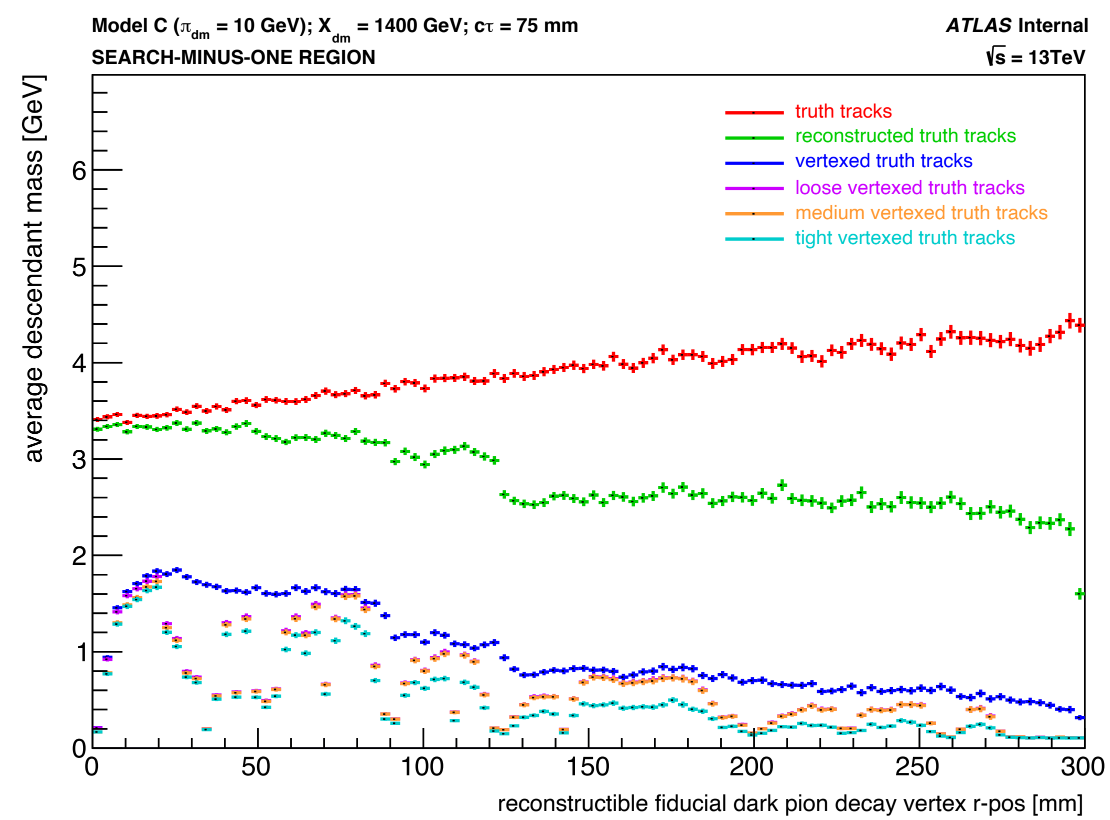
- Ability to select EJs signal events directly depends on efficiency of reconstructing displaced vertices from dark pion decays in event
- Overall VSI performance depends on tracking efficiency for reconstructing charged particles as tracks and vertexing efficiency for reconstructing tracks as vertices → study efficiencies separately to extract individual effects of tracking and vertexing algorithms
- → illustrated through average vertex track multiplicity and descendant mass as functions of radial position
  - Average N truth tracks = maximum N available tracks from LLP decay to be reconstructed into DV
  - Average N reconstructed truth tracks = maximum N tracks reconstructed by tracking algorithms and available for use in vertexing

    → decrease from truth = efficiency loss due to track reco
  - Average N vertexed truth tracks = average N tracks per reconstructed DV

    → decrease from reco = efficiency loss due to vertex reco
  - Average N Loose, Medium, Tight vertexed truth tracks = average N tracks per reconstructed signal DV

    → decrease from vertexed = efficiency loss due to DV signal selections



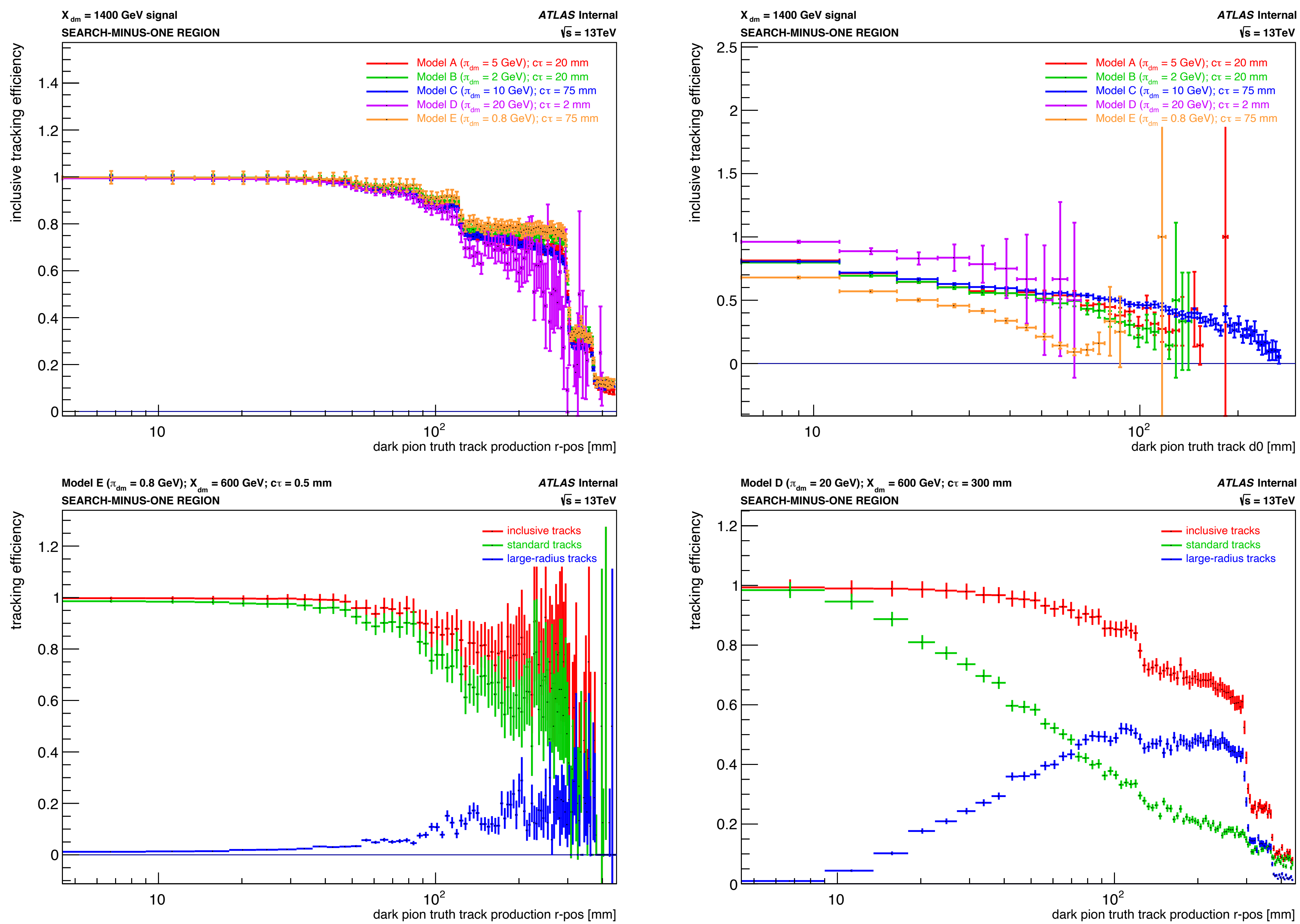


DV Tracking Efficiency

$$\text{reconstruction efficiency} = \frac{\text{N signal truth particles matched to reconstructed tracks}}{\text{N signal truth particles}}$$

$p_T > 1$ GeV
$|\eta| < 2.5$
$r_{\text{prod}} < 440$ mm
charge $= \pm 1$
in LLP decay chain

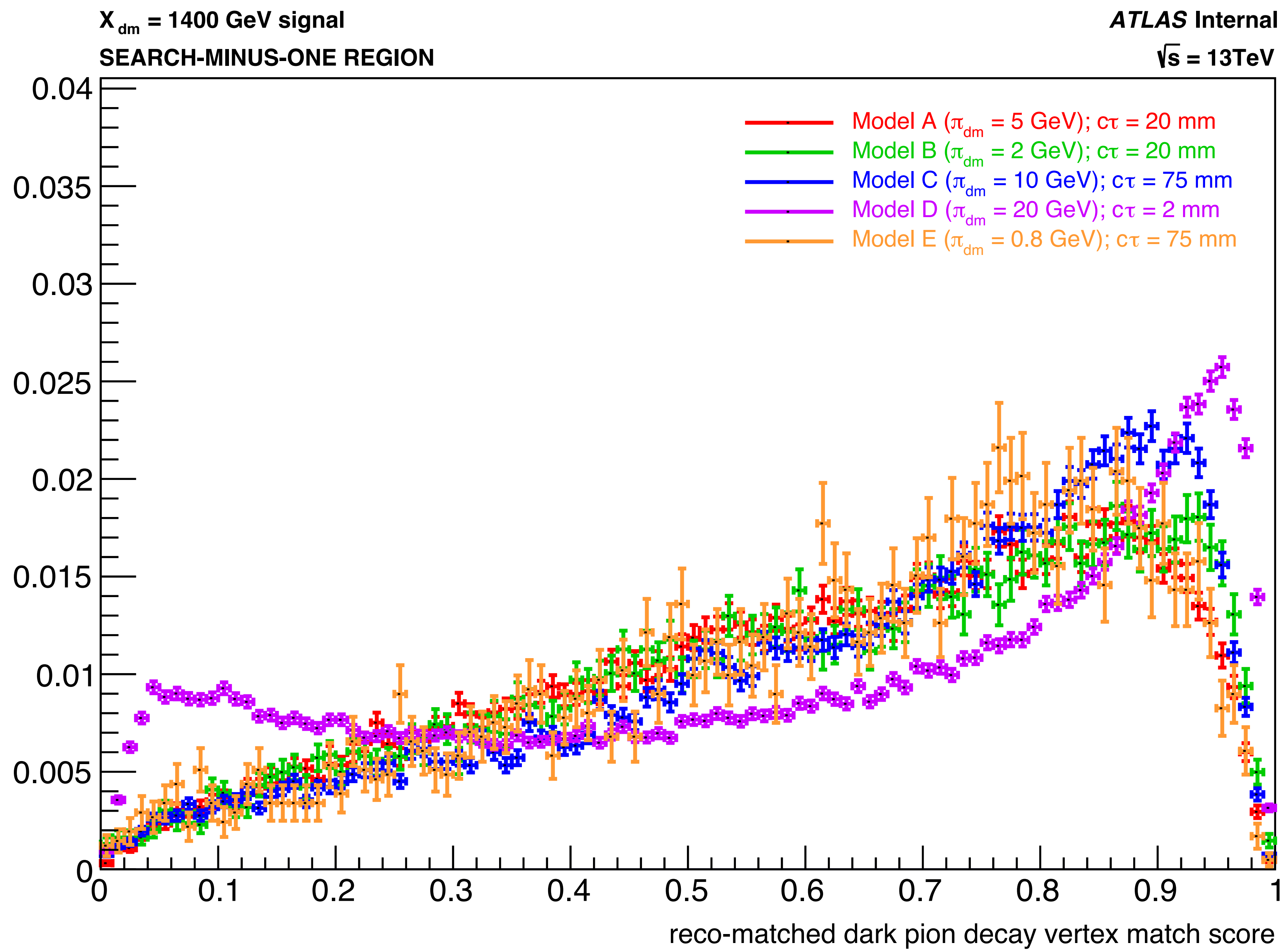
- Tracking efficiency deteriorates with displacement from IP
  - Fully efficient up to r ≈ 50 mm, then starts to degrade → significantly drops around r ≳ 300 mm
  - Efficiency higher for heavier models and shorter lifetimes, as more likely to produce higher-p_T and less displaced tracks
- Standard tracking highest for prompt tracks and decreases with r → opposite for LRT
  - Short lifetimes dominated by standard tracking, and LRT efficiency low across r
  - As lifetime increases, standard tracking efficiency sharply decreases, and LRT compensates



DV Match Score

Require match score **s > 0.5**
→ most vertices satisfy criteria

$$s(v,l) \equiv \frac{\sum_{i\,\in\,\text{tracks}\,\in\,v} \left(p_{\mathrm{T}}^{(i)} \mid \text{descendant of LLP decay } l \right)}{\sum_{i\,\in\,\text{tracks}\,\in\,v} p_{\mathrm{T}}^{(i)}}$$



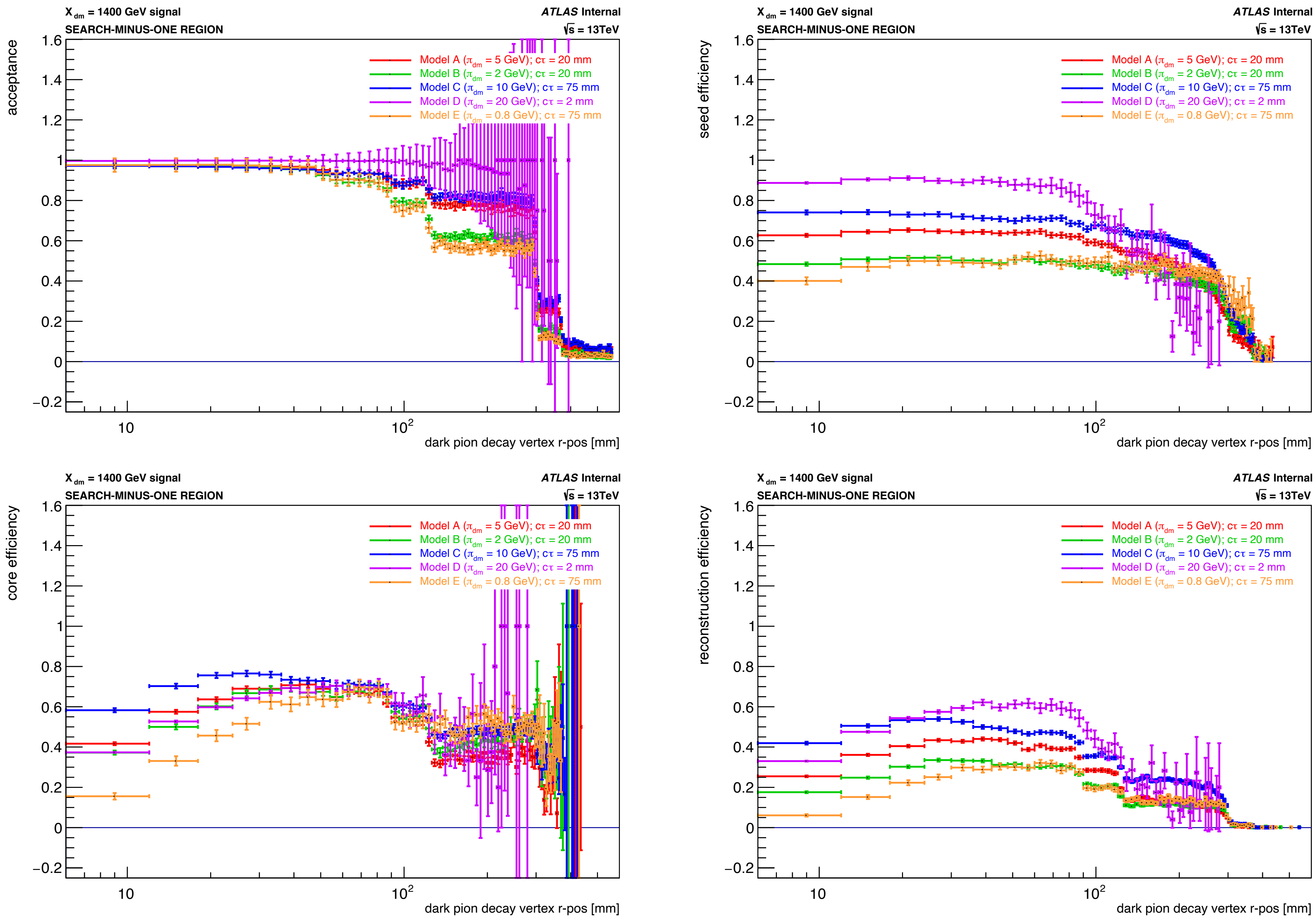
Vertexing Efficiency

$$\text{reconstruction efficiency} = \frac{\text{N signal truth LLP decays matched to reconstructed DVs}}{\text{N signal truth LLP decays}}$$

→

r < 563 mm
|z| < 2720 mm
≥ 2 charged descendants w/ p_T > 1 GeV

$$\varepsilon^{\text{tot}} = \mathcal{A} \cdot \varepsilon^{\text{alg}} = \mathcal{A} \cdot \varepsilon^{\text{seed}} \cdot \varepsilon^{\text{core}}$$

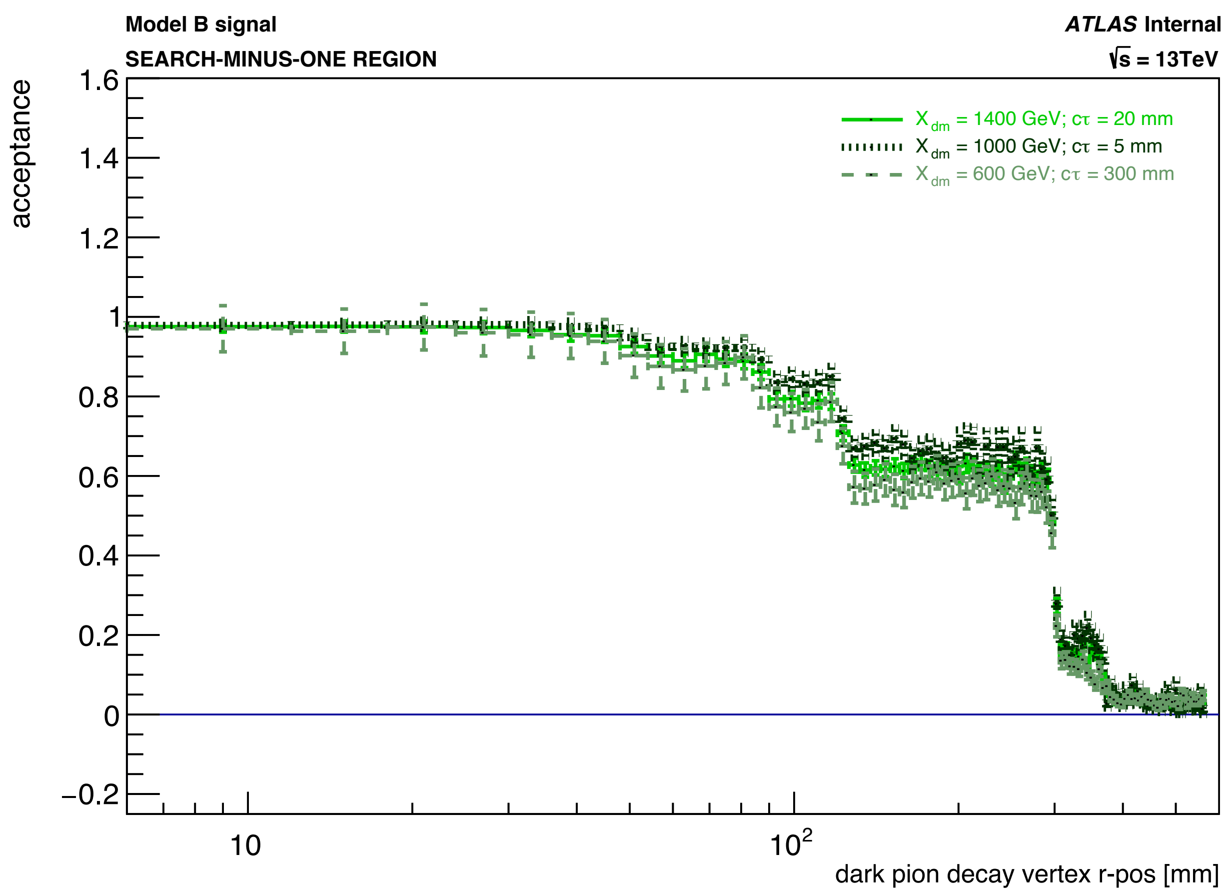


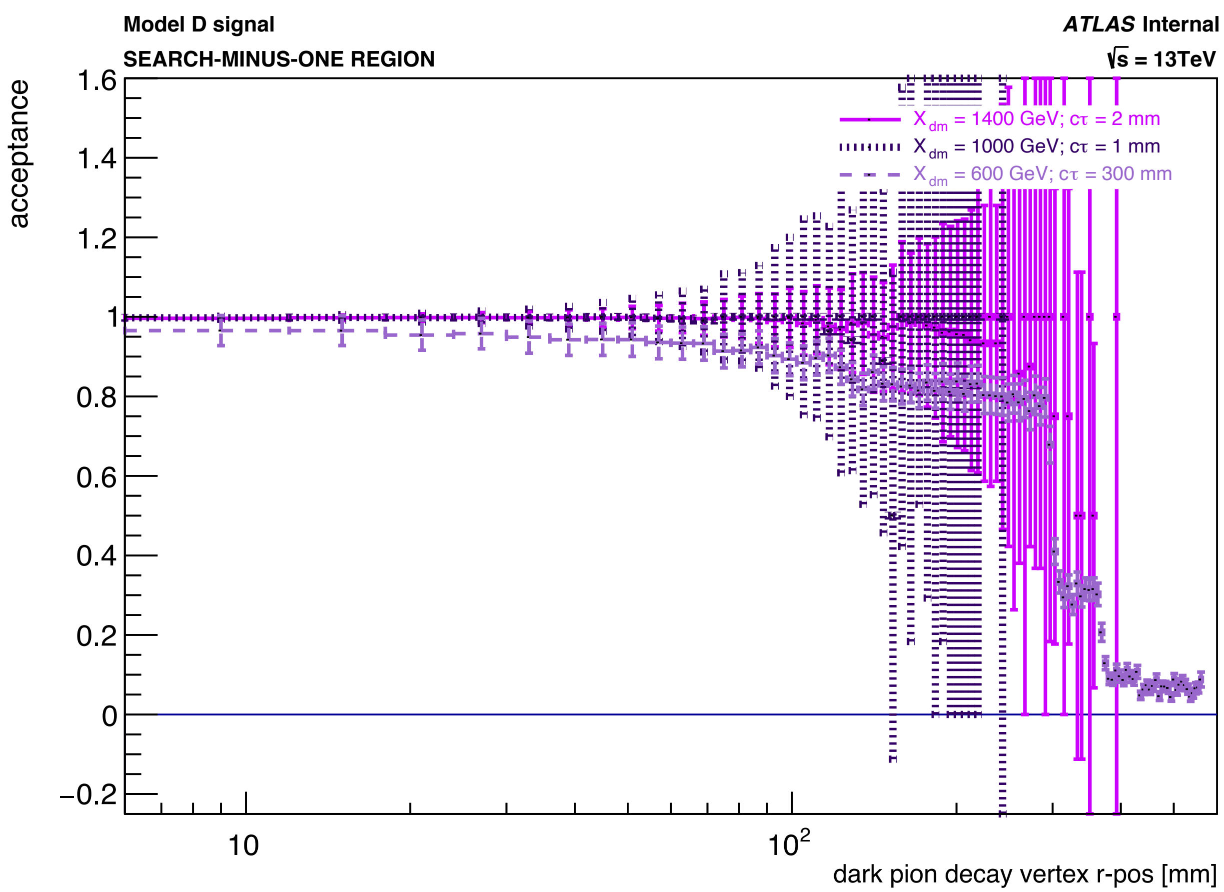
- Ability to select EJs signal events directly depends on efficiency of reconstructing displaced vertices from dark pion decays in event
- DV reco efficiency decreases with increasing radial displacement and is impacted by all EJs signal parameters: model mass, mediator mass, and dark pion lifetime
- → seed efficiency primarily dependent on model mass, as heavier models produce higher-multiplicity DVs, increasing probability of ≥ 2 tracks being successfully reconstructed and selected for vertexing
- Overall vertexing efficiency considerably lower for EJs search compared to other DV searches
- → algorithmic efficiency suffers for EJs DVs, which tend to be low-mass and low-multiplicity

DV Acceptance

$$A = \frac{\text{N signal truth LLP decays with} \geq 2 \text{ reconstructed tracks}}{\text{N signal truth LLP decays}}$$

- Acceptance = measure of tracking algorithm impact on total vertex reconstruction efficiency
  - Minimally dependent on dark pion lifetime → slightly lower for larger lifetimes, as track efficiency degrades somewhat with lifetime
  - Stronger dependency on model mass → lighter model acceptance degrades faster with r compared to heavier models, since lighter models produce DVs with lower track multiplicities

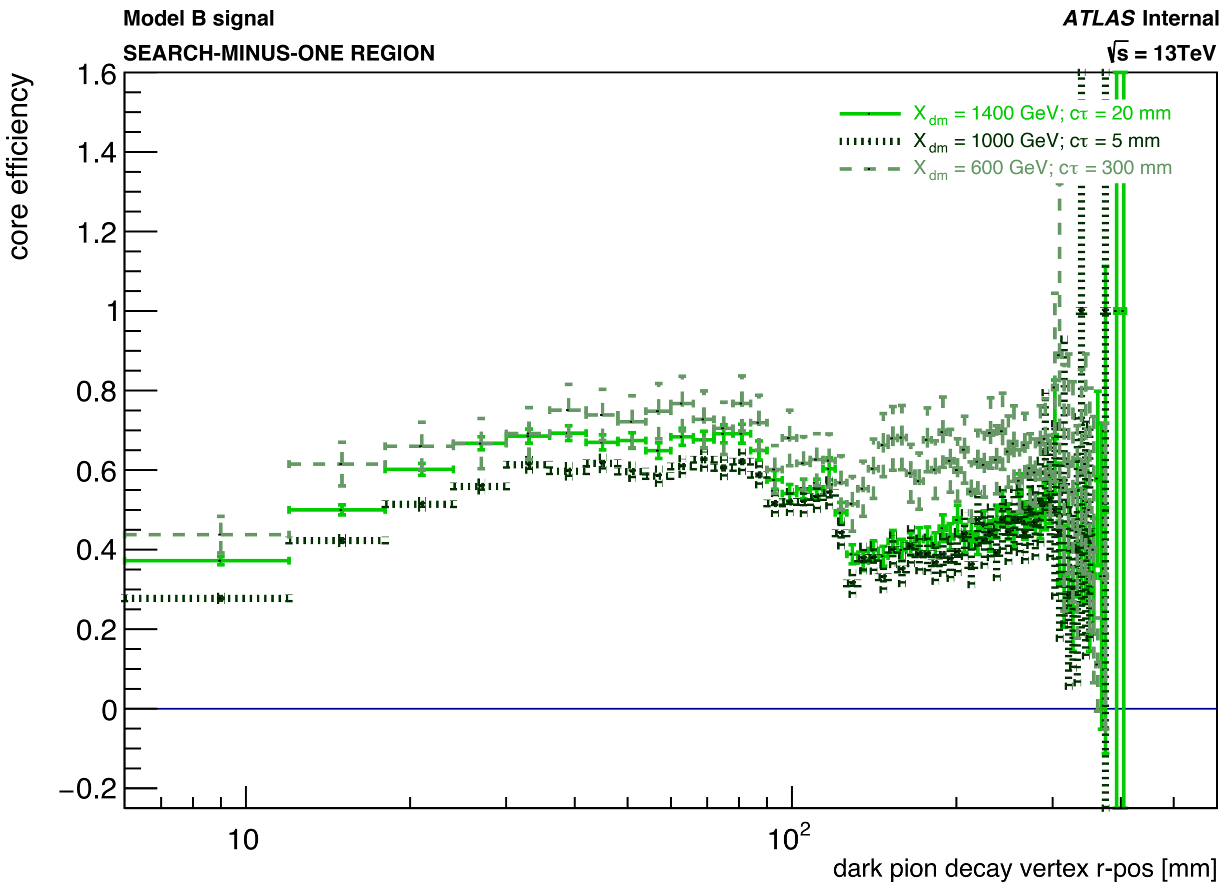


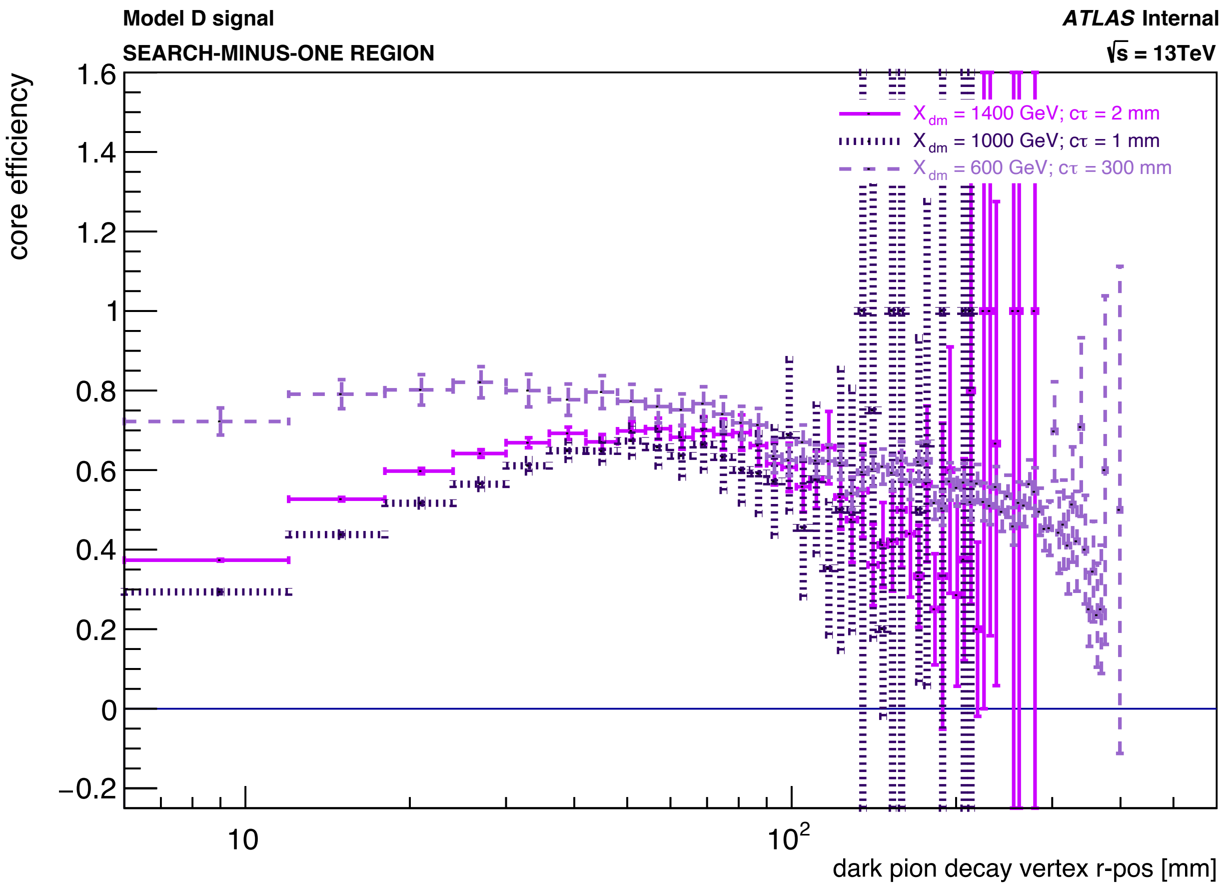


DV Core Efficiency

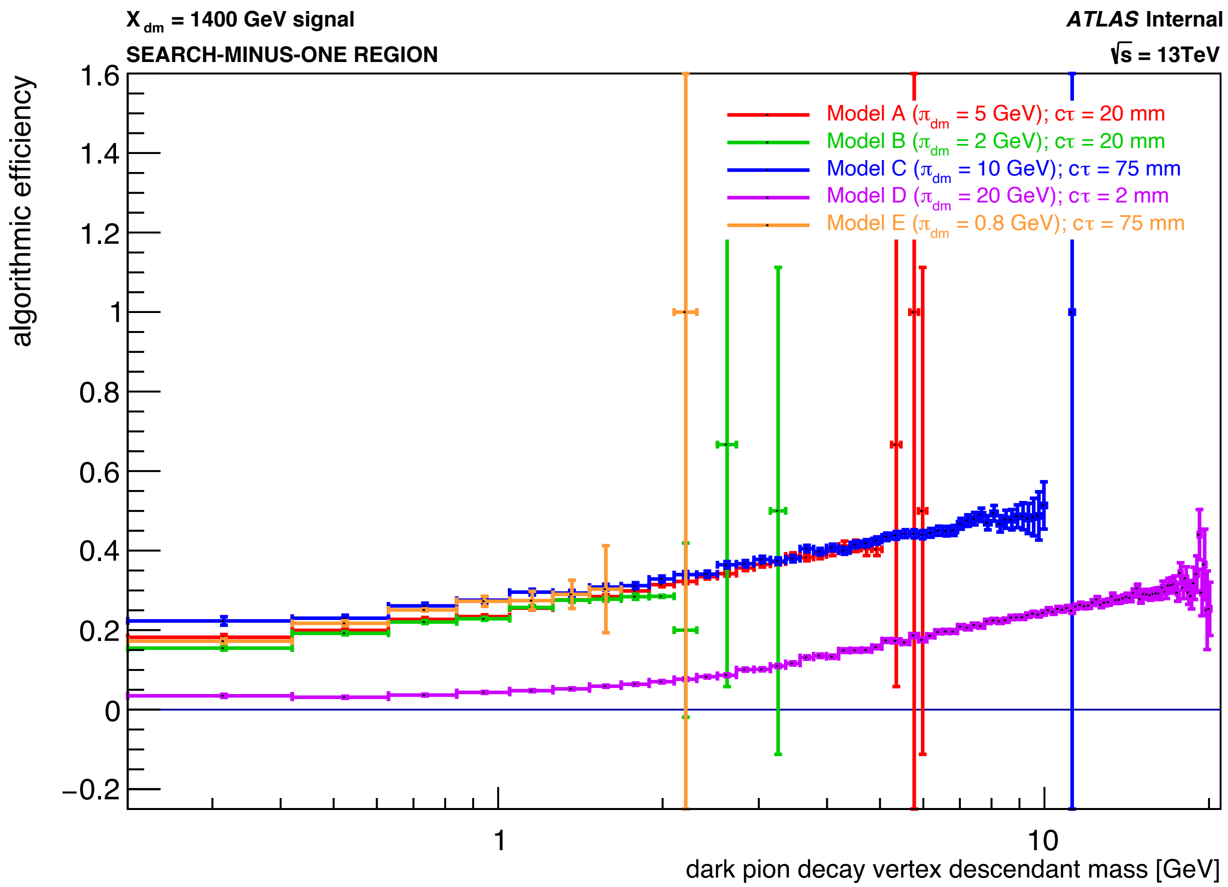
$$\epsilon^{core} = \frac{\text{N signal truth LLP decays matched to reconstructed DVs}}{\text{N signal truth LLP decays with} \geq 2 \text{ selected tracks}}$$

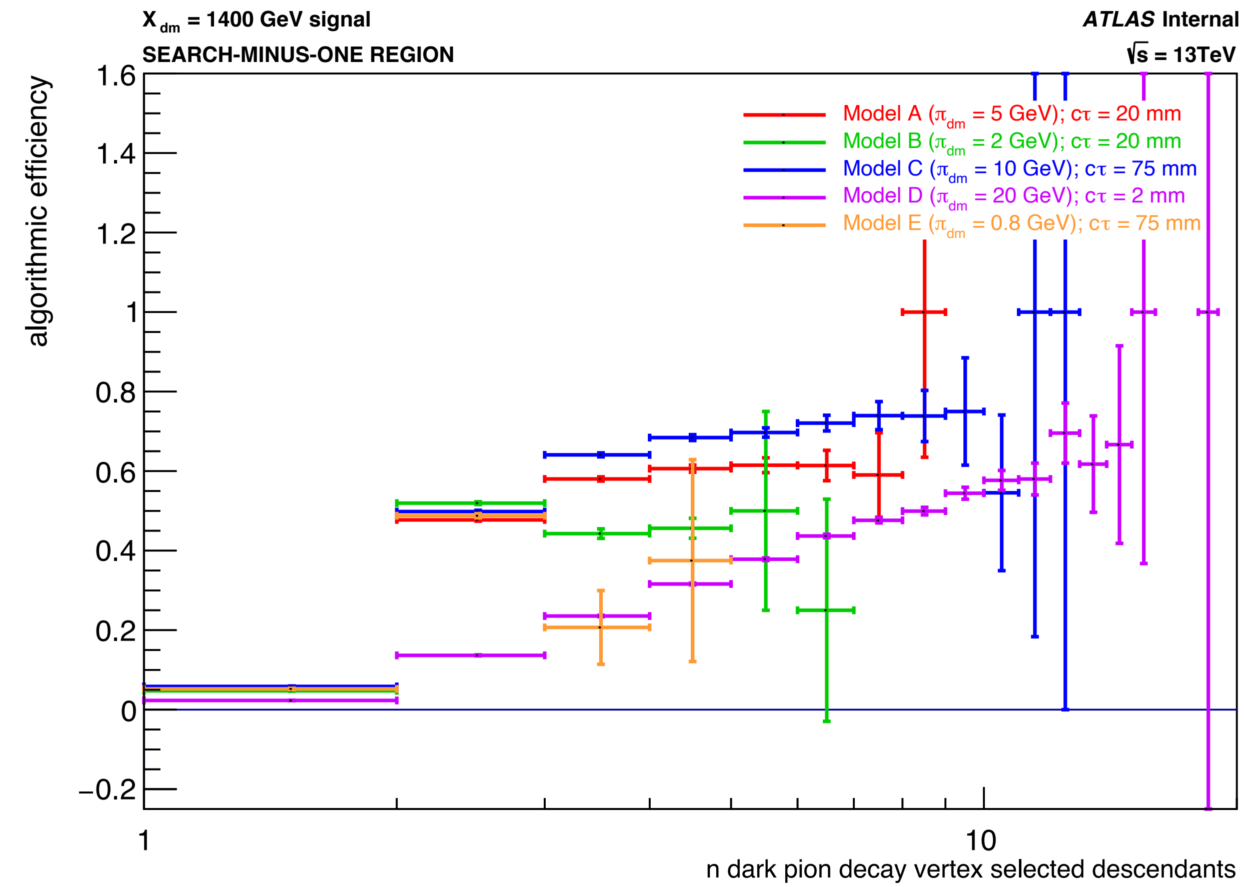
- Core efficiency = measure of vertex fitting efficiency given set of constituent tracks
  - Low efficiency at low r and for short lifetimes → consequence of d_0 cut during seed track selection and kinematic cuts during two-track seed-finding
  - Efficiency degrades with increasing r due to LRT inefficiency and average LLP boost → fewer tracks available for vertexing and more collimated decay products (making DVs more difficult to reconstruct) at large r





DV Algorithmic Efficiency

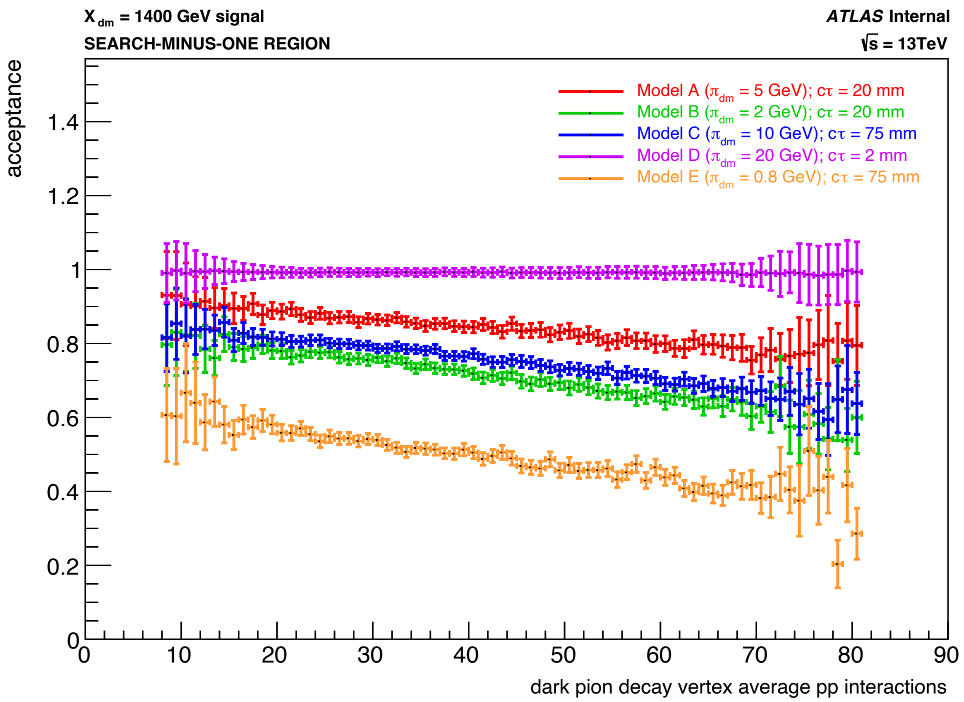


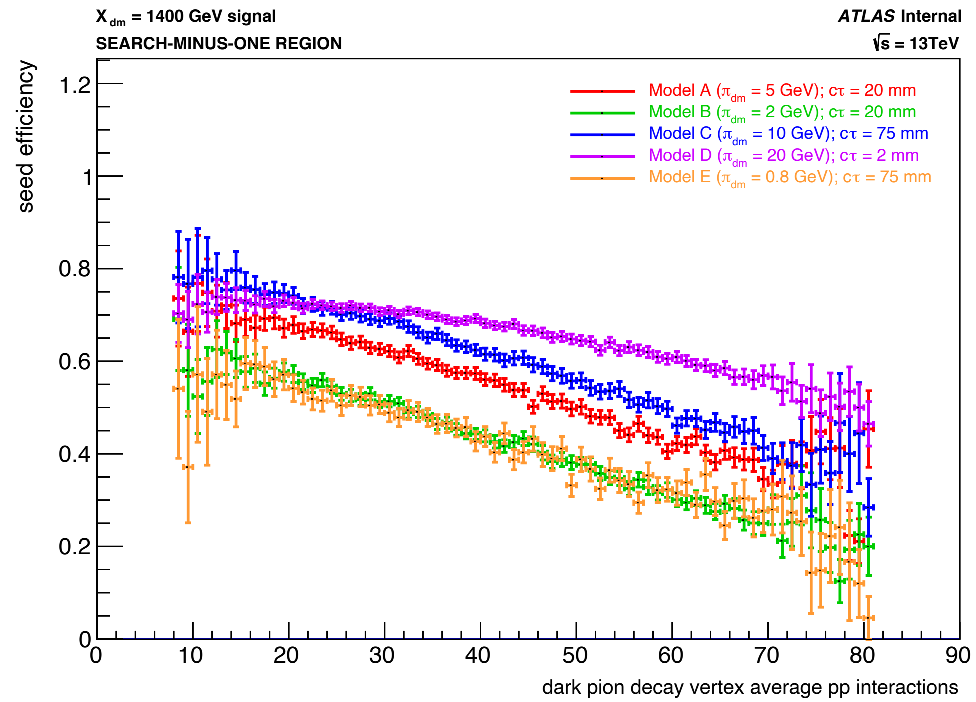


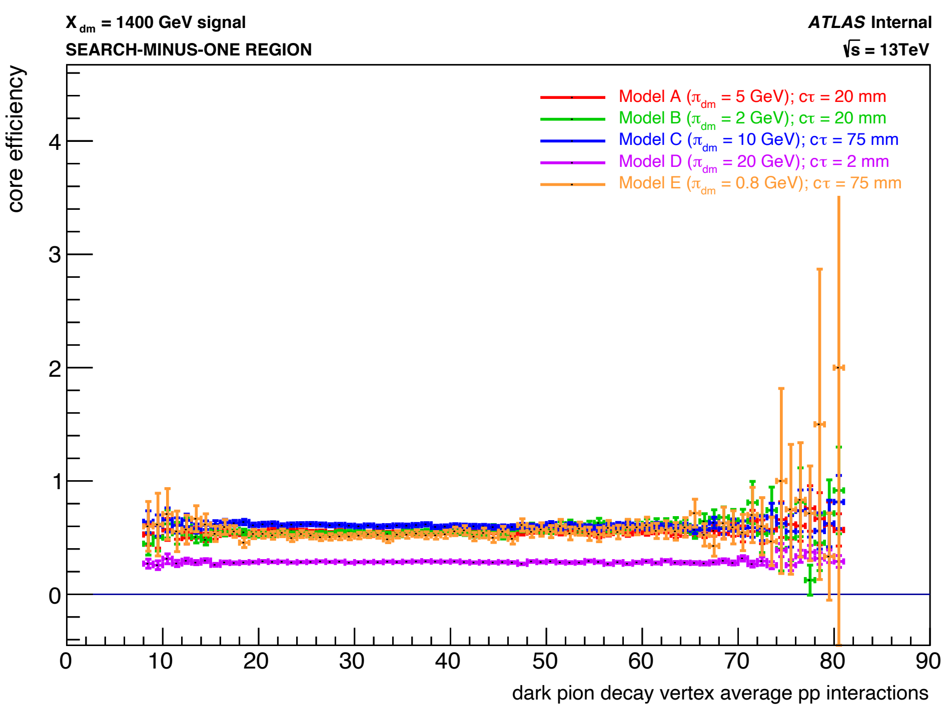
$$\epsilon^{alg} = \frac{\text{N signal truth LLP decays matched to reconstructed DVs}}{\text{N signal truth LLP decays with} \geq 2 \text{ reconstructed tracks}}$$

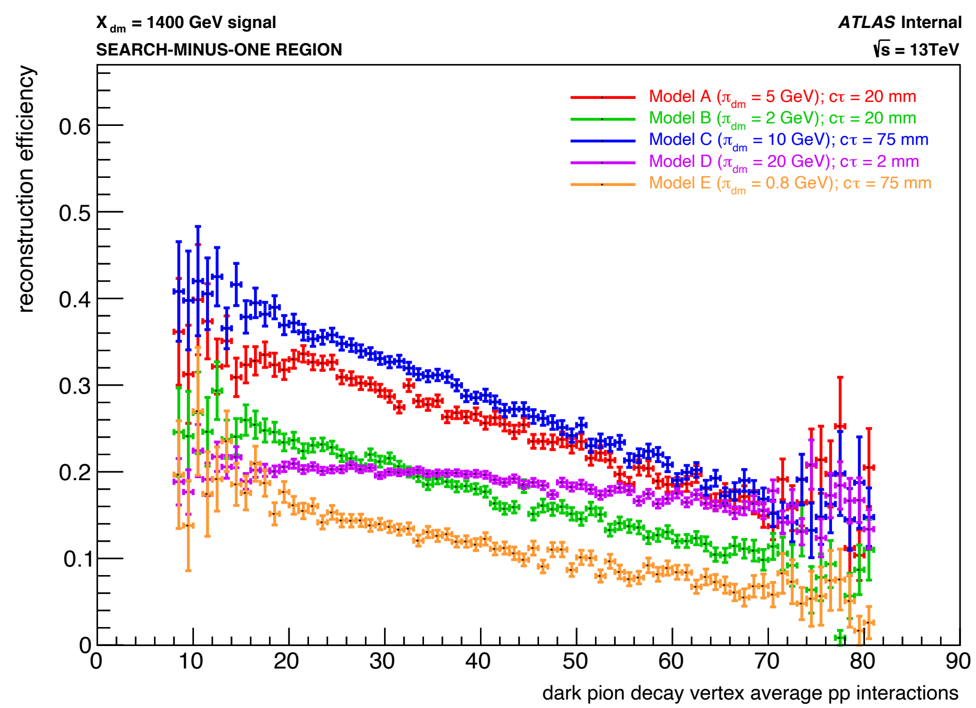
- Algorithmic efficiency = measure of vertexing algorithm performance decoupled from tracking reconstruction efficiency
  - Directly dependent on vertex mass and selected track multiplicity → primary cause of signal efficiency loss in EJs signal

DV Efficiency vs Pileup









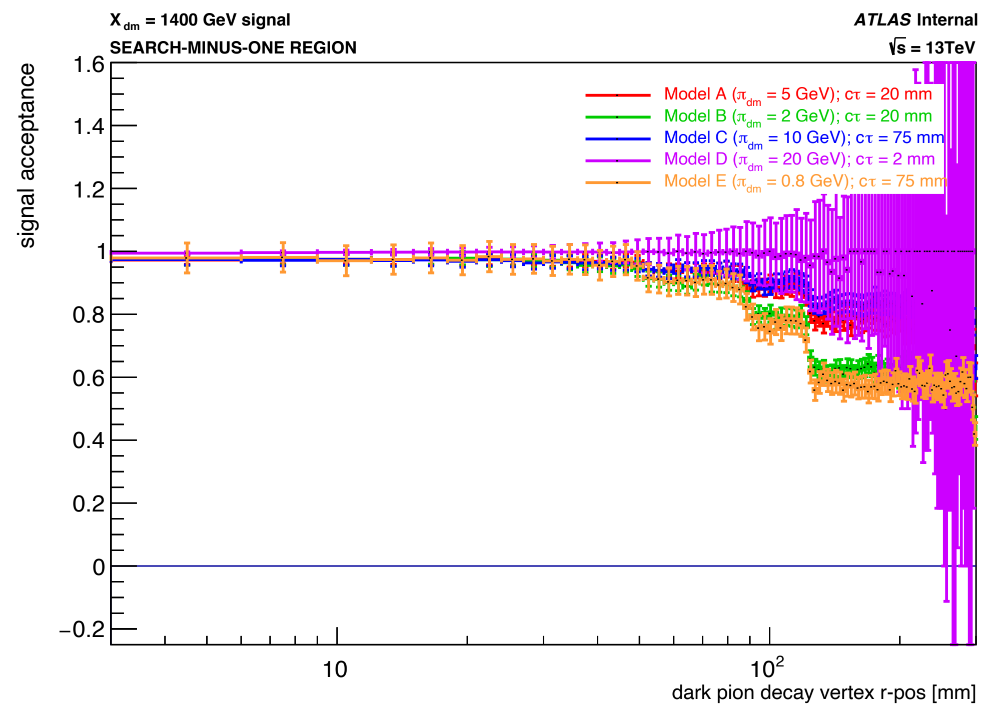
- Acceptance and seed efficiency degrade with increased <μ> → consequence of LRT inefficiency at high pileup
- Acceptance loss minimal → ~10-20% for dark pion decays with lifetime $c\tau_{\pi d} \geq 5$ mm
- Seed efficiency loss more significant, as it depends on efficiently reconstructed selected tracks → ~25-40% for all dark pion lifetimes

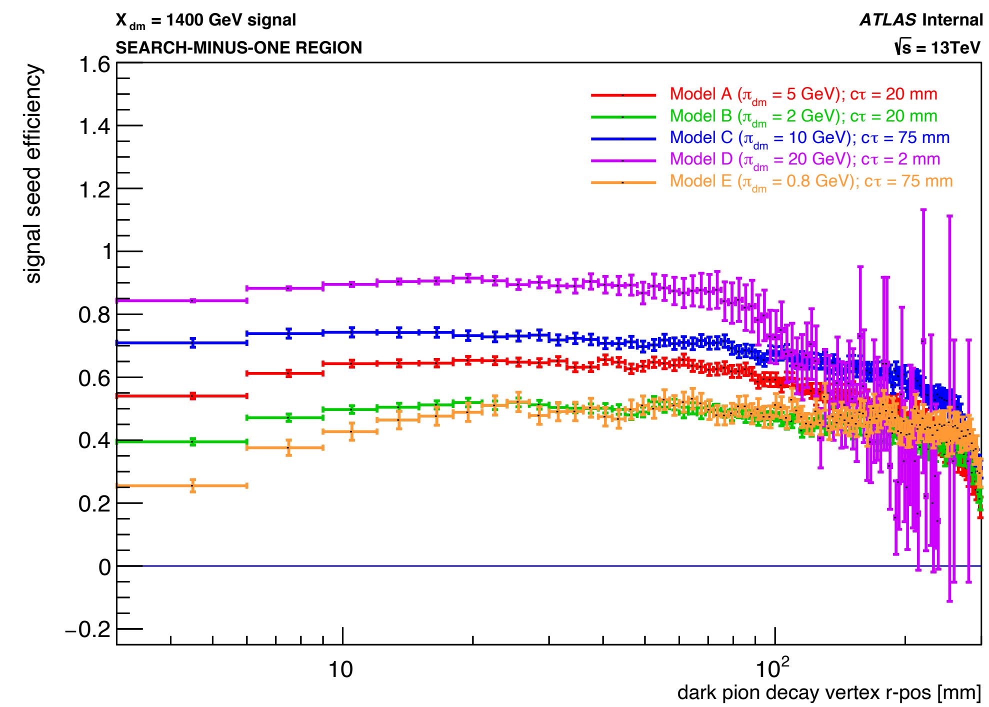
DV Signal Efficiency

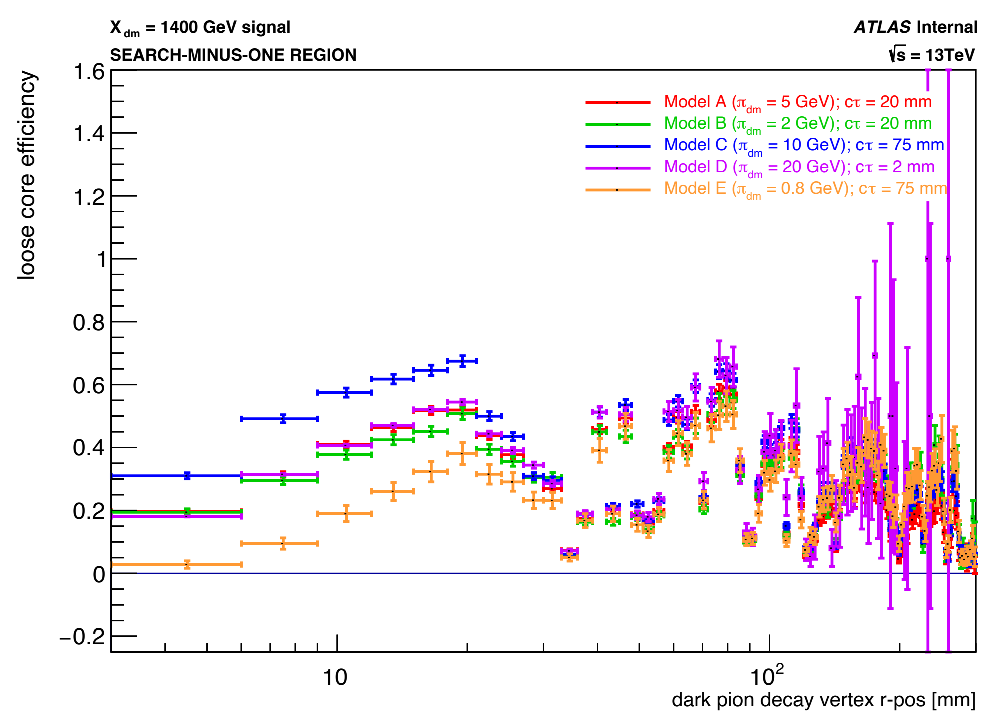
$$\text{signal reconstruction efficiency} = \frac{\text{N fiducial signal truth LLP decays matched to reconstructed signal DVs}}{\text{N fiducial signal truth LLP decays}}$$

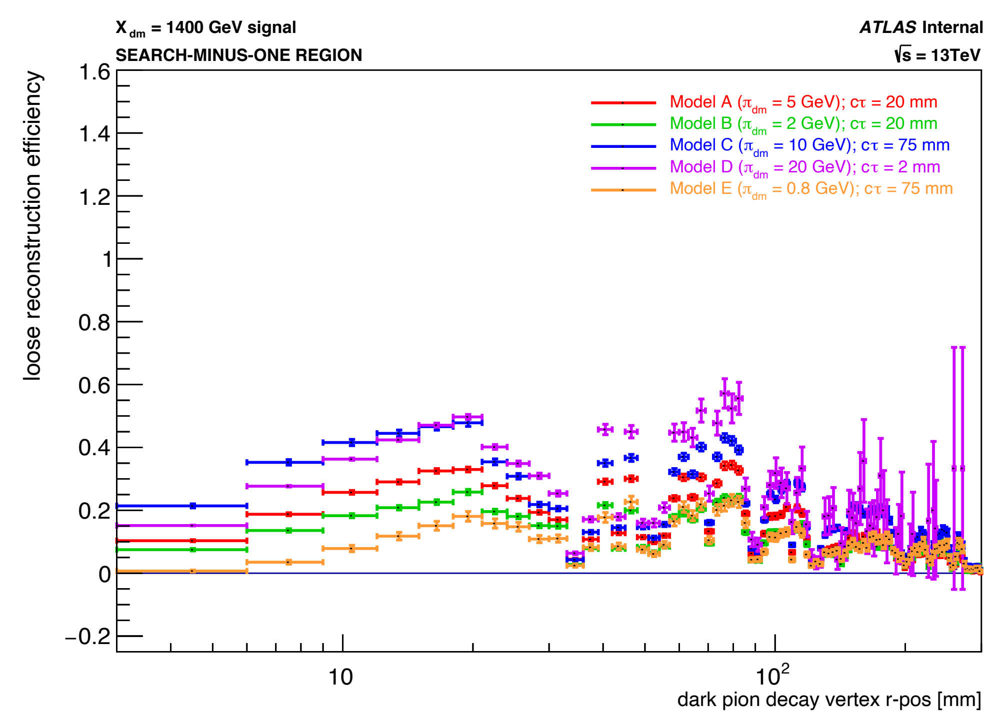
- Truth vertices required to be in fiducial volume

- Matched reconstructed vertices required to pass signal DV WP



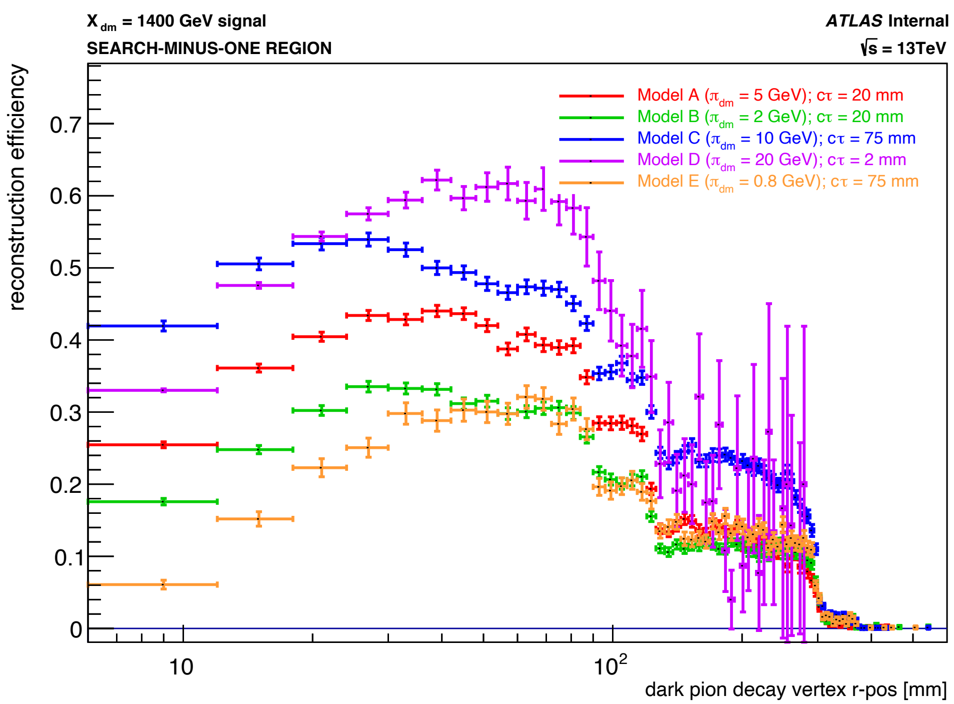


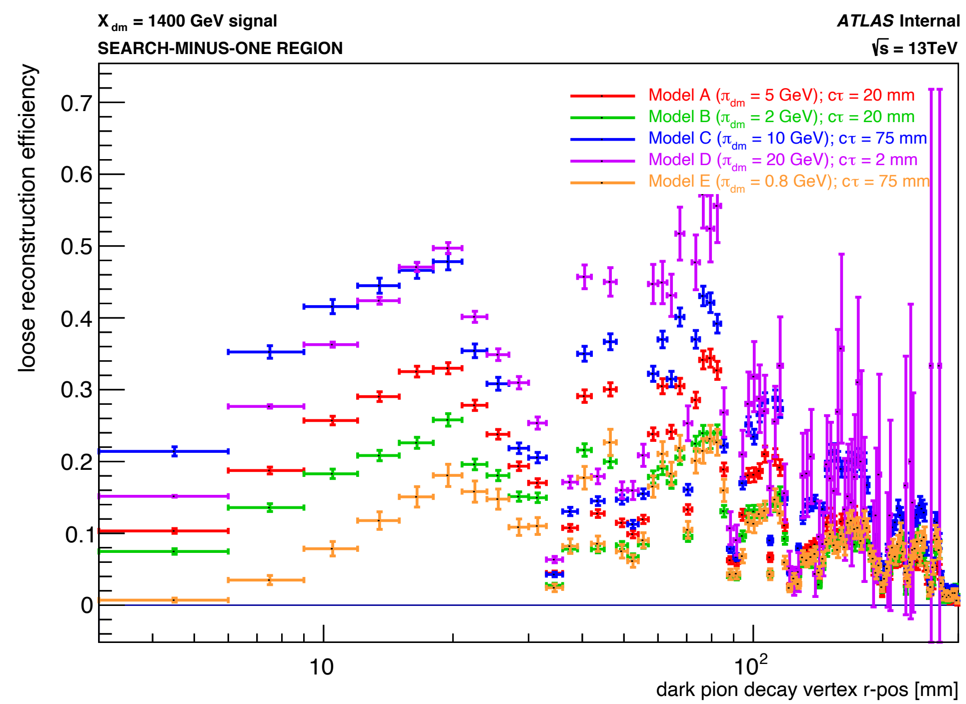


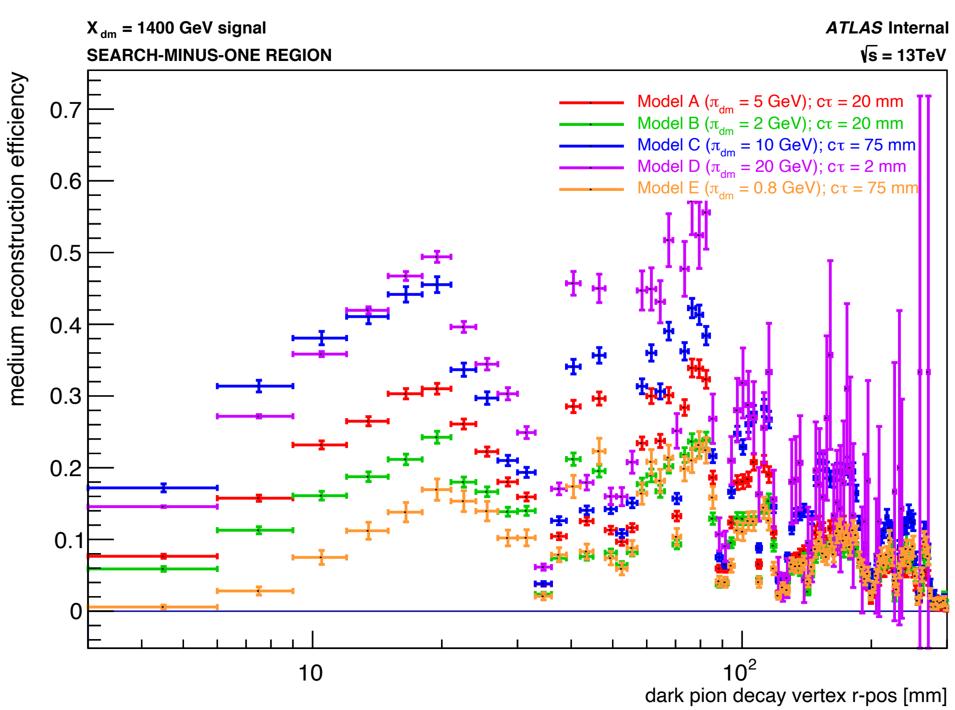


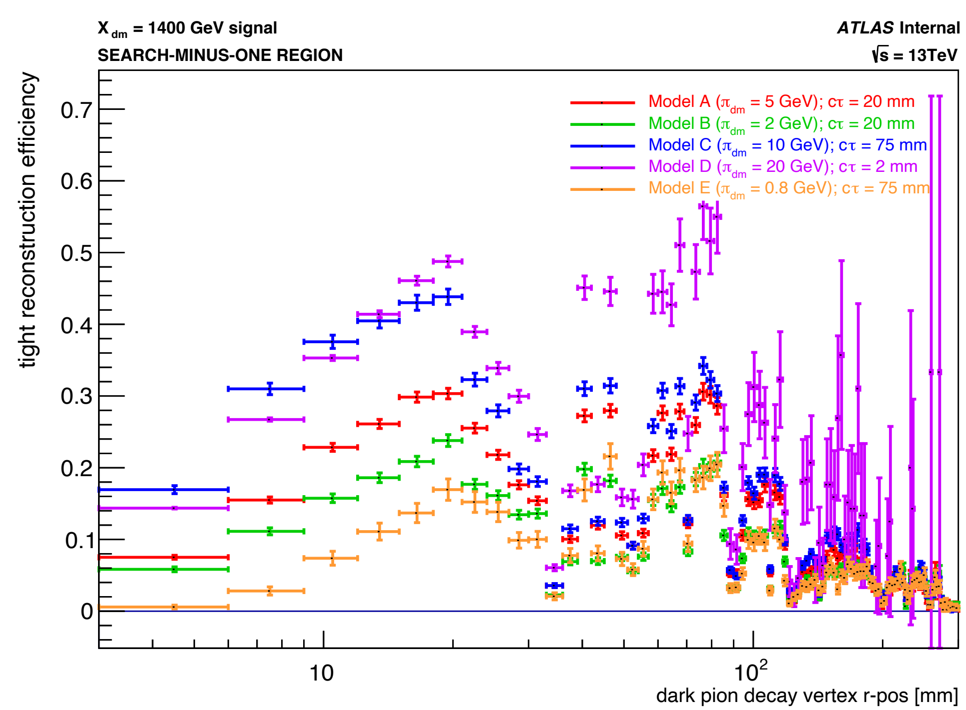
DV Efficiency: Inclusive vs Signal

- Loose DV selection results in ~10% loss of total efficiency compared to inclusive vertex reconstruction → Medium + Tight DV selections decrease efficiency buy only a few more percent









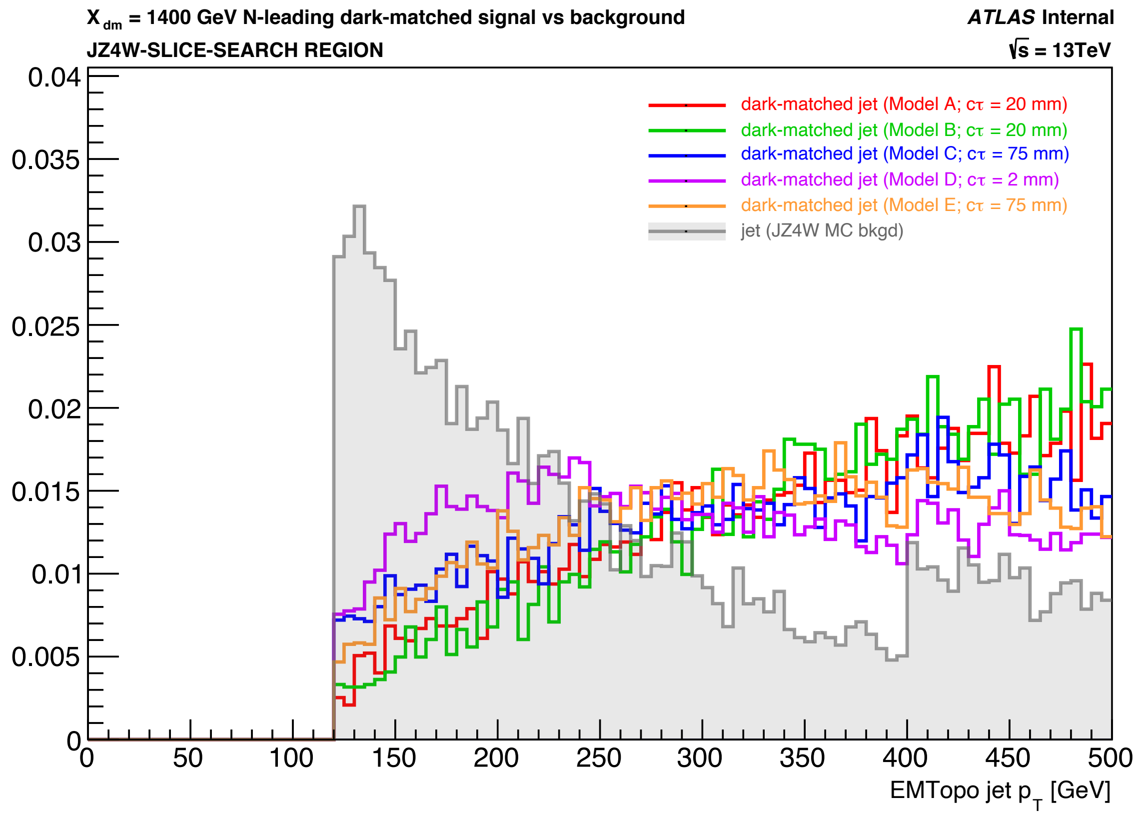
EJ Selection

- Signal events required to contain some number of emerging-like jets satisfying set of criteria designed to discriminate from SM QCD jets: **$N_{EJ} \geq$ 0, 1, or 2**
  - N_{EJ} = 0 would maintain model independence of search and increased efficiency of overall signal, but nonzero NEJ cut would further remove large QCD background, which proves challenging to reduce

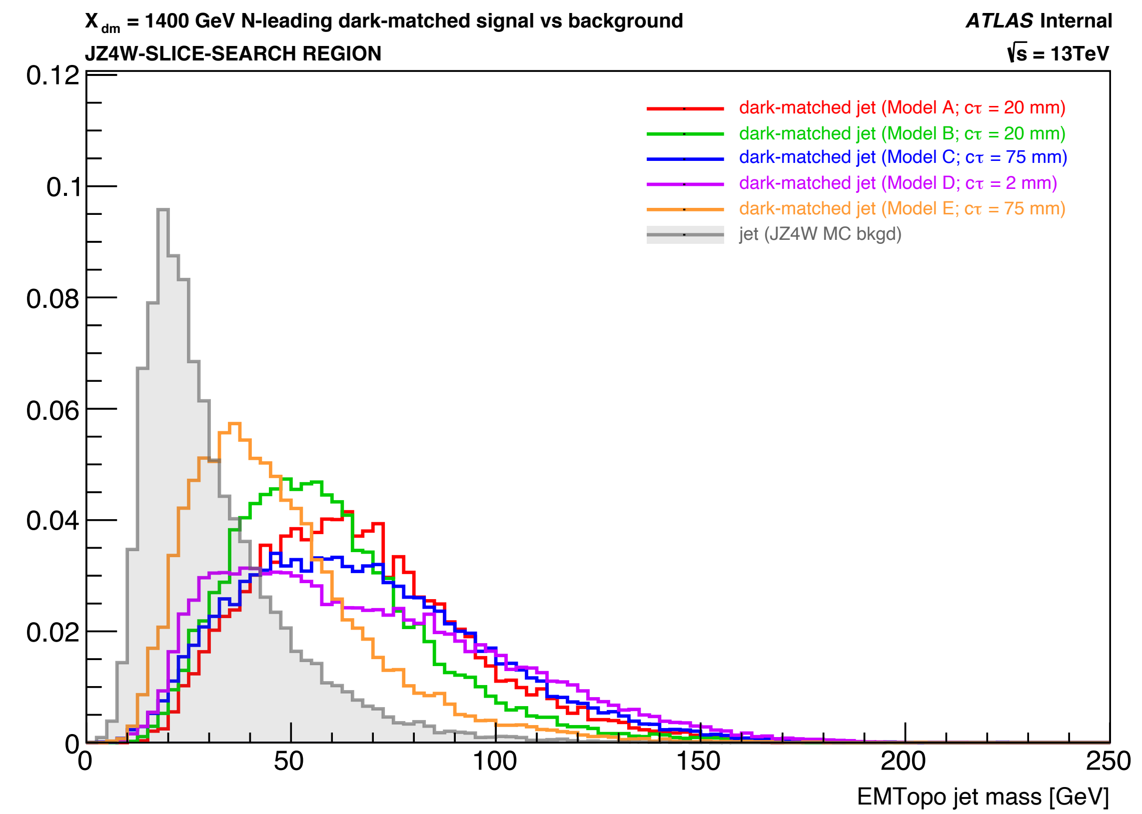
- Two tentative WPs defined to optimally select signal EJs: **Loose** + **Tight**
  - Selections designed to target signal jets originating from dark pion decays within dark jets and separate them from QCD jets from SM quarks in order to maximize background rejection with minimal signal loss
  - Jets matched to truth dark jets in signal MC are compared to unmatched jets in background MC to identify discriminating variables and investigate effects of EJ WP selections on overall signal and background
  - Jet matched to truth dark jet if within **ΔR < 0.3** → dark jet quality criteria: pT > 30 GeV, ΔR-matched to same truth jet matched to corresponding reco jet, and lead ghost parton = down quark

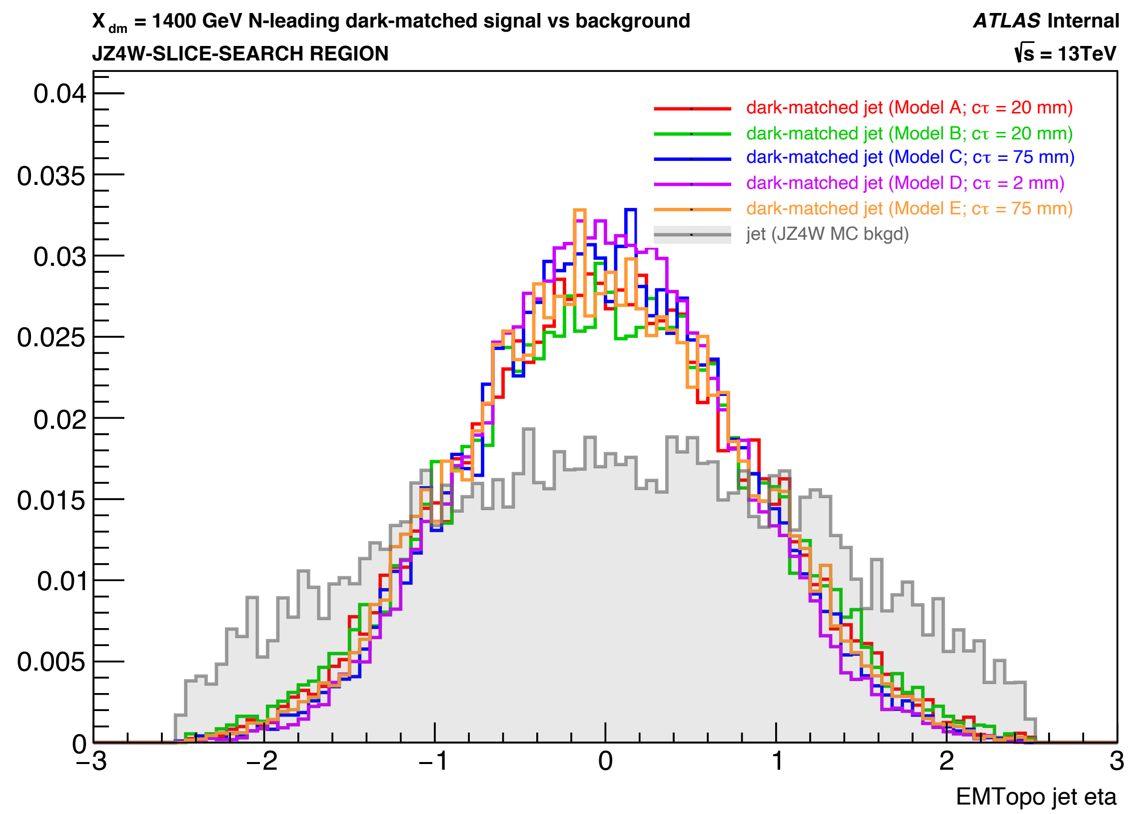
Loose
$p_{\mathrm{T}} > 200$ GeV
$\lvert\eta\rvert < 2.0$
$m > 25$ GeV
$p_{\mathrm{T}}^{N_{\mathrm{SV}}} > 5$ GeV
$N_{\mathrm{jet-trk}}^{N_{\mathrm{SV}}} > 2$
Tight
Loose selections
$p_{\mathrm{T}}^{N_{\mathrm{SV}}} > 7.5$ GeV
$N_{\mathrm{jet-trk}}^{N_{\mathrm{SV}}} > 4$

Hard Jet Selections



- Emerging-like signal jets observed to be somewhat harder and more central than SM background jets
- → EJs required to have
  - **p_T > 200 GeV**
  - **$|\eta|$ < 2.0**
  - **m > 25 GeV**



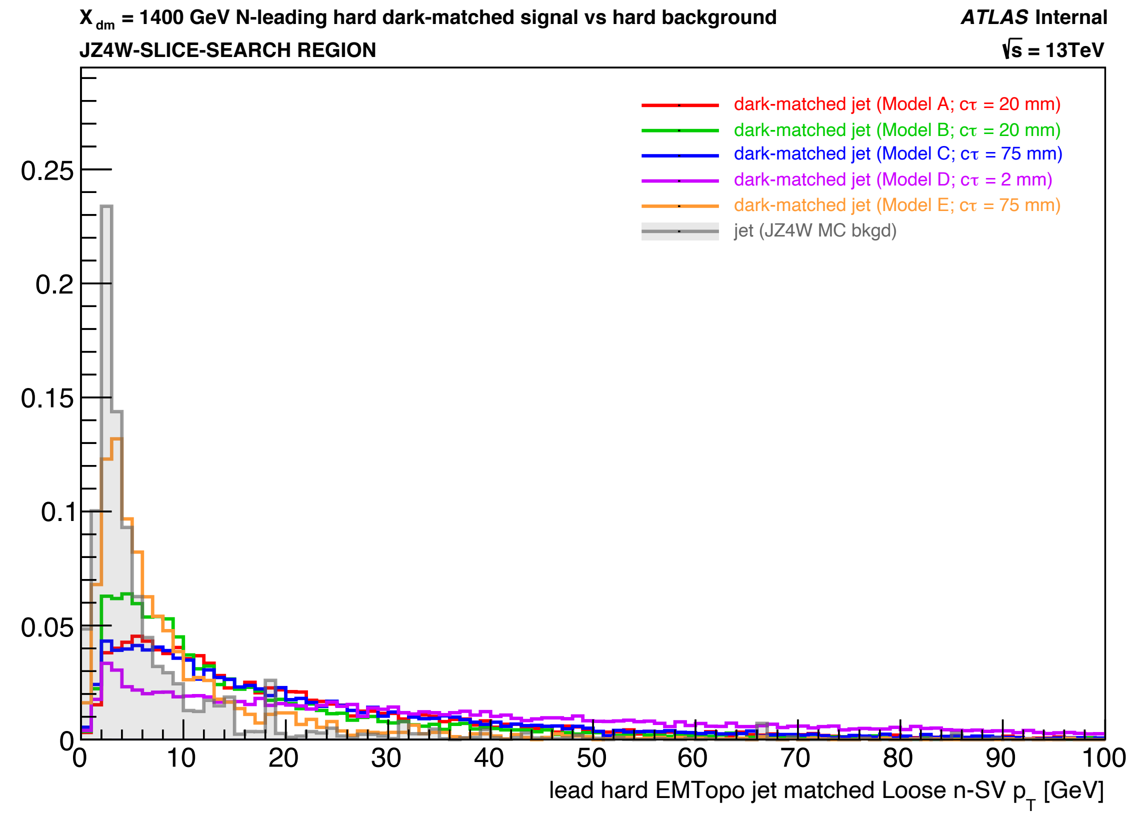


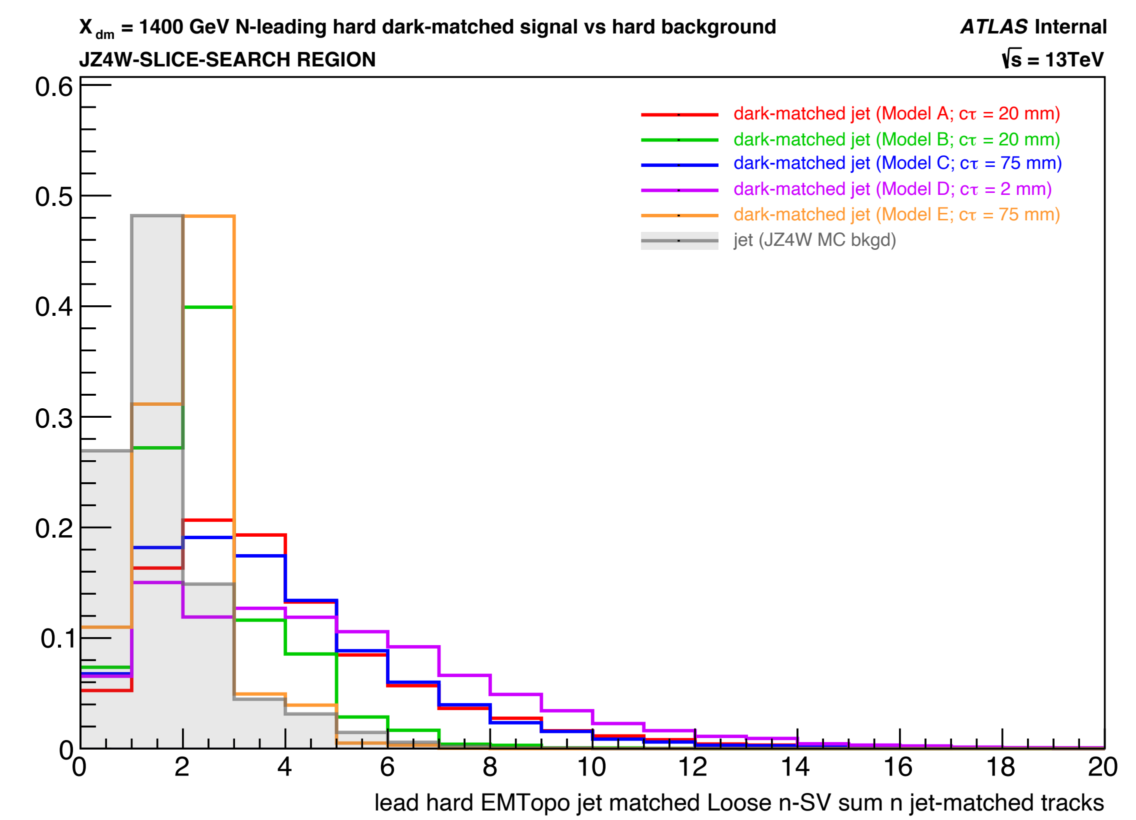
SV Jet Selections

- EJs expected to contain some number associated secondary vertices, corresponding to multiple dark pion decays within jet cone, characteristic of EJs topology

- EJ cut on associated SV multiplicity redundant on top of event-level NDV cut → instead apply alternate cuts that target jets with inherently larger SV multiplicities while providing more flexibility in selections…

- → EJs required to have
  - $\mathbf{p_T^{NSV}}$ = four-momentum-pT of matched SVs within $\Delta R < 0.6$
  - $\mathbf{N_{jet\text{-}trk}^{NSV}}$ = number of SV jet tracks, where SV jet tracks are those associated to matched SVs and independently matched to jet within $\Delta R <$ 0.6
  - → Loose EJs: $\mathbf{p_T^{NSV} > 5}$ **GeV** and $\mathbf{N_{jet\text{-}trk}^{NSV} > 2}$
  - → Tight EJs: $\mathbf{p_T^{NSV} > 7.5}$ **GeV** and $\mathbf{N_{jet\text{-}trk}^{NSV} > 4}$
  - → SV cuts applied separately for for each DV WP, defining Loose-SV, Medium-SV, and Tight-SV EJ selections





NJetX Selection

- Signal events required to pass some leading **NJetX** observable threshold to further reduce QCD background
- Number of potential NJetX variables related to **leading four-jet system** or **dijet system** under consideration, with range of cut values proposed for each → chosen to target signal-like jet systems originating from pair-produced scalar mediators and distinguish them from background QCD multijet events
  - Distributions compared between signal and background MC samples to determine observables with large discrimination and selection thresholds maximizing background rejection with minimal effect on signal

- Leading four-jet system:
  - **H_T^{NJet} > 1200 - 1500 GeV**
  - **p_T^{NJet} > 100 - 200 GeV**
  - **m_{sum}^{NJet} > 120 - 175 GeV**
- Nominal dijet system = average of two dijet pairs among four leading jets with smallest difference in invariant mass between them
  - **p_T^{NJetJJ} > 350 - 700 GeV**
- Maximum-dR dijet system = average of two dijet pairs among four leading jets with largest angular distance between them
  - **$p_T^{NJetJJ,max\text{-}dR}$ > 300 - 600 GeV**
- Average dijet system = average over all possible jet pairs among four leading jets
  - **$p_T^{NJetJJ,avg}$ > 350 - 525 GeV**
- Maximum / minimum dijet system = dijet system with maximum / minimum invariant mass among four leading jets
  - **$H_T^{NJetJJ,max}$ > 750 - 1200 GeV**
  - **$m_{sum}^{NJetJJ,max}$ > 60 - 125 GeV**
  - **$H_T^{NJetJJ,min}$ > 500 - 800 GeV**
  - **$m_{sum}^{NJetJJ,min}$ > 50 - 100 GeV**
- Maximum-pT dijet system = dijet system with maximum pT among four leading jets
  - **$p_T^{NJetJJ,max\text{-}pT}$ > 500 - 900 GeV**
  - **$H_T^{NJetJJ,max\text{-}pT}$ > 600 - 1000 GeV**
  - **$m_{sum}^{NJetJJ,max\text{-}pT}$ > 50 - 100 GeV**

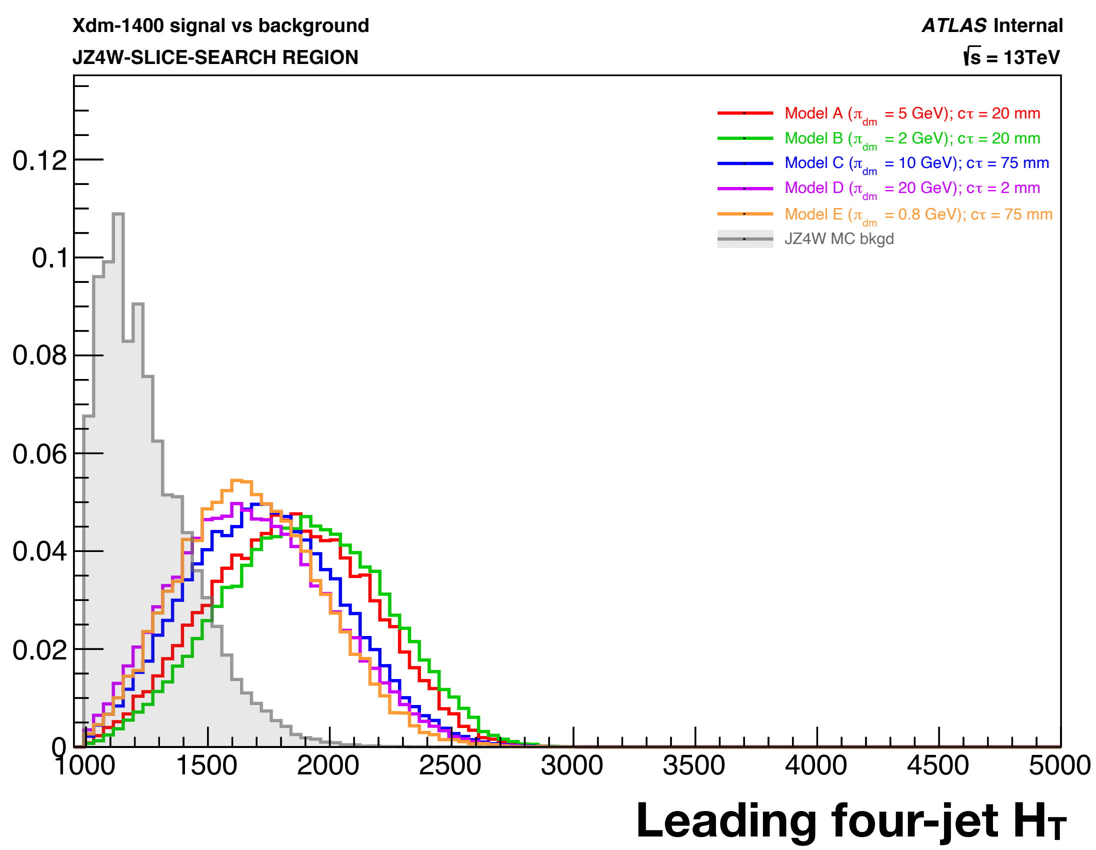
NJetX Selection

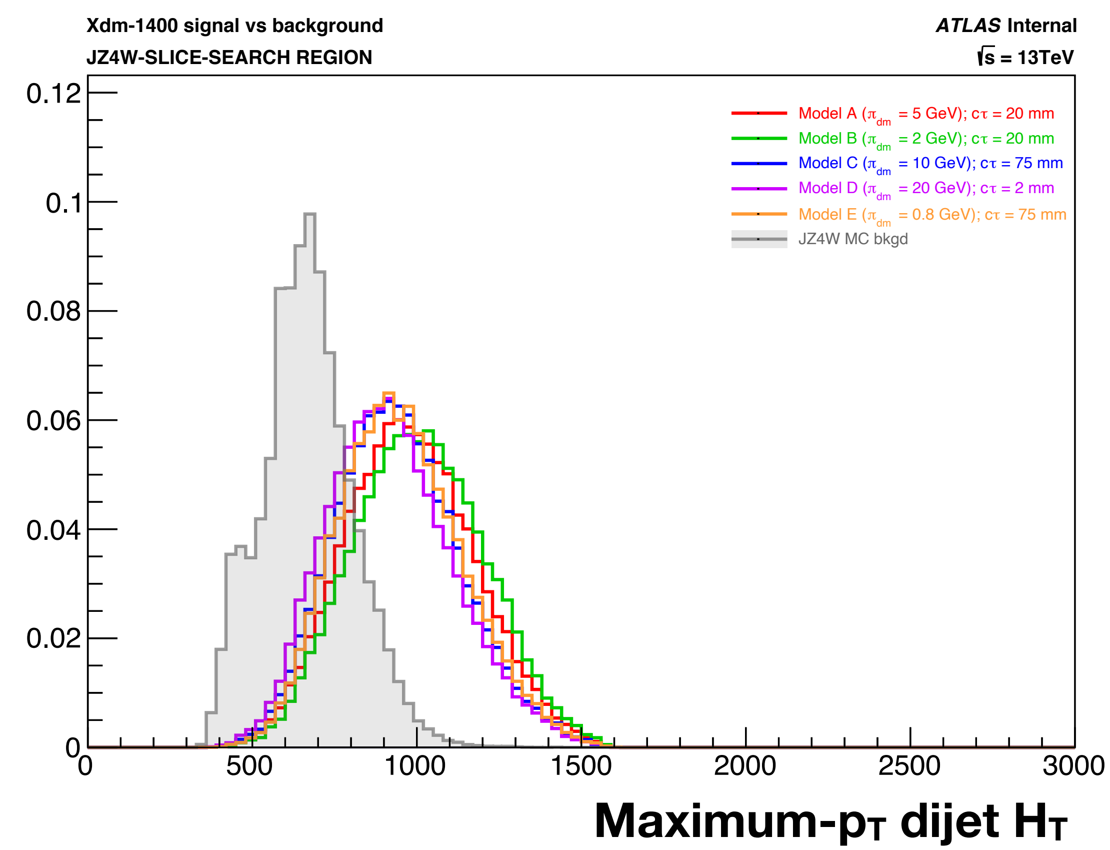
- Potential NJetX variables and cut values related to **leading four-jet system** or **dijet system** chosen to target signal-like jet systems originating from pair-produced scalar mediators and distinguish them from background QCD multijet events

- NJetX variable distributions peak at higher values and differ widely from QCD background in signal samples with large mediator masses, but discrimination in NJetX can be very small for lower mediator masses, making it difficult to remove background without heavily impacting signal

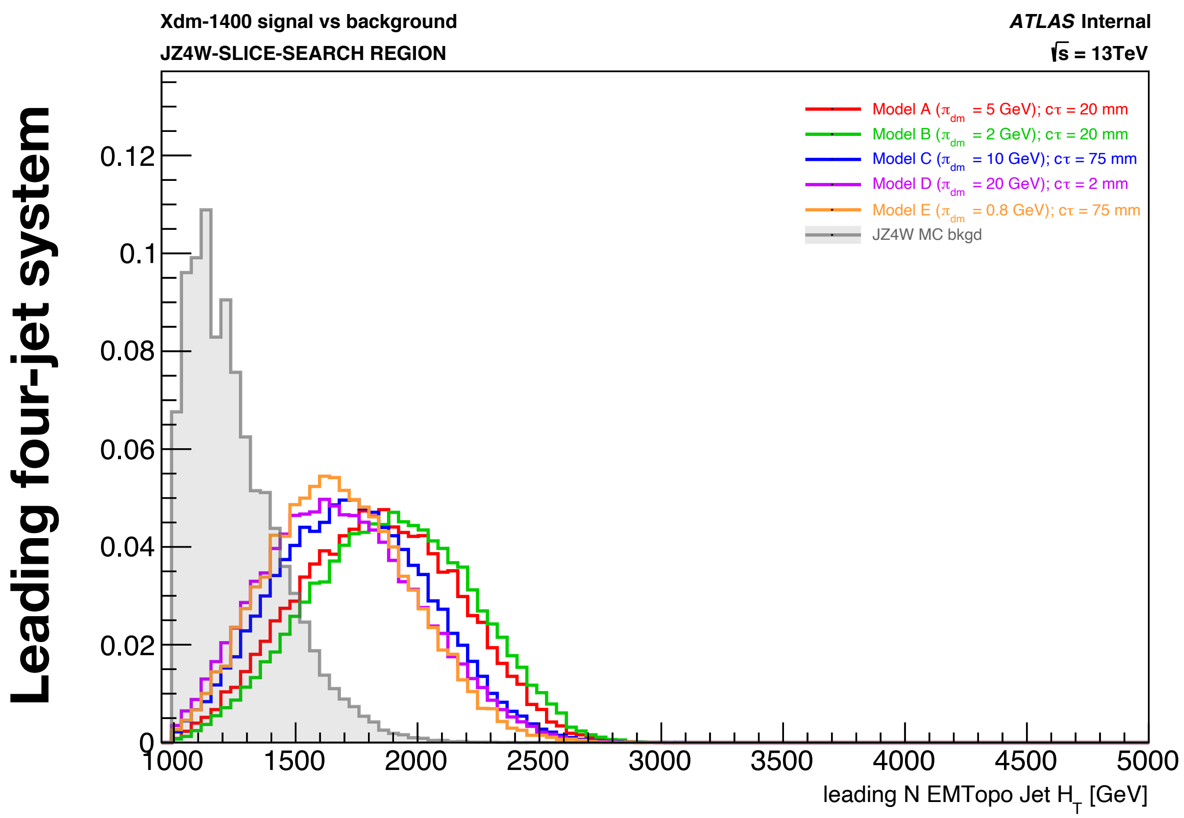
- → used in background estimation since independent of NDV → still to compare all potential selections against one another to identify most promising means of reducing background without removing too much signal…

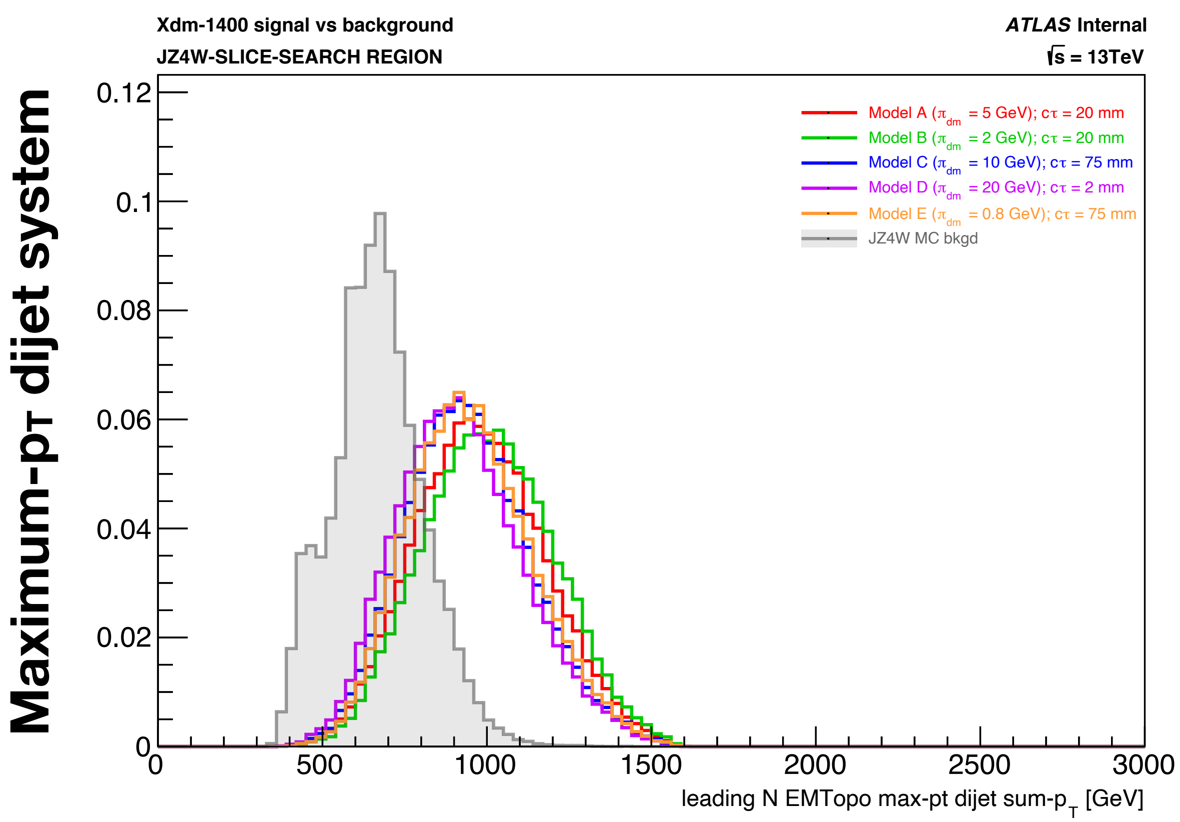
> - Leading four-jet system H_T: **H_T^{NJet} > 1200 - 1500 GeV**
> - Maximum-pT dijet system H_T: **$H_T^{NJetJJ,max\text{-}pT}$ > 600 - 1000 GeV**





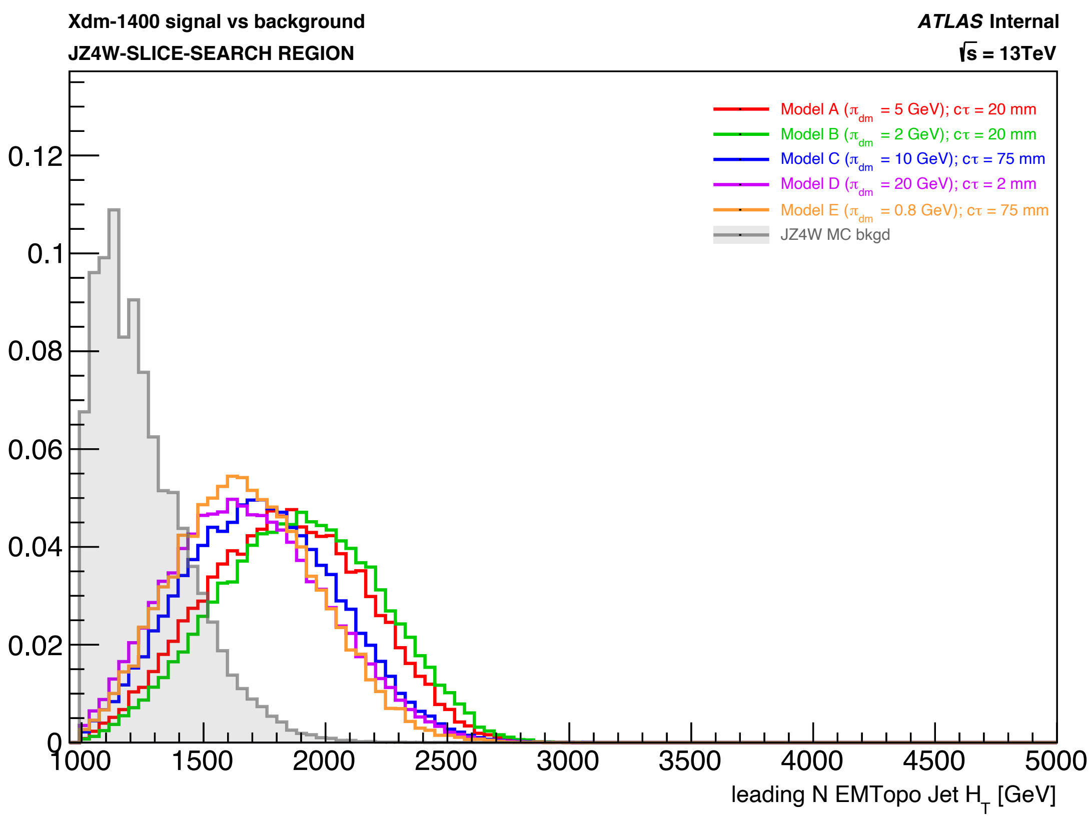
NJet Scalar Sum-p_T



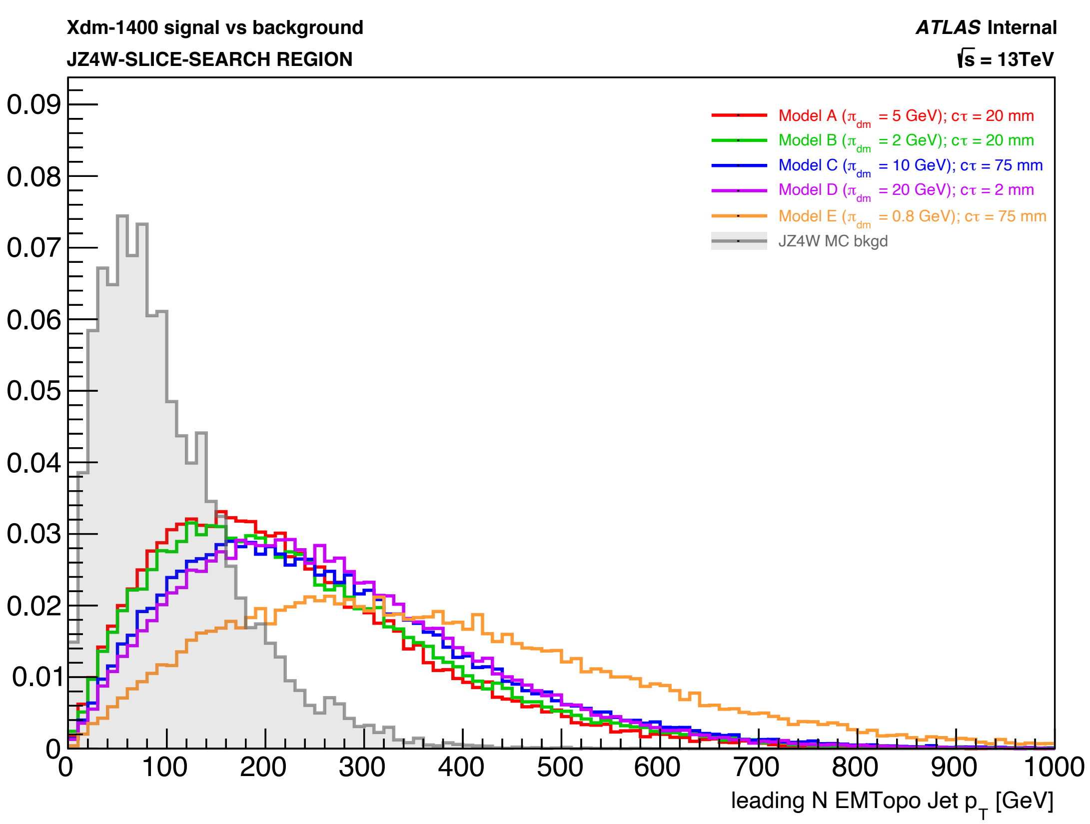


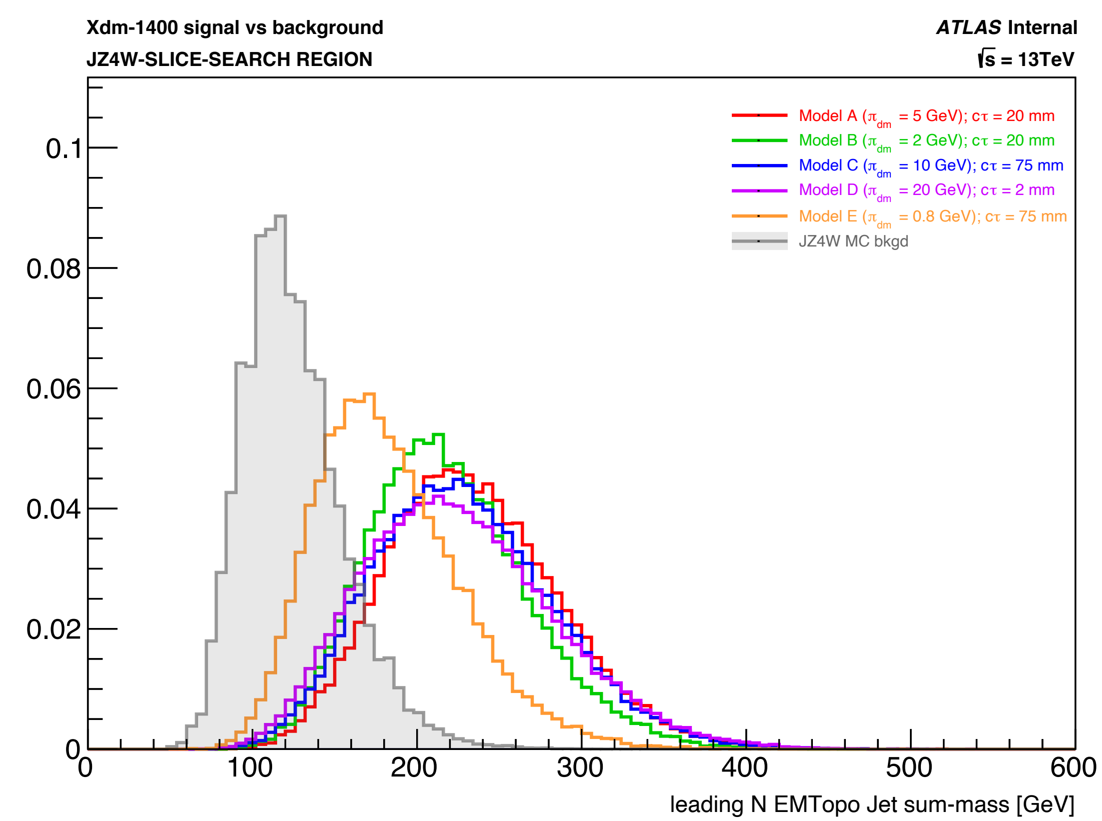
- NJetX variable distributions peak at higher values and differ more widely from QCD background in signal samples with large mediator masses
- → discrimination in NJetX can be very small for lower mediator masses, making it difficult to remove background without heavily impacting signal
- Dijet systems can provide similar discrimination as (or potentially slightly more than) leading four-jet system NJetX observables
- → still to compare all potential selections against one another to identify most promising means of reducing background without removing too much signal…

Leading Four-Jet System

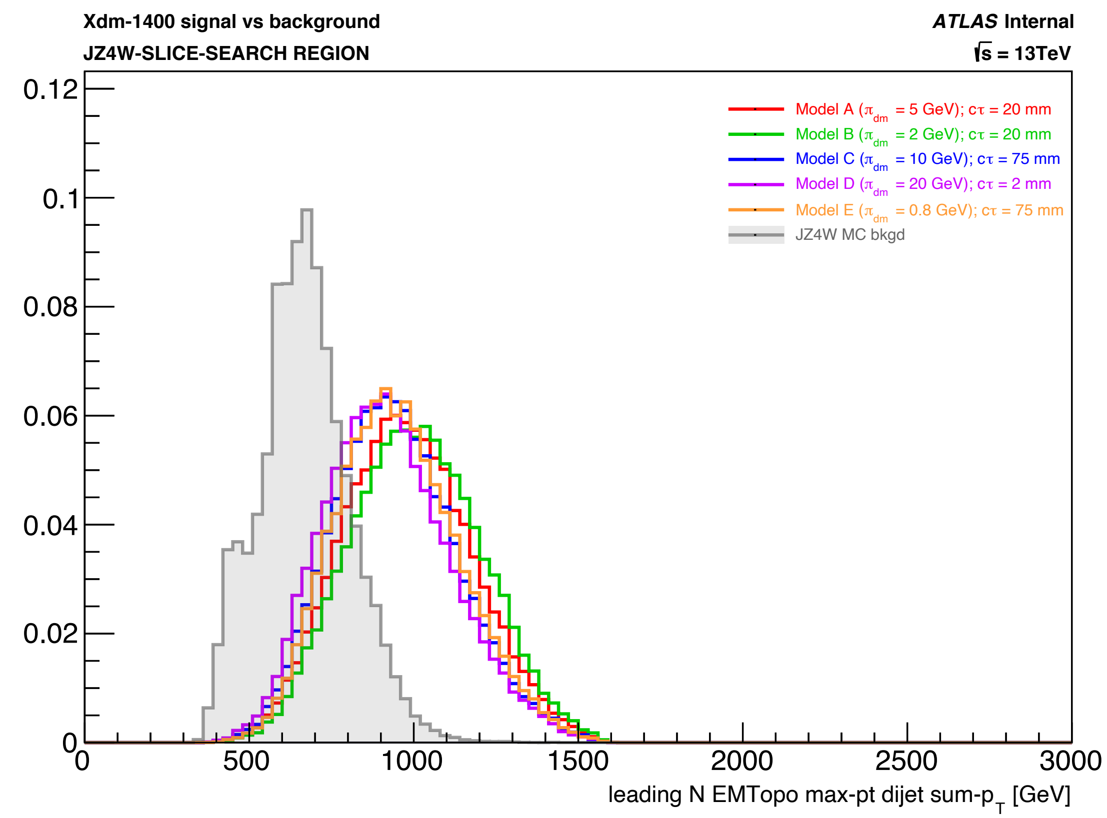


- NJetX variable distributions peak at higher values and differ more widely from QCD background in signal samples with large mediator masses
- → background discrimination in NJetX can be very small for lower mediator masses, making it difficult to remove background without heavily impacting signal

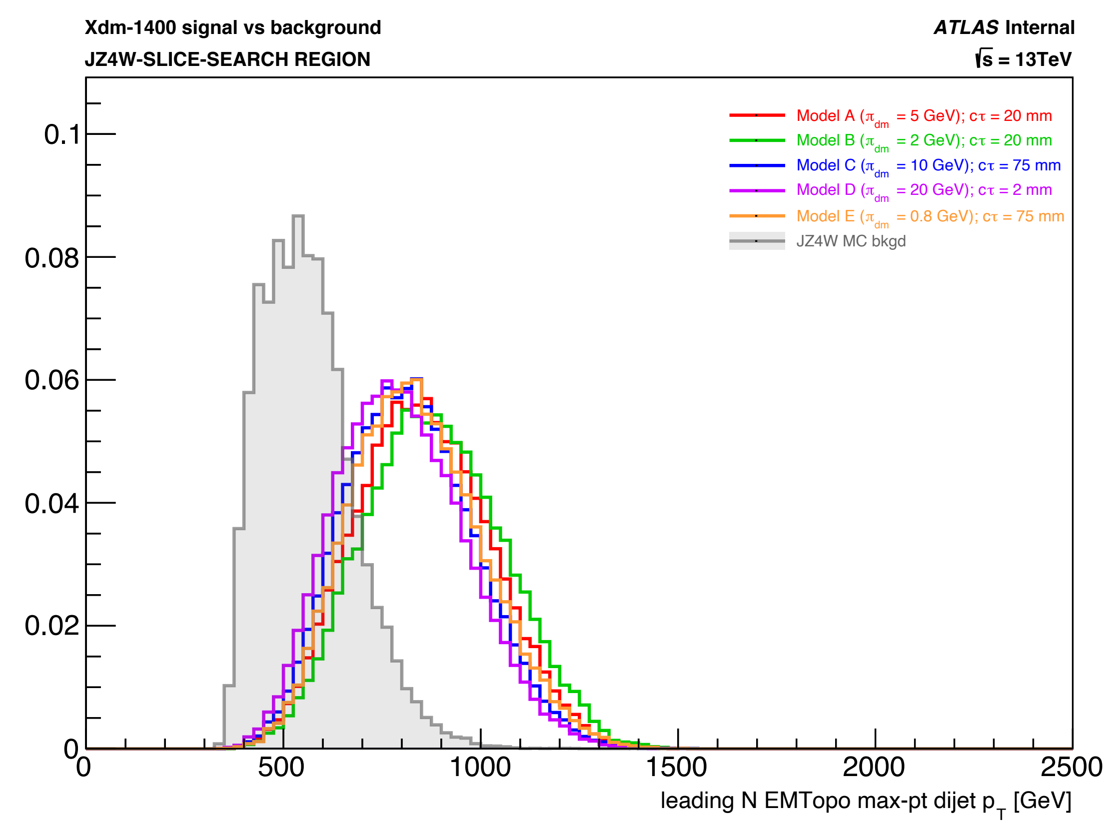


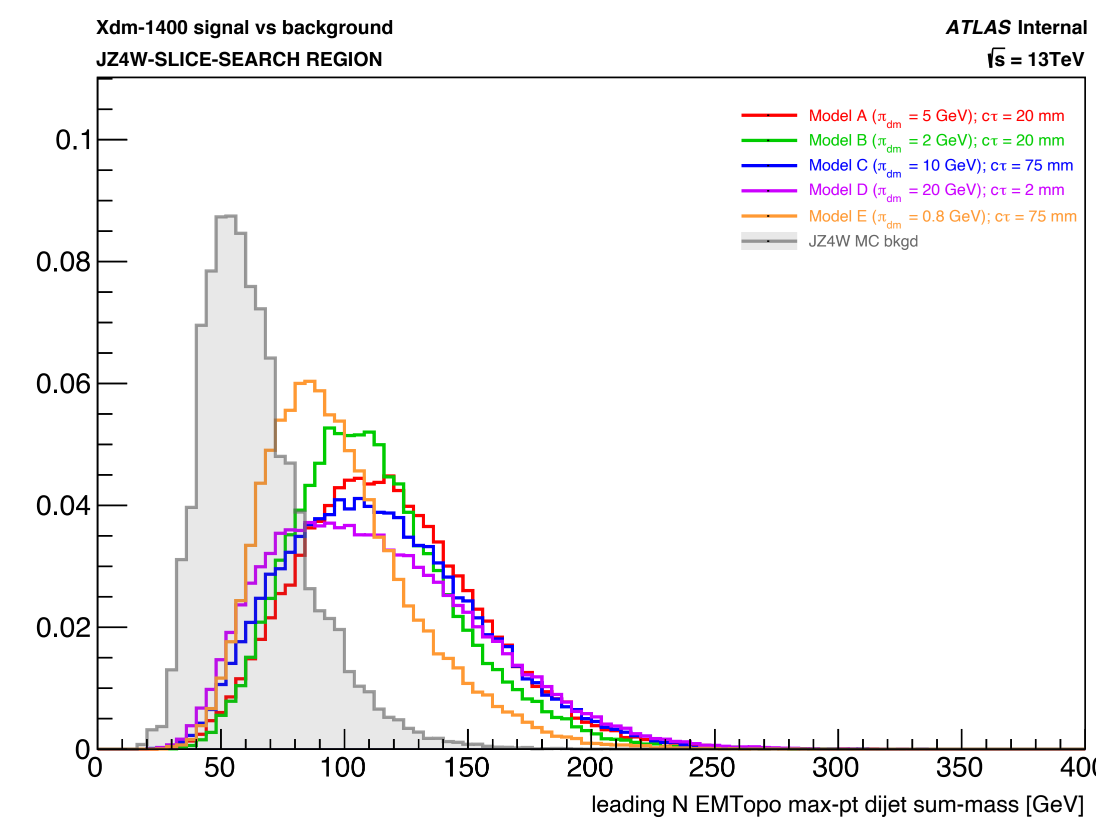


Maximum-p_T Dijet System

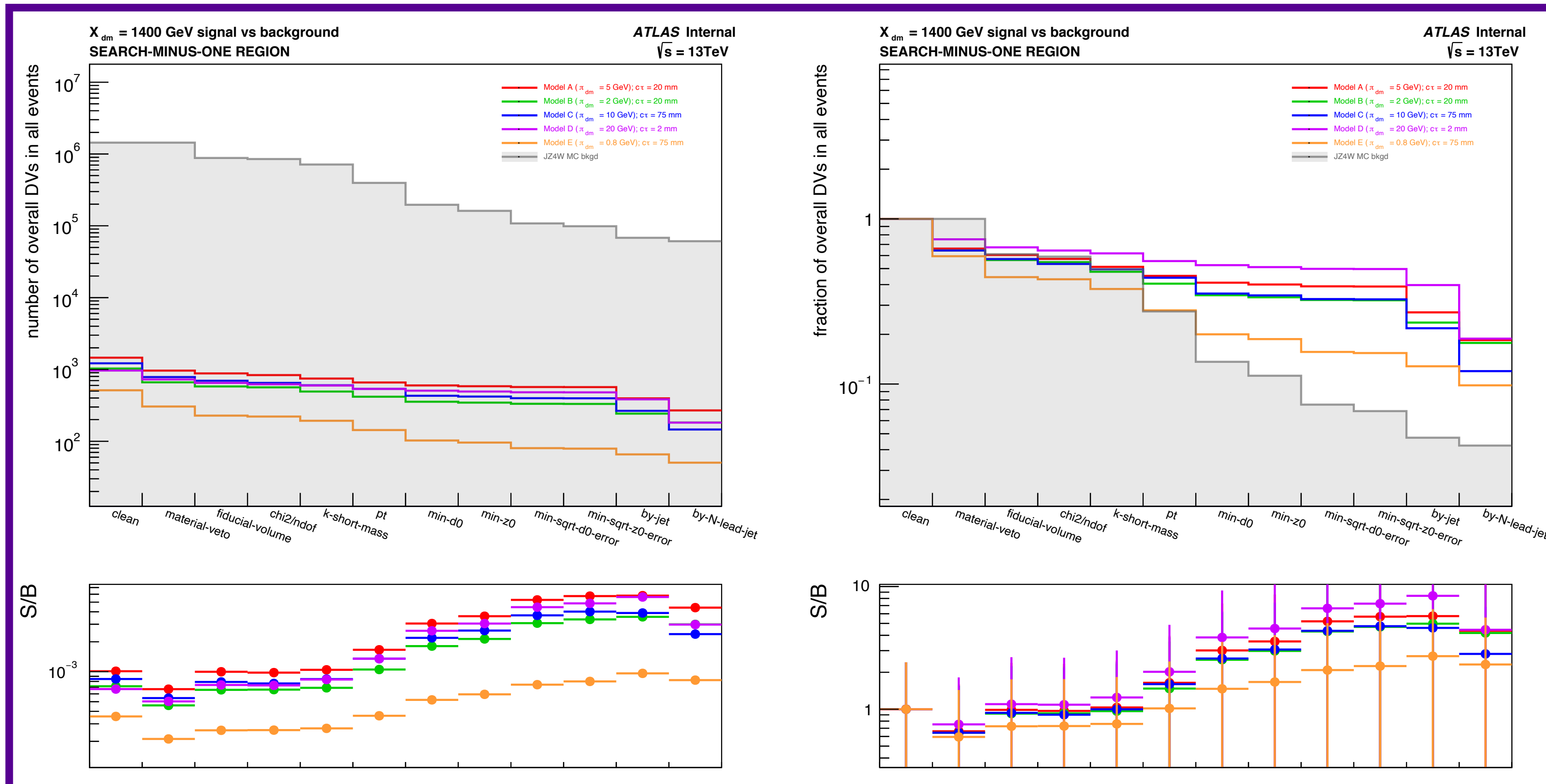


- Similar discrimination as (or potentially slightly more than) leading four-jet system NJetX observables
- → still to compare all potential selections against one another to identify most promising means of reducing background without removing too much signal…



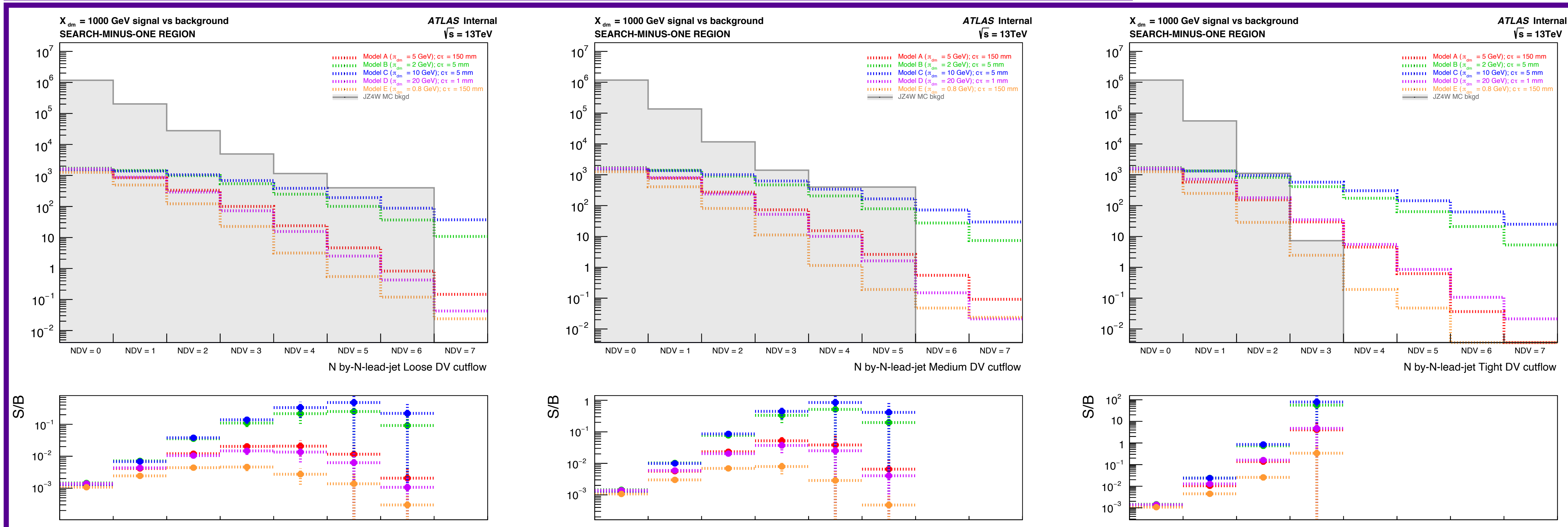


DV Cutflows



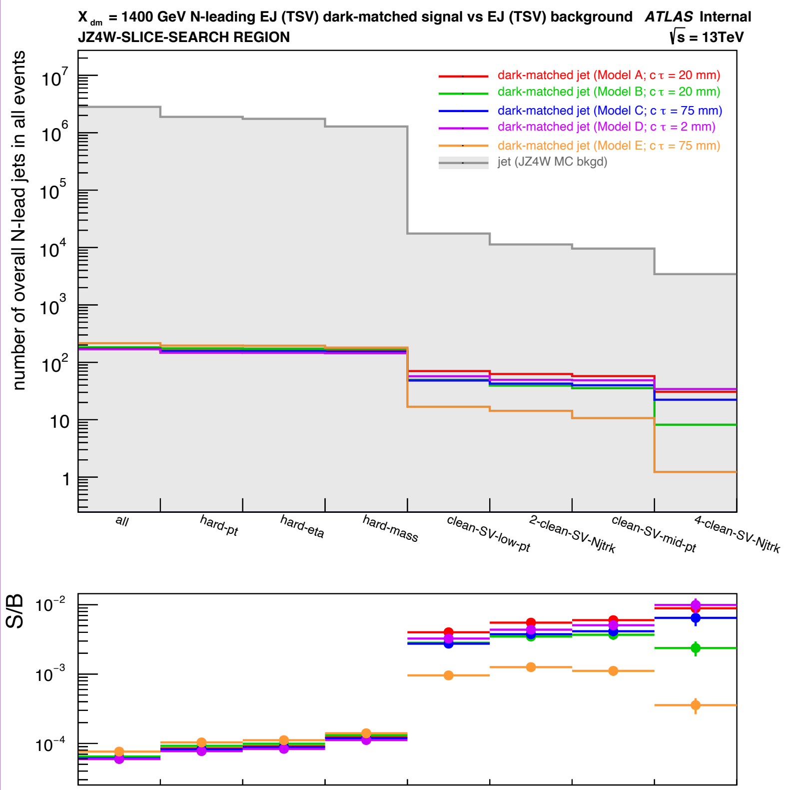
Number of fraction of total DVs in all events passing DV selections

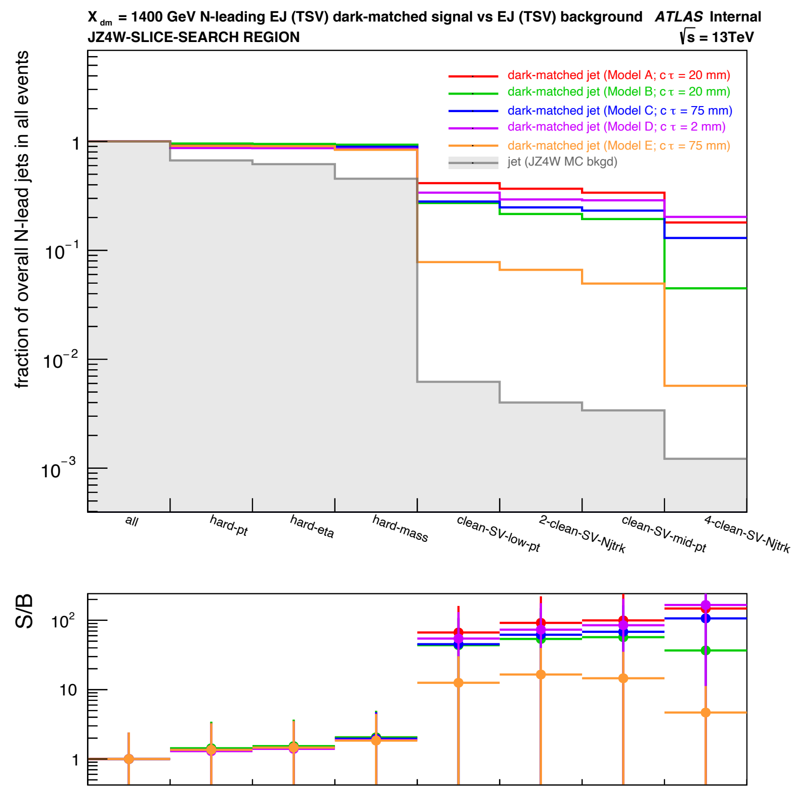
- DV cuts target background DVs more often than signal
- S/B increases with each cut but final leading jet association selection → can increase signal, at risk of lower background rejection, by loosening jet association requirement to any jet in event (instead of leading jets)
- MC background effectively goes to zero for **N_{DV} > 6, 5, and 3** with Loose, Medium, and Tight WPs
- Signal yield suffers in low-mass models and for extreme dark pion lifetimes, and signal efficiency decreases as DV WP tightens



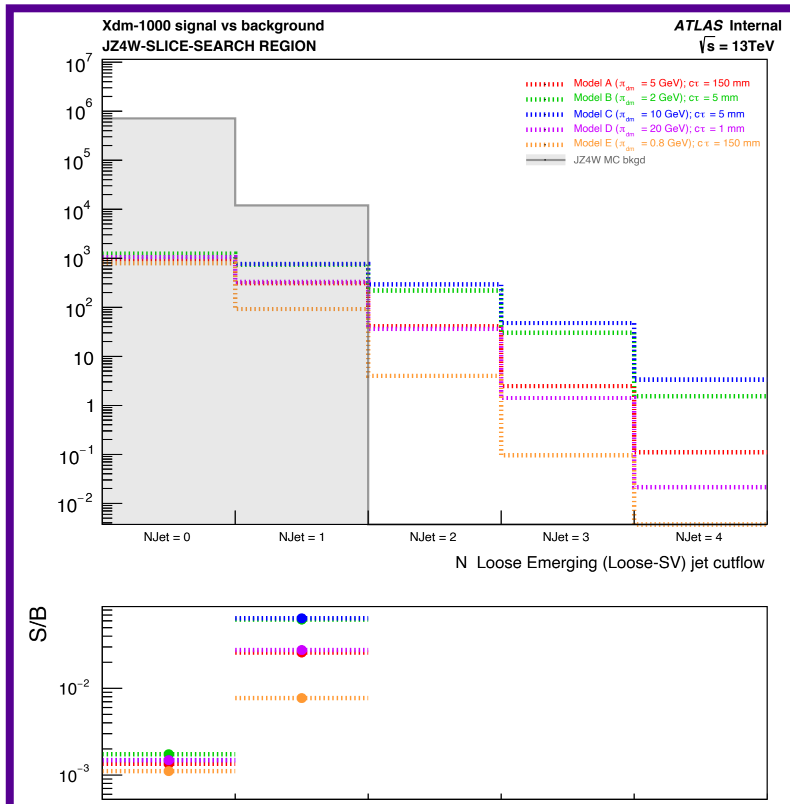
Number of events passing NDV cut

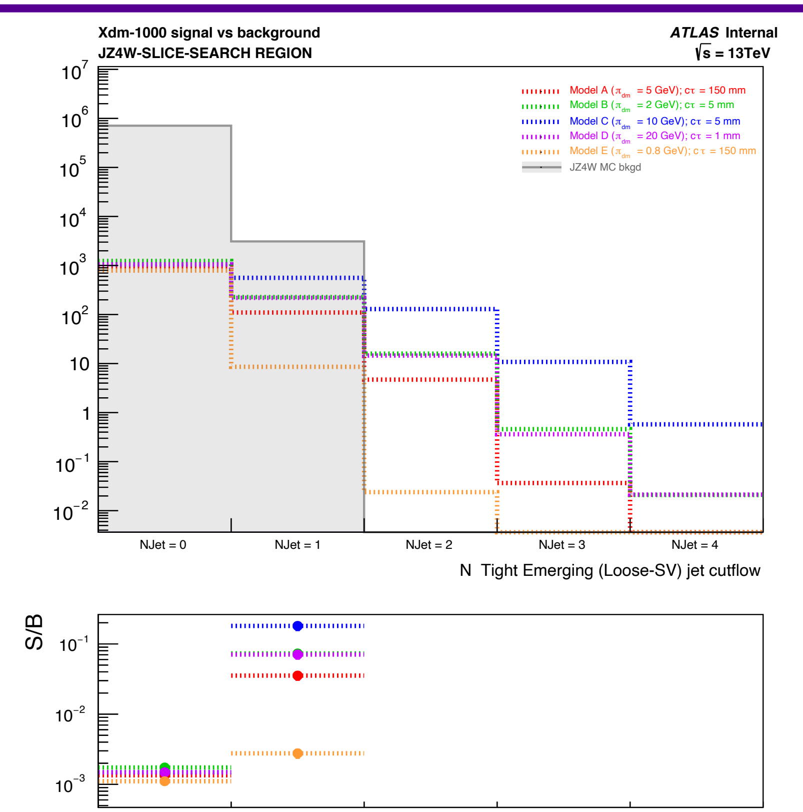
EJ Cutflows





Number of fraction of total jets in all events passing EJ selections



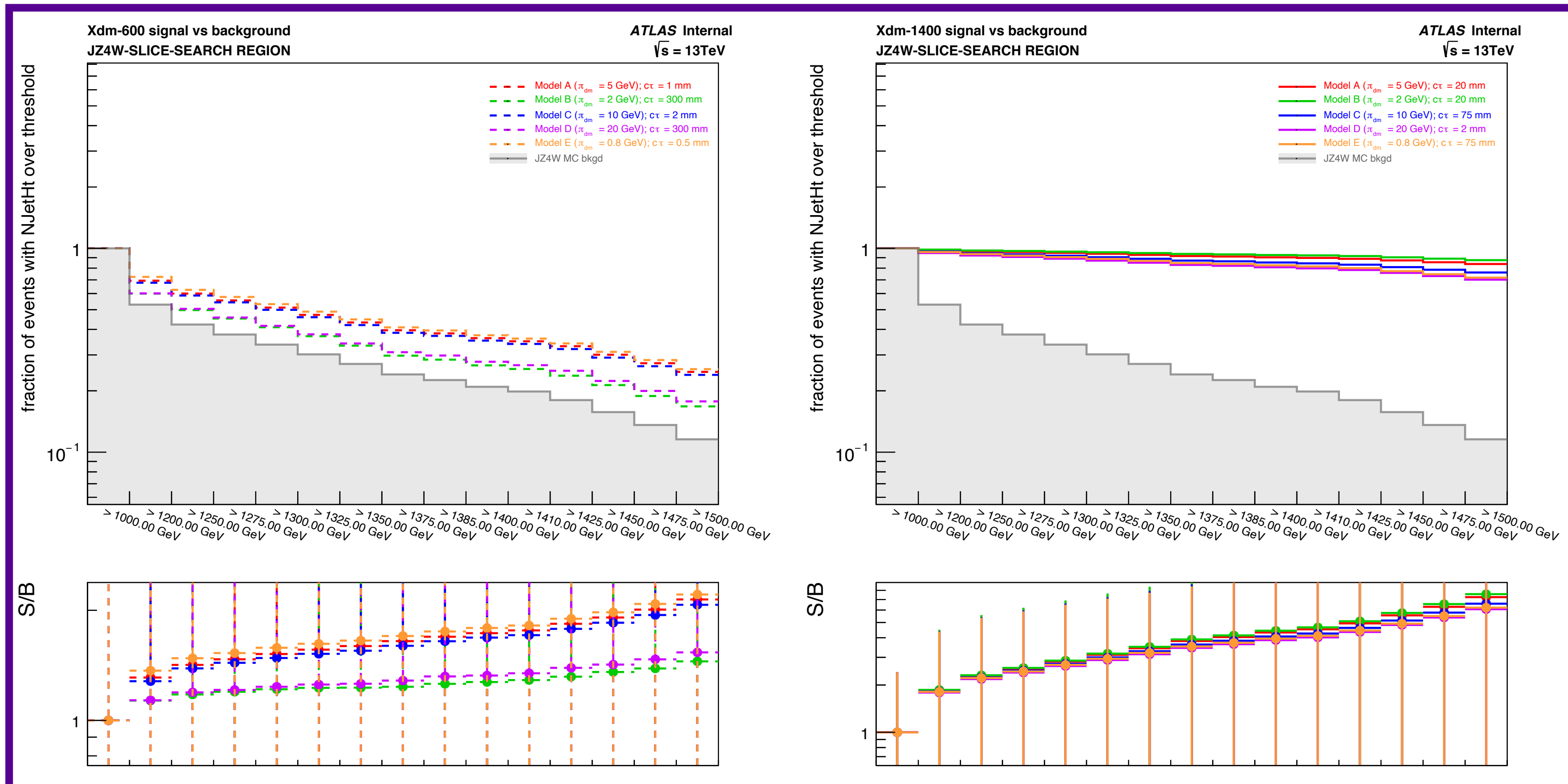


Number of events passing NEJ cut

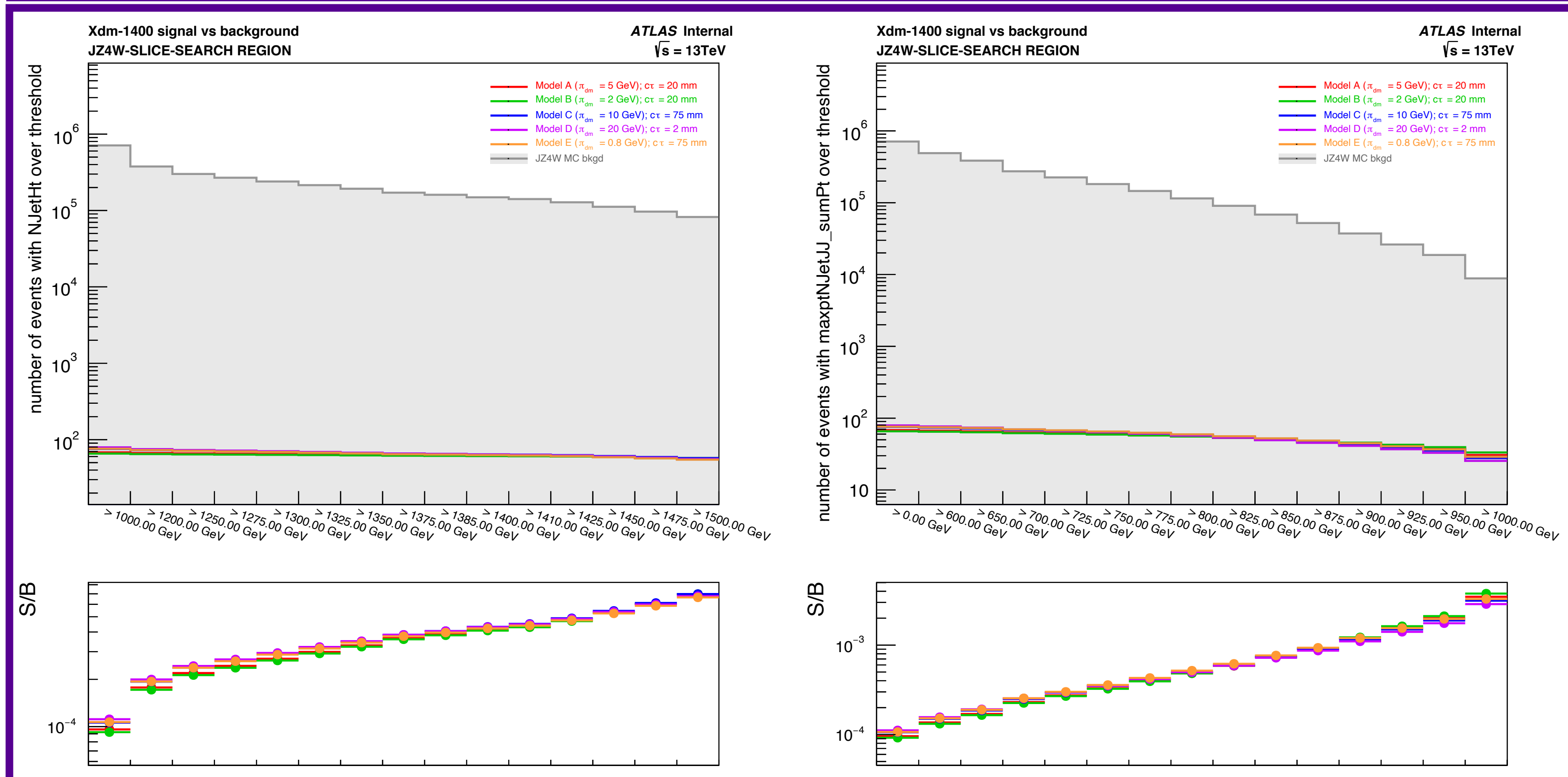
- Jet-level selection efficiency decreases more slowly for signal than background, indicating cuts target background QCD jets more often than signal EJs
- Background efficiency decreases significantly with first SV-related cut
- S/B increases with each cut, except for final cut in low mass models → Tight EJ WP may too tight…
- MC background effectively goes to zero for **N$_{EJ}$ > 1** → Loose EJ WP boasts higher signal efficiency and appears to sufficiently reduce background, but further studies needed, i.e. with more background statistics / slices

NJetX Cutflows

- NJetX selections do well at targeting background events, with signal efficiency remaining high even with tightest thresholds, but lower mediator masses suffer more signal loss (although still manageable)
- → S/B increases with each cut, so tightest NJetX cuts can be implemented to maximize background reduction without complete loss of sensitivity to signal
- More background rejection with maximum-pT dijet system, but also more signal loss → delicate balance to find NJetX selection that maximizes S/B in conjunction with other signal selections…



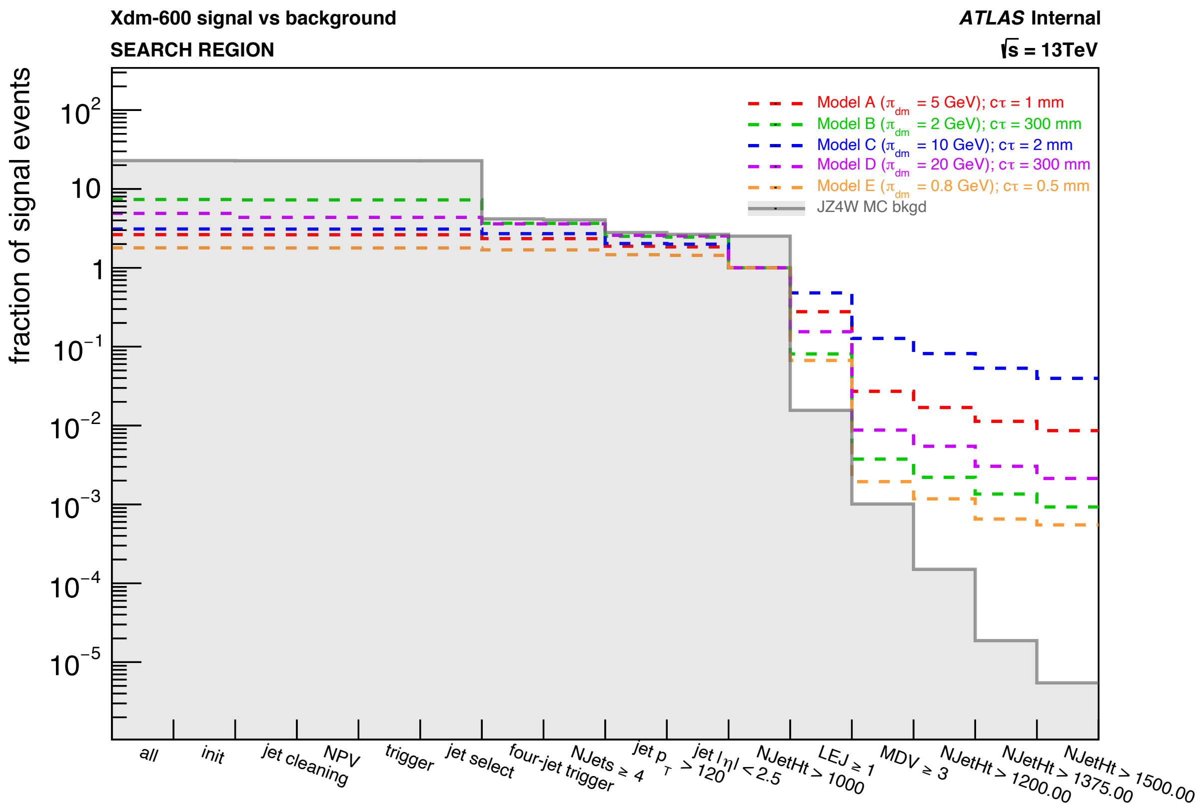
Fraction of events passing NJetX cut



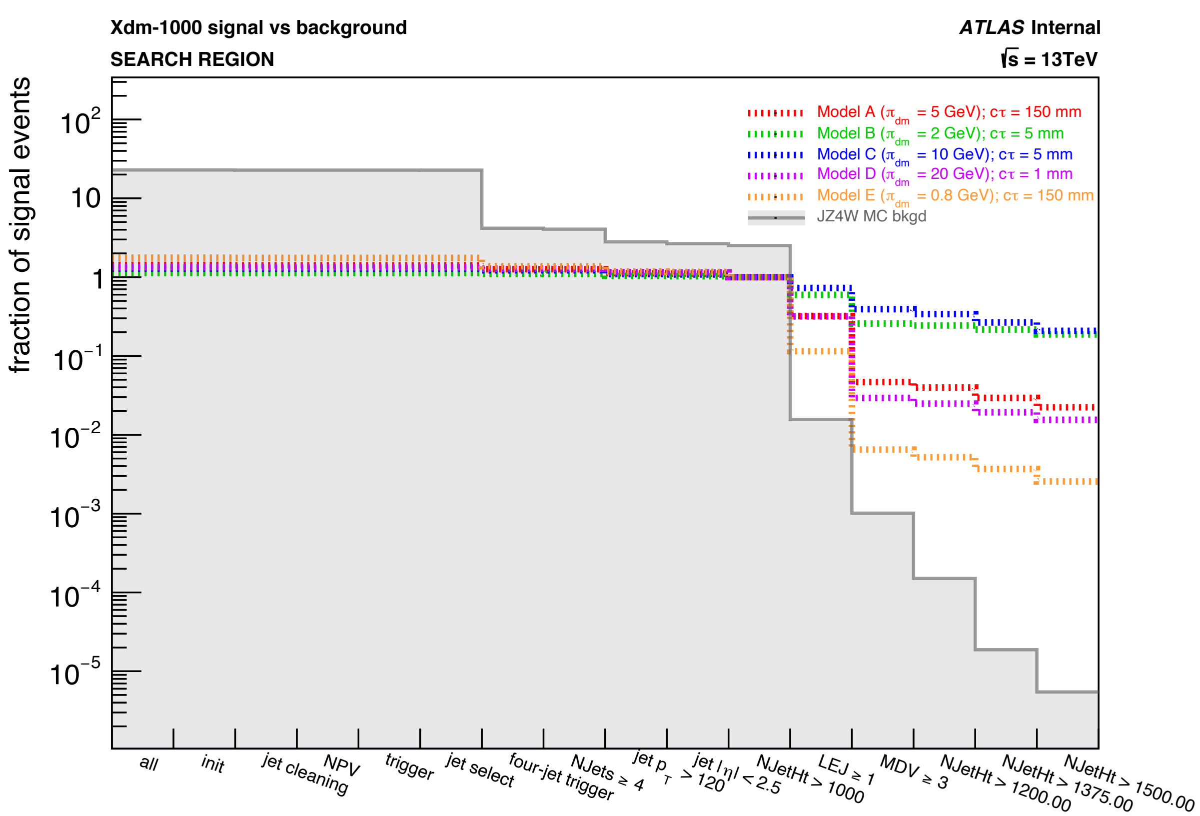
Number of events passing NJetX cut

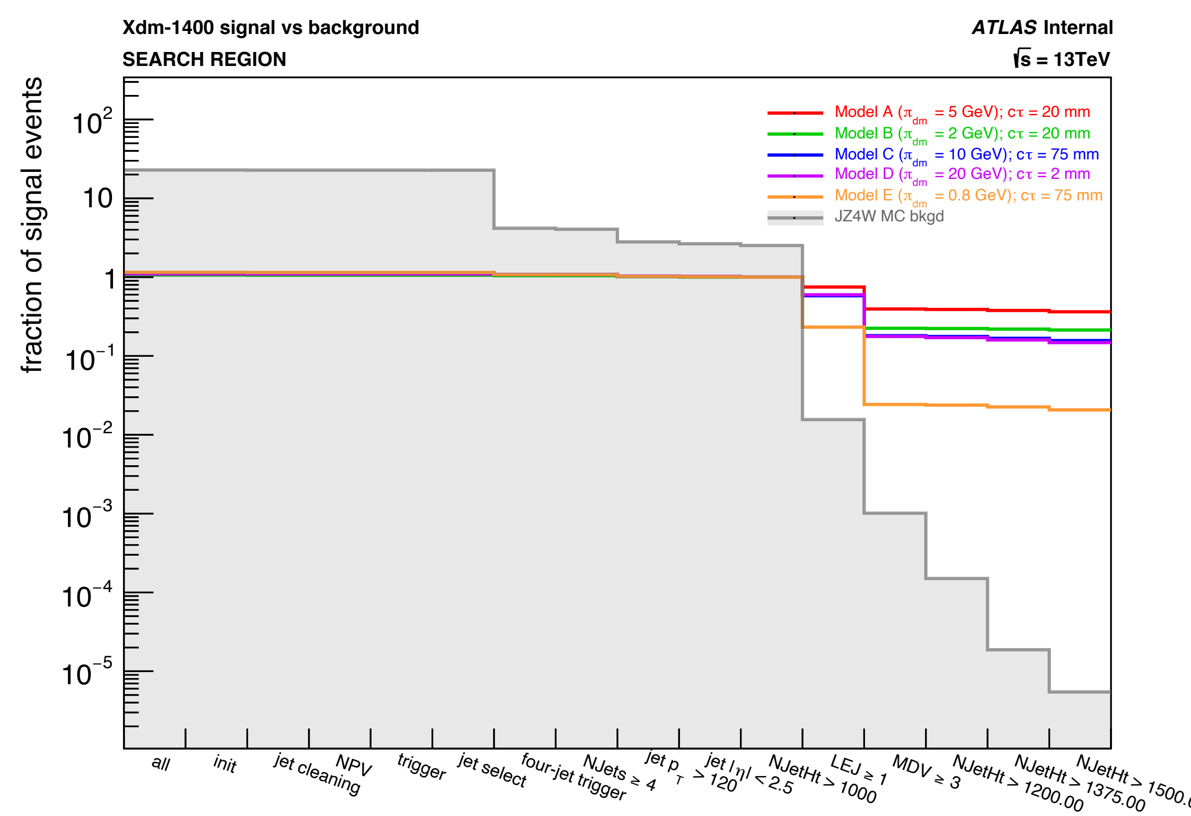
Full Signal Region Cutflow

$N_{LEJ} \geq 1 + N_{MDV} \geq 3 + H_T^{NJet} > 1200\text{-}1500$



- Final criteria defining potential signal regions = $N_{EJ} + N_{DV} + X^{NJet}$ → still to be optimized to identify most promising signal regions resulting in maximum signal efficiency and background rejection
- Signal minimally impacted by initial preselection and search region selections → only 600-GeV samples suffer any real signal loss with initial H_T cut
- → final $N_{DV} + N_{EJ} + X^{NJet}$ signal region selections drastically reduce background with smaller, but non-negligible, effect on signal
- → signal efficiencies and background rejections vary between individual and combined N_{DV}, N_{EJ}, and X^{NJet} selections, which differ in dependencies on signal model parameters
- → generally, overall signal efficiency decreases with decreasing mediator mass and model mass and with extreme dark pion lifetime





Signal Event Cutflow + Efficiency

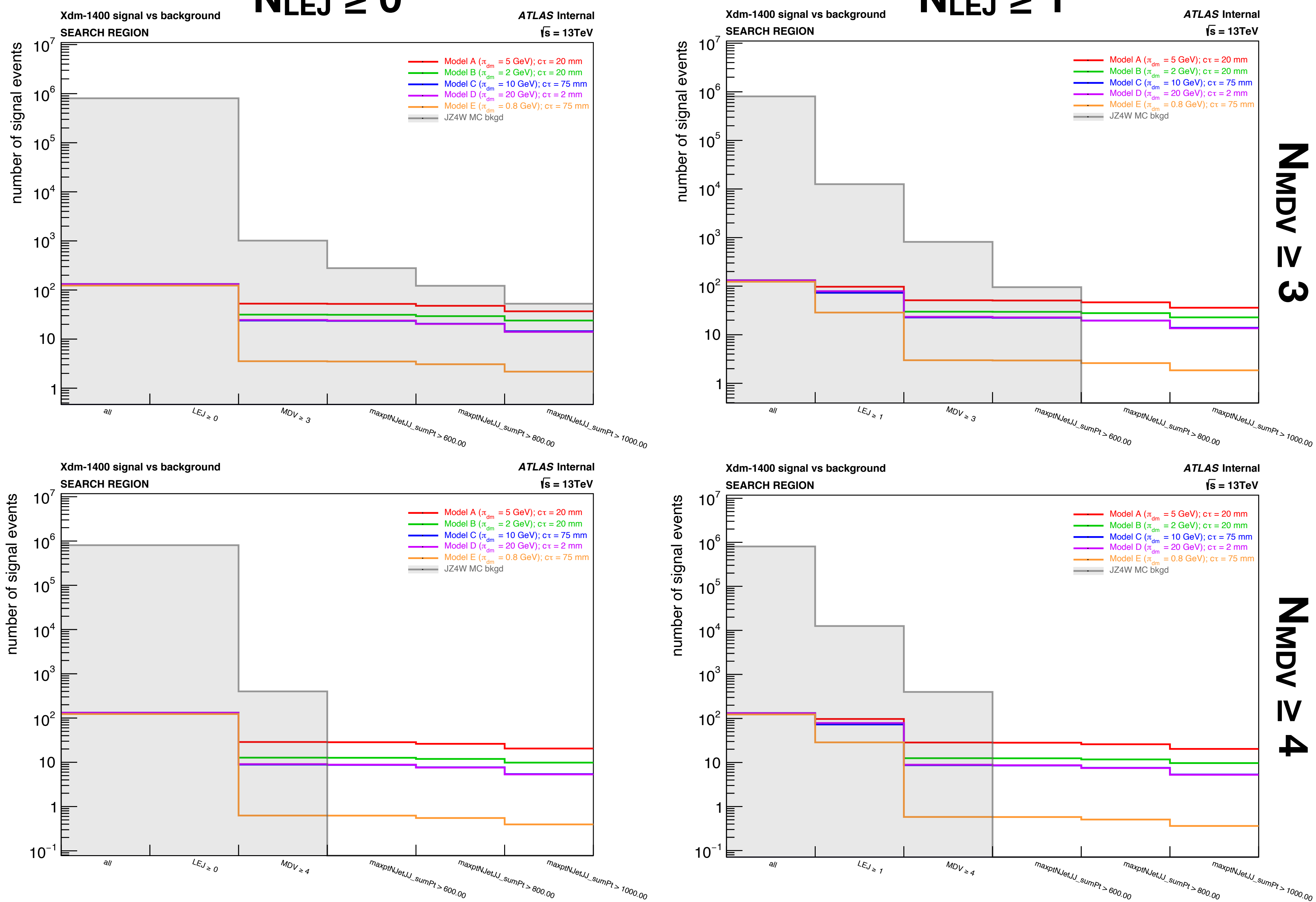
- Loose EJ cut of $N_{EJ} \geq 1$ minimally reduces overall event count in signal but significantly reduces → background effectively goes to zero earlier or when it otherwise would not in sequence of signal event selections versus $N_{EJ} = 0$
- For $N_{DV} \geq 4$ with Medium and Tight WPs, NEJ cut not required to effectively eliminate background → Tight $N_{DV} \geq 3$ cut can also sufficiently remove MC background without NEJ requirement, depending on NJetX selection
- Low-mass-model and extreme-lifetime points most heavily impacted by NEJ selection → lower signal efficiency also observed for lower mediator masses

$H_T^{NJetJJ,max\text{-}pT}$

$N_{LEJ} \geq 0$ | $N_{LEJ} \geq 1$

$N_{MDV} \geq 3$

$N_{MDV} \geq 4$

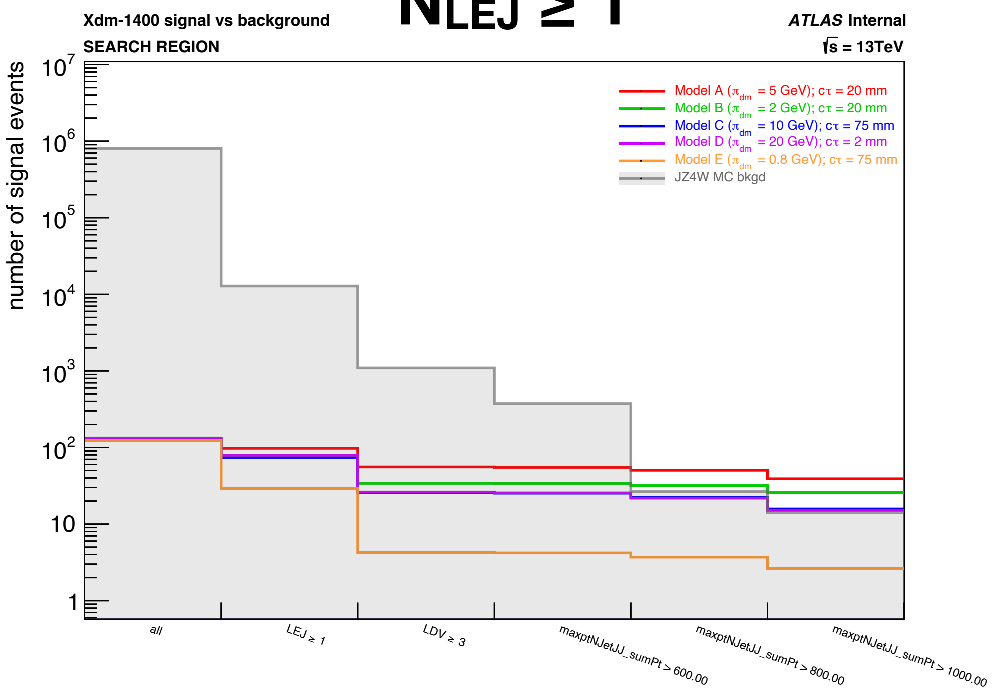


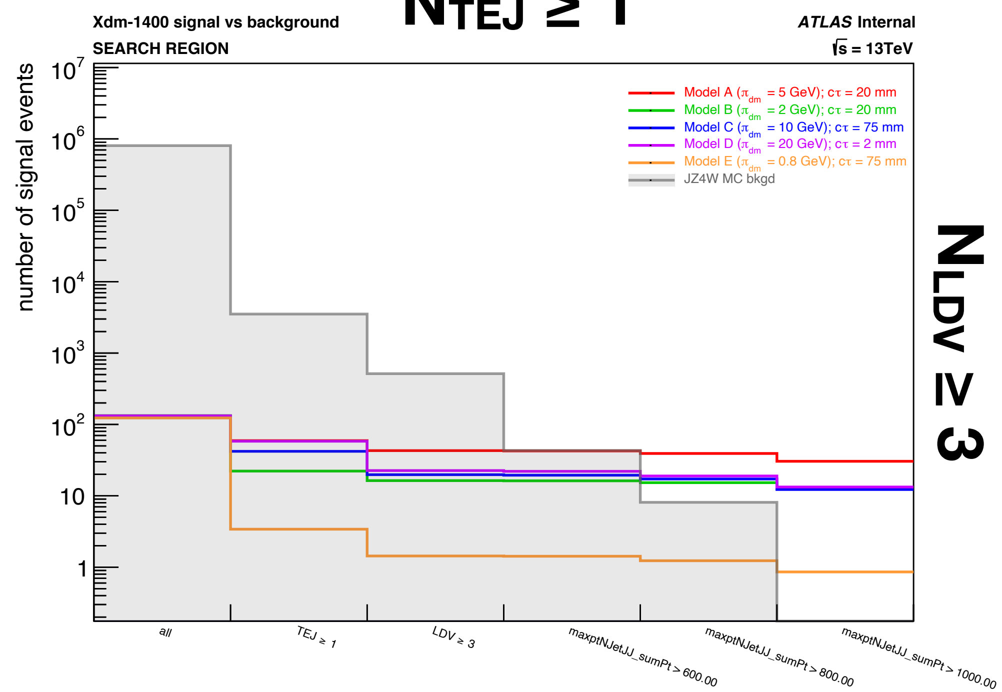
Signal Event Cutflow + Efficiency

$H_T^{\text{NJetJJ,max-pT}}$

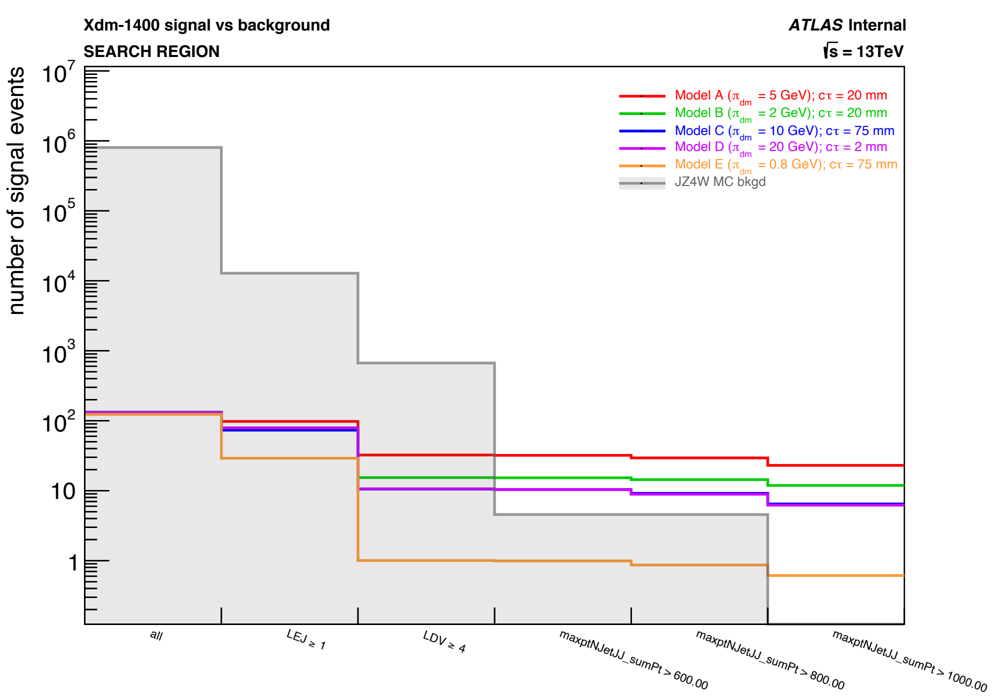
$N_{LEJ} \geq 1$ | $N_{TEJ} \geq 1$

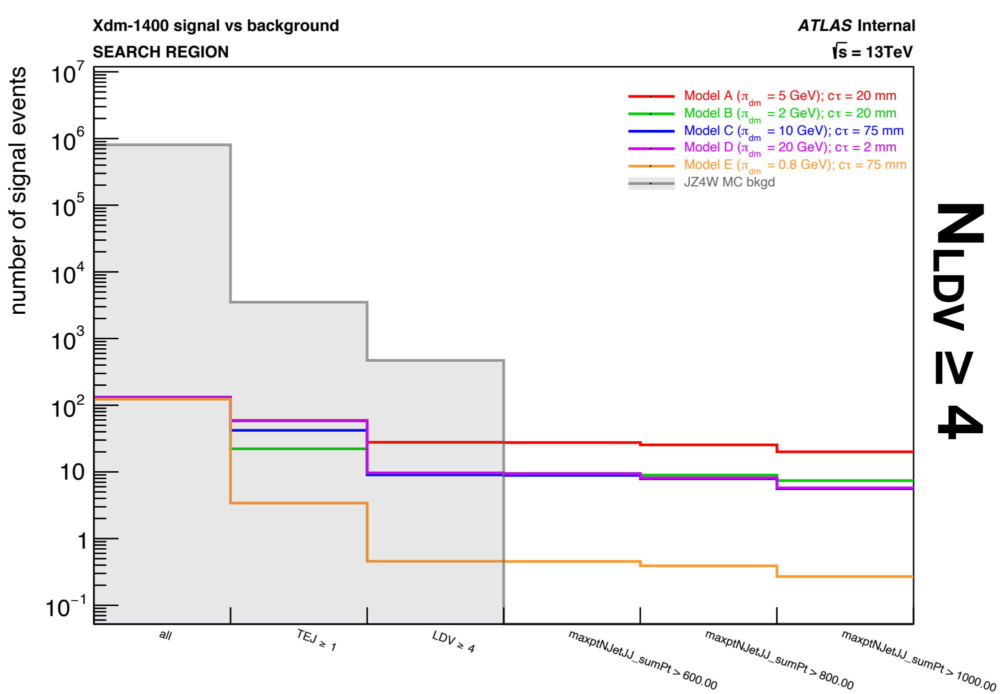
$N_{LDV} \geq 3$





$N_{LDV} \geq 4$

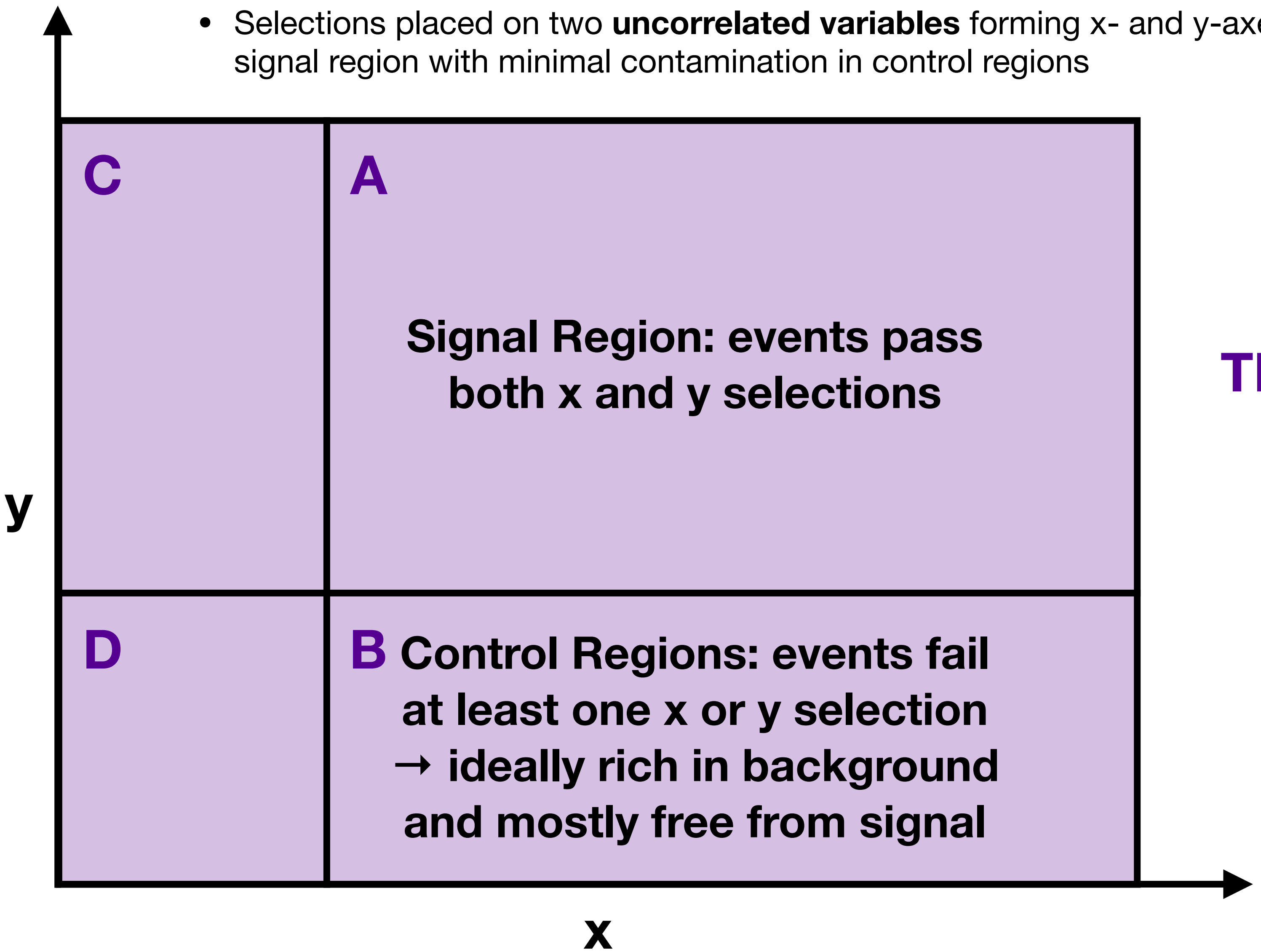




- Tightening EJ selection WP can help reduce or even eliminate background with very little impact on overall signal efficiency → signal points with low mediator mass, low model mass, and extreme lifetime suffer most signal loss with Tight EJ WP
- N_{EJ} = 2 cut likely too stringent → eliminates MC background regardless of DV or EJ WP but with immediate and severe loss of signal efficiency → other combinations of cuts can also effectively reduce background without as much effect on signal

Background Estimation: ABCD Method

- Single inclusive background process contributing to EJs signal region: **SM QCD multijets** → can yield four high-p_T jets and signal-like DVs that mimic EJs signal
- Background estimation performed using **data-driven ABCD method** to predict number of non-signal events in signal region from **transfer factor** measured from event yields in control regions
- 2D ABCD plane where signal and background separated into four regions: A, B, C, and D → defined by two independent selections forming part of final signal region
    - Selections placed on two **uncorrelated variables** forming x- and y-axes of ABCD plane → cuts chosen to isolate signal events in signal region with minimal contamination in control regions



ABCD method:

$$N_A^{\mathrm{bkgd}} = N_B^{\mathrm{bkgd}} \times N_C^{\mathrm{bkgd}} / N_D^{\mathrm{bkgd}}$$

Assumption:

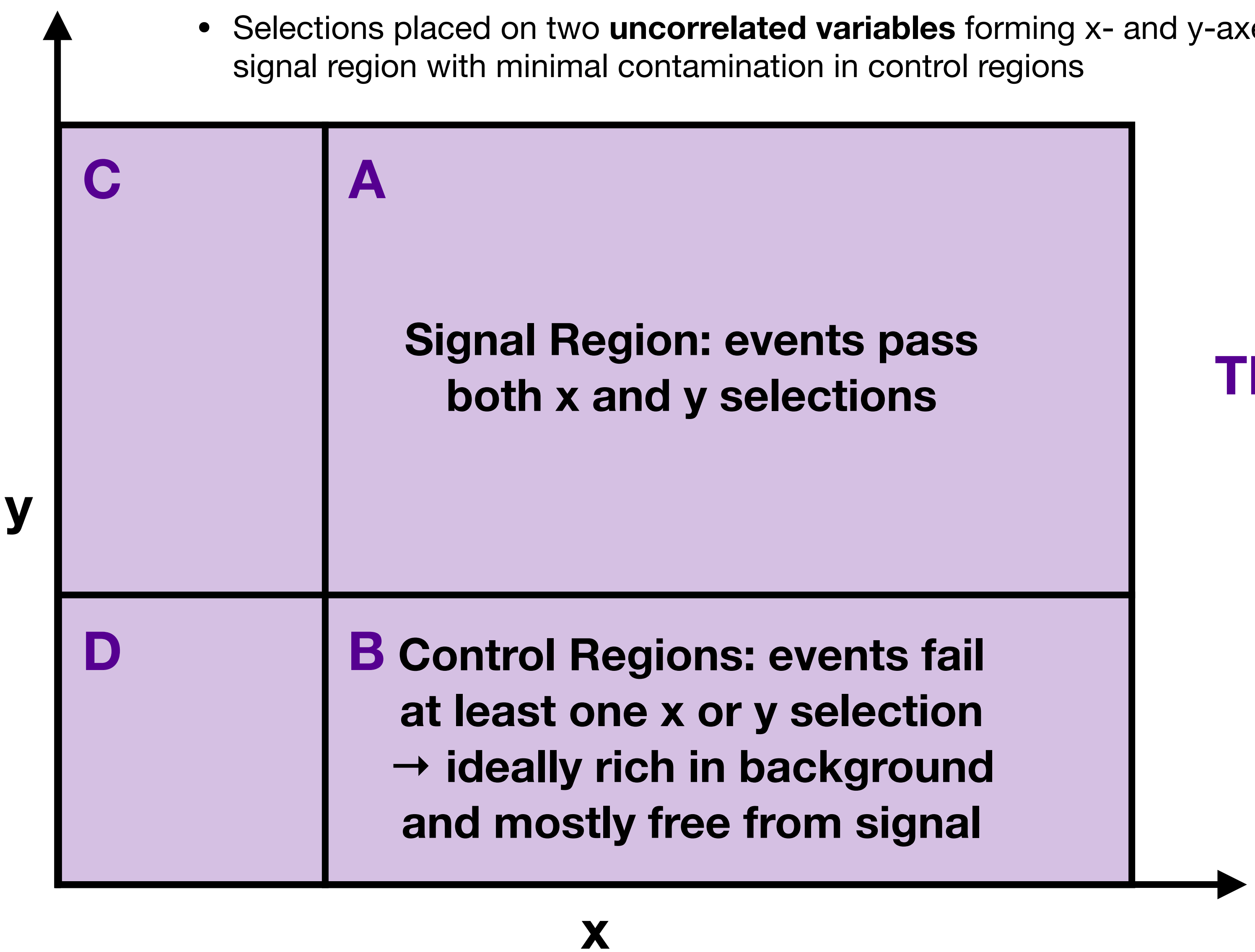
TF

$$\underline{N_C^{\mathrm{bkgd}} / N_D^{\mathrm{bkgd}}} = N_A^{\mathrm{bkgd}} / N_B^{\mathrm{bkgd}}$$

- → if ABCD observables sufficiently uncorrelated, transfer factor for one selection variable does not change as function of second
- TF ideally measured in region of pure background → basic method assumes no signal contamination in CRs

Background Estimation: ABCD Method

- Single inclusive background process contributing to EJs signal region: **SM QCD multijets** → can yield four high-p_T jets and signal-like DVs that mimic EJs signal
- Background estimation performed using **data-driven ABCD method** to predict number of non-signal events in signal region from **transfer factor** measured from event yields in control regions
- 2D ABCD plane where signal and background separated into four regions: A, B, C, and D → defined by two independent selections forming part of final signal region
    - Selections placed on two **uncorrelated variables** forming x- and y-axes of ABCD plane → cuts chosen to isolate signal events in signal region with minimal contamination in control regions



ABCD method:

$$N_A^{\text{bkgd}} = N_B^{\text{bkgd}} \times N_C^{\text{bkgd}} / N_D^{\text{bkgd}}$$

Assumption:

TF

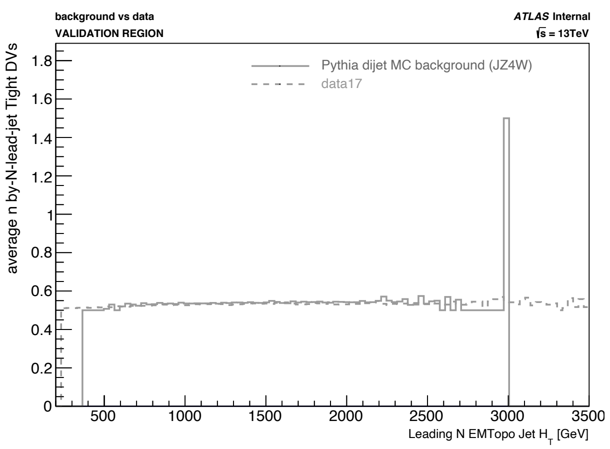
$$\underline{N_C^{\text{bkgd}} / N_D^{\text{bkgd}}} = N_A^{\text{bkgd}} / N_B^{\text{bkgd}}$$

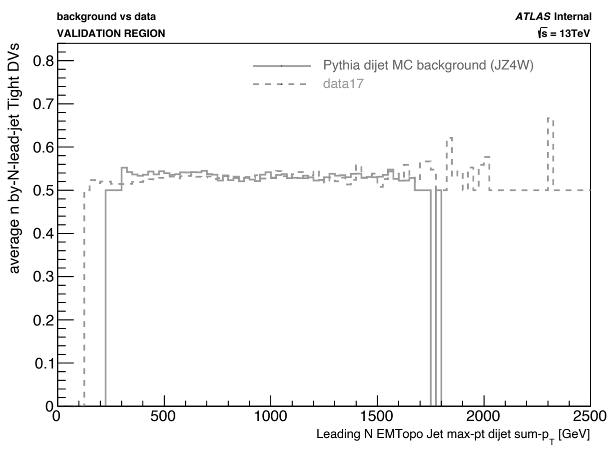
- → if ABCD observables sufficiently uncorrelated, transfer factor for one selection variable does not change as function of second
- TF ideally measured in region of pure background → basic method assumes no signal contamination in CRs

Potential ABCD Planes

Initial studies performed using MC background to aid in designing and validate discriminating power of signal selections and to test background estimation method in potential signal regions

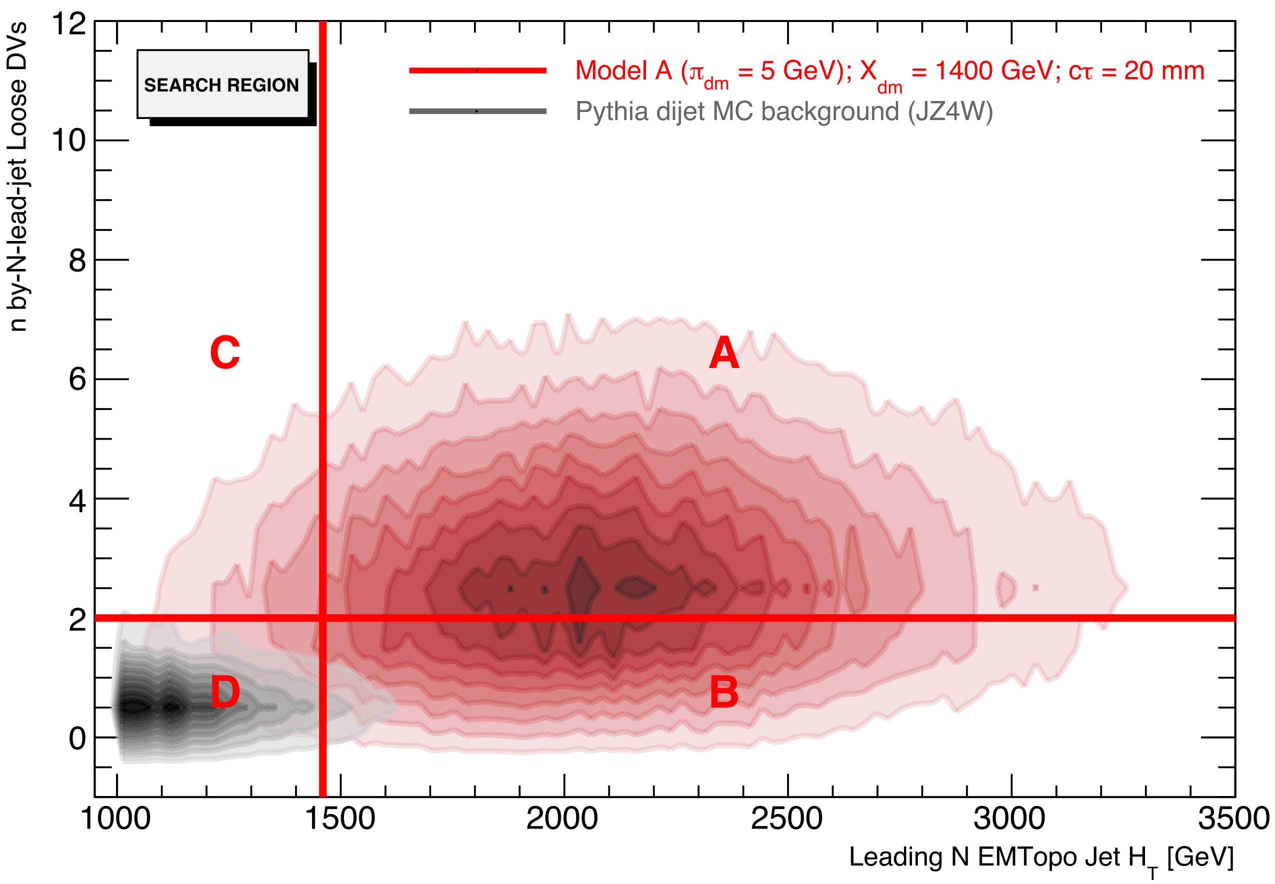
→ N_{DV} and X^{NJet} observables sufficiently uncorrelated



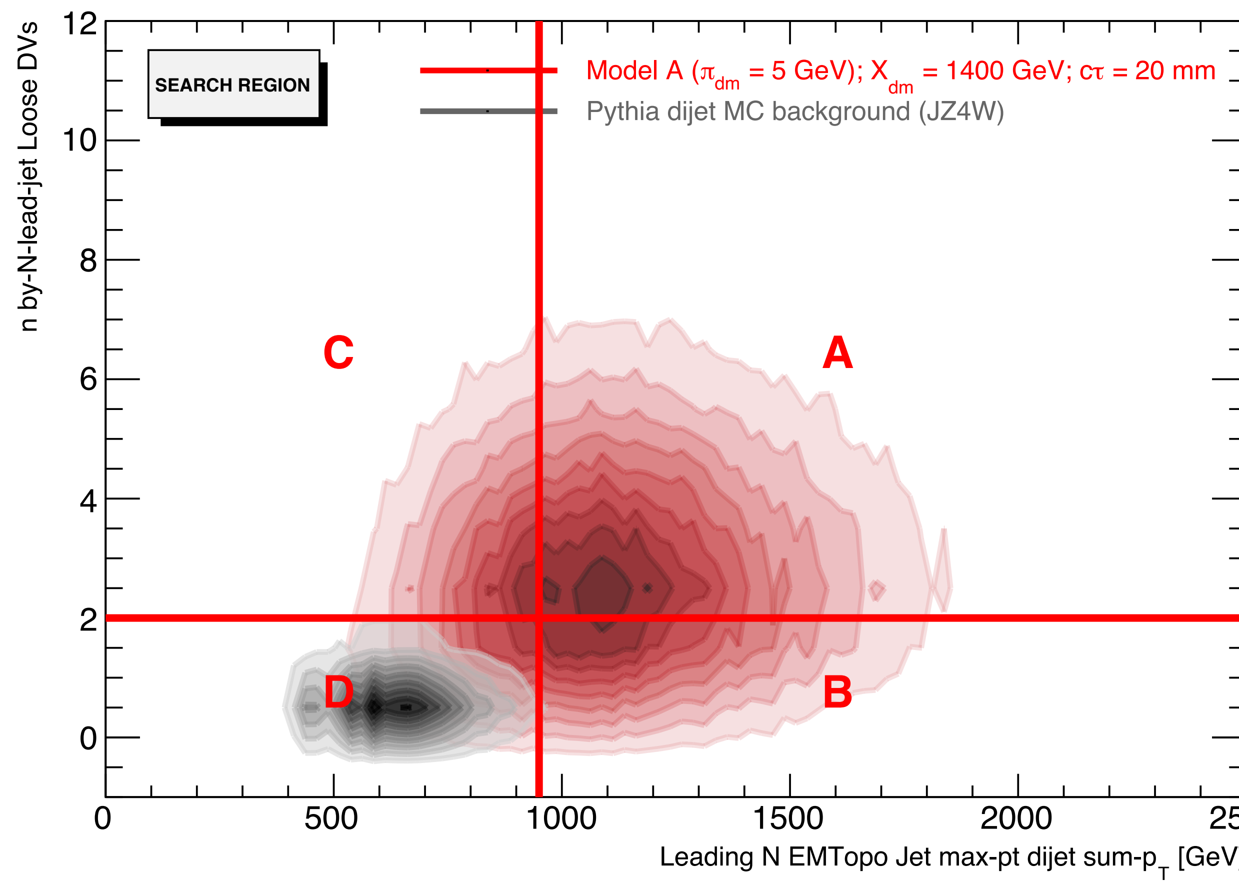


- Distributions in ABCD plane show obvious signal contamination in CRs, particularly for low mediator mass models, where discrimination between signal and background is less significant
  - → signal contamination larger and background isolation worse for some DV × NJetX planes than others
  - → nonzero NEJ cut could improve signal and background isolation

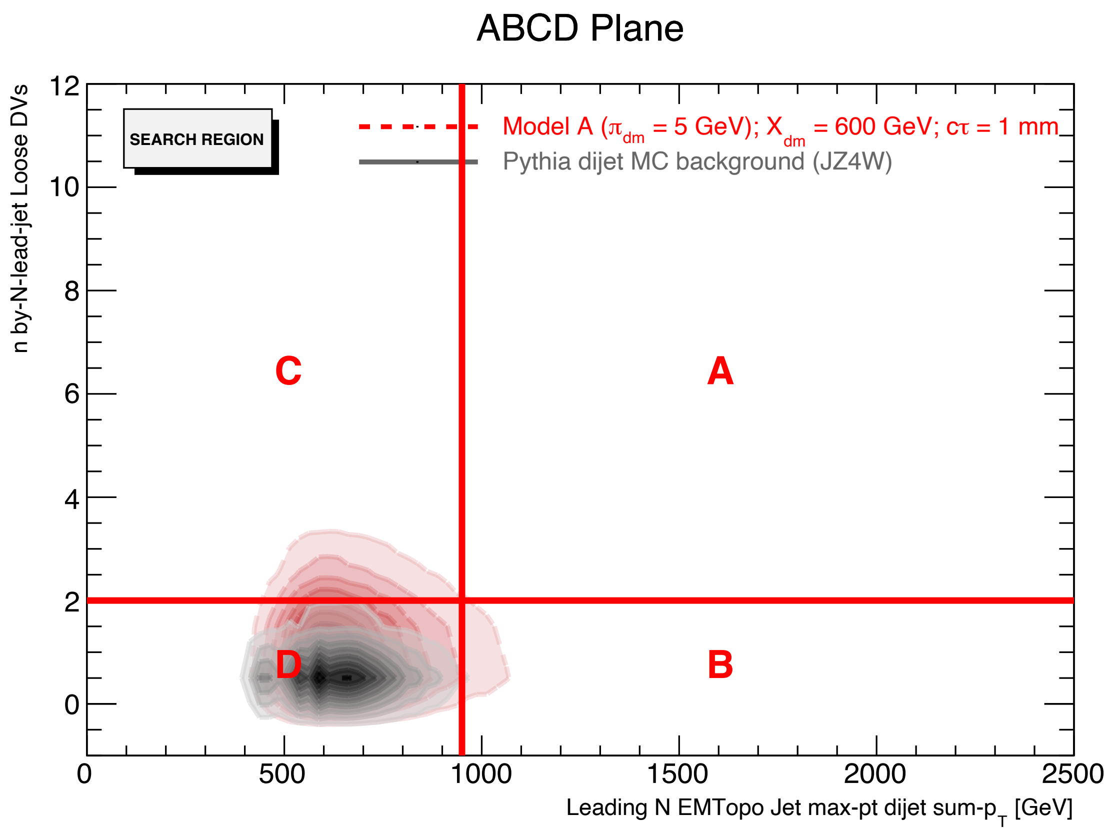
ABCD Plane

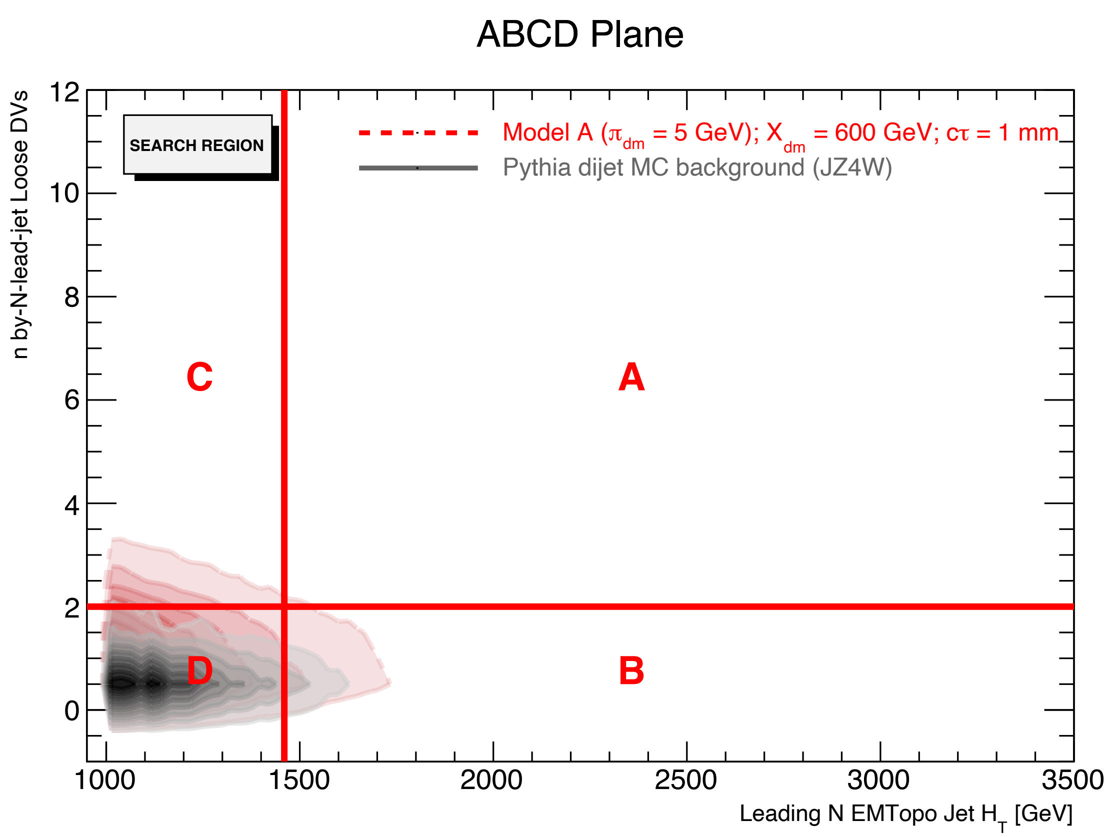


ABCD Plane



Initial Tests with Basic Method





- Initial background estimates and validation studies performed with basic ABCD method and JZ4W MC dijet slice (and subset of data for validation)
  - Basic method may provide inaccurate estimate in face of seemingly large signal contamination → signal strength small compared to background, so may be safe to regard contamination as negligible
  - Agreement between predicted and observed background yields in A' vary widely between different signal selections, with basic ABCD method performing very well for some proposed signal regions and very poorly for others → ultimate validity of method will depend on final signal region selections
  - Background contribution to signal region decreases as selections tightened, with Tight DV WP yielding few to zero MC background events
  - Tightened selections also decrease overall signal efficiency, so expected signal significance will need consideration before finalizing signal region
  - Current MC estimates likely inaccurate representations (underestimates) of expected backgrounds, but provide rough approximations of size of background in various potential signal regions to aid in optimizing signal selections for use in final analysis → dedicated shifted VR method can be applied to full data in validation region as secondary rough estimate before determining final analysis cuts
- Full background estimates and validation still to be performed with full (blinded) ATLAS Run 2 dataset

ABCD Likelihood-Fit Method

- In case of non-negligible signal contamination in ABCD background regions, likelihood-based approach can be implemented
  - → fitting of statistic model constructed with underlying assumption of basic ABCD method about relationship of background distribution between different regions

$$\tilde{N}_A = \tilde{m}\tilde{N}_B, \tilde{N}_C = \tilde{m}\tilde{N}_D$$

$$L(\{N_A, N_B, N_C, N_D\}|\tilde{N}_B, \tilde{N}_D, \tilde{m}) = \text{Poisson}(N_A + N_B + N_C + N_D|\tilde{N}_{\text{total}}) \times \prod_{N_A} \frac{\tilde{m}\tilde{N}_B}{\tilde{N}_{\text{total}}} \prod_{N_B} \frac{\tilde{N}_B}{\tilde{N}_{\text{total}}} \prod_{N_C} \frac{\tilde{m}\tilde{N}_D}{\tilde{N}_{\text{total}}} \prod_{N_D} \frac{\tilde{N}_D}{\tilde{N}_{\text{total}}}$$

$$\tilde{N}_{\text{total}} = \tilde{m}\tilde{N}_B + \tilde{N}_B + \tilde{m}\tilde{N}_D + \tilde{N}_D$$

$$L(\{N_B, N_C, N_D\}|\tilde{N}_B, \tilde{N}_D, \tilde{m}) = \text{Poisson}(N_B + N_C + N_D|\tilde{N}_{\text{blind total}}) \times \prod_{N_B} \frac{\tilde{N}_B}{\tilde{N}_{\text{blind total}}} \prod_{N_C} \frac{\tilde{m}\tilde{N}_D}{\tilde{N}_{\text{blind total}}} \prod_{N_D} \frac{\tilde{N}_D}{\tilde{N}_{\text{blind total}}}$$

$$\tilde{N}_{\text{blind total}} = \tilde{N}_B + \tilde{m}\tilde{N}_D + \tilde{N}_D$$

Shifted VR Method

- Current background estimates using MC dijets not wholly accurate (MC mis-modeling, single jet slice) → need to cross check rough estimates before finalizing signal region to ensure sufficiently free from background

- Shifted validation region (VR) ABCD method = data-driven method to estimate background without partially unblinding analysis → requires looking inside region D only

  1. Shift data in ABCD validation plane by applying scale factor SF that effectively reproduces expected search region distribution → SF determined by comparing known MC background distributions in search and validation regions

     $SF = \mu^{search}{}_{NJetX} / \mu^{validation}{}_{NJetX}$

  2. Count number of events in A', B', C', and D' regions of shifted validation plane

  3. Count number of events in region D of nominal search plane

  4. Calculate ratio between number of events in region D and number of shifted events in region D':

     $N_D{}^{ratio} = N_D{}^{search} / N_{D'}{}^{shifted\ validation}$

     → assume regions B' and C' scale to B and C similarly as D' to D between shifted and validation regions, since x- and y-axis observables uncorrelated

  5. Multiply number of shifted events in B' and C' by NDratio to approximate number of events in B and C:

     $N_{B/C}{}^{search} = N_{B/C'}{}^{shifted\ validation} \times N_D{}^{ratio}$

  6. Estimate number of events in signal region by applying ABCD method using approximated numbers from shifted validation plane:

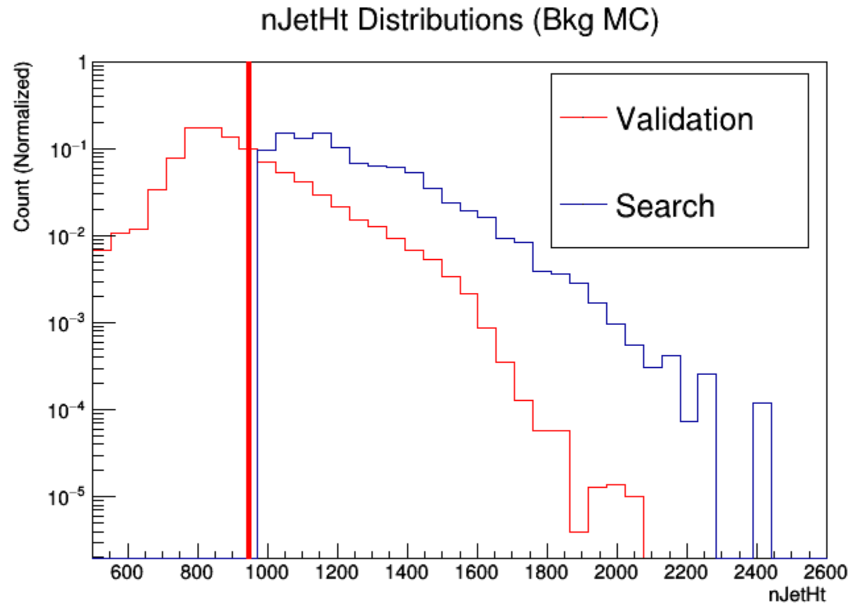
     $N_A{}^{search} = N_B{}^{search} \times N_C{}^{search} / N_D{}^{search}$

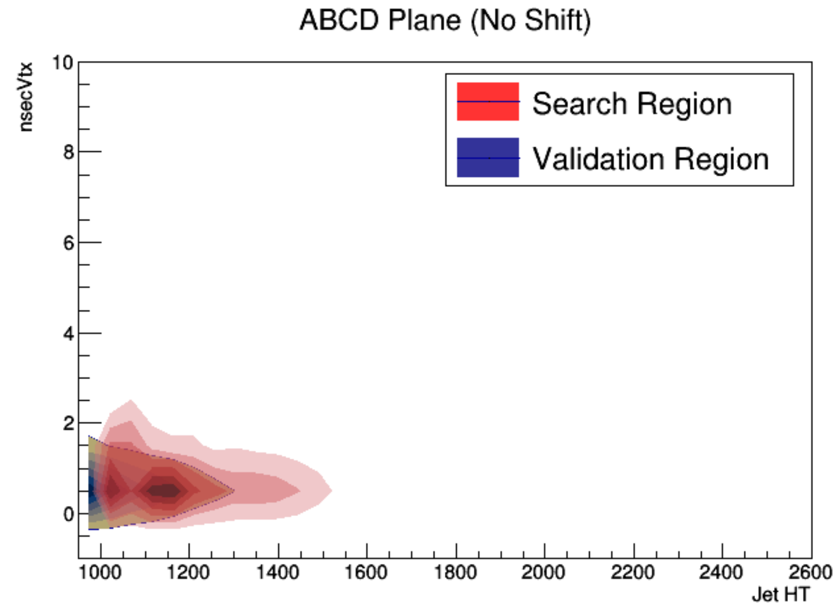
     → $N_D{}^{search}$ counted directly from observation in search region, and $N_B{}^{search}$ and $N_C{}^{search}$ approximated from shifted events in B' and C' and shifted VR ration $N_D{}^{ratio}$

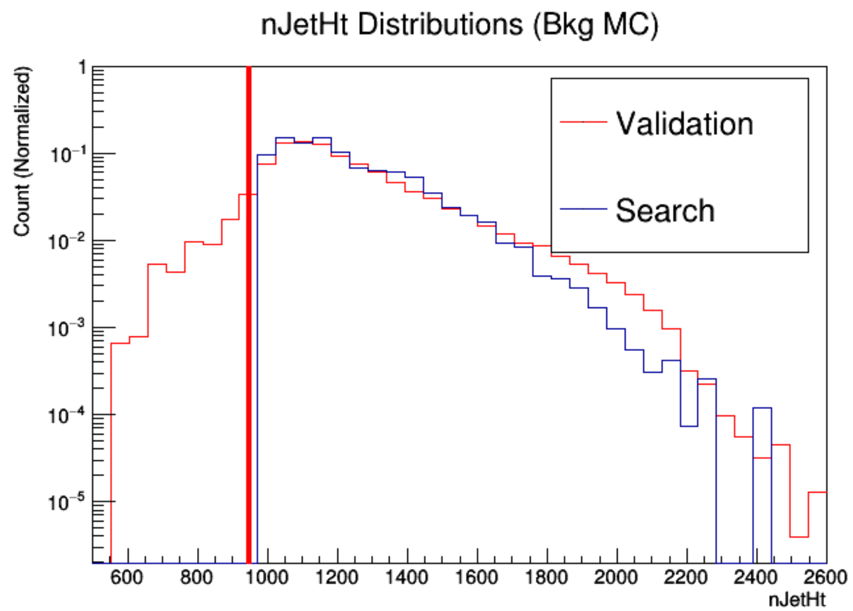
- Scale factor $SF = \mu^{search}{}_{NJetX} / \mu^{validation}{}_{NJetX}$ applied to NJetX observable, since it differs significantly between validation and search regions because of jet multiplicities in two regions

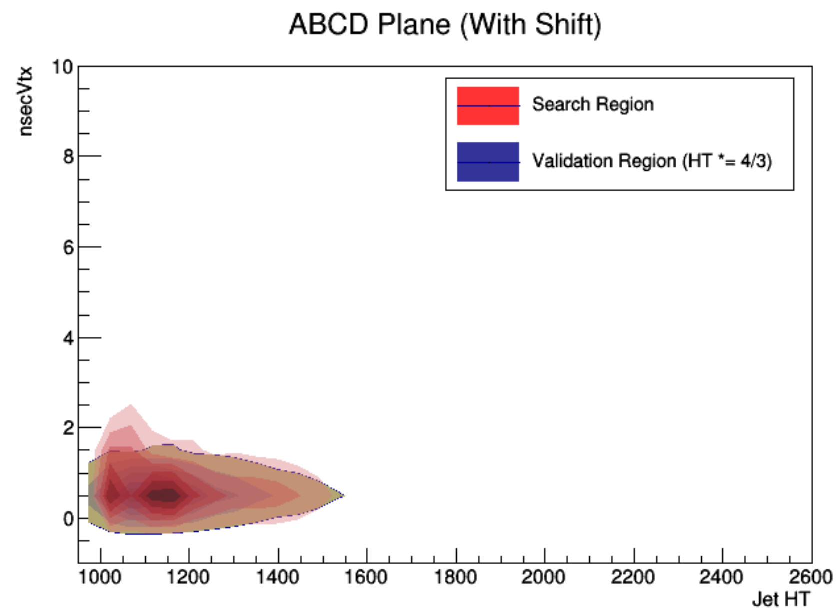
Shifted VR Method

- Scale factor SF = $\mu^{search}{}_{NJetX} / \mu^{validation}{}_{NJetX}$ applied to NJetX observable, since it differs significantly between validation and search regions because of jet multiplicities in two regions









ABCD Validation

- ABCD background estimation method validated in dedicated ABCD validation plane (free from signal) by comparing background prediction to observed data
  - Validation plane (regions A', B', C', and D') defined by inverting one analysis cut to exclude signal events → EJs validation plane = validation region, defined by inverting jet multiplicity cut
  - Background yield in A' estimated using three other primed regions in ABCD method, and prediction compared to direct observation
  - → relative uncertainty between predicted and observed values of $N_{A'}$ taken as systematic uncertainty on background estimate if values do not agree
- Initial validation studies performed with MC Pythia dijet background and subset of data → full validation still to be performed over all Run 2 data
  - Agreement between predicted and observed background yields in A' vary widely between different signal selections, with basic ABCD method performing very well for some proposed signal regions and very poorly for others
  - No realistic measure of ABCD method validity yet, but obvious trend observed → tighter signal selections provide for more accurate background estimates, due to better discrimination between signal and background and better isolation to respective regions

ABCD Initial Validation

- Measure agreement in terms of statistical difference, or **z-score**
- → quantifies how far away estimate is from background yield in terms of number of standard deviations between two values

z = (x - μ) / σ

x = data point of interest = measured background yield from ABCD estimate

μ = mean of distribution = nominal background yield from direct observation

σ = standard deviation = square-root quadrature sum of statistical and systematic uncertainties of background estimate

- Relative uncertainty = difference between predicted and observed background yields divided by background prediction
- → can be taken as systematic uncertainty on overall background estimate in case of prediction being vastly different from observation

MC

N_{DV} Selection	$H_{\mathrm{T}}^{\mathrm{NJet}}$ Selection	Statistical Difference	Relative Uncertainty
≥ 2 (Loose)	> 1200 GeV	-61.98	0.19
≥ 2 (Loose)	> 1375 GeV	-58.51	0.23
≥ 2 (Loose)	> 1500 GeV	-44.66	0.23
≥ 3 (Loose)	> 1200 GeV	-11.06	0.09
≥ 3 (Loose)	> 1375 GeV	-20.44	0.21
≥ 3 (Loose)	> 1500 GeV	-9.65	0.13
≥ 4 (Loose)	> 1200 GeV	-14.51	0.42
≥ 4 (Loose)	> 1375 GeV	-27.45	1.05
≥ 4 (Loose)	> 1500 GeV	-20.43	0.99
≥ 2 (Medium)	> 1200 GeV	-23.59	0.11
≥ 2 (Medium)	> 1375 GeV	-26.10	0.16
≥ 2 (Medium)	> 1500 GeV	-15.42	0.12
≥ 3 (Medium)	> 1200 GeV	11.73	0.18
≥ 3 (Medium)	> 1375 GeV	-5.76	0.12
≥ 3 (Medium)	> 1500 GeV	-6.22	0.17
≥ 4 (Medium)	> 1200 GeV	-30.31	2.97
≥ 4 (Medium)	> 1375 GeV	-5.29	0.51
≥ 4 (Medium)	> 1500 GeV	-12.77	1.64
≥ 2 (Tight)	> 1200 GeV	-6.60	0.09
≥ 2 (Tight)	> 1375 GeV	-0.18	0.00
≥ 2 (Tight)	> 1500 GeV	-11.28	0.27
≥ 3 (Tight)	> 1200 GeV	-6.57	2.23
≥ 3 (Tight)	> 1375 GeV	-14.80	7.24
≥ 3 (Tight)	> 1500 GeV	-0.16	0.07
≥ 4 (Tight)	> 1200 GeV	0.00	0.00
≥ 4 (Tight)	> 1375 GeV	0.00	0.00
≥ 4 (Tight)	> 1500 GeV	0.00	0.00

Data

N_{DV} Selection	$H_{\mathrm{T}}^{\mathrm{NJet}}$ Selection	Statistical Difference	Relative Uncertainty
≥ 2 (Loose)	> 1200 GeV	-238.92	0.84
≥ 2 (Loose)	> 1375 GeV	-188.58	0.71
≥ 2 (Loose)	> 1500 GeV	-156.79	0.64
≥ 3 (Loose)	> 1200 GeV	-86.77	0.96
≥ 3 (Loose)	> 1375 GeV	-69.64	0.82
≥ 3 (Loose)	> 1500 GeV	-50.28	0.64
≥ 4 (Loose)	> 1200 GeV	-38.52	1.26
≥ 4 (Loose)	> 1375 GeV	-28.30	0.97
≥ 4 (Loose)	> 1500 GeV	-4.73	0.17
≥ 2 (Medium)	> 1200 GeV	-167.15	0.97
≥ 2 (Medium)	> 1375 GeV	-144.38	0.90
≥ 2 (Medium)	> 1500 GeV	-112.70	0.76
≥ 3 (Medium)	> 1200 GeV	-51.62	1.17
≥ 3 (Medium)	> 1375 GeV	-34.80	0.82
≥ 3 (Medium)	> 1500 GeV	-18.44	0.46
≥ 4 (Medium)	> 1200 GeV	-30.19	2.17
≥ 4 (Medium)	> 1375 GeV	-21.72	1.56
≥ 4 (Medium)	> 1500 GeV	-0.15	0.01
≥ 2 (Tight)	> 1200 GeV	-61.74	0.99
≥ 2 (Tight)	> 1375 GeV	-66.93	1.18
≥ 2 (Tight)	> 1500 GeV	-44.64	0.83
≥ 3 (Tight)	> 1200 GeV	0.00	0.00
≥ 3 (Tight)	> 1375 GeV	0.00	0.00
≥ 3 (Tight)	> 1500 GeV	0.00	0.00
≥ 4 (Tight)	> 1200 GeV	0.00	0.00
≥ 4 (Tight)	> 1375 GeV	0.00	0.00
≥ 4 (Tight)	> 1500 GeV	0.00	0.00

Initial Background Estimates

- Initial estimates of background yield in A performed with JZ4W MC dijet slice
    - → inaccurate (likely underestimate), but provide rough approximations of size of background in various potential signal regions to aid in optimizing signal selections for use in final analysis
- Background contribution to signal region decreases as selections tightened, with Tight DV WP yielding very few to zero MC background events
- Tightened selections also decrease overall signal efficiency, so expected signal significance will need to be considered before finalizing signal region
- MC estimates likely inaccurate representations of expected backgrounds, so dedicated shifted VR method can be applied to full data in validation region as secondary rough estimate before determining final analysis cuts
- Full background estimates still to be performed with full blinded ATLAS Run 2 dataset

ABCD Initial Background Estimates

N_{DV} Selection	$H_{\mathrm{T}}^{\mathrm{NJet}}$ Selection	Predicted Events $\pm$ stat. $\pm$ syst.	Observed Events $\pm$ stat.
≥ 2 (Loose)	> 1200 GeV	$22239.61 \pm 149.13 \pm 160.57$	22629.08 ± 150.43
≥ 2 (Loose)	> 1375 GeV	$9943.33 \pm 99.72 \pm 58.27$	10323.77 ± 101.61
≥ 2 (Loose)	> 1500 GeV	$4917.48 \pm 70.12 \pm 28.19$	5264.61 ± 72.56
≥ 3 (Loose)	> 1200 GeV	$4029.20 \pm 63.48 \pm 67.00$	3175.12 ± 56.35
≥ 3 (Loose)	> 1375 GeV	$1690.07 \pm 41.11 \pm 22.99$	1315.21 ± 36.27
≥ 3 (Loose)	> 1500 GeV	$812.86 \pm 28.51 \pm 10.53$	636.93 ± 25.24
≥ 4 (Loose)	> 1200 GeV	$1310.68 \pm 36.20 \pm 38.10$	330.12 ± 18.17
≥ 4 (Loose)	> 1375 GeV	$416.20 \pm 20.40 \pm 11.32$	159.90 ± 12.65
≥ 4 (Loose)	> 1500 GeV	$178.14 \pm 13.35 \pm 4.86$	159.90 ± 12.65
≥ 2 (Medium)	> 1200 GeV	$8502.14 \pm 92.21 \pm 97.65$	10698.07 ± 103.43
≥ 2 (Medium)	> 1375 GeV	$4246.50 \pm 65.17 \pm 36.92$	4558.70 ± 67.52
≥ 2 (Medium)	> 1500 GeV	$2145.85 \pm 46.32 \pm 17.60$	2082.39 ± 45.63
≥ 3 (Medium)	> 1200 GeV	$1377.83 \pm 37.12 \pm 39.06$	781.96 ± 28.00
≥ 3 (Medium)	> 1375 GeV	$518.95 \pm 22.78 \pm 12.65$	337.75 ± 18.38
≥ 3 (Medium)	> 1500 GeV	$240.73 \pm 15.52 \pm 5.65$	196.52 ± 14.02
≥ 4 (Medium)	> 1200 GeV	$884.38 \pm 0.00 \pm 0.00$	0.00 ± 0.00
≥ 4 (Medium)	> 1375 GeV	$245.63 \pm 0.00 \pm 0.00$	0.00 ± 0.00
≥ 4 (Medium)	> 1500 GeV	$105.18 \pm 0.00 \pm 0.00$	0.00 ± 0.00
≥ 2 (Tight)	> 1200 GeV	$1200.92 \pm 34.65 \pm 36.45$	904.94 ± 30.08
≥ 2 (Tight)	> 1375 GeV	$416.73 \pm 20.41 \pm 11.32$	632.77 ± 25.15
≥ 2 (Tight)	> 1500 GeV	$236.19 \pm 15.37 \pm 5.60$	194.15 ± 13.94
≥ 3 (Tight)	> 1200 GeV	$0.00 \pm 0.00 \pm 0.00$	14.63 ± 3.82
≥ 3 (Tight)	> 1375 GeV	$0.00 \pm 0.00 \pm 0.00$	14.63 ± 3.82
≥ 3 (Tight)	> 1500 GeV	$0.00 \pm 0.00 \pm 0.00$	14.63 ± 3.82
≥ 4 (Tight)	> 1200 GeV	$0.00 \pm 0.00 \pm 0.00$	0.00 ± 0.00
≥ 4 (Tight)	> 1375 GeV	$0.00 \pm 0.00 \pm 0.00$	0.00 ± 0.00
≥ 4 (Tight)	> 1500 GeV	$0.00 \pm 0.00 \pm 0.00$	0.00 ± 0.00

Statistical uncertainty = standard Poisson uncertainty = square root of estimated background yield in A

Systematic uncertainty calculated from standard propagation of statistical uncertainties of observed background yields in B, C, and D =

$$N_A \times \sqrt{(\sqrt{N_B}/N_B)^2 + (\sqrt{N_C}/N_C)^2 + (\sqrt{N_D}/N_D)^2}$$