# Tagging Emerging Jets using Graph Neural Networks

SFU

## MSc Defence


**Candidate**
Paras Pokharel
MSc student @ SFU
16 April 2024


**Committee**
Dr. Andrei Frolov (Chair)
Dr. Dugan O'Nei (Internal examiner)
Dr. Matthias Danninger (Supervisor, Member
Dr.  Bernd Stelzer (Member)

# Search for new Physics?

- No strong indications of new physics at the modern collider experiments.
  - Indicate two possibilities: either the new physics is **above the energy scale accessible to LHC** - the largest particle collider, or we have been looking at the "wrong places".
  - Wrong places?
    - Most BSM physics searches have been performed with the assumption that the particles decay (promptly) near the primary interaction point of collider experiments

# Long Lived Particles (LLPs)

- LLPs: Particles that travel an observable distance from the primary collision point in particle detectors. Will have macroscopic proper lifetimes.
- Long-lived particle signatures : Unexplored phase space for BSM physics search, and requires a dedicated search
- As SM has LLPs (muons) no reason to exclude BSM searches with LLP signatures!

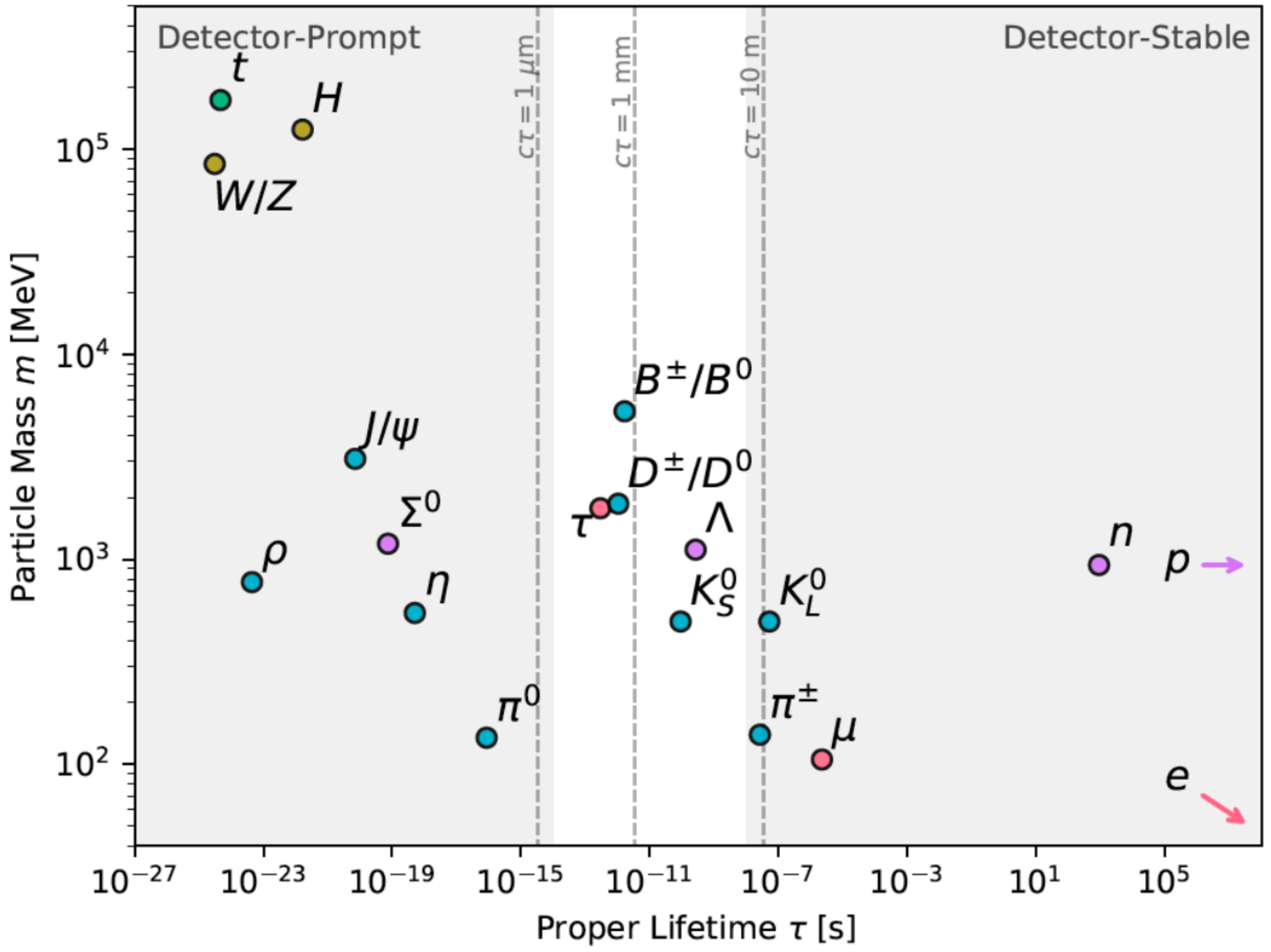
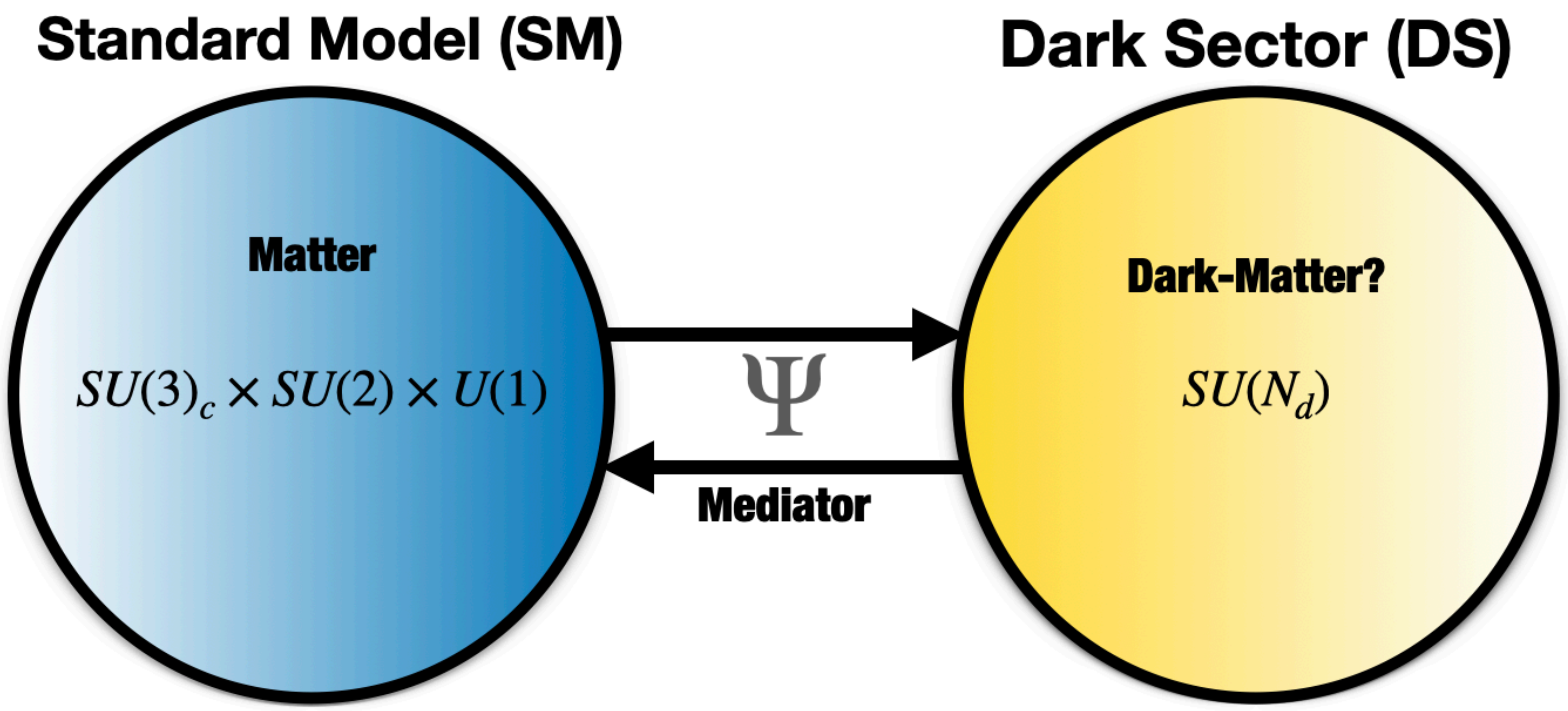



Image from Ref[1]

[1] Lawrence Lee, Christian Ohm, Abner Soffer, and Tien-Tien Yu. Collider searches or long-lived particles beyond the standard model. Progress in Particle and Nuclear Physics, 106:210–255, may 2019.65

# Theoretical Motivation for BSM LLPs



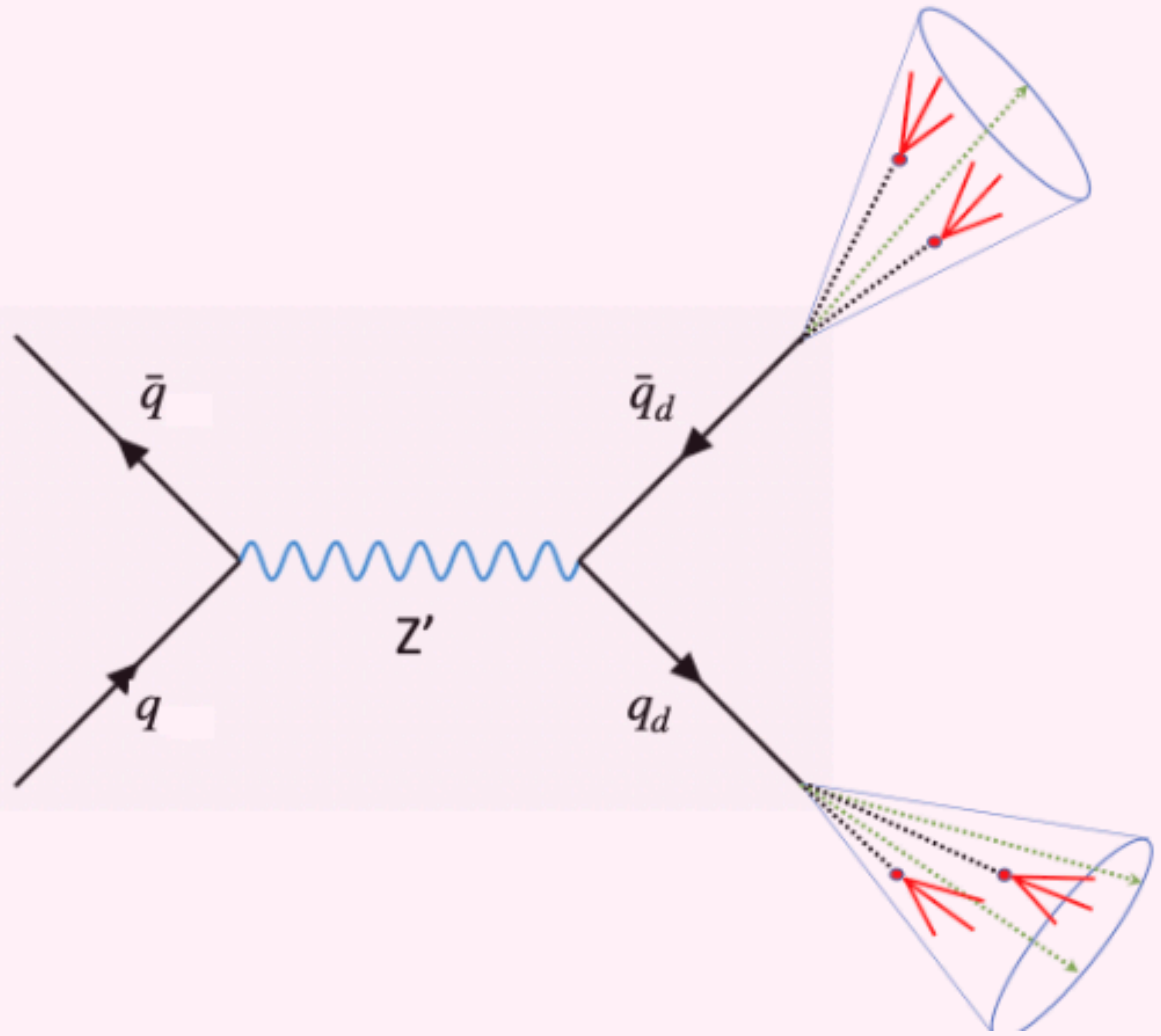

- Extended SM with additional particles and forces collectively referred as dark sector(DS).
- Weak coupling between SM and DS can give rise to LLPs

# Benchmark Model

$$\mathcal{L}_{\text{med}} = -\frac{1}{4}Z'^{\mu\nu}Z'_{\mu\nu} - \frac{1}{2}M_{Z'}^2 Z'^{\mu}Z'_{\mu} + Z'_{\mu}(\bar{q}'_i\gamma^\mu q'_i + \bar{q}_j\gamma^\mu q$$

## BSM physics processes with <u>LLP signature</u>

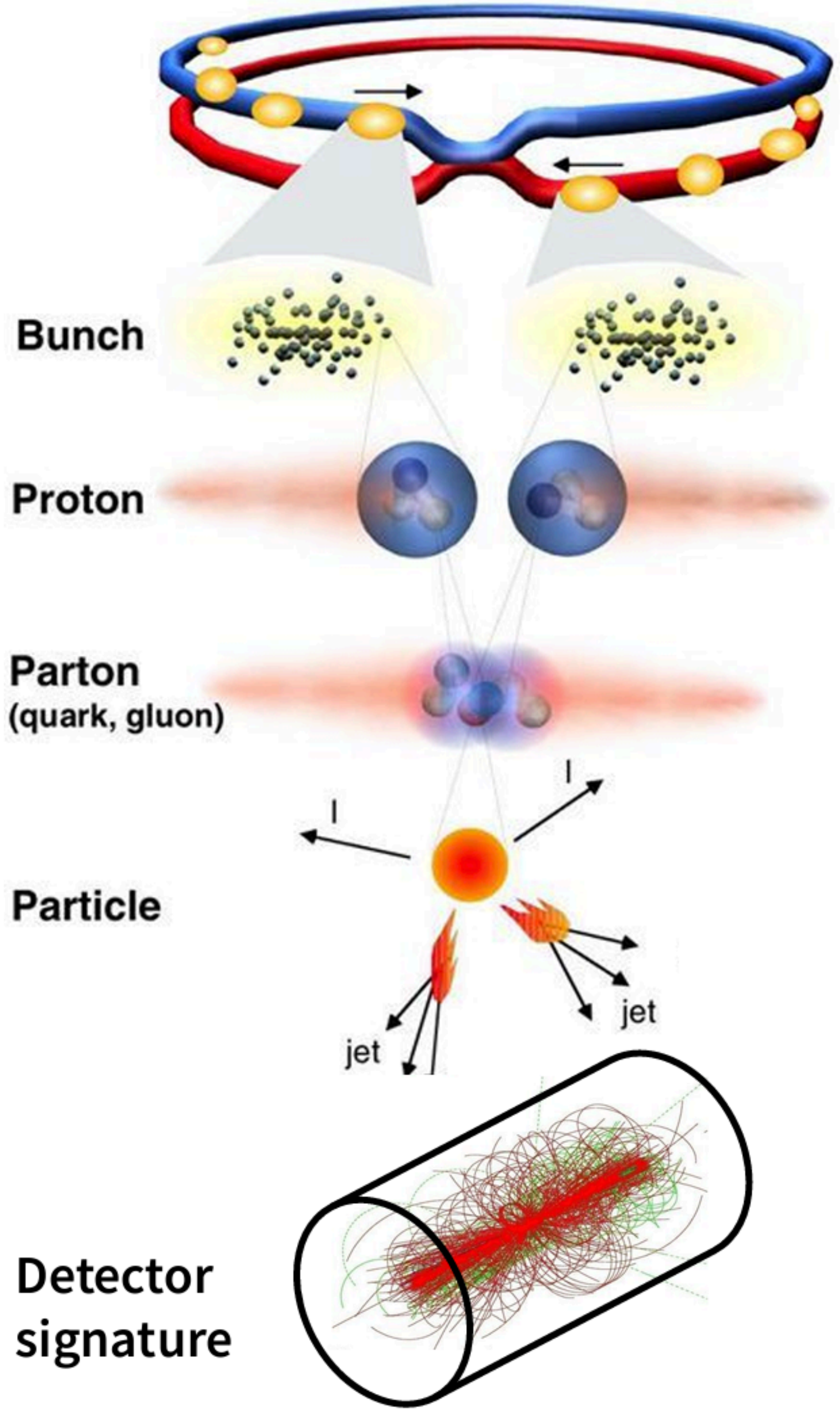
- Production of dark quarks via $Z'$ (vector) mediator.
- Dark mesons travel sizeable distances (5mm-50mm) before decaying back to SM
- Leads to exotic LLP signature known as Emerging Jets (EJs) with with unique signature → smoking gun for BSM physics.



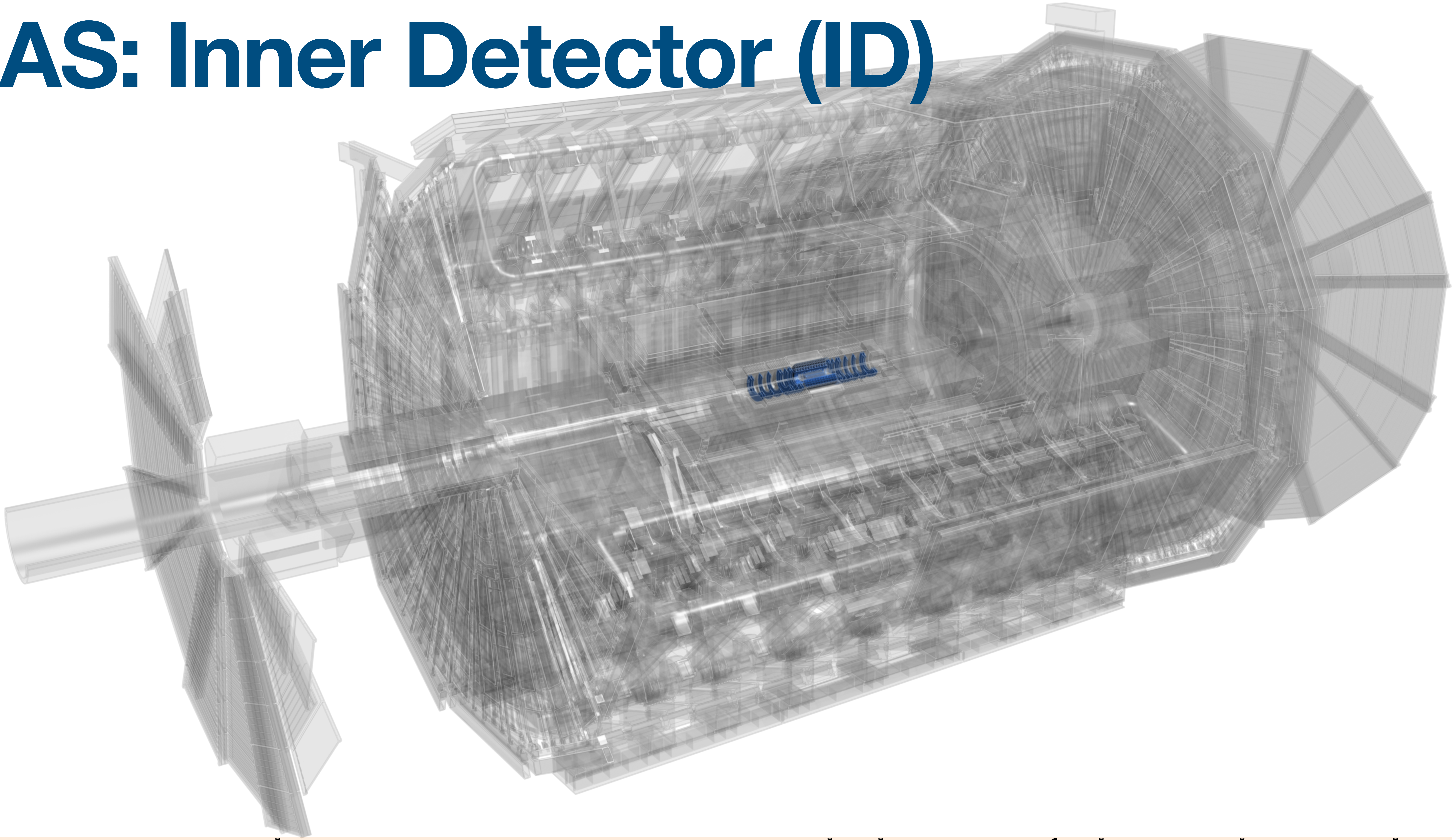
# ATLAS Detector

## Intro

- Bunches of protons are accelerated almost at the speed of light and collided at LHC, such at there are 40,000,000 interactions per second.
- A general purpose detector at LHC, ATLAS “detects” collision remnants.

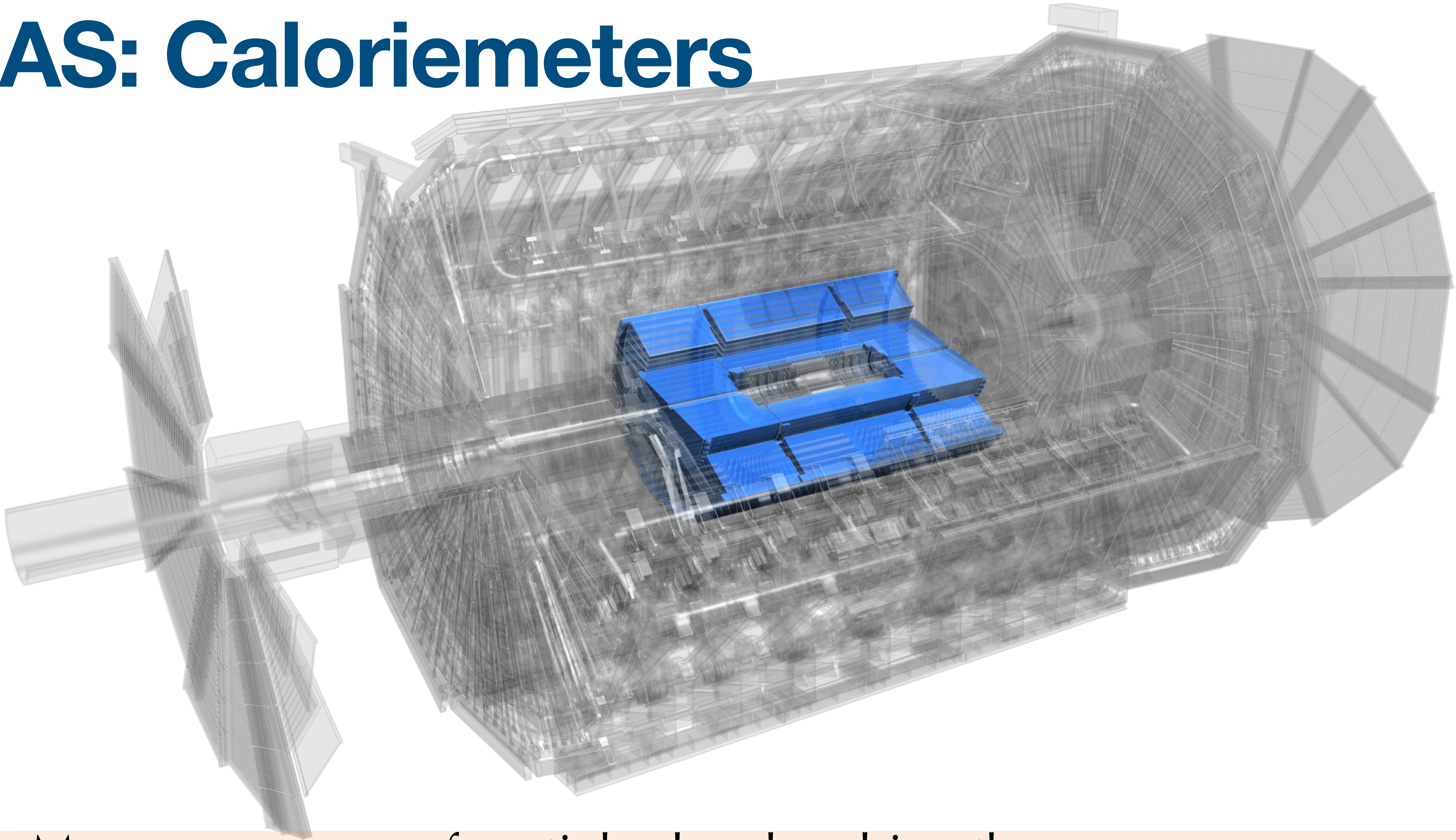


# ATLAS: Inner Detector (ID)



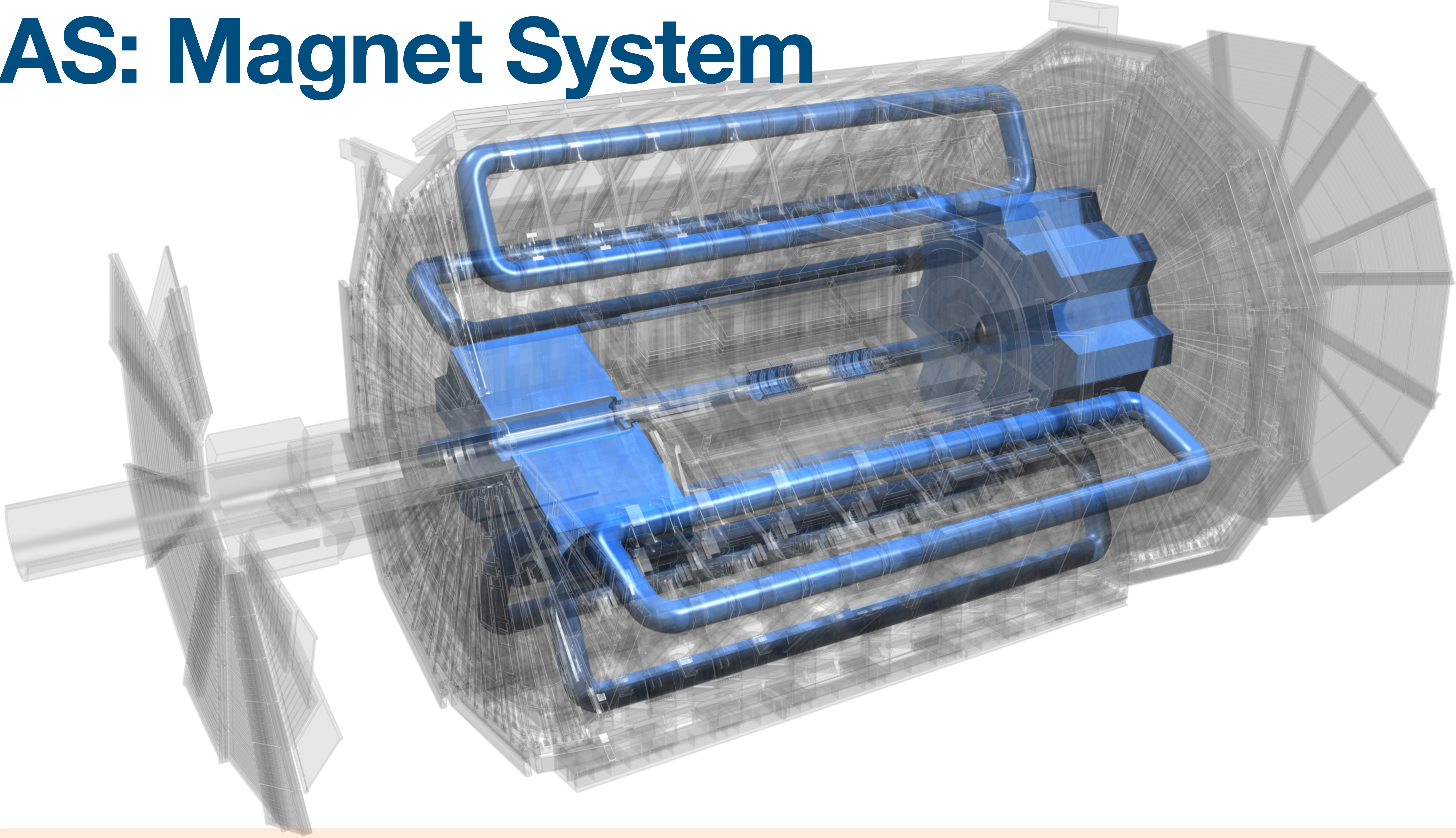
- Measures direction, momentum and charge of charged particles.
- Is made up of Pixel Detector, Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT)

# ATLAS: Caloriemeters



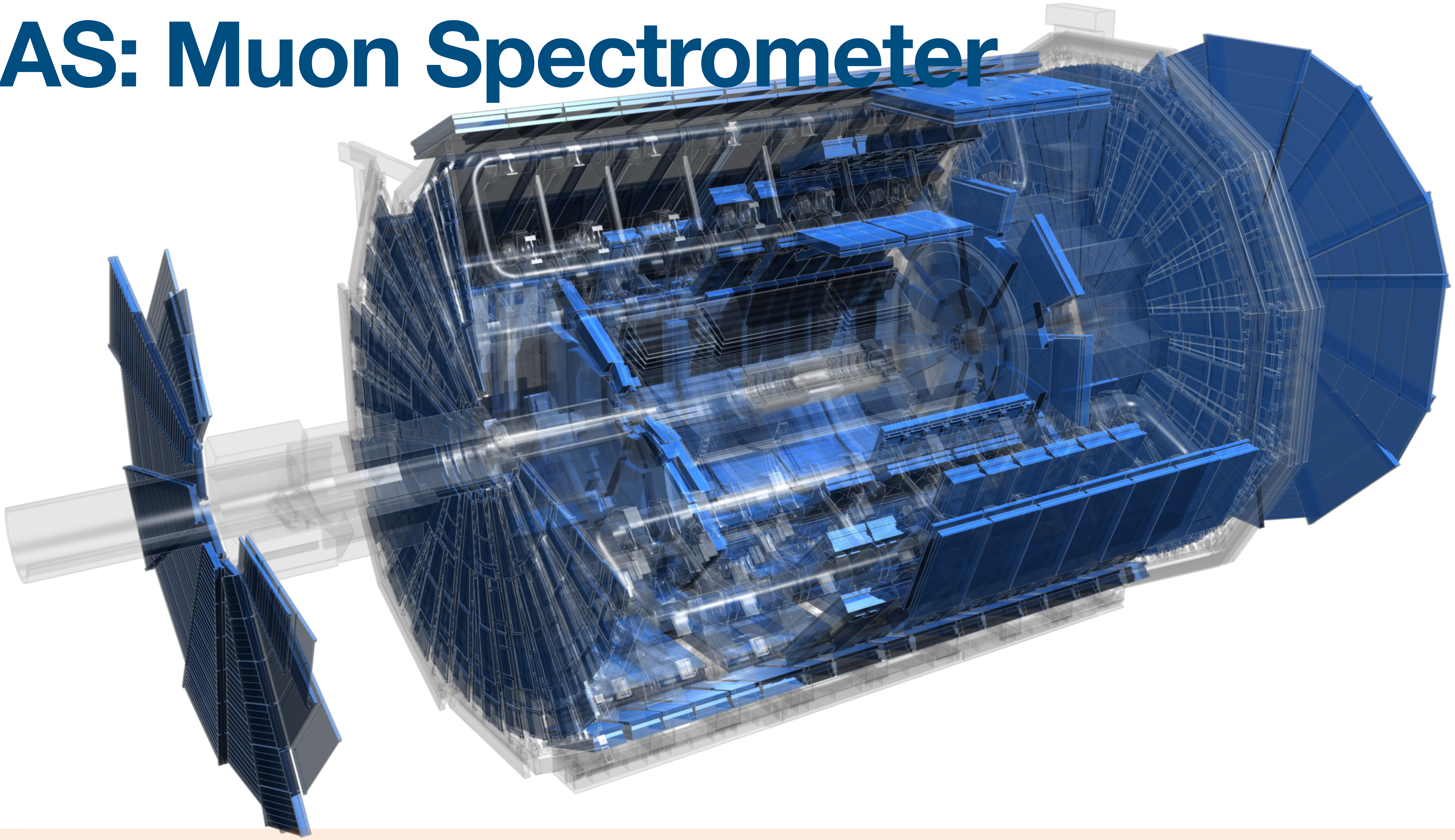
- Measures energy of particles by absorbing them.
- Is made up of Electronic Calorimeters(ECAL) and Hadronic Calorimeters (HCAL)

# ATLAS: Magnet System



- Magnet system bends the trajectory of charged particles to measure momentum and charge.

# ATLAS: Muon Spectrometer



- Measures momentum of muons as they escape the calorimeter without being absorbed.

# Particle Signatures in ATLAS

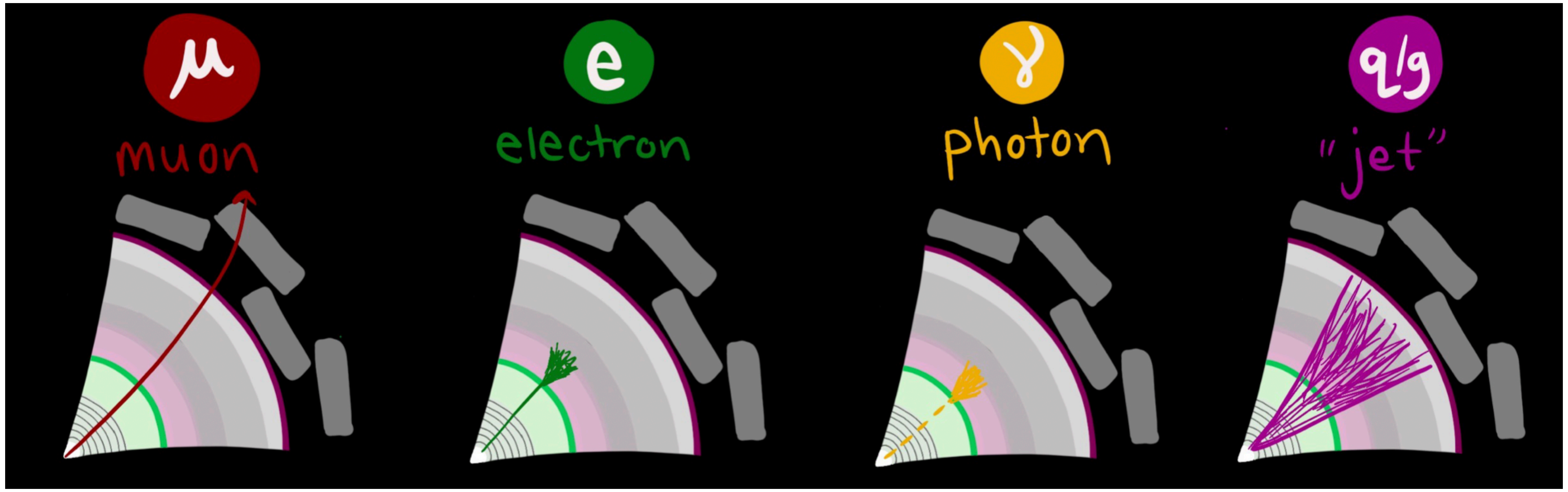



Image: Heather Russel

- Particle identification involves reconstructing the trajectory of charged particles in the ID, reconstruction of the jets from the calorimeters .. and so on.
- **Tracks** are the paths traced by charged particles as they move through the ATLAS detector.
- **Jets** are collimated sprays of particles produced when quarks and gluons (partons), ejected from the proton-proton collisions, undergo hadronization.

# Tracking: Charged particle trajectory reconstruction

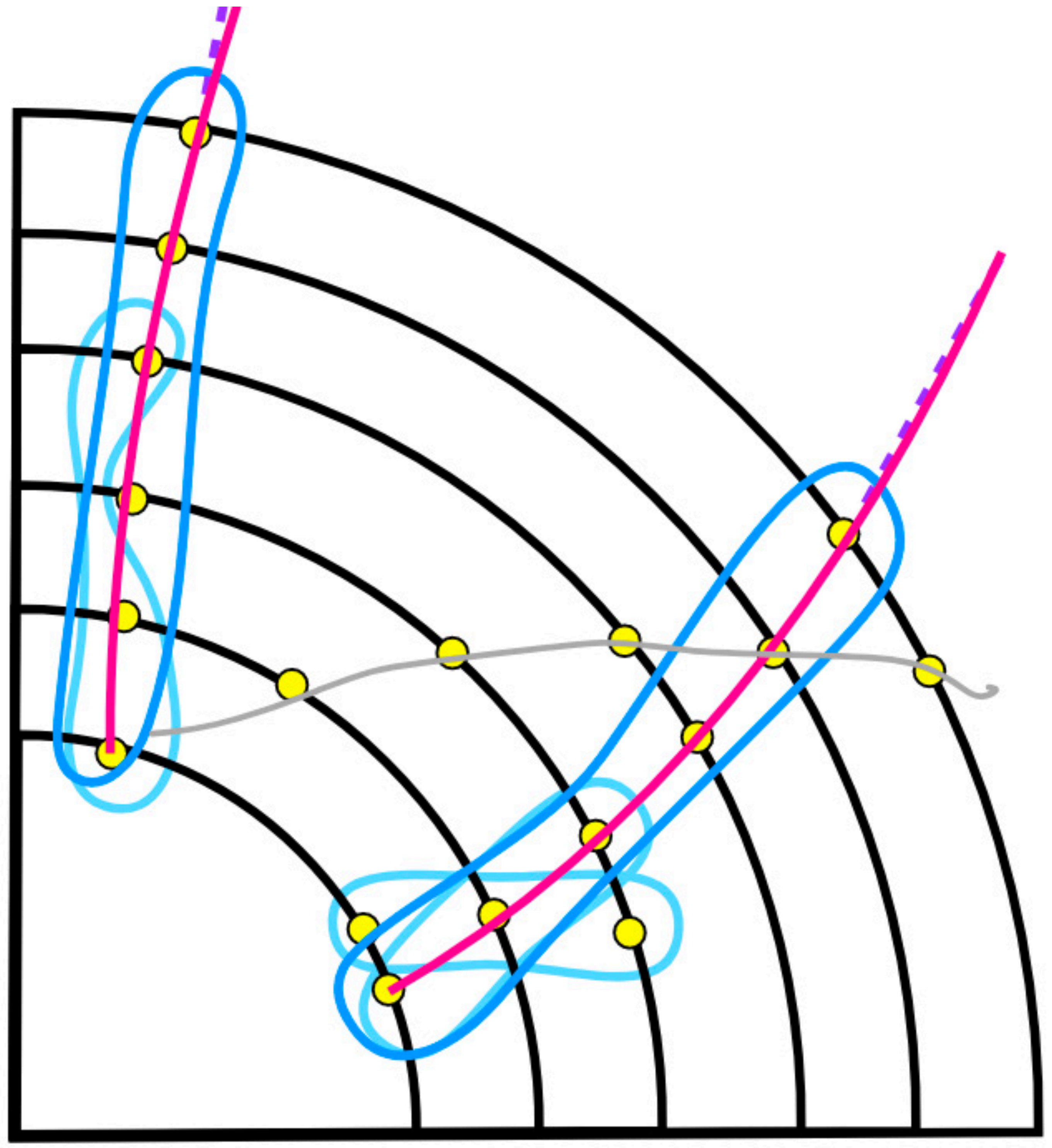
- Form seeds using three hit groups (space points)
- Extend the seeds with additional space points using recursive algorithm
- Fit and score the track candidates using $\chi^2$ and other metrics. Discard bad track candidates based on the score.
- Extrapolate the track candidates to TRT
- Refit with all points and score the track candidate. Also discard candidate with bad score

→ **Inside-Out tracking**

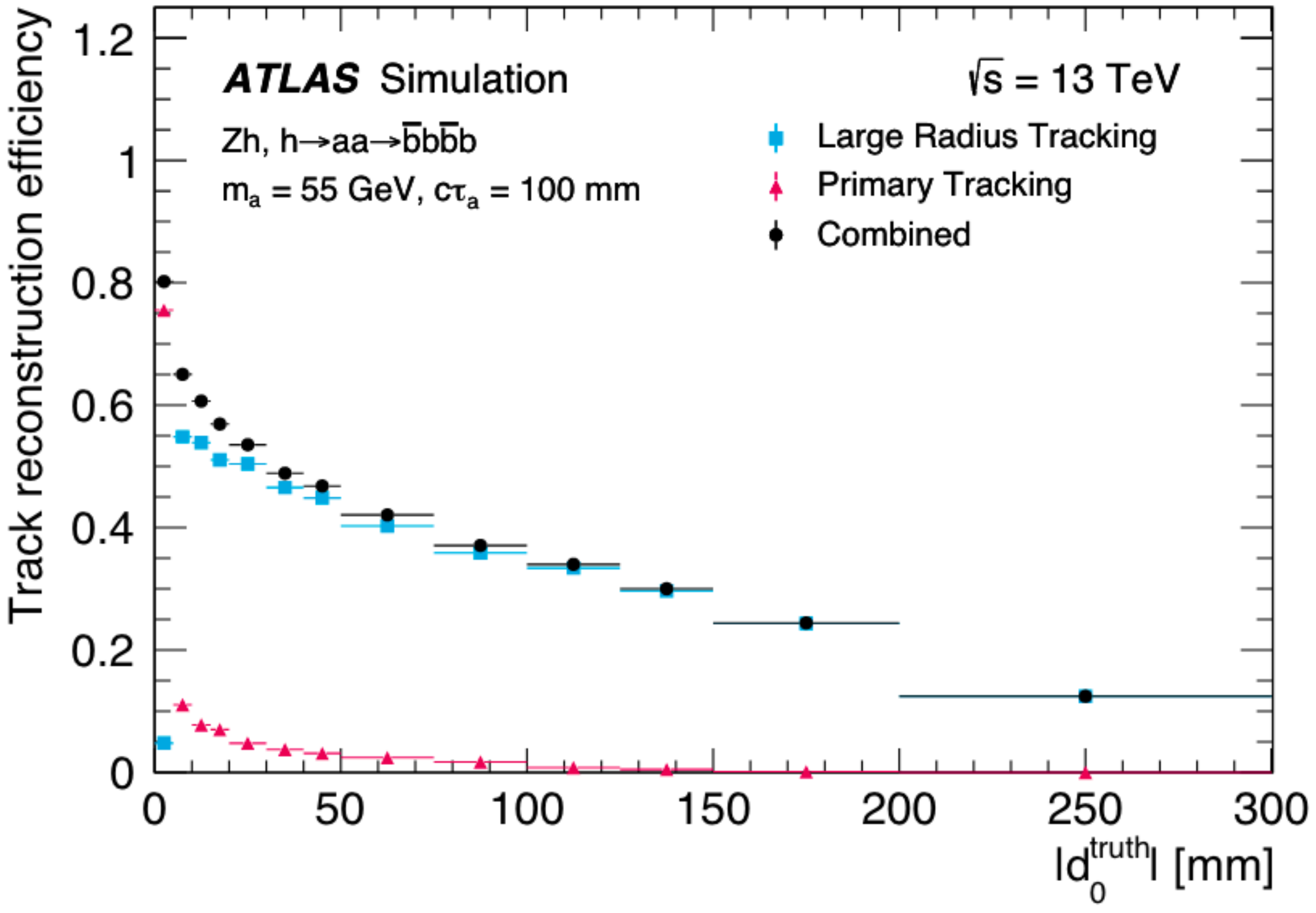
- Form TRT track segments are that extended back to the silicon layers

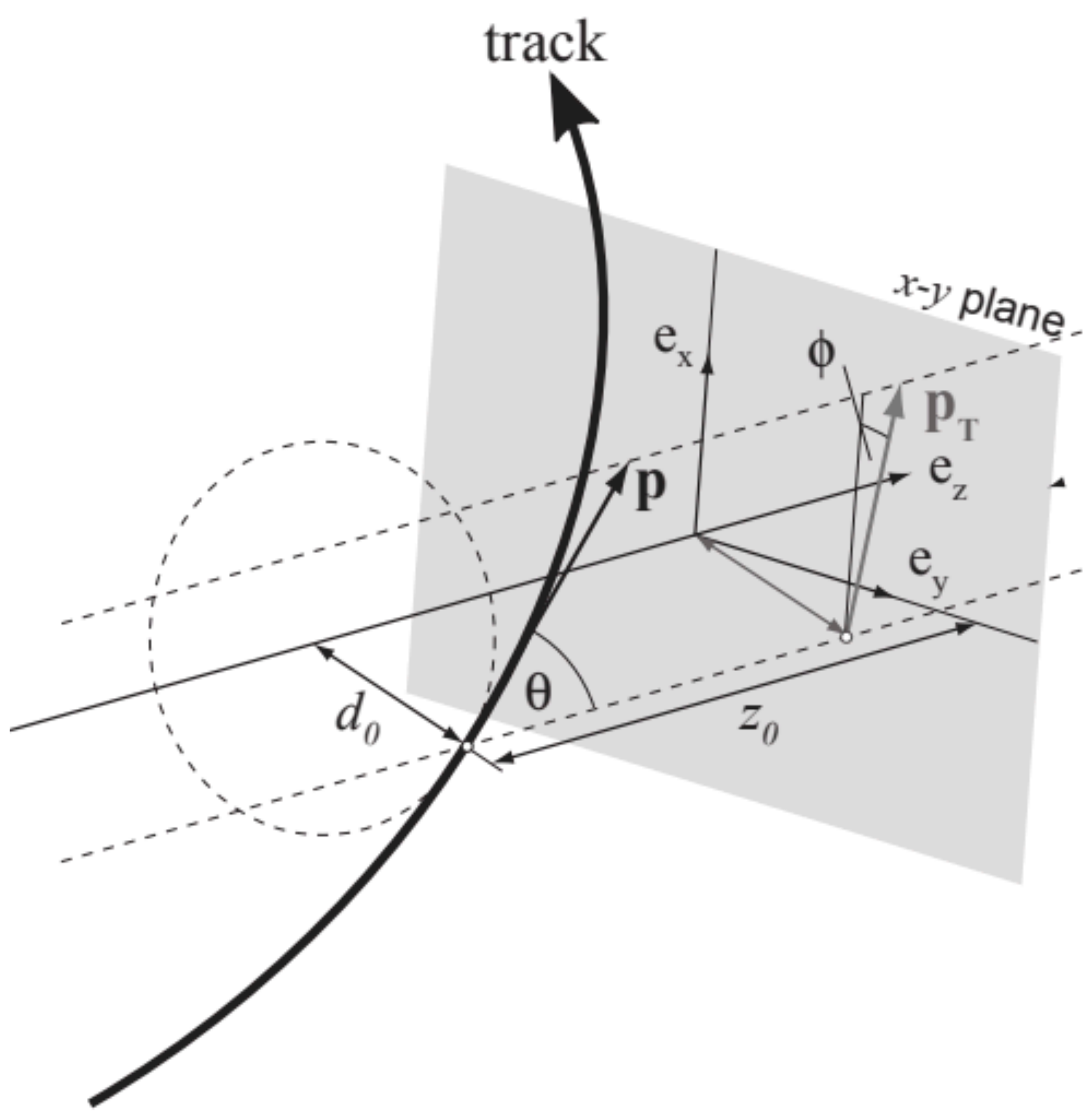
→ **Outside-In tracking**

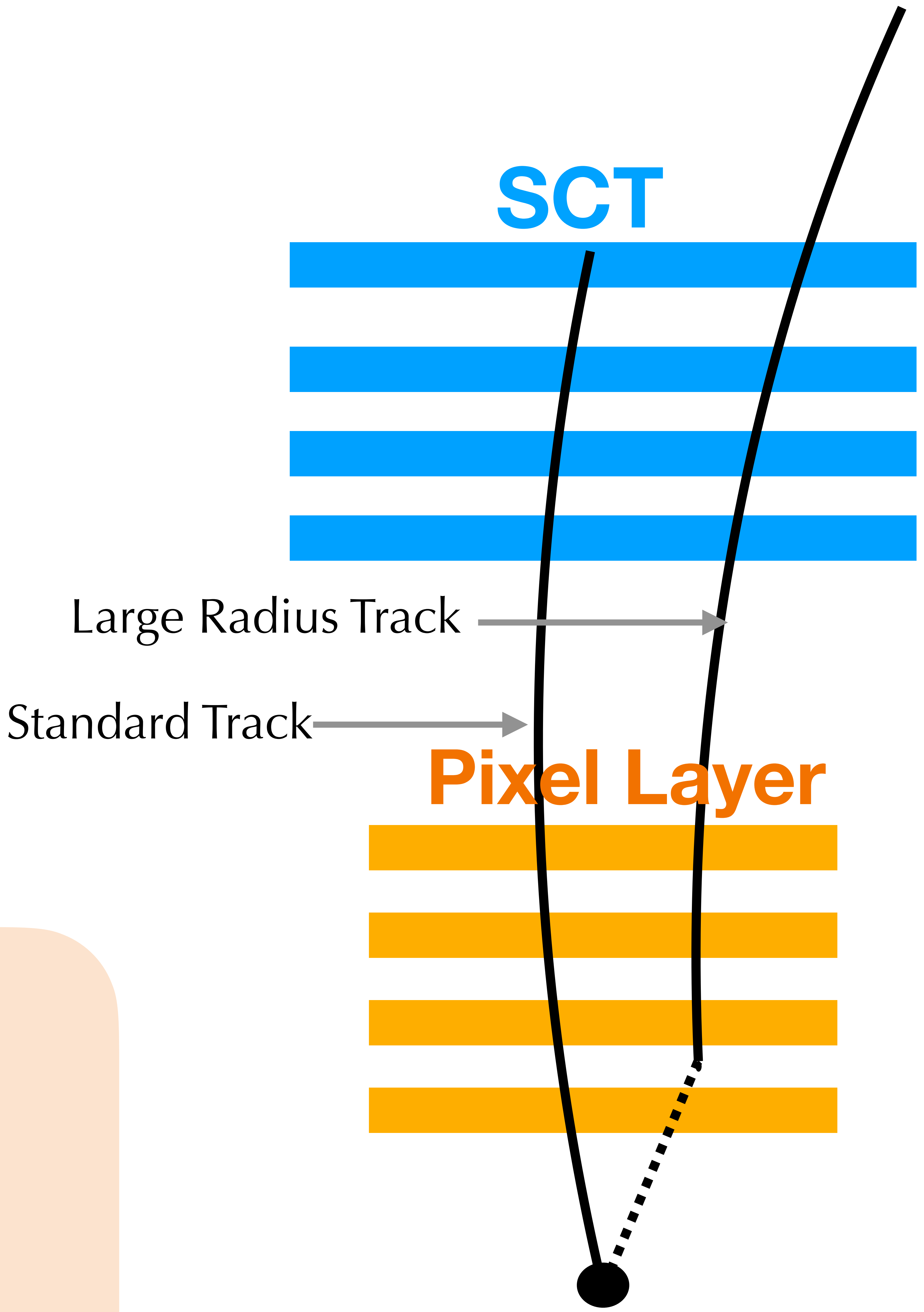
Pierfrancesco Butti



# Large Radius Tracking (LRT)




- Tracks inside EJs are from LLP particle decay. Standard tracking (ST) cannot reconstruct those tracks efficiently.
- LRT run after ST and manages to retain substantial efficiency unto transverse impact parameter < 300 mm

# Jets Reconstruction

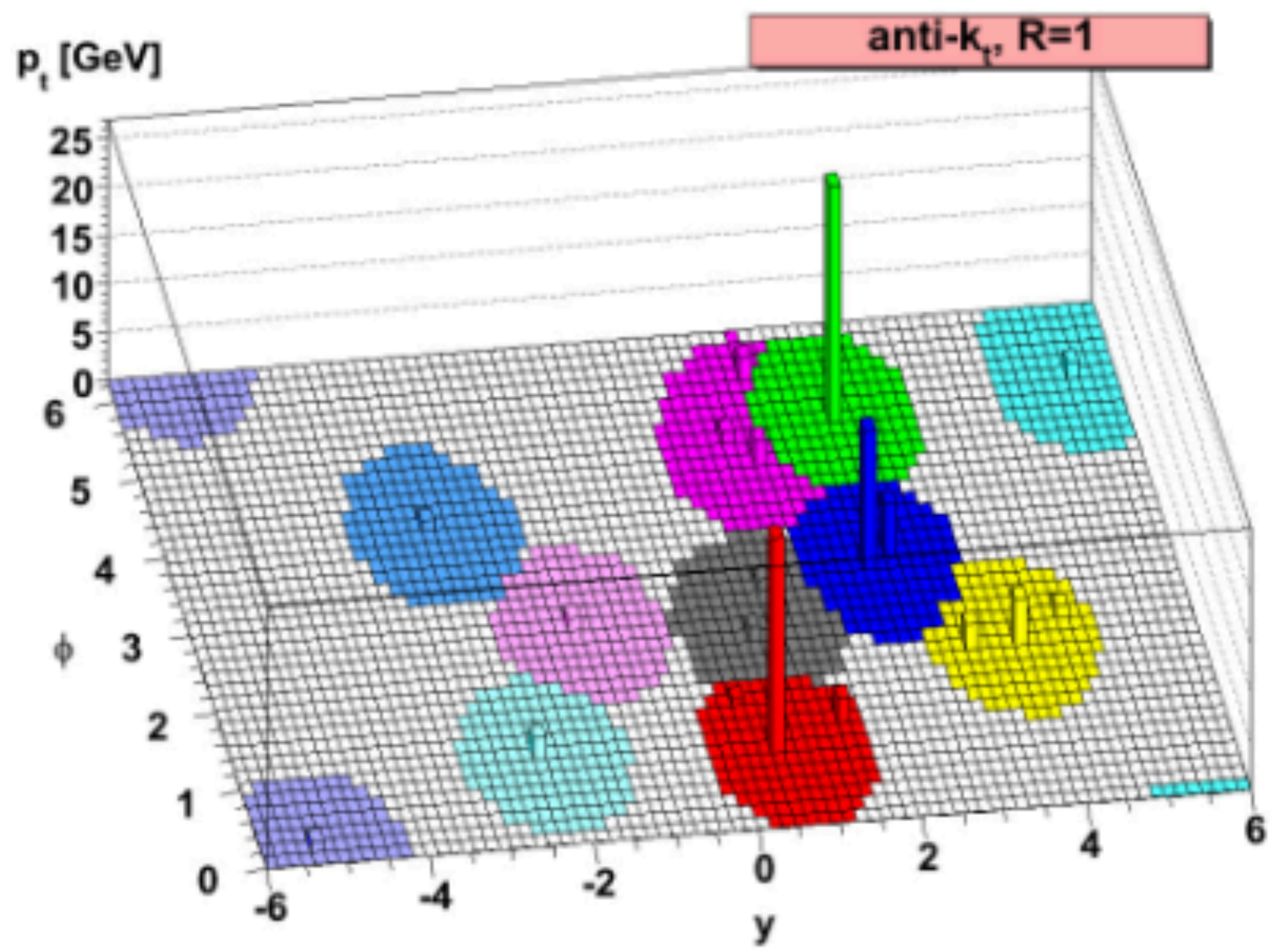



Figure A sample parton-level event clustered with the anti-kt algorithm used to reconstruct jets with R=1.

- First, calorimeter cells are grouped into three-dimensional clusters (topo-clusters) using the nearest-neighbour algorithm.

- Then clusters are merged based sequential recombination algorithm (anti-kt), meaning it builds jets by iteratively merging particles based on a specific distance metric.

# Emerging Jets in ATLAS detector

- ◦ EJ's are BSM LLP signature!
- **EJs are jets with many displaced tracks and displaced vertices.**
- Difficult to identify!
    - Calorimeter signature looks similar to a QCD jet
    - Need to use the displaced tracks and vertices to identify the EJ using conventional methods

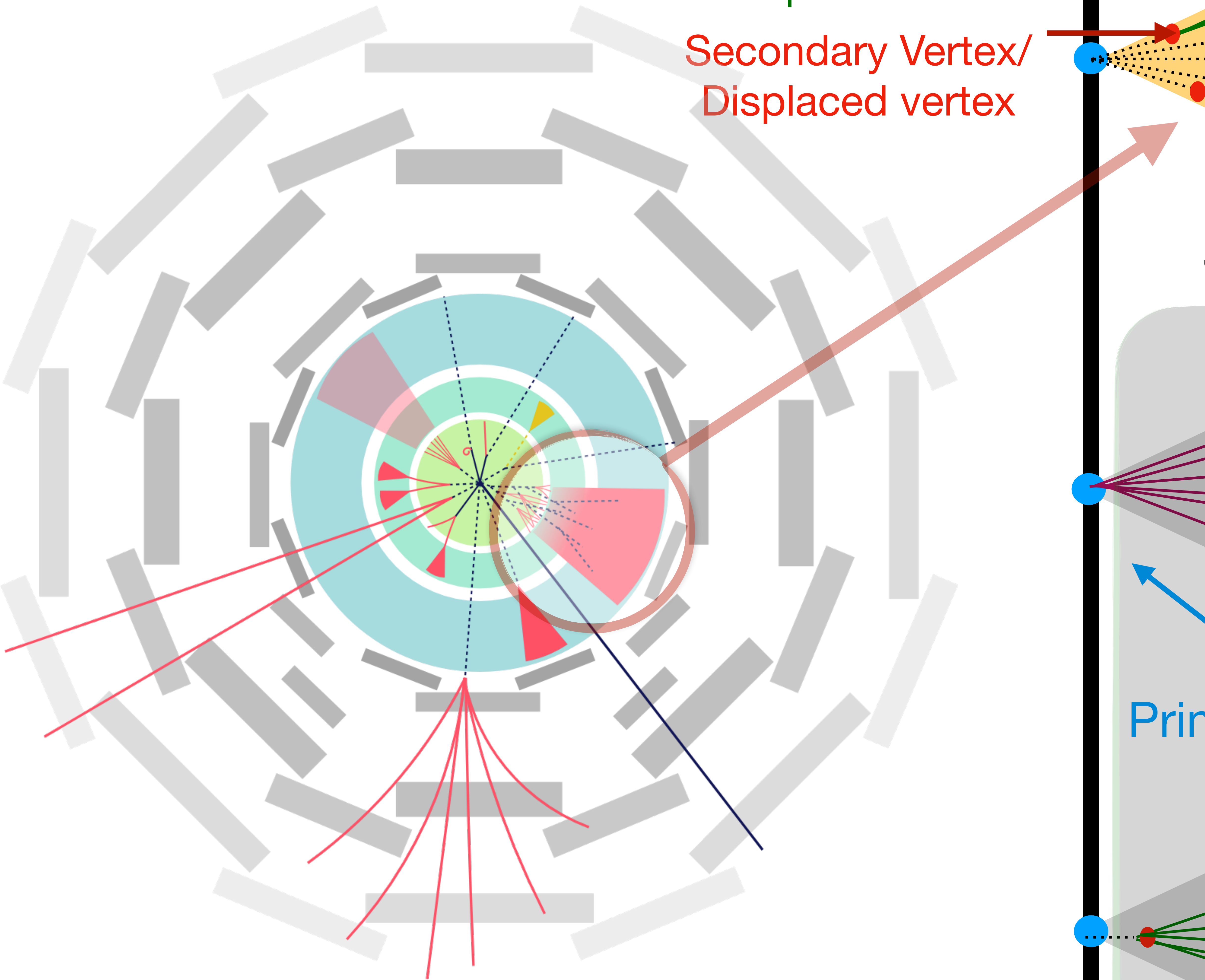
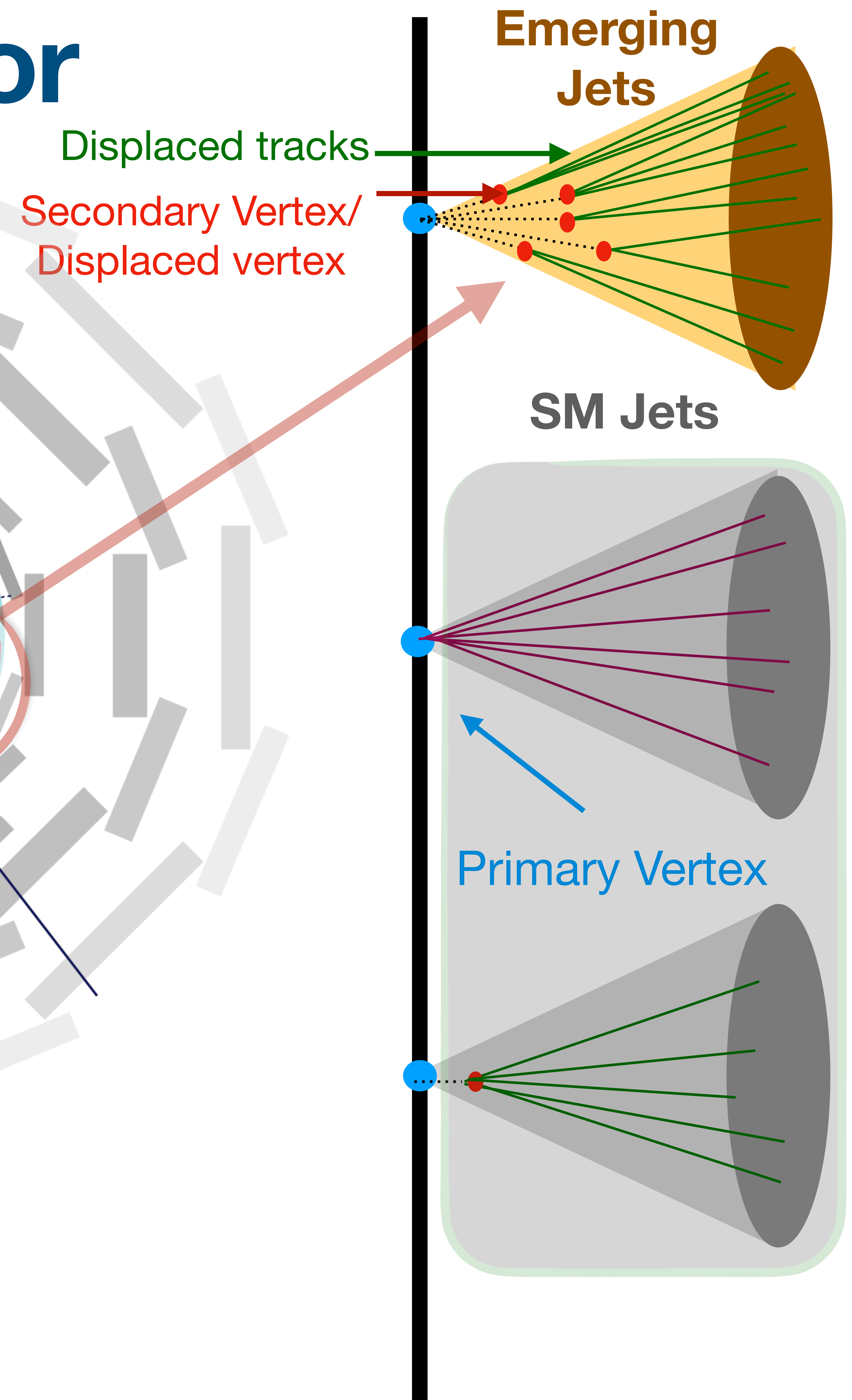
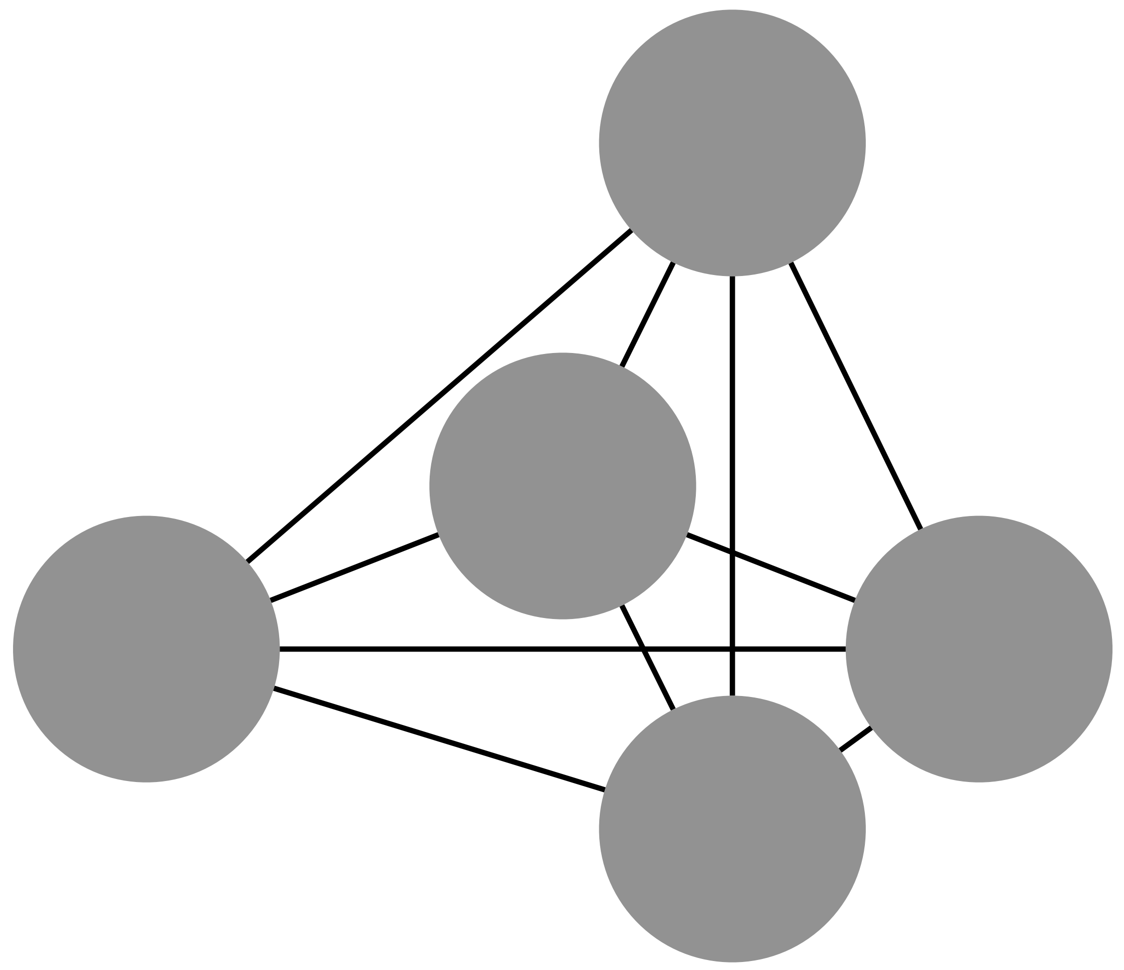


Image: Heather Russel

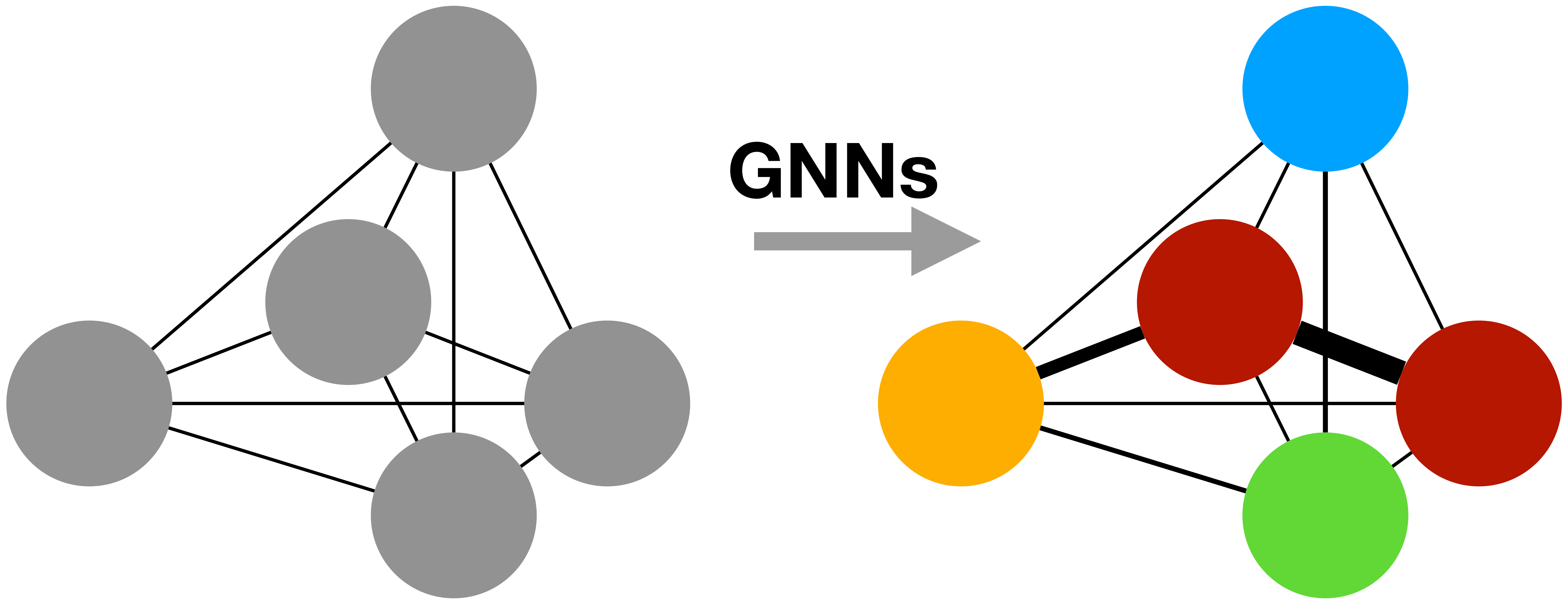


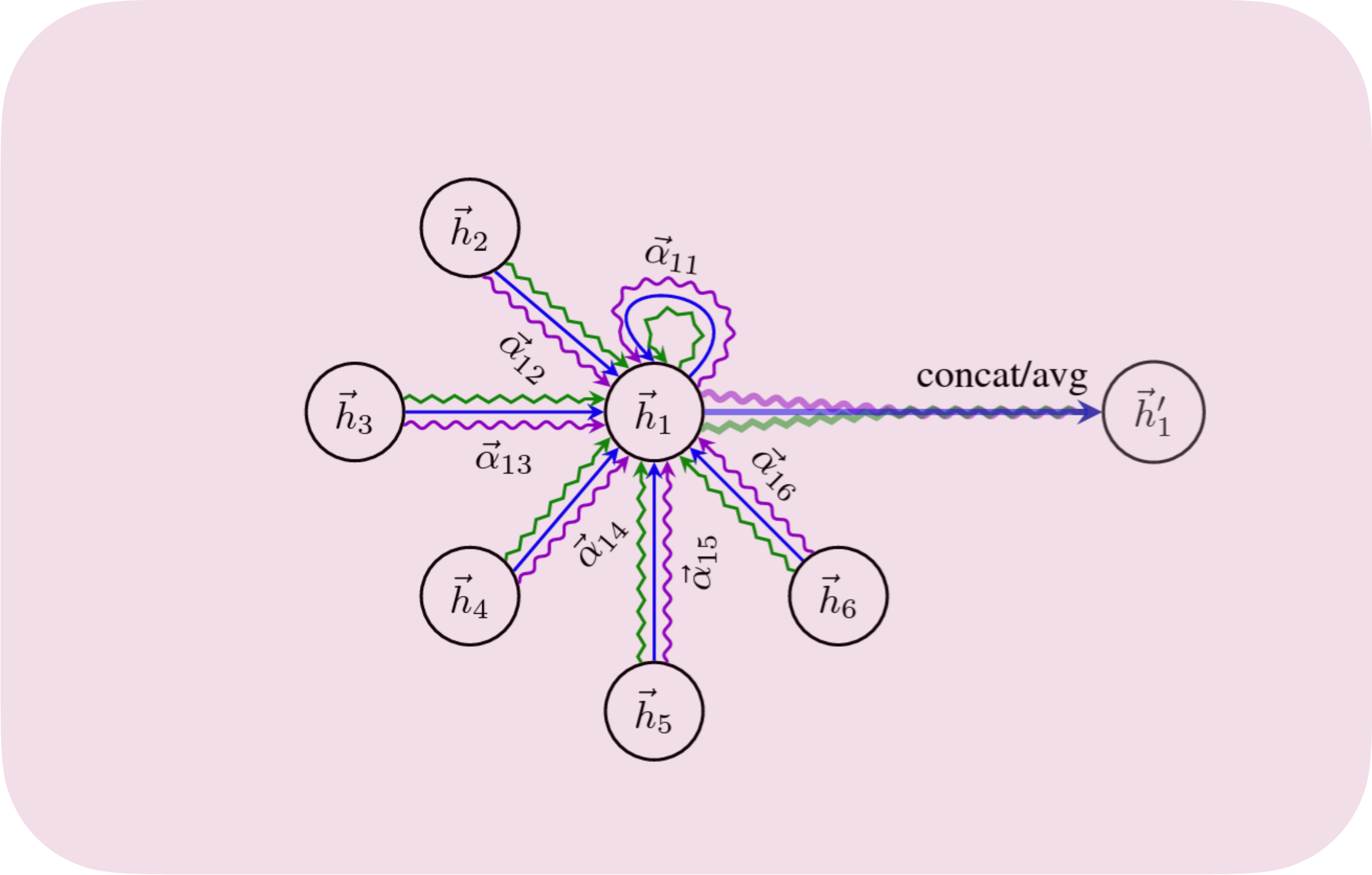
# Graphs



- A graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is COLLECTION of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$and edges$\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$.
- Nodes are often used to represent multi-dimensional feature-vectors. Feature vectors are numerical representations of data entities and denoted as $\mathbf{x_u}$ for u $\in \mathcal{V}$

# GNNS

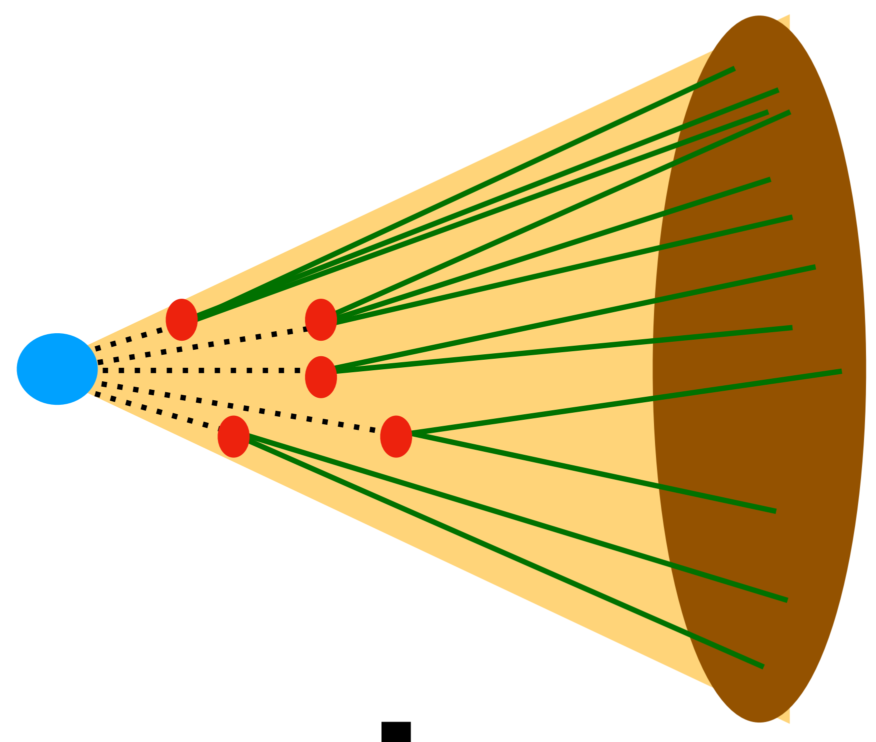


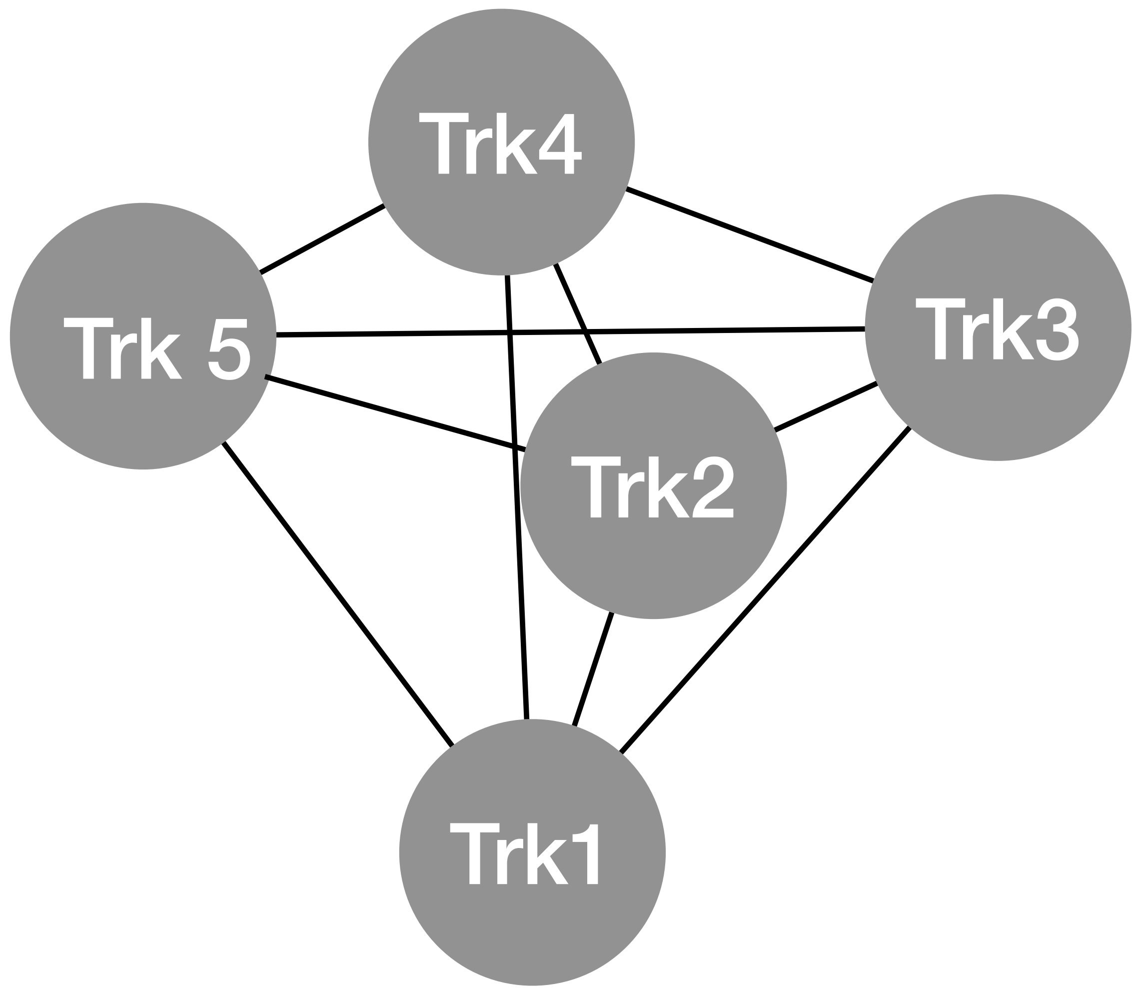




Transformation of node feature vector $\overrightarrow{h}_1$ into $\overrightarrow{h}'_1$ by using neighbourhood feature vectors .

- Optimizable transformation on graph attributes such as nodes and edges.
- For example, transformation's of node representation $\overrightarrow{h}_1$ to $\overrightarrow{h}'_1$ through a weighted aggregation of its neighbour's representation, where the weights are derived from attention mechanism, $a(\mathbf{x}_u, \mathbf{x}_v)$.
  - $\mathbf{h}_u = \phi\left(\mathbf{x}_u, \square_{v \in \mathcal{N}_u} a(\mathbf{x}_u, \mathbf{x}_v)\psi(\mathbf{x}_v)\right)$, where $\psi(\mathbf{x}_v) = \mathbf{W}\mathbf{x}_v$

# GNNs for Emerging Jets Analysis (Run 03)



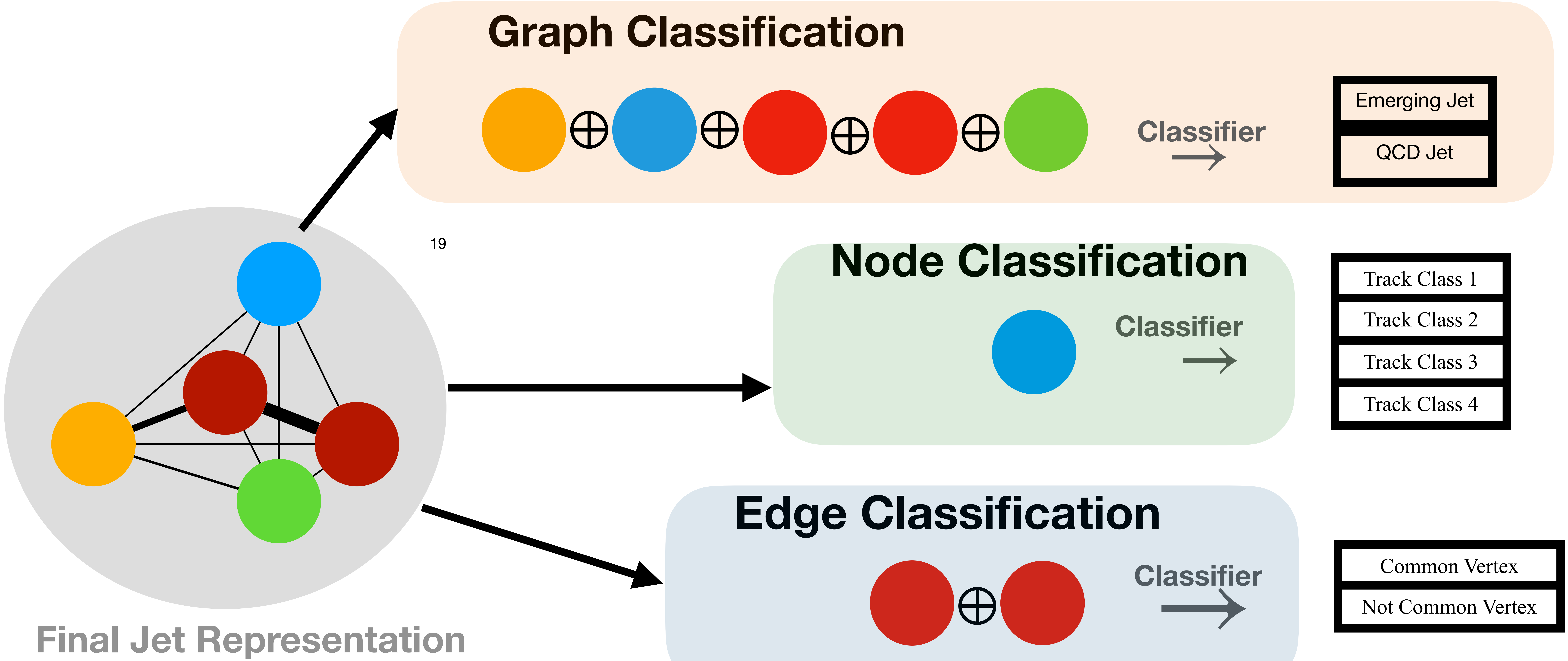



- GNNs can handle large sized inputs because of permutation symmetry.
- EJs have large number of tracks inside them
- As EJs are difficult to identify using conventional methods, GNNS facilitate the use of several low level track input variables.

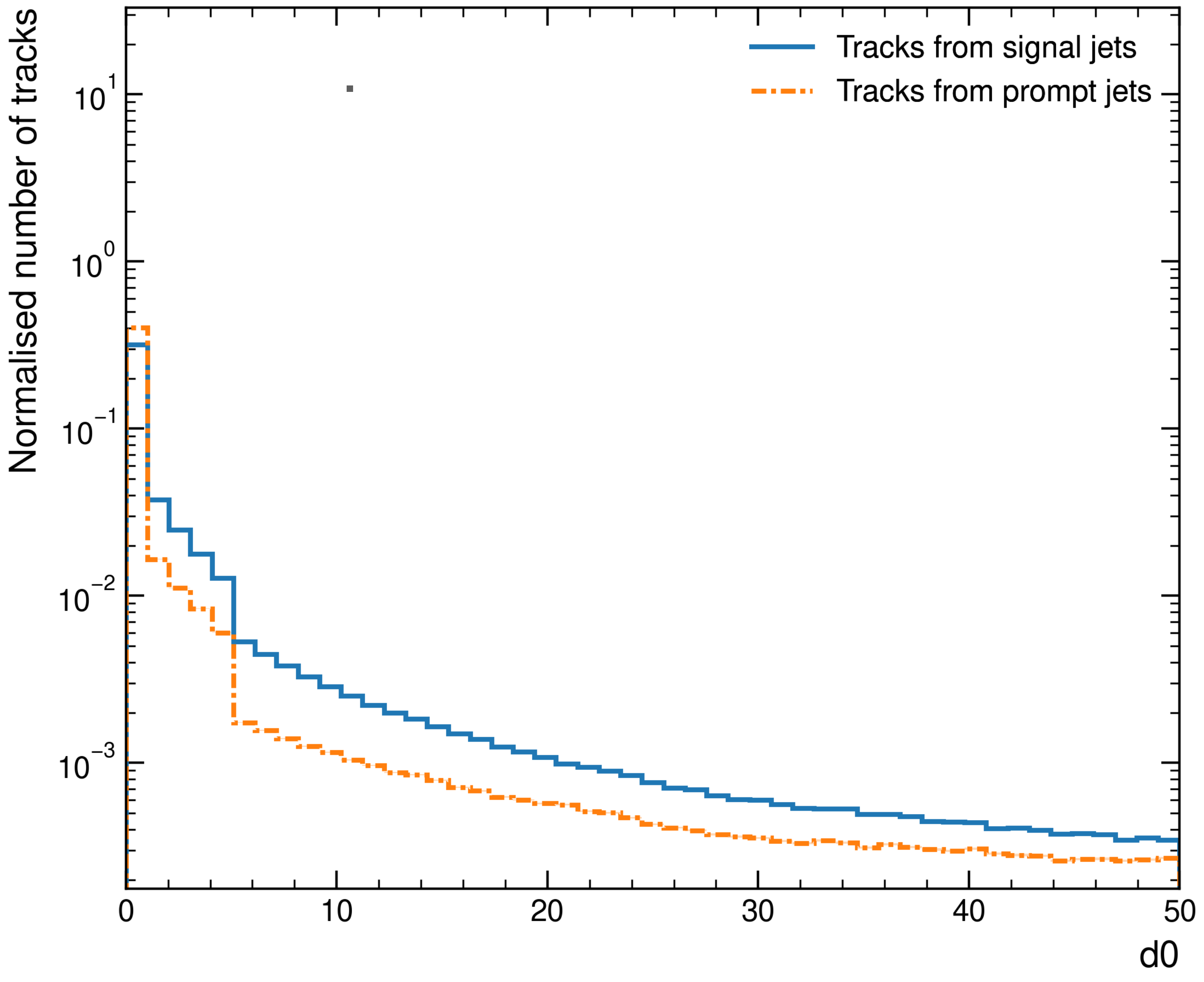
- GNNs handle irregular sized inputs
- Number of tracks in EJ is not fixed, therefore well suited

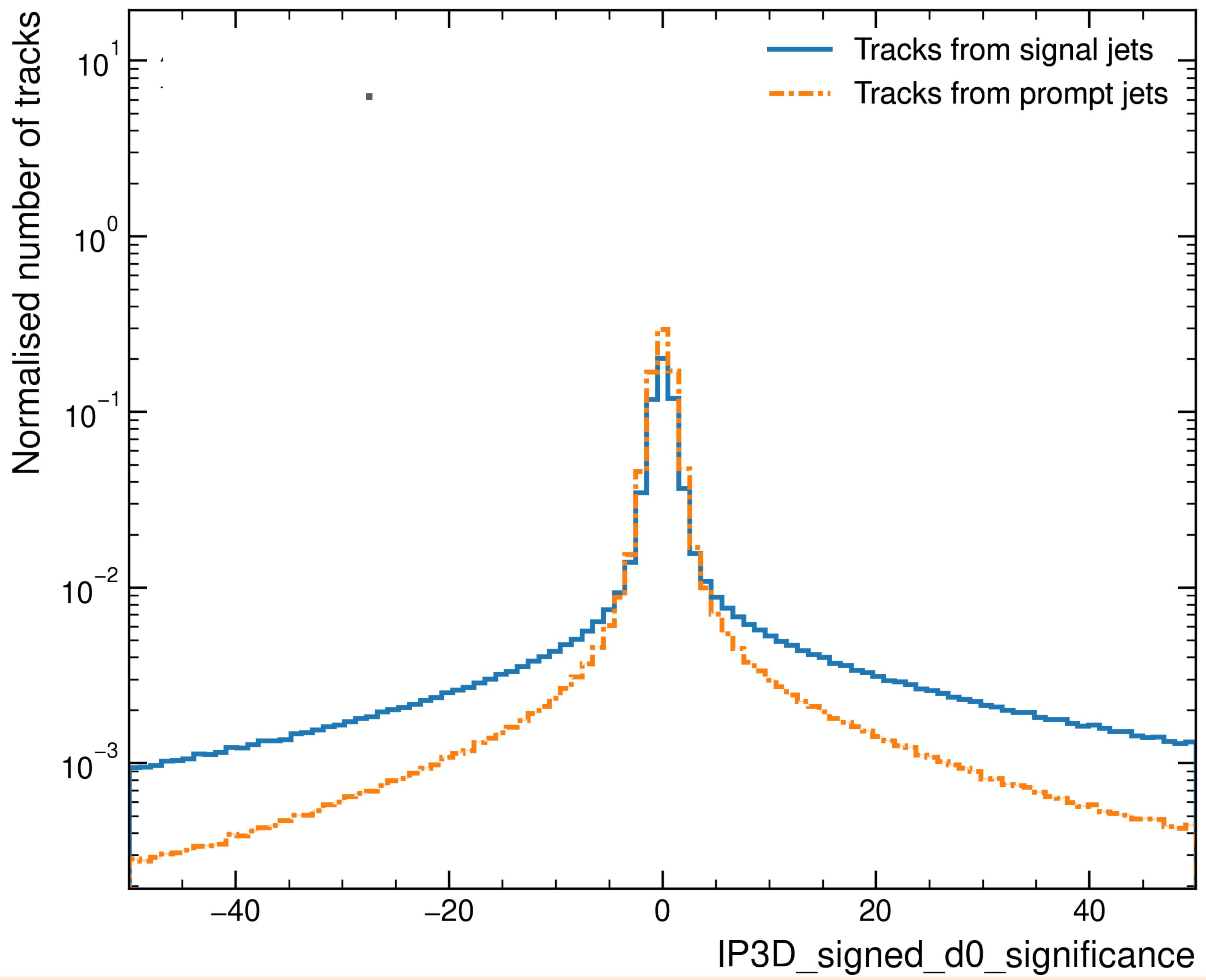
- GNNs exploit relationship between data entity
- Tracks in EJ exhibit rich relations due to the presence of multiple displaced vertex and displaced tracks

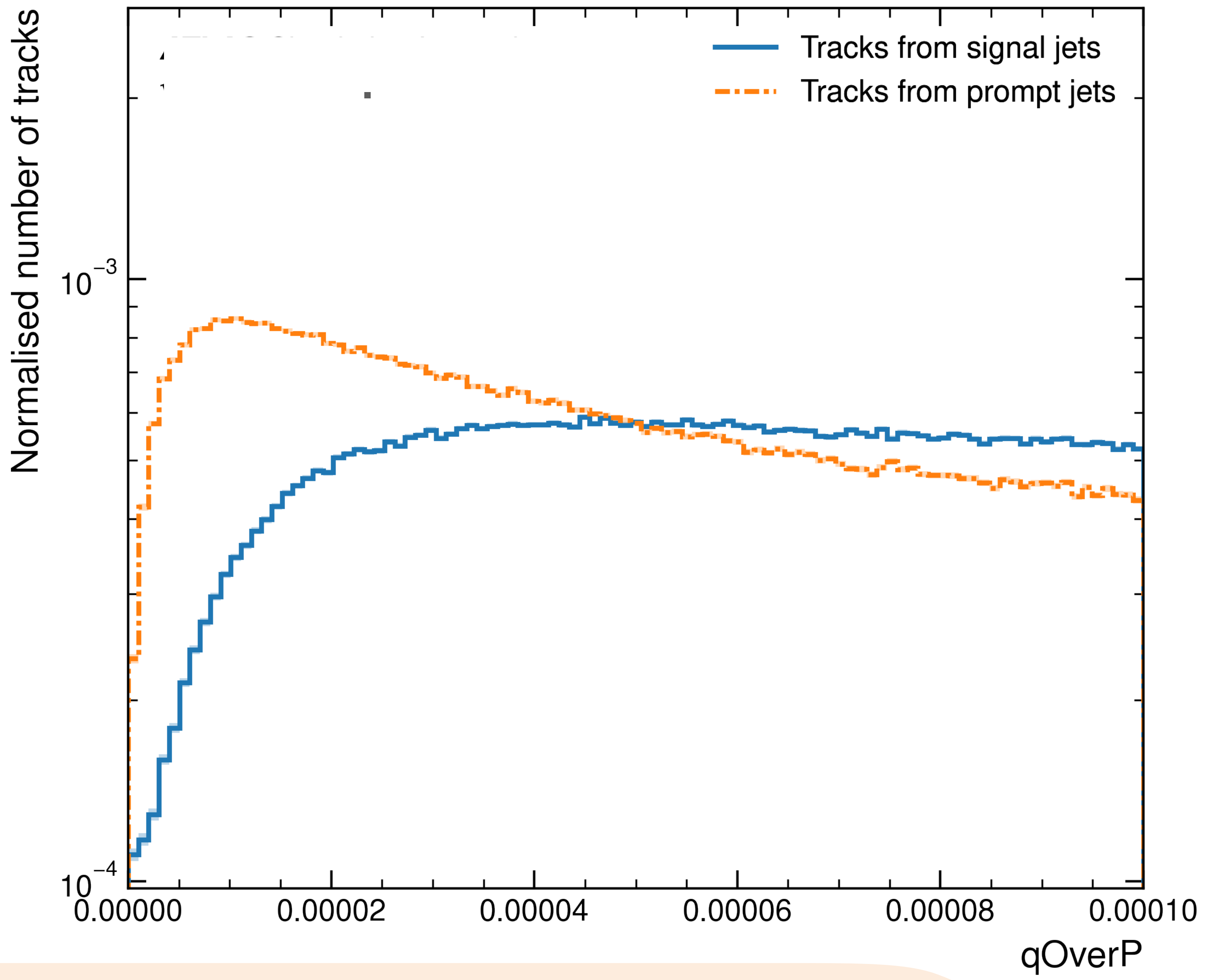
# Classification tasks



# Input Variables




- 16 track variables including track parameters in ATLAS tracking system, detector hits and holes variables, uncertainty in track parameters … (detailed in backup slides)

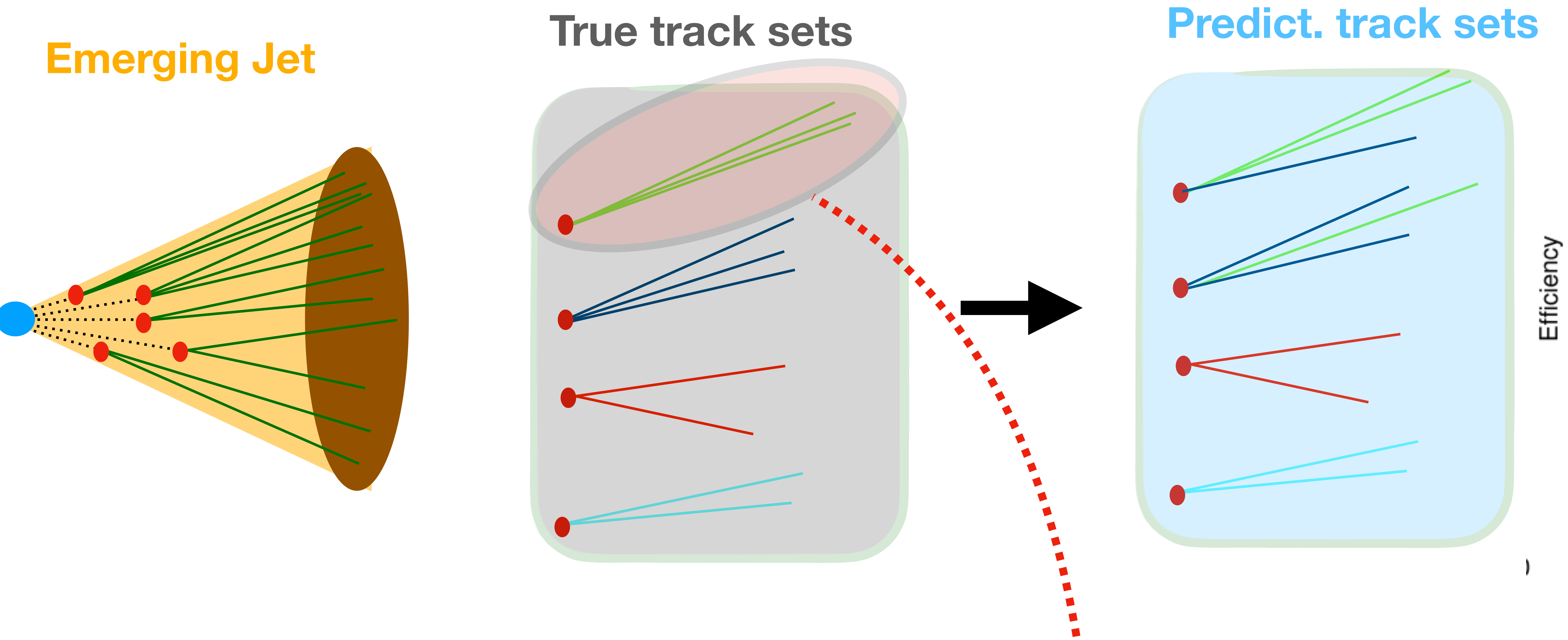
- Most discriminating ones include

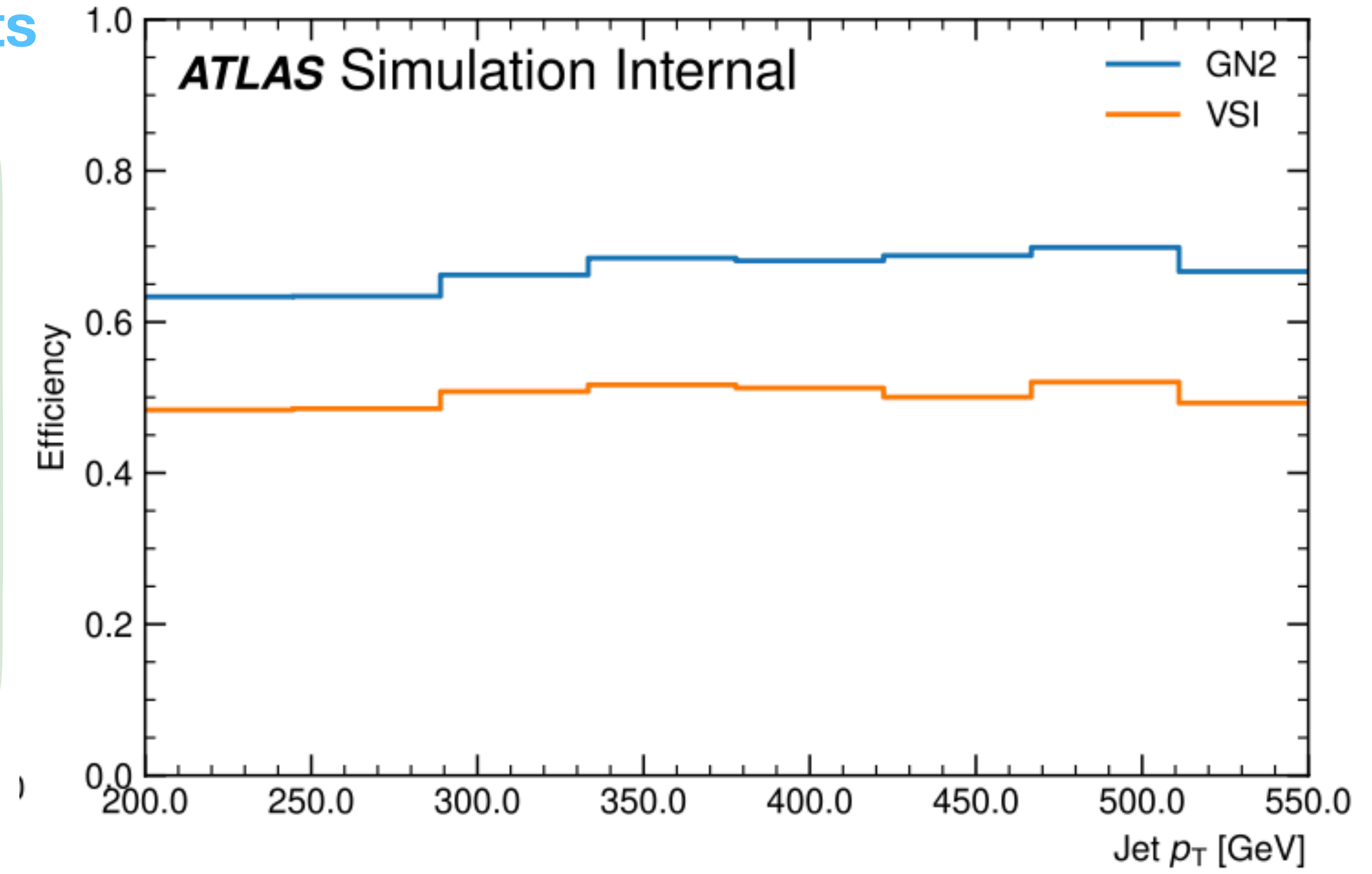
- \- $d_0$: Distances of closest approach between the track
  \- IP3D_signed_d0_significance:  Ratio of $d_0$ and $\sigma(d_0)$ defined for both positive and negative scale with reference to the primary interaction point of the ATLAS detector
  \- $\frac{q}{p}$ Track charge divided by momentum (measure of curvature)

# Results from Performance of GNNS in Vertex Classification


Probability of having common vertex

# Vertex Performance: Efficiency

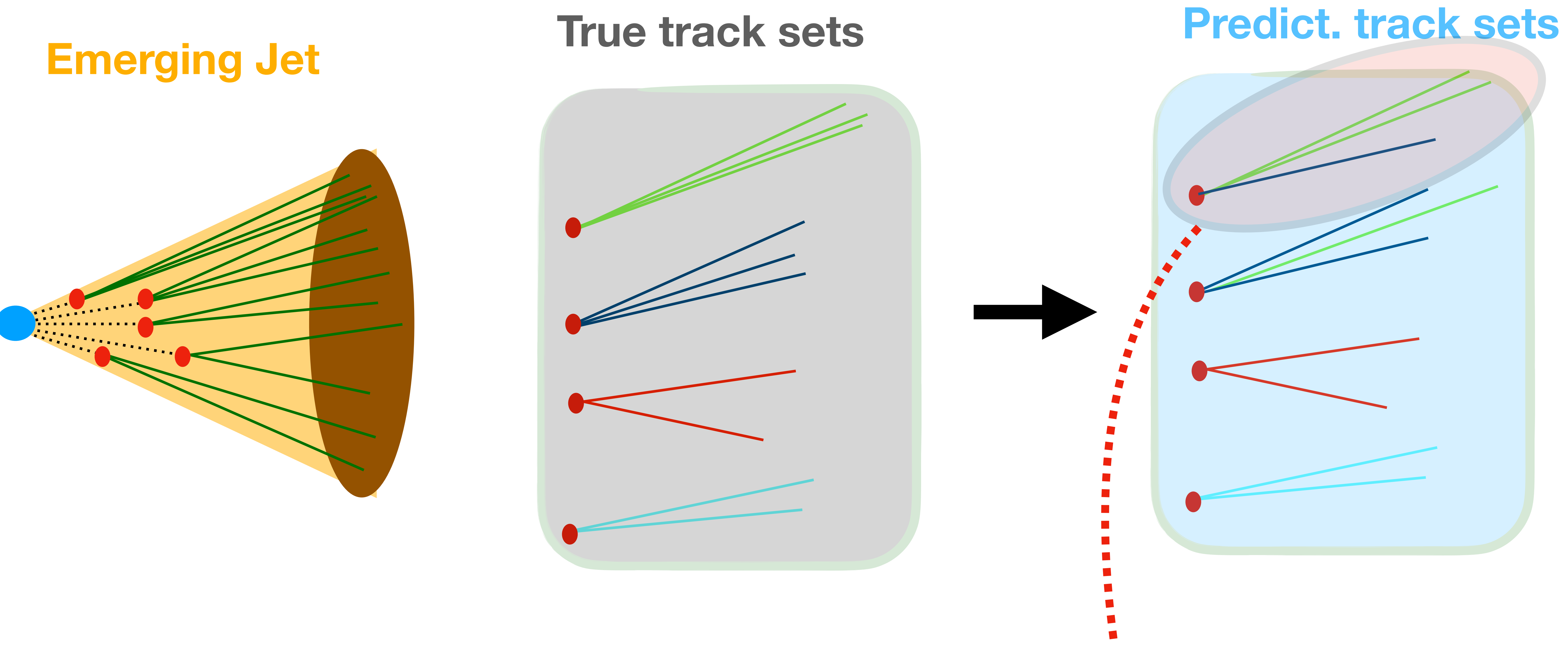


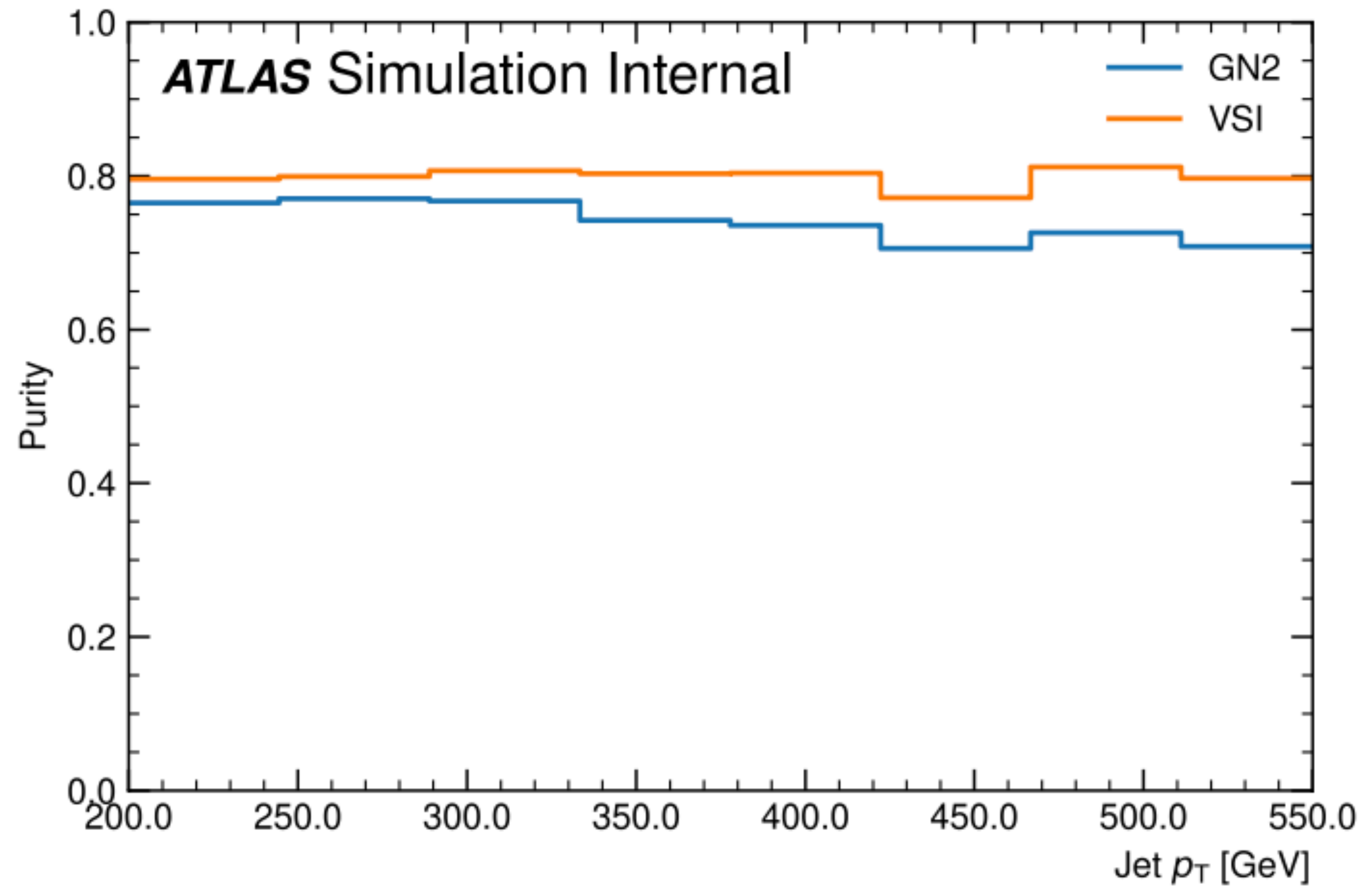




*Are these vertex efficiently reconstructed?*

- Efficiency: Per-vertex fraction of tracks in the truth-vertex which are included in a common reco-vertex!
- GNNs have higher efficiency then VSI

# Vertex Performance: Purity

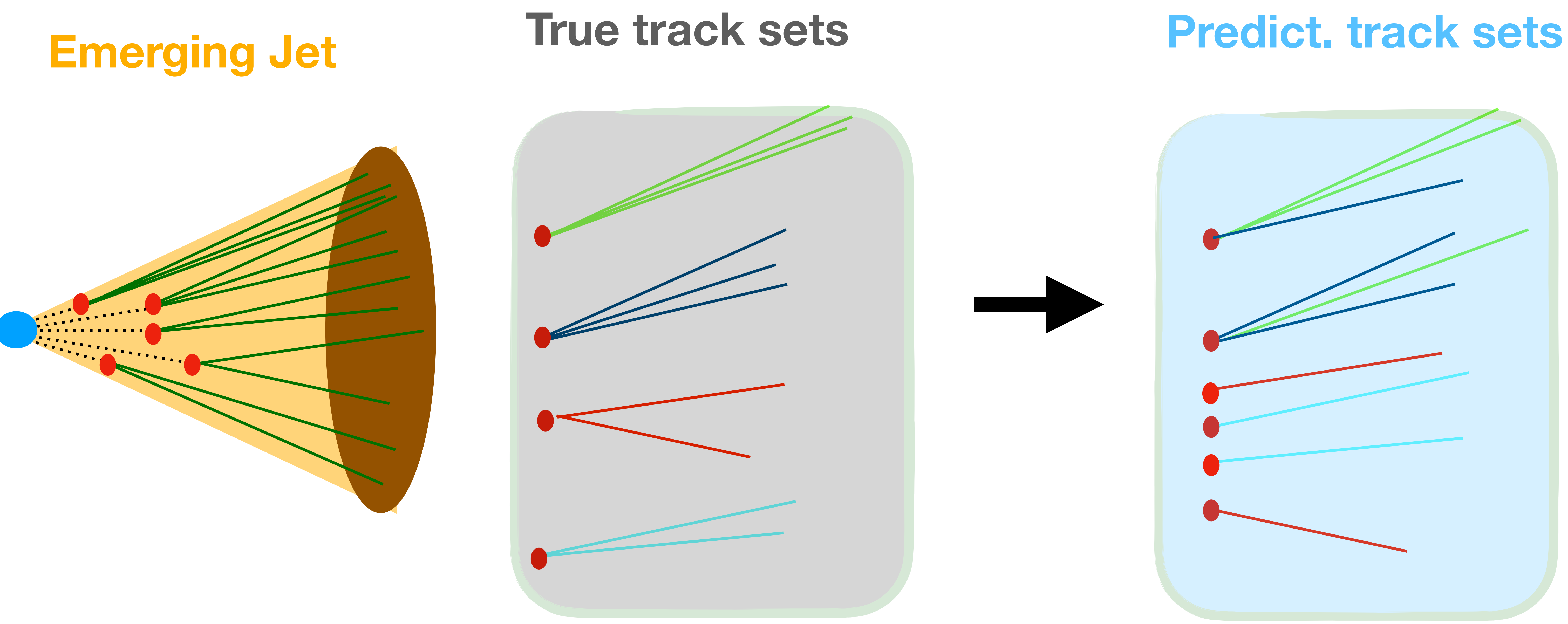


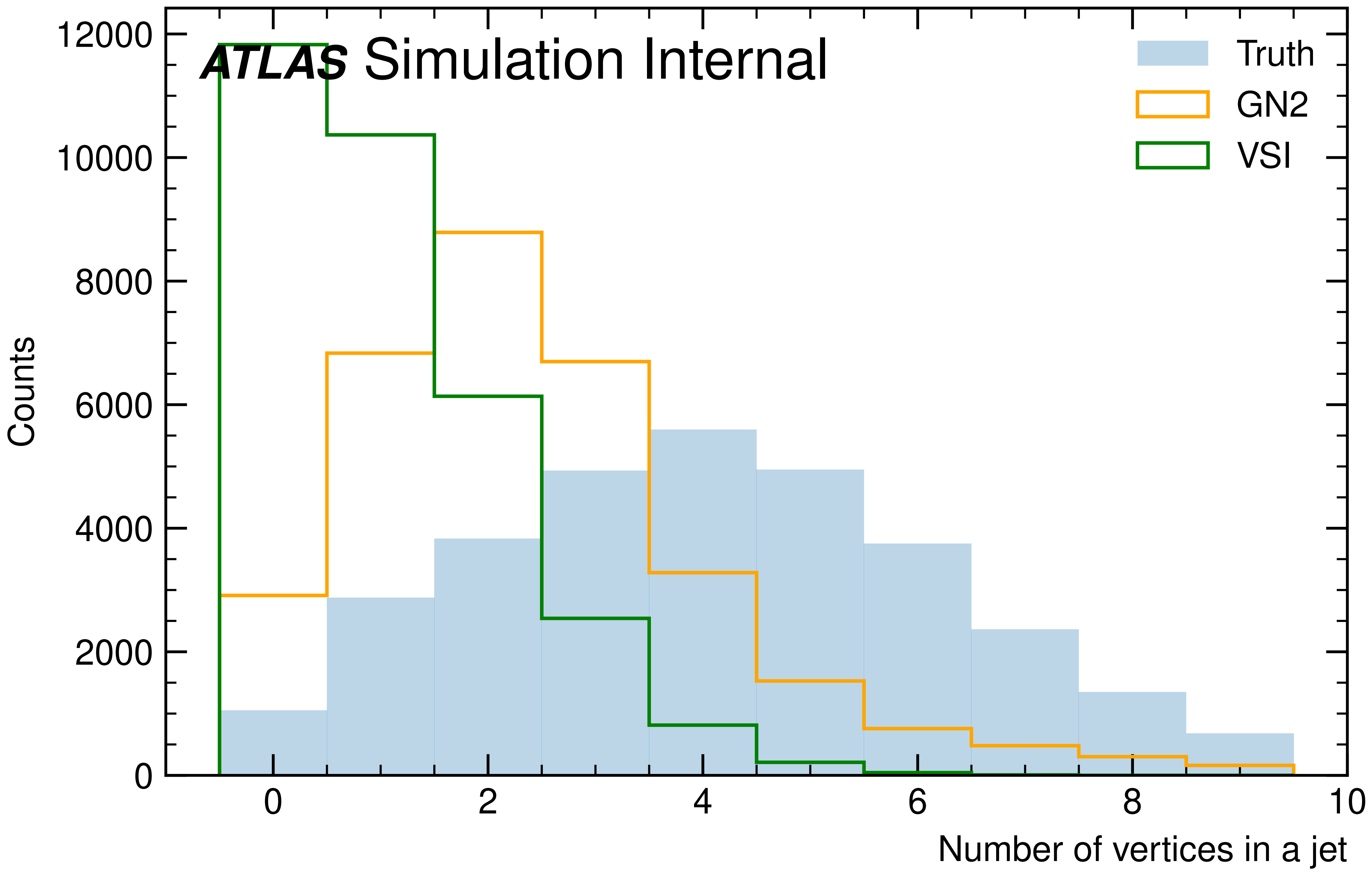




*How "PURE" are these groupings?*

- Purity: Per-vertex fraction of tracks in the reconstructed vertex which are from the same truth vertex.
- GNN predicted vertex and VSI have similar purity.

# Vertex Performance: NumVertex Dist.

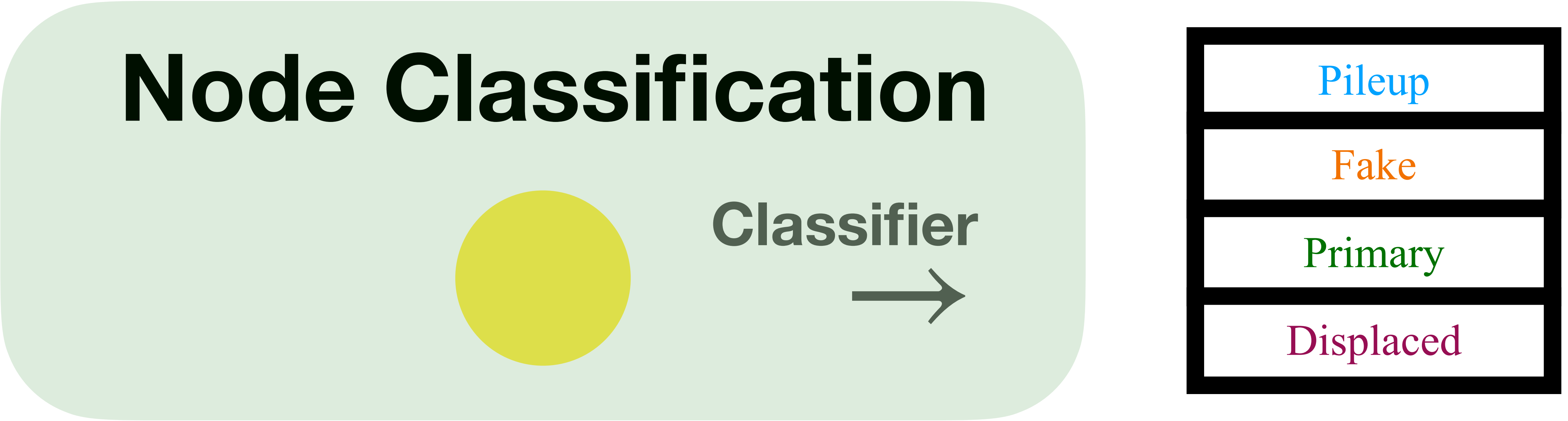



*How many vertex identified?*

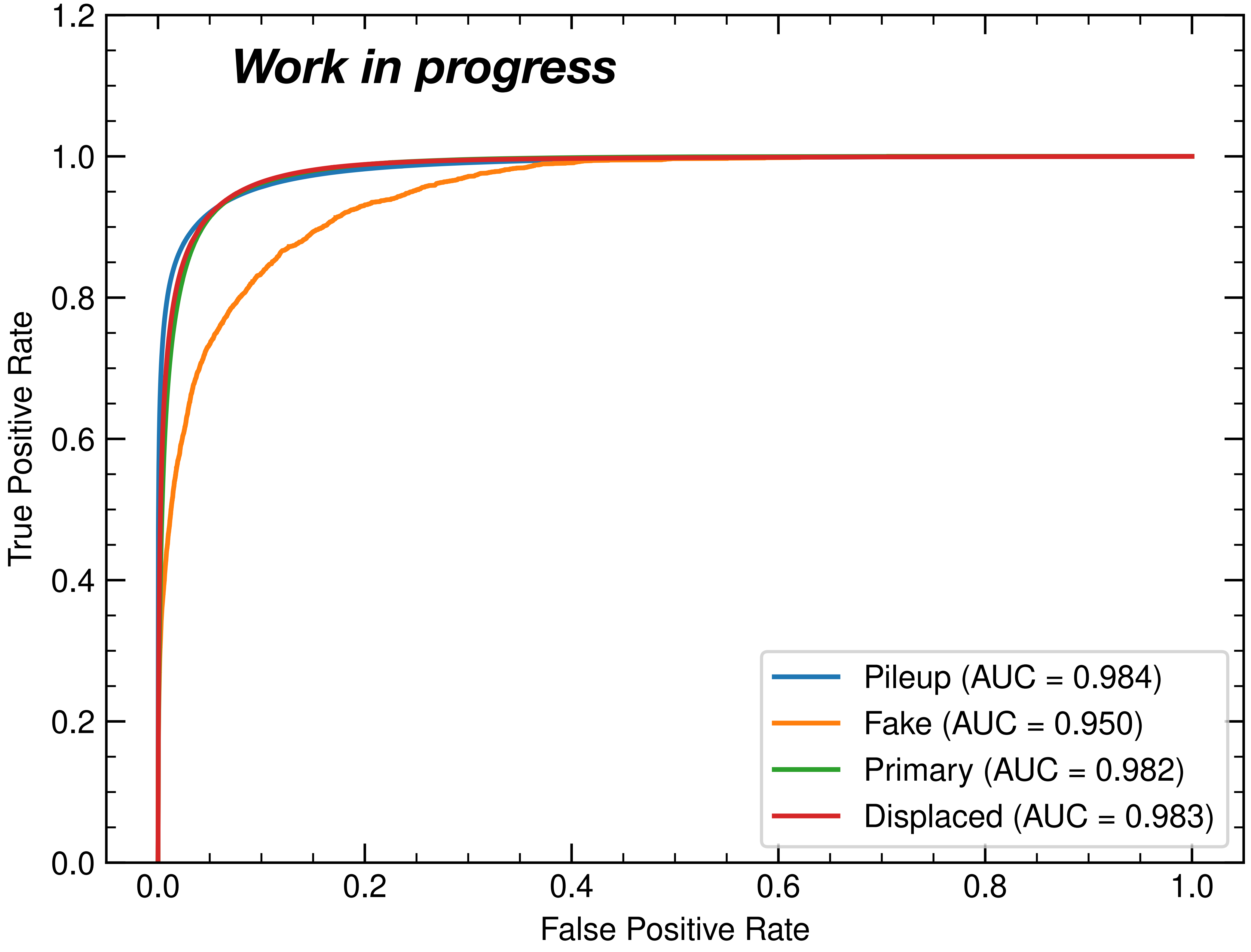
- Emerging jets, by definition, has multiple vertices in a jet.
- Number of vertex in per jet distribution shows jet topology identified by GNN closer to the truth.

# Results from Performance of GNNS in Track Classification




- Pileup: From additional proton-proton interactions that occur within the same bunch crossing
- Fake: From purely combinatorial collections of hits
- Primary: From Primary Vertex
- Displaced: From Secondary vertices

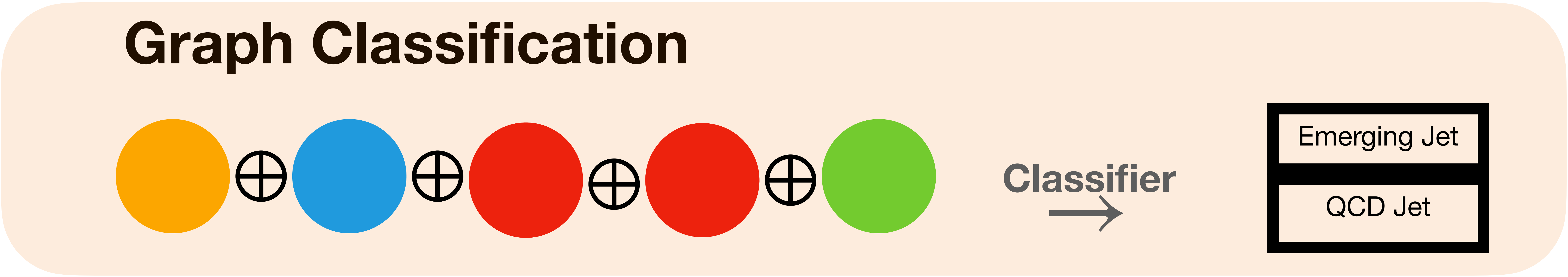
# GNNs Performance: Track Origin Classification (ROC)



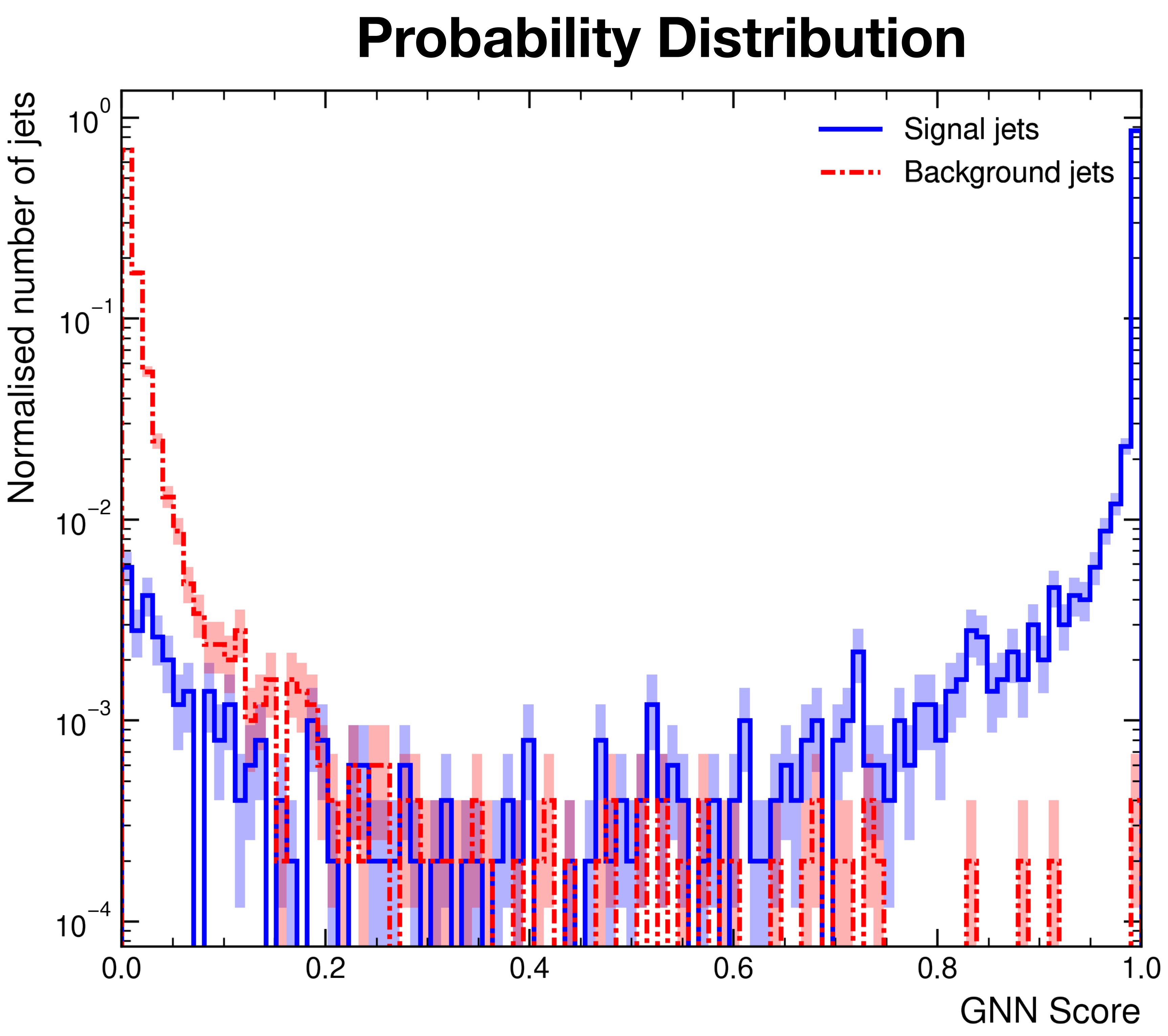

- FPR: proportion of actual negatives that are incorrectly identified as positives
- TPR: proportion of actual positives that are correctly identified

- Highly effective in classifying tracks!
- Displaced tracks classification AUC: 0.983!

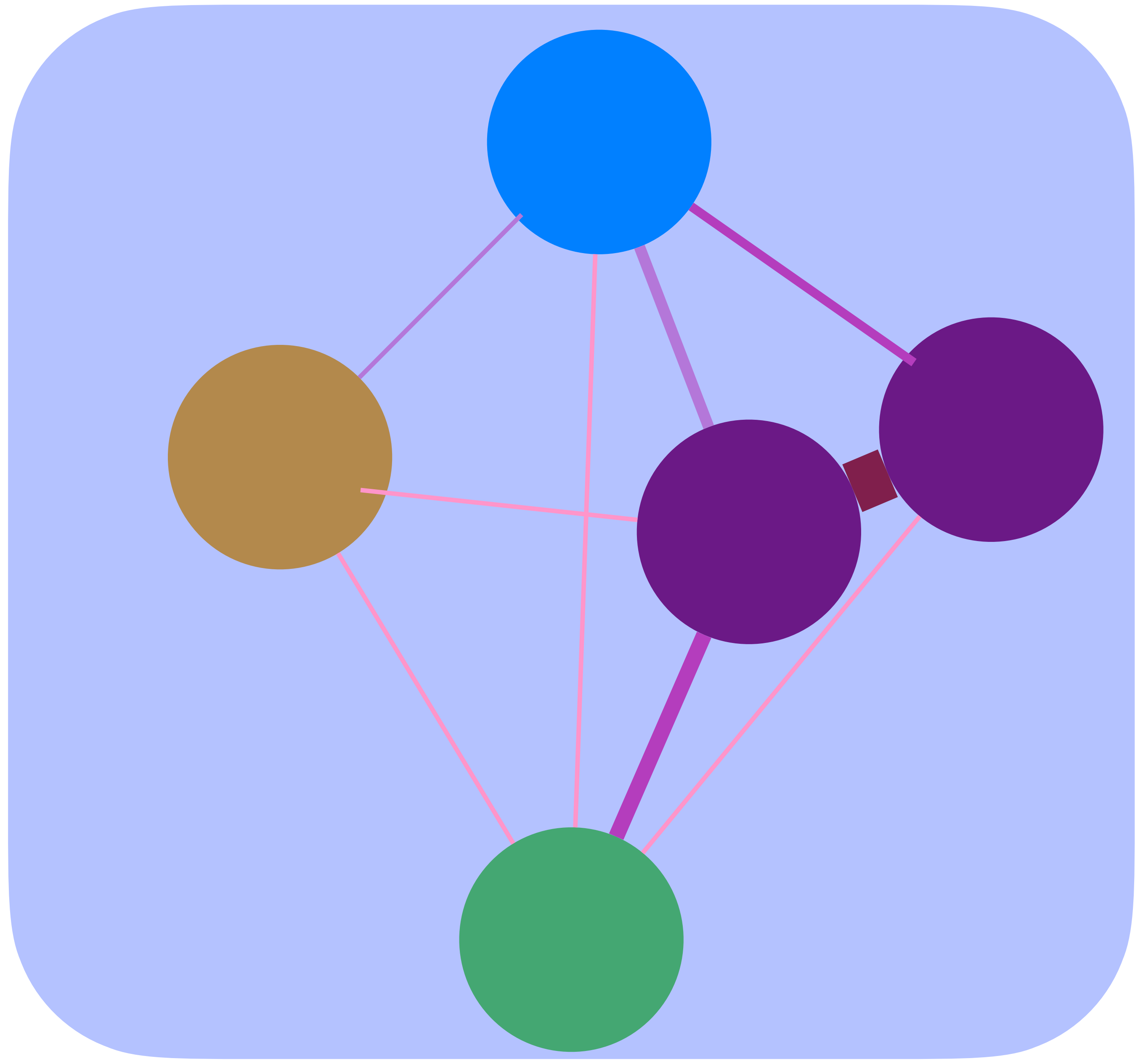
# Results from Performance of GNNS in EJ Classification



# Jet Classification: Probability Distribution

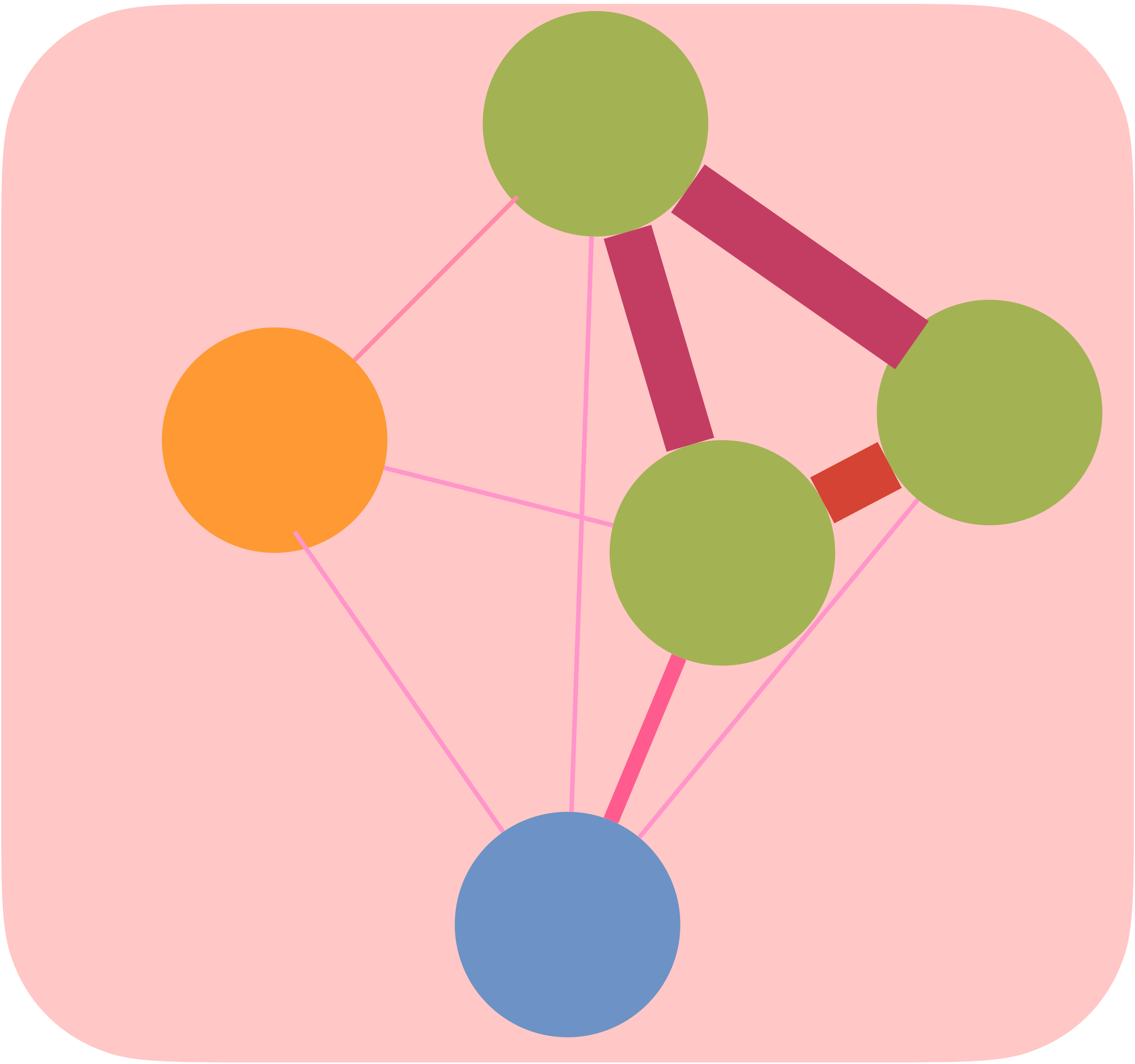



True EJ



Probability EJ = 0.9

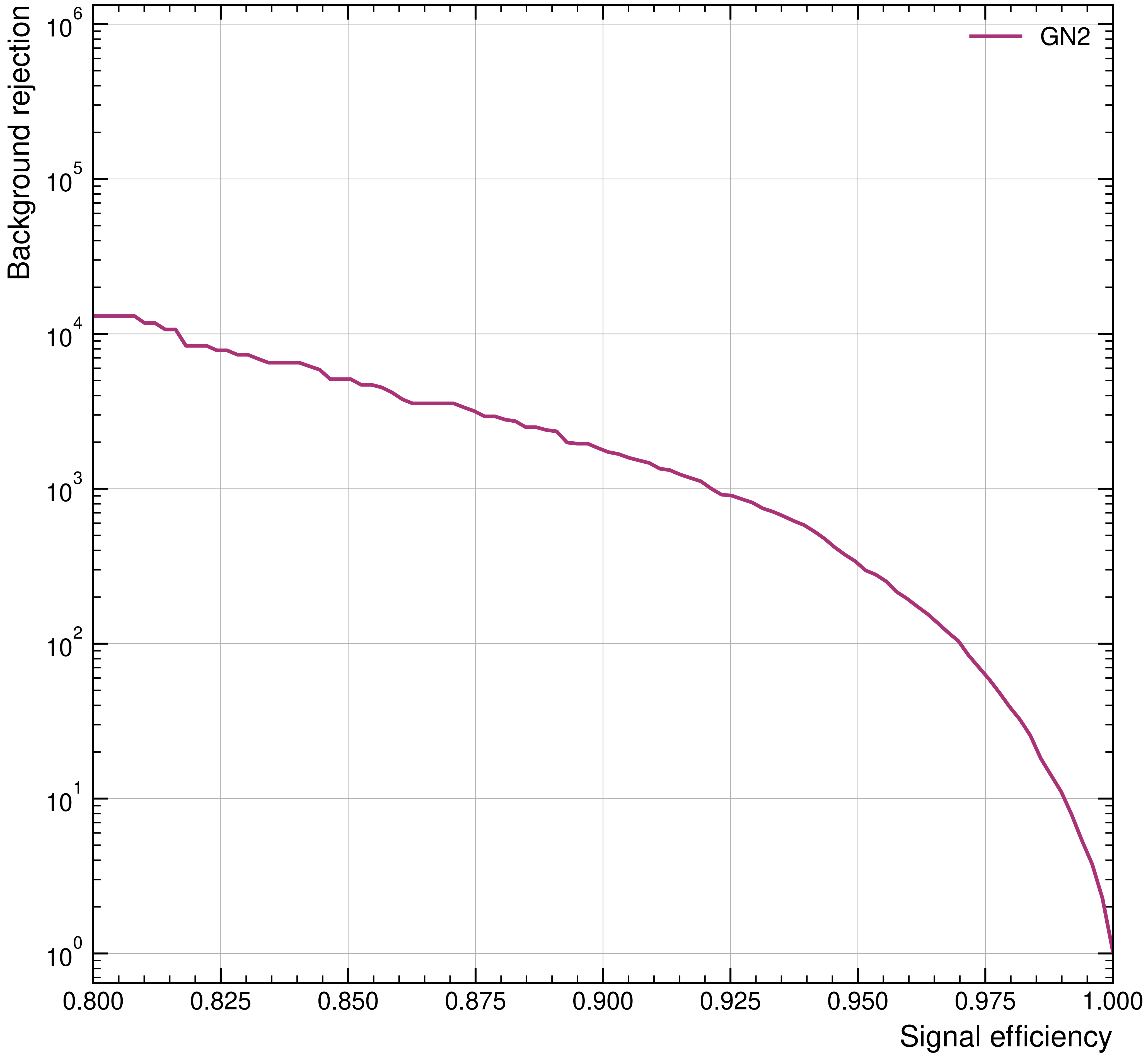
True QCD jet



Probability EJ = 0.2

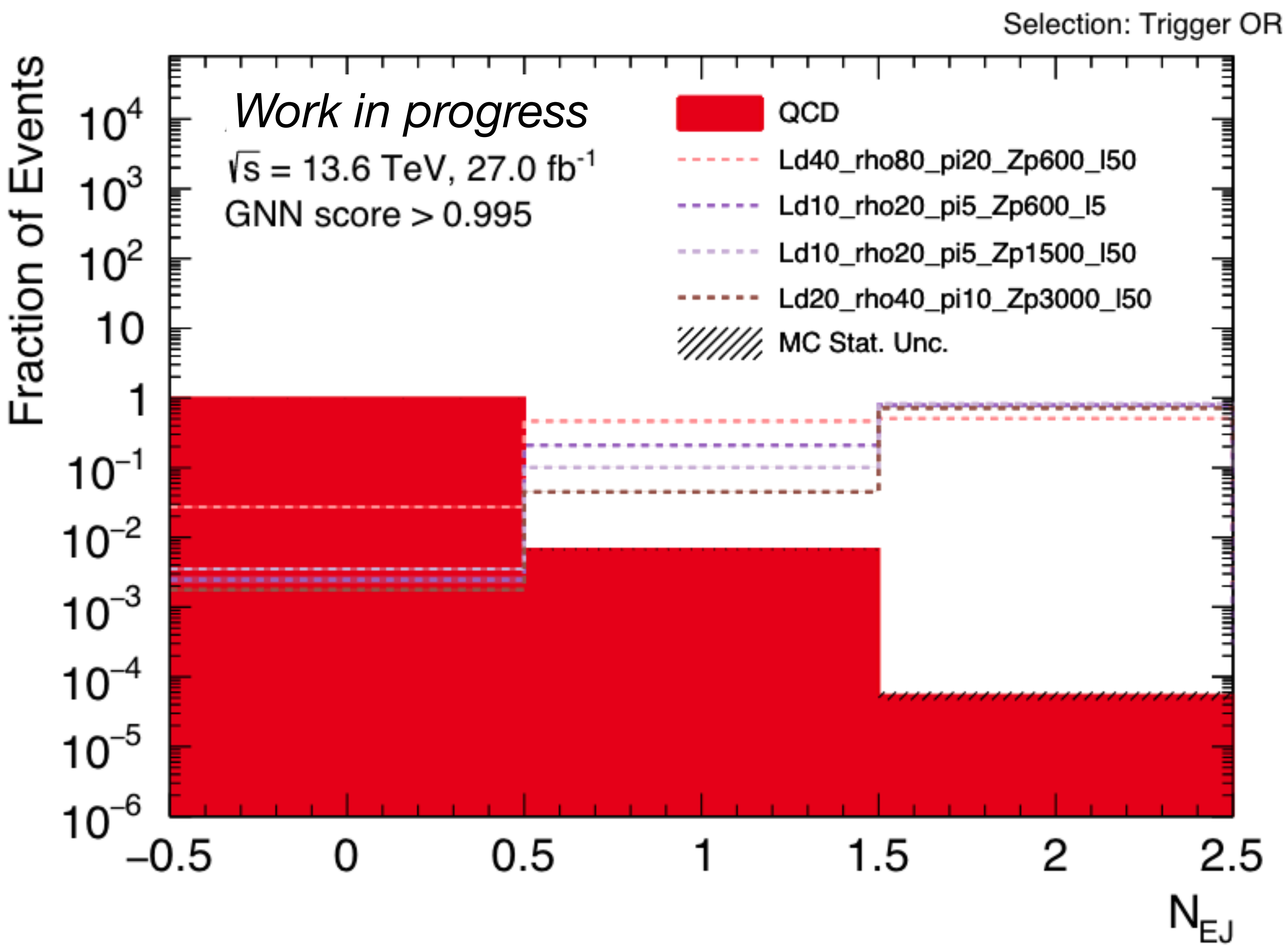
- Two categories: Signal Jets (*EJs*) from long lived dark mesons and background Jets from QCD background process!
- EJ probability(GNNScore) distribution.
- Signal jets peaks at last bin suggesting extremely high likelihood for majority of signal jets to be correctly identified!

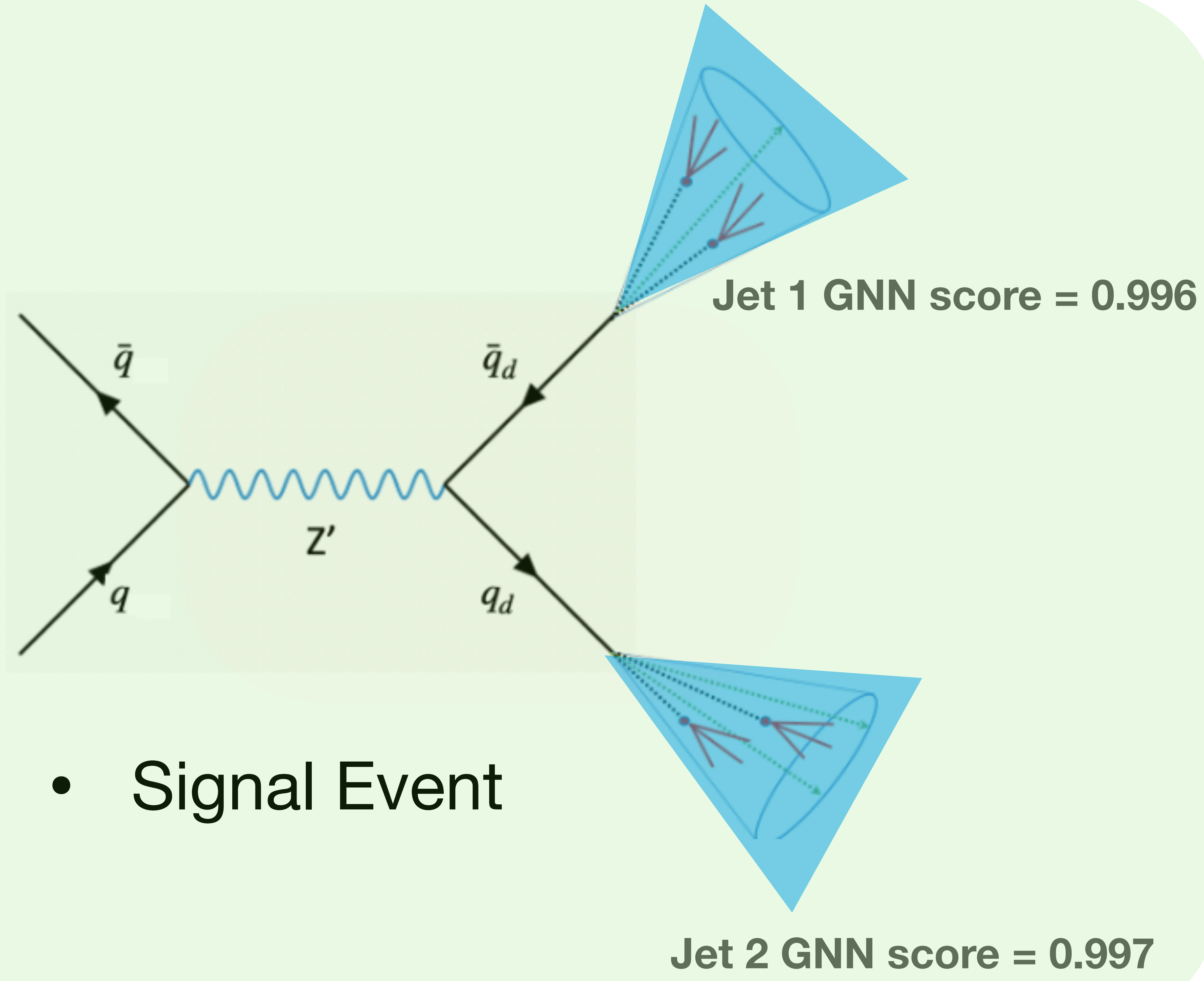
# Jet Classification: ROC




- Extremely good classifier with great background rejection while retaining majority of the signal.
- This implies that within a threshold where $80\ \%$ of the signal jets are accurately identified, there is a misclassification of 1 jet for every $\sim 10^4$ jets.

# GNN in EJ (Run 03) Analysis




- Signal Event

- Requiring two jets to have GNN score > 0.995 gives significant background reduction with high signal efficiency!

# Conclusion

- Use of GNNs are very effective in classifying atypical LLP signature- emerging jets.
    - Additionally GNNs were also able to perform classification of tracks inside the jet and find of pair of tracks belong to the same vertex.

Thank You!

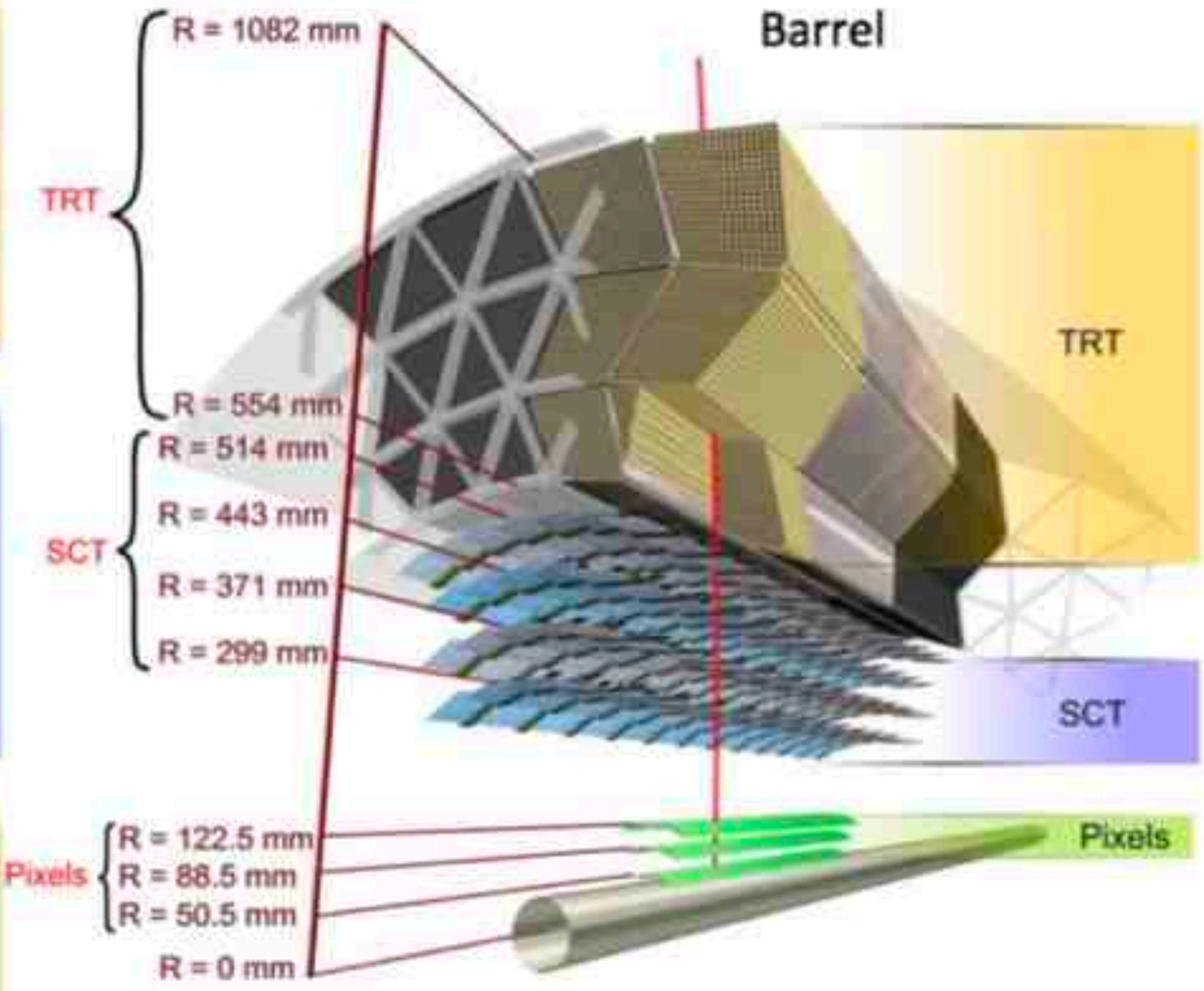
# Backup

# ATLAS Detector

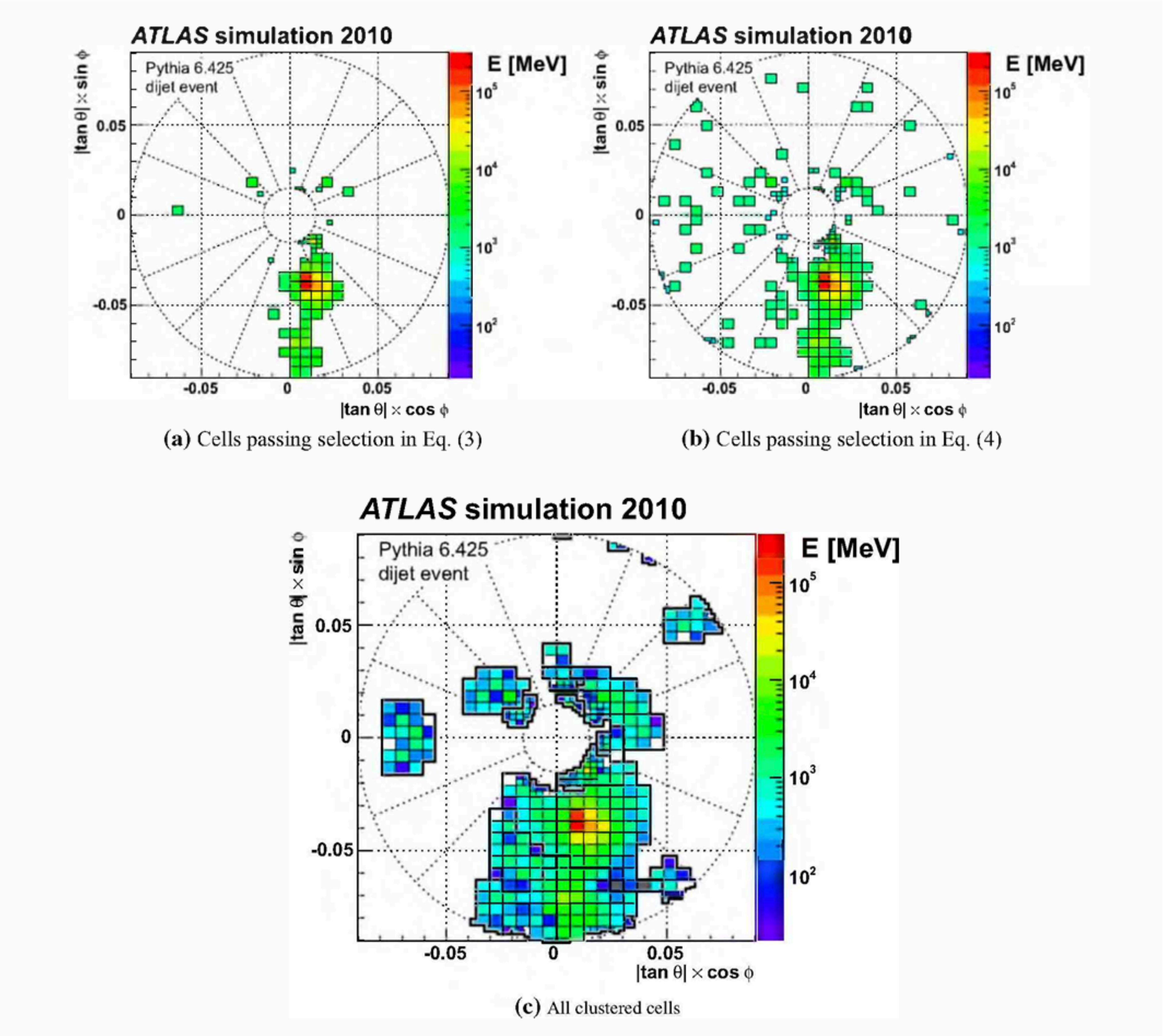
- Straw tracker + Transition Radiation
- 4mm diameter straws with 35 μm anode wire
- Layers: 73 in Barrel (axial) 2x160 in Endcap (radial)

- 4(9) double layers in Barrel/Endcap
- 4088 modules, 6M chan., strips 80 μm
- Resolution 17 x 580 μm

- 3 layers in Barrel and Endcap
- Pixel size 50 x 400 μm
- Resolution 10 x110 μm
- 80 M channels



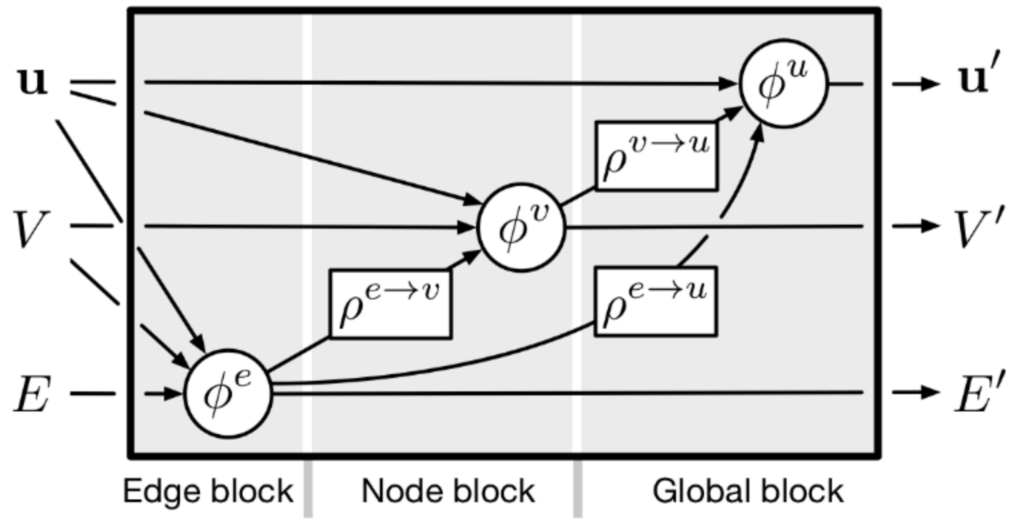
# Jet cluster



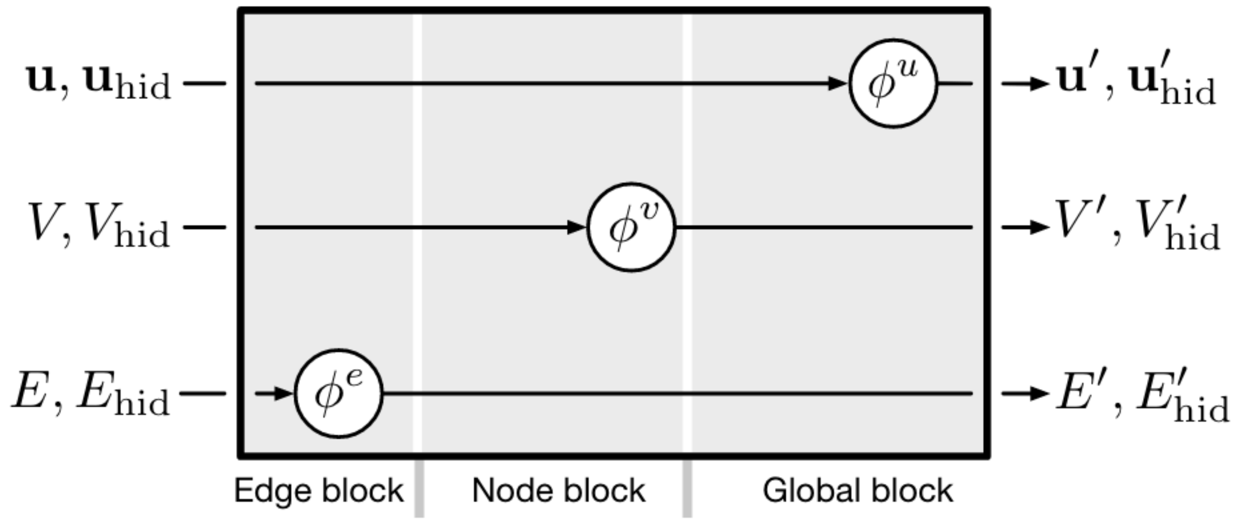
(a) Cells passing selection in Eq. (3)
(b) Cells passing selection in Eq. (4)
(c) All clustered cells

- Image: https://arxiv.org/abs/1603.02934

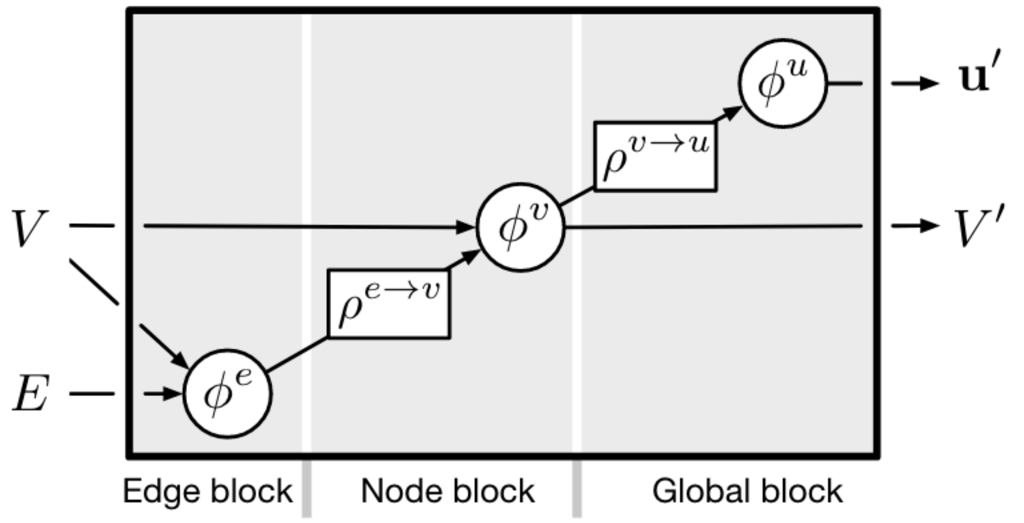
# GNN generalizations



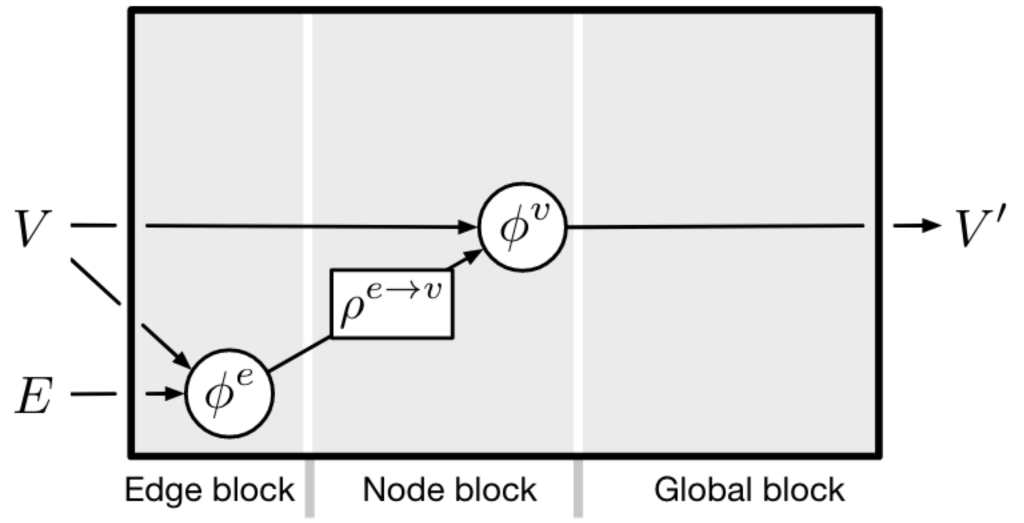

(a) Full GN block



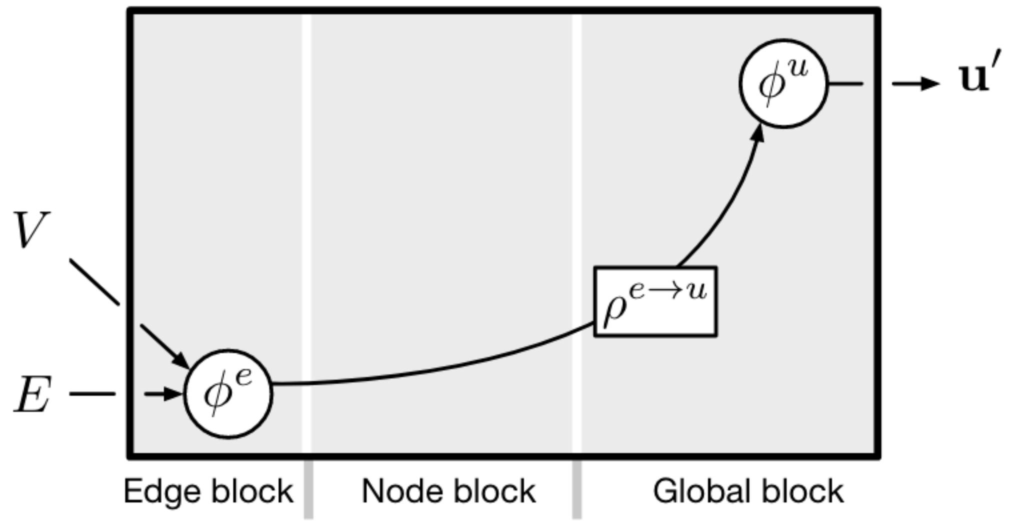

(b) Independent recurrent block



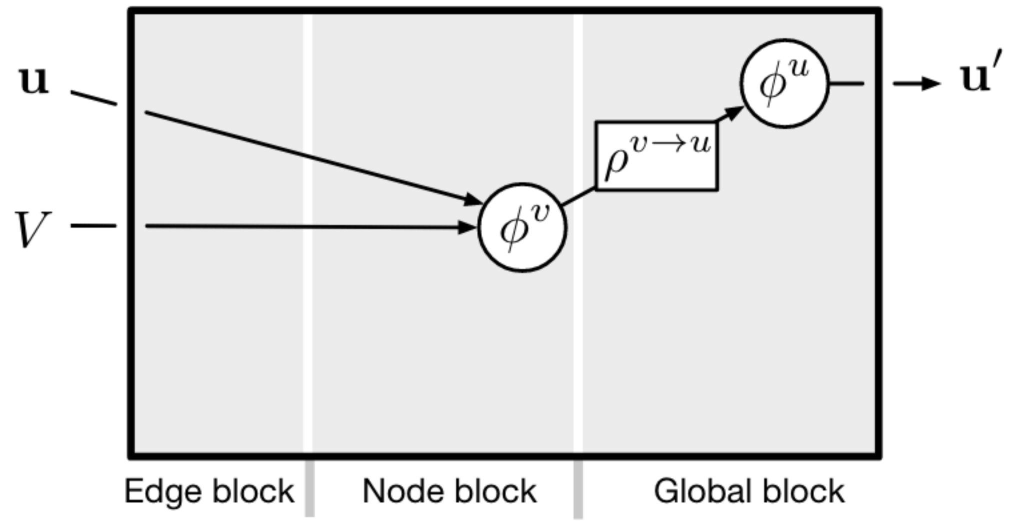

(c) Message-passing neural network



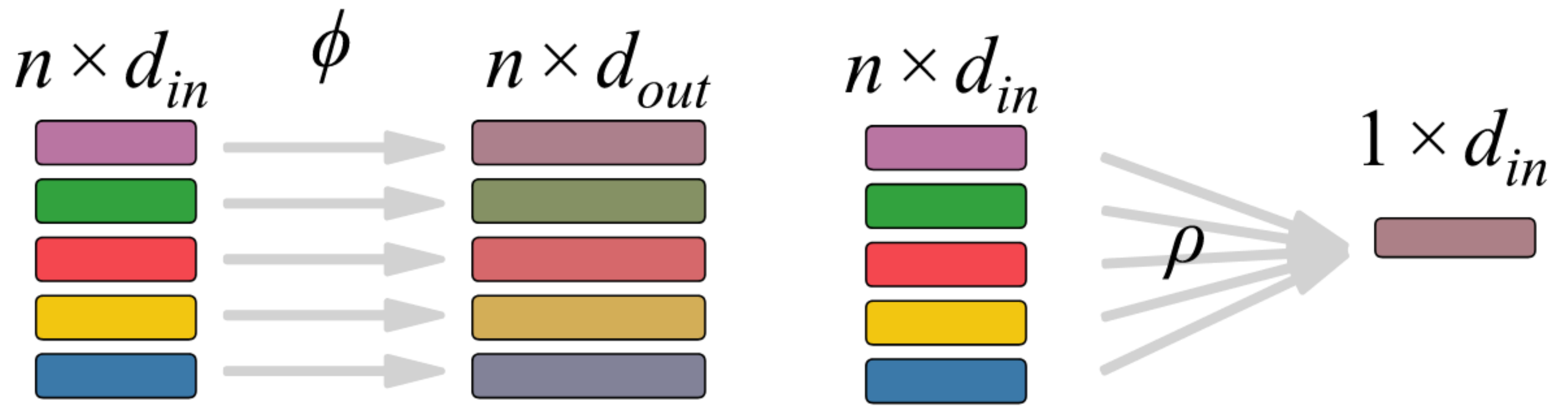

(d) Non-local neural network




(e) Relation network




(f) Deep set



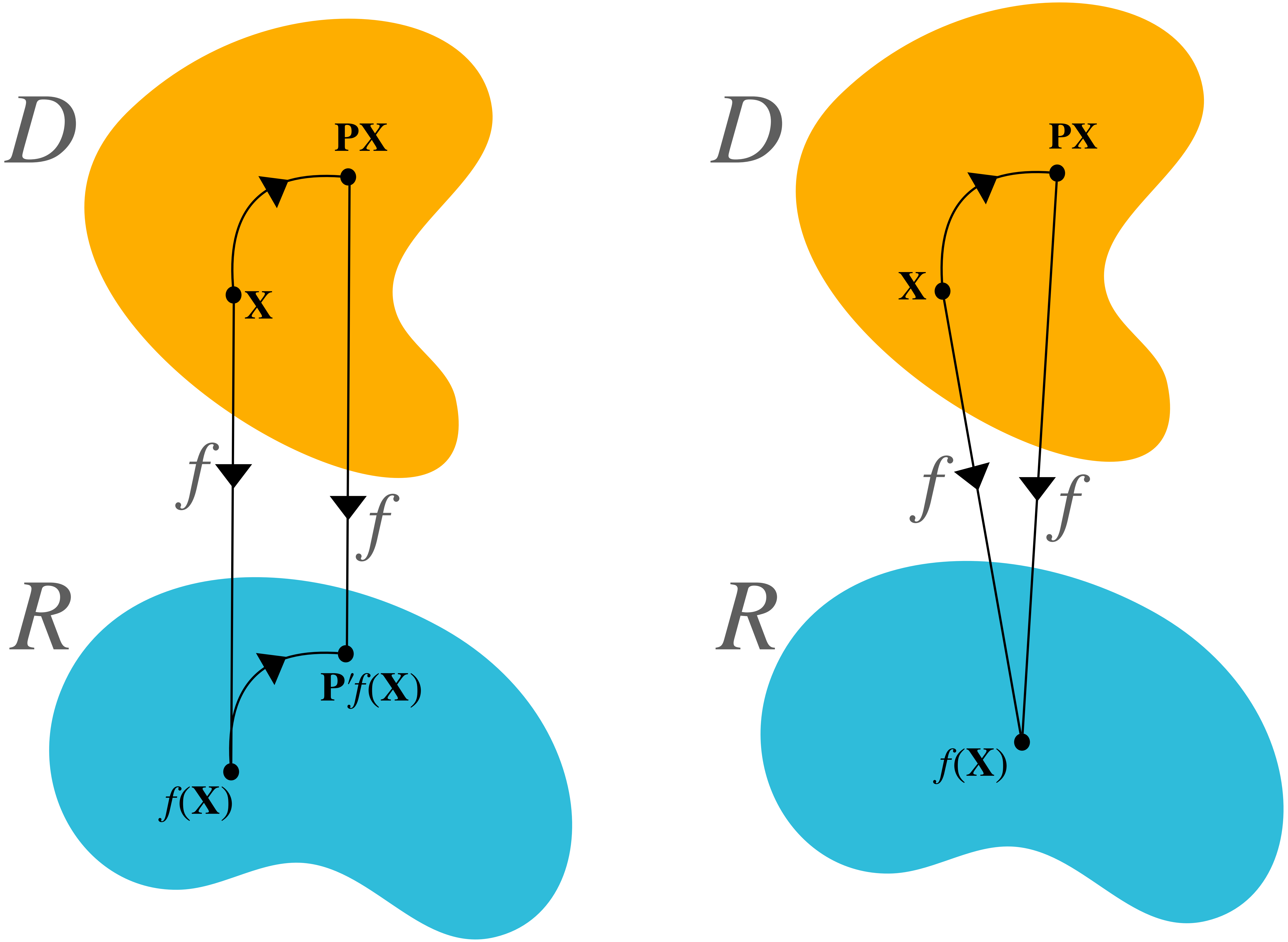

## Internal component of GNNs

## Equivariance

$$f(\mathbf{PX}) = \mathbf{P}'f(\mathbf{X})$$

## Invariance

$$f(\mathbf{PX}) = f(\mathbf{X})$$



# GNN Architecture

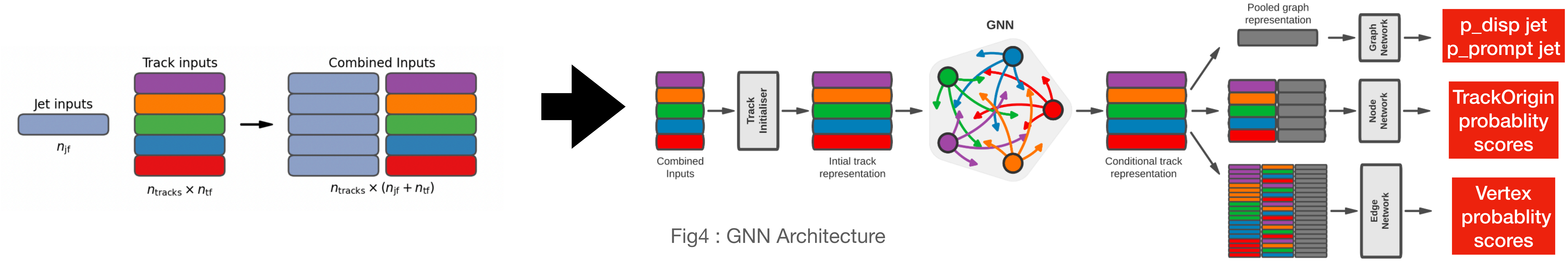


Fig4 : GNN Architecture

- Combined input prepared and fed into network architecture (2 jet variables 16 track variables)
- Initial latent representation for each track created. These representations are then used to populate the node features of a fully connected graph network
- Message passing graph neural network's loss function  also accounts node and vertex classification loss function.
- After the graph network, the resulting node representations used to predict Track Label (truthOriginLabel), JetLabel (isDisplaced) probability score.
- Architecture based on the ATLAS Flavour tagging software!

4: http://cds.cern.ch/record/2811135/files/ATL-PHYS-PUB-2022-027.pdf

# Samples Used for Training EJ classifier

QCD

Ld40_rho80_pi20_Zp600_l50

Ld10_rho20_pi5_Zp600_l5

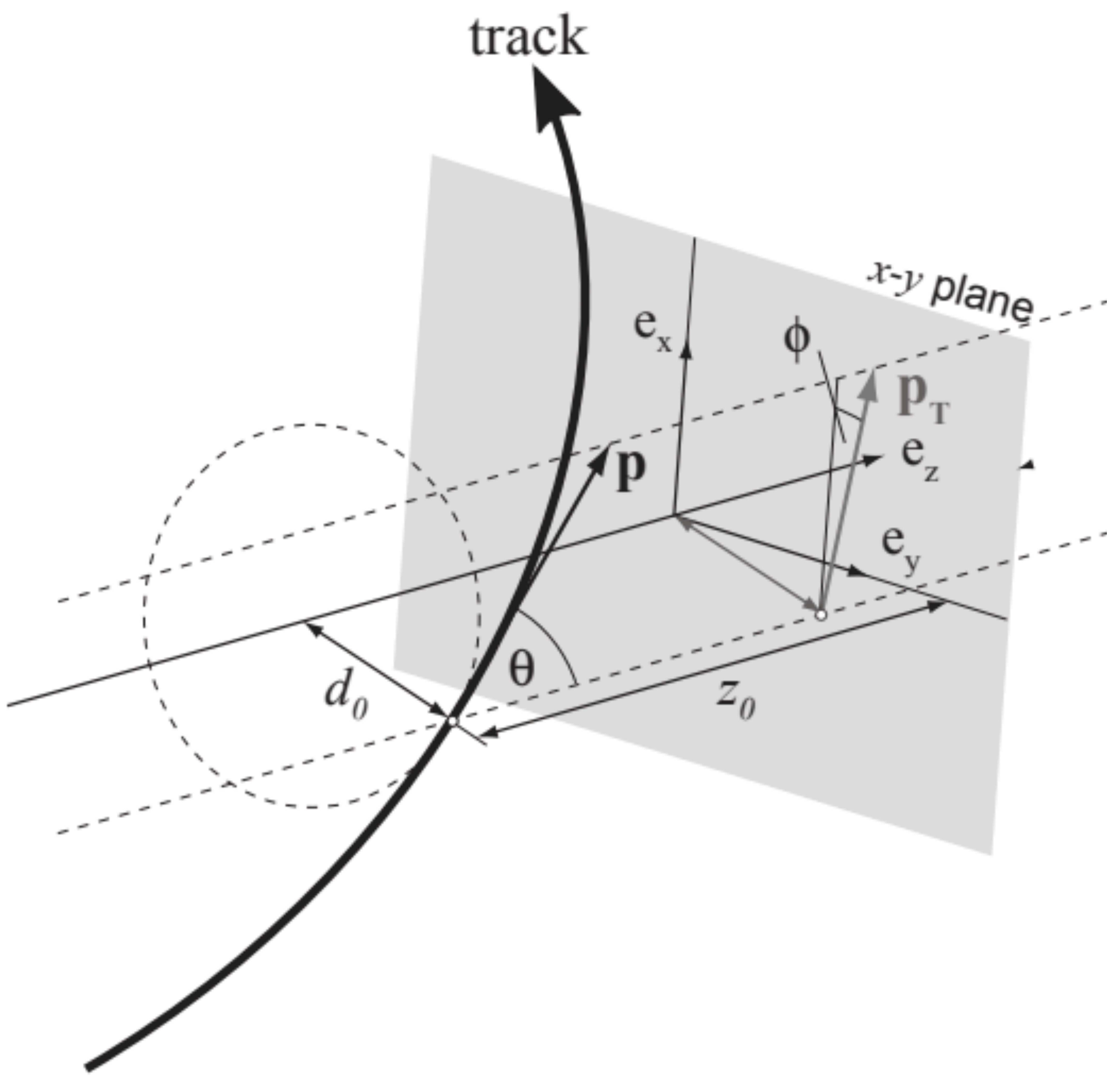
Ld10_rho20_pi5_Zp1500_l50

Ld20_rho40_pi10_Zp3000_l50

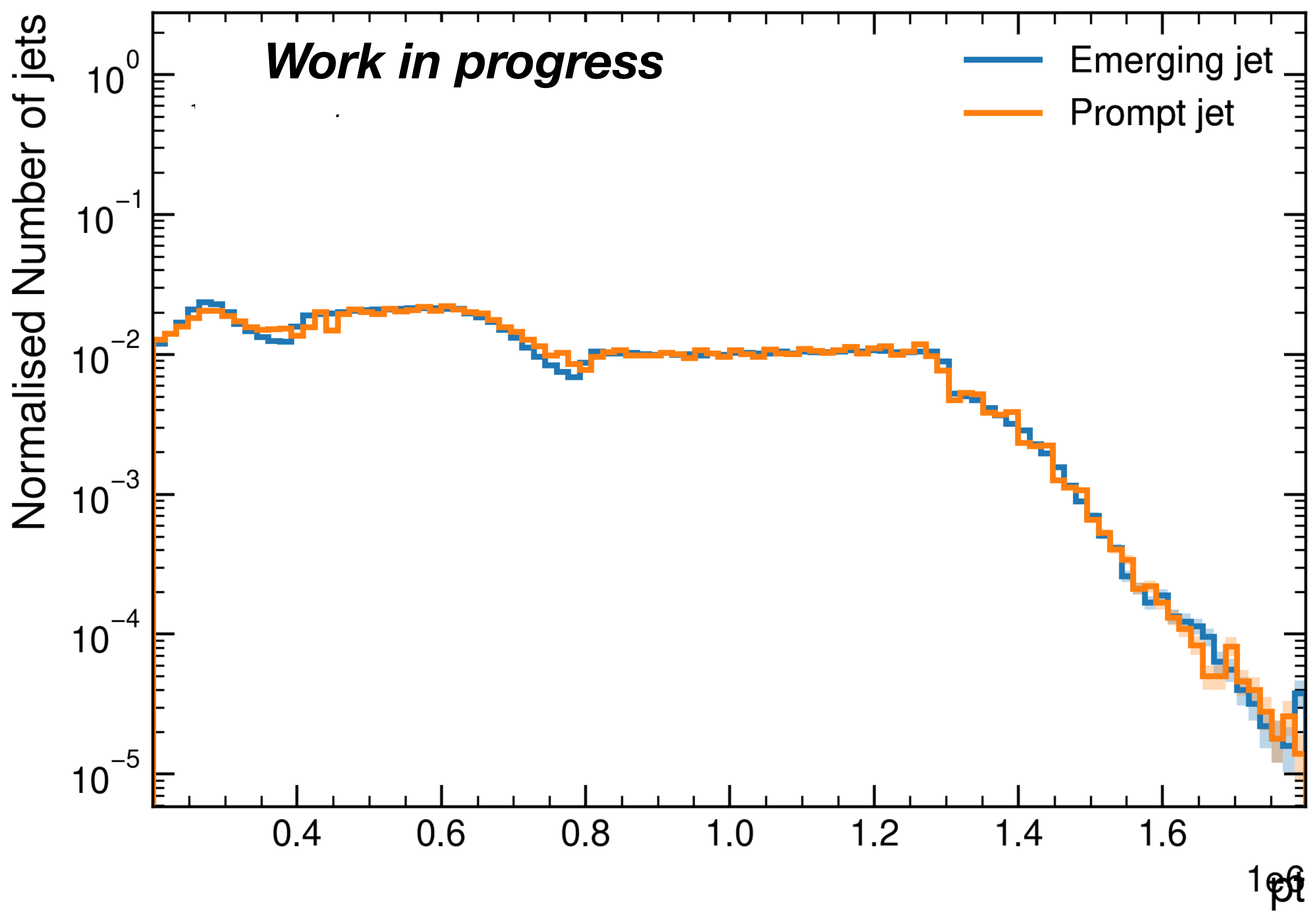
- Ld = dark confinement scale [GeV]
- rho = mass of rho meson [GeV]
- pi = mass of dark pion [GeV]
- Zp = mass of Z' [GeV]
- l = lifetime [mm]

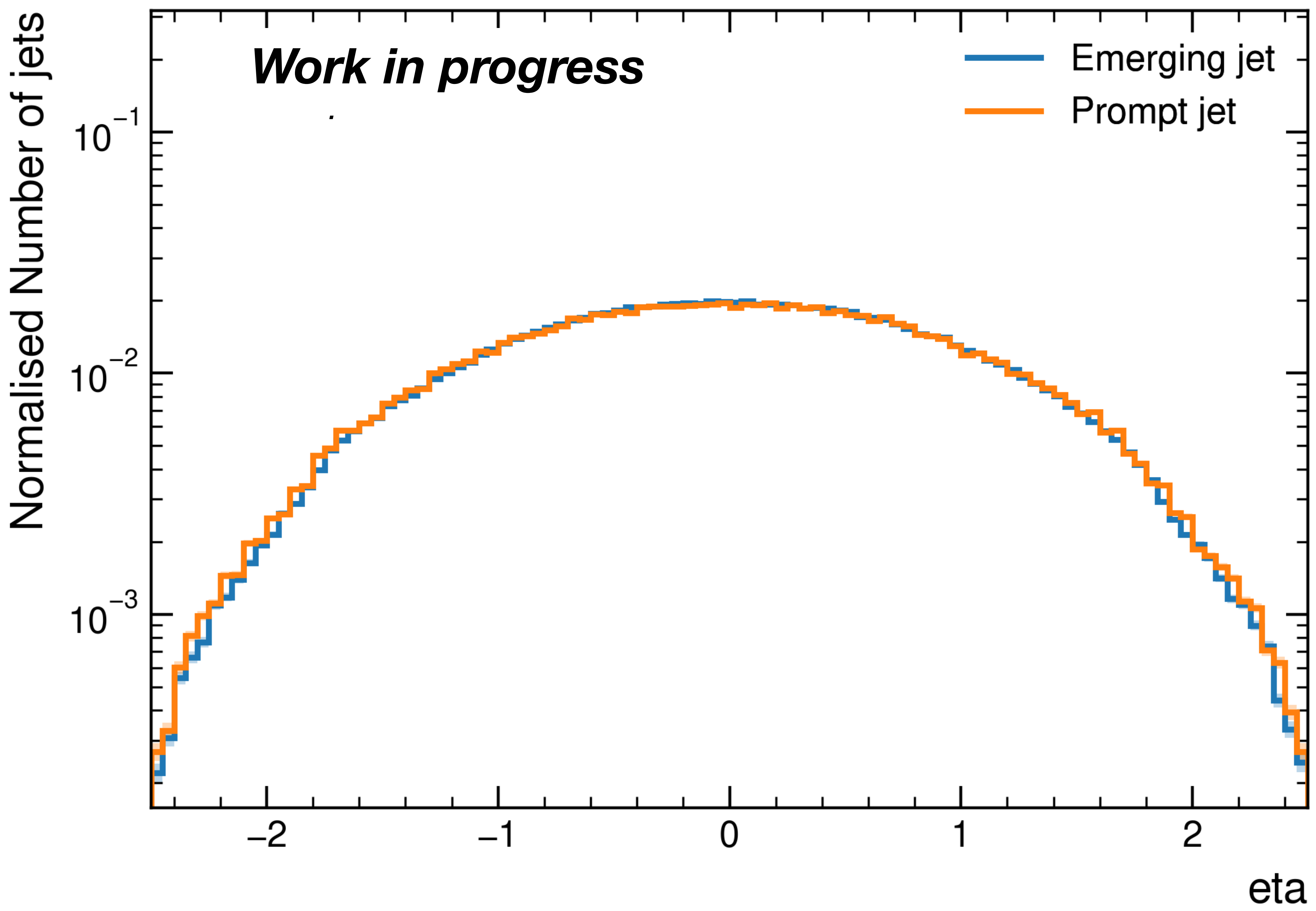
# Jet-Track Inputs

| Category | Variable | Description |
|---|---|---|
| Jet | $p_T$ | Jet transverse momentum |
| | $\eta$ | Signed jet pseudorapidity |
| Track | $d_0$ | Distances of closest approach between the track and beamline in the transverse plane |
| | $z_0 \sin\theta$ | Closest distance from the track to the primary interaction point in the transverse plane |
| | $d_\phi$ | Azimuthal angle of the track, relative to the jet $\phi$ |
| | $\mathrm{d}\eta$ | Pseudorapidity of the track, relative to the jet $\eta$ |
| | $\frac{q}{p}$ | Track charge divided by momentum (measure of curvature) |
| | $\sigma(\phi)$ | Uncertainty on track azimuthal angle $\phi$ |
| | $\sigma(\theta)$ | Uncertainty on track polar angle $\theta$ |
| | $\sigma(\frac{q}{p})$ | Uncertainty on $\frac{q}{p}$ |
| | nPixHits | Number of pixel hits |
| | nSCTHits | Number of SCT hits |
| | nPixShared | Number of shared pixel hits |
| | nSCTShared | Number of shared SCT hits |
| | nPixHoles | Number of pixel holes |
| | nSCTHoles | Number of SCT holes |
| | IP3D_signed_d0_significance | Ratio of $d_0$ and $\sigma(d_0)$ defined for both positive and negative scale with reference to the primary interaction point. |
| | IP3D_signed_z0_significance | Ratio of $z_0 \sin(\theta)$ and $\sigma(z_0 \sin(\theta))$ defined for both positive and negative scale with reference to the primary interaction point |



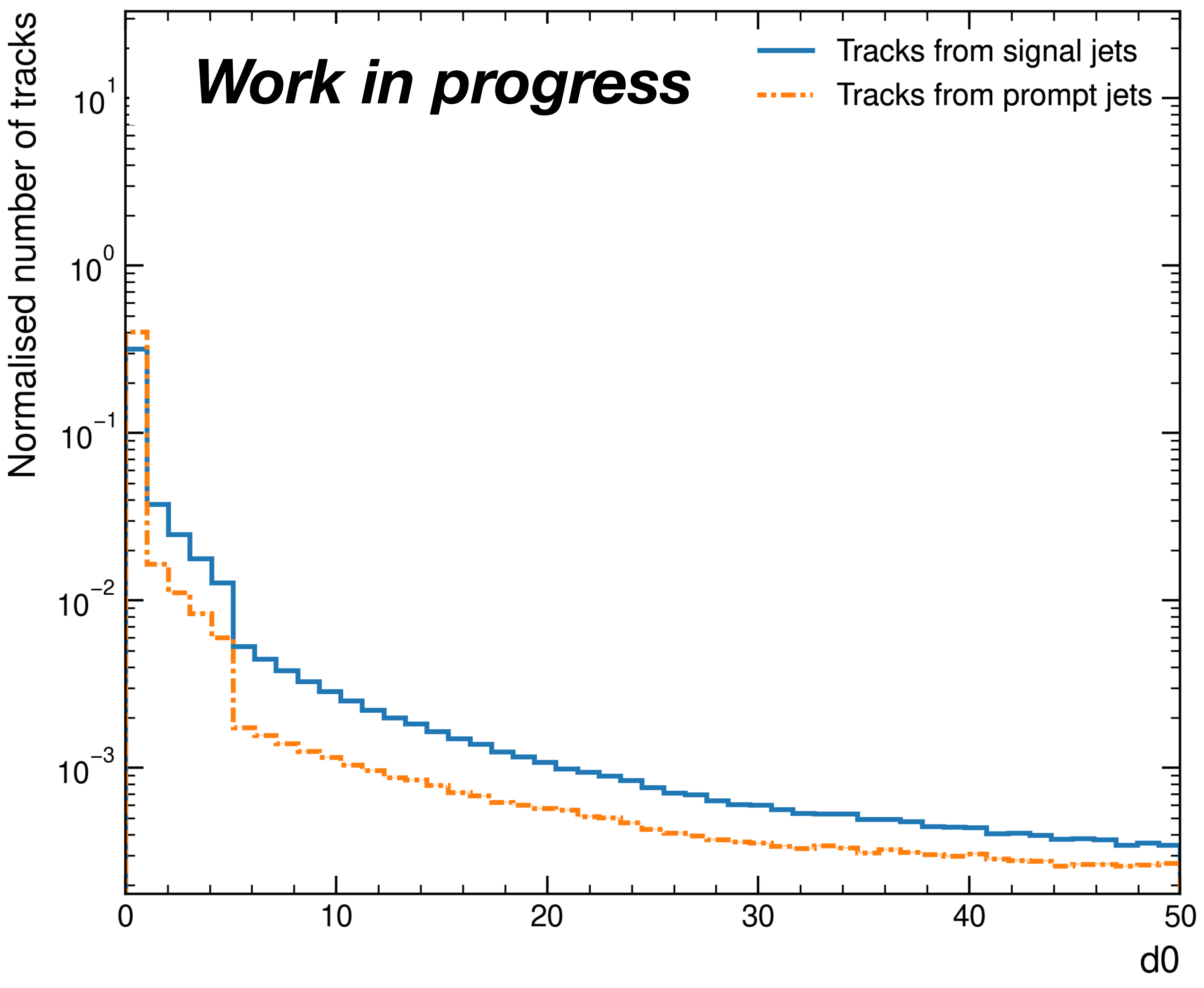
# Input Variables: Jets

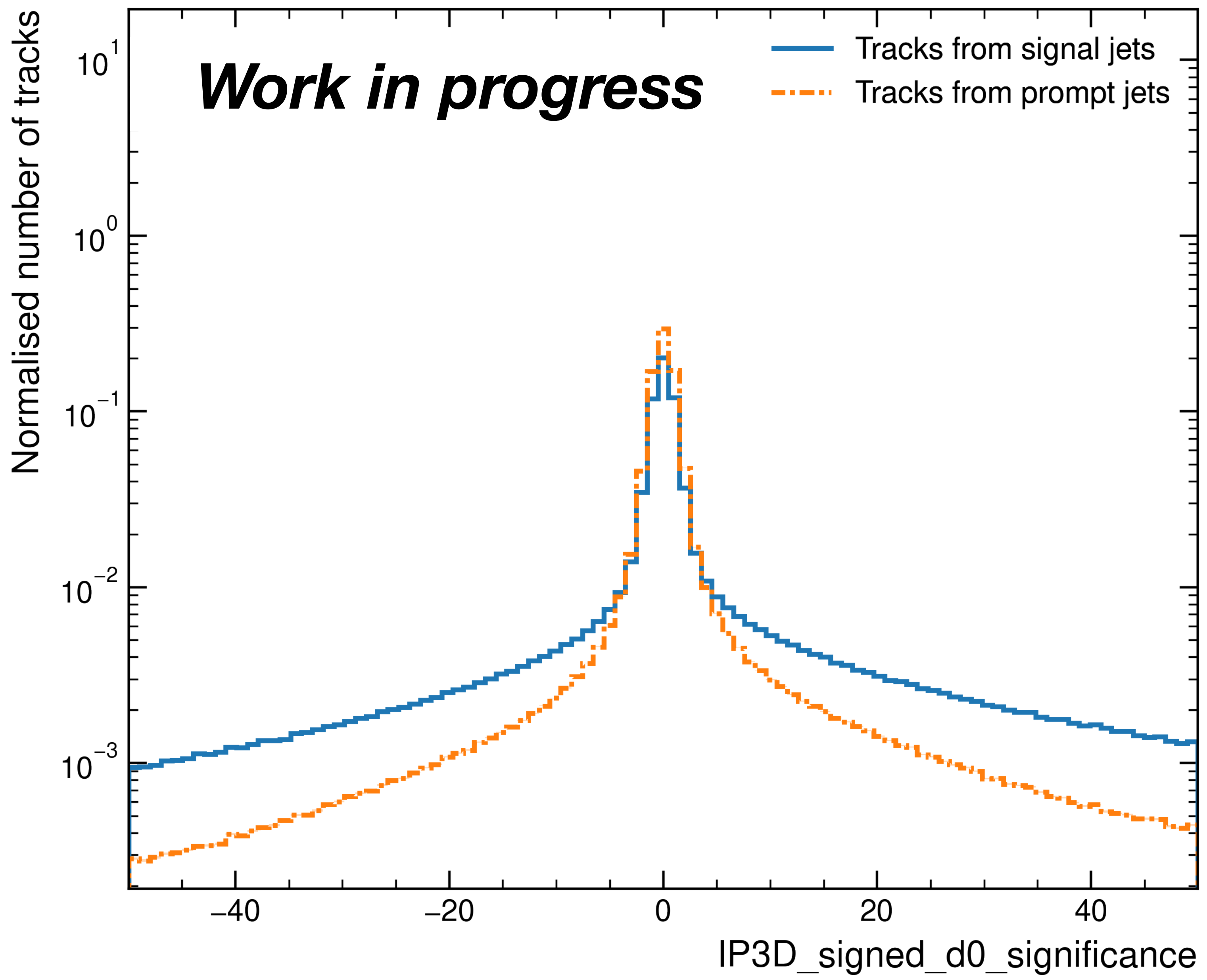


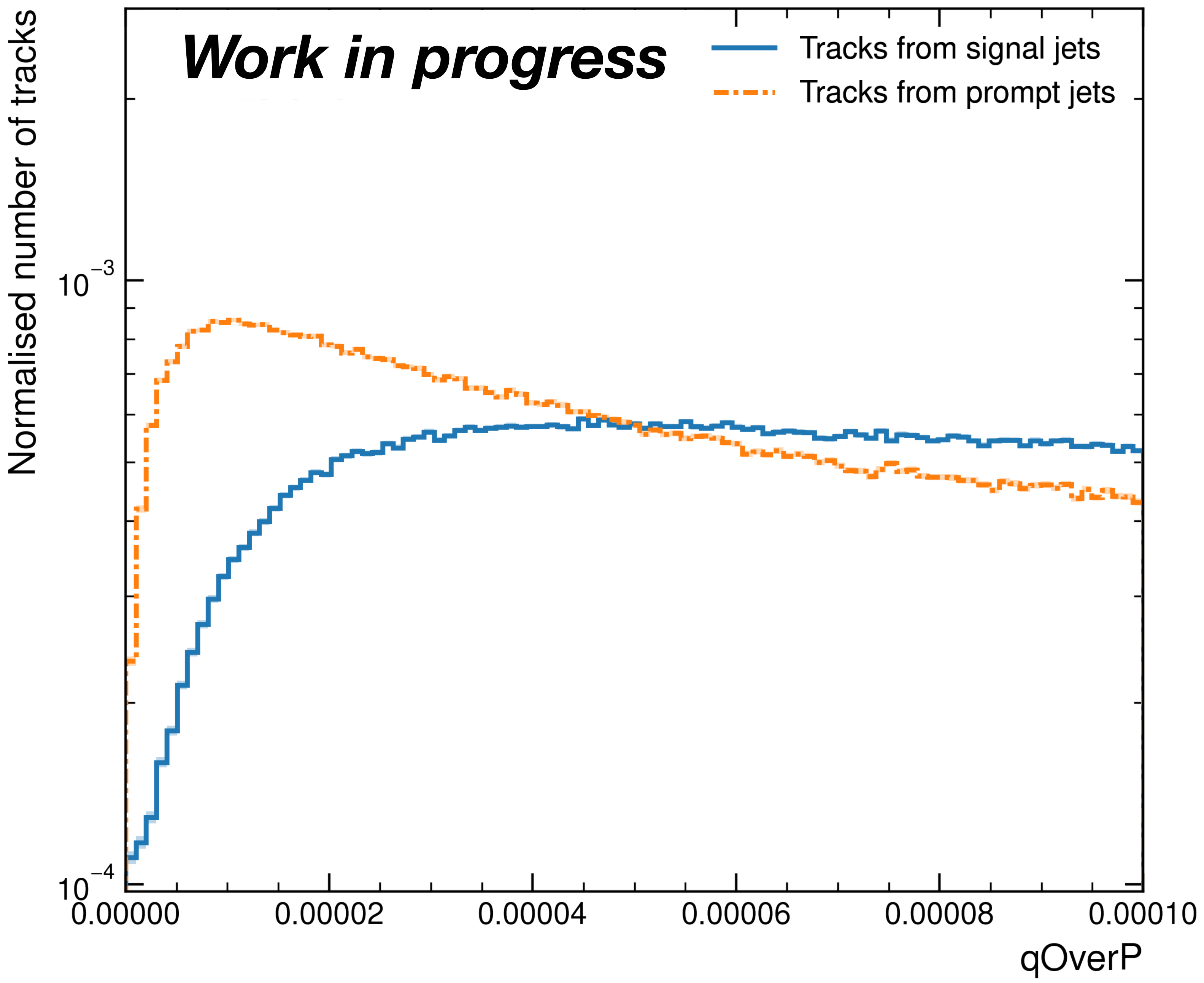




- Two jet variables that constitute the basic kinematics of a jet $p_T, \eta$
- To avoid avoid kinematic biases for jet tagger, the distributions are "resampled", i.e ensure uniformity in the kinetic distribution!

# Input Variables: Tracks



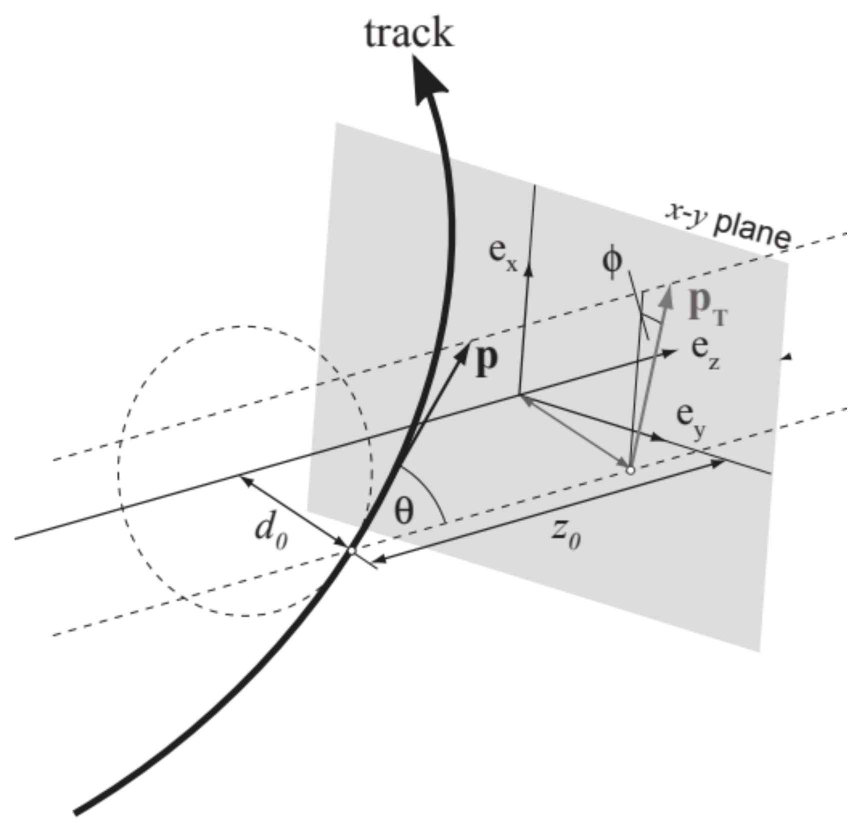




- 16 track variables including track parameters in ATLAS tracking system, detector hits and holes variables, uncertainty in track parameters … (detailed in backup slides)

- Most discriminating ones include

- - $d_0$: Distances of closest approach between the track
  - IP3D_signed_d0_significance: Ratio of $d_0$ and $\sigma(d_0)$ defined for both positive and negative scale with reference to the primary interaction point of the ATLAS detector
  - $\frac{q}{p}$ Track charge divided by momentum (measure of curvature)

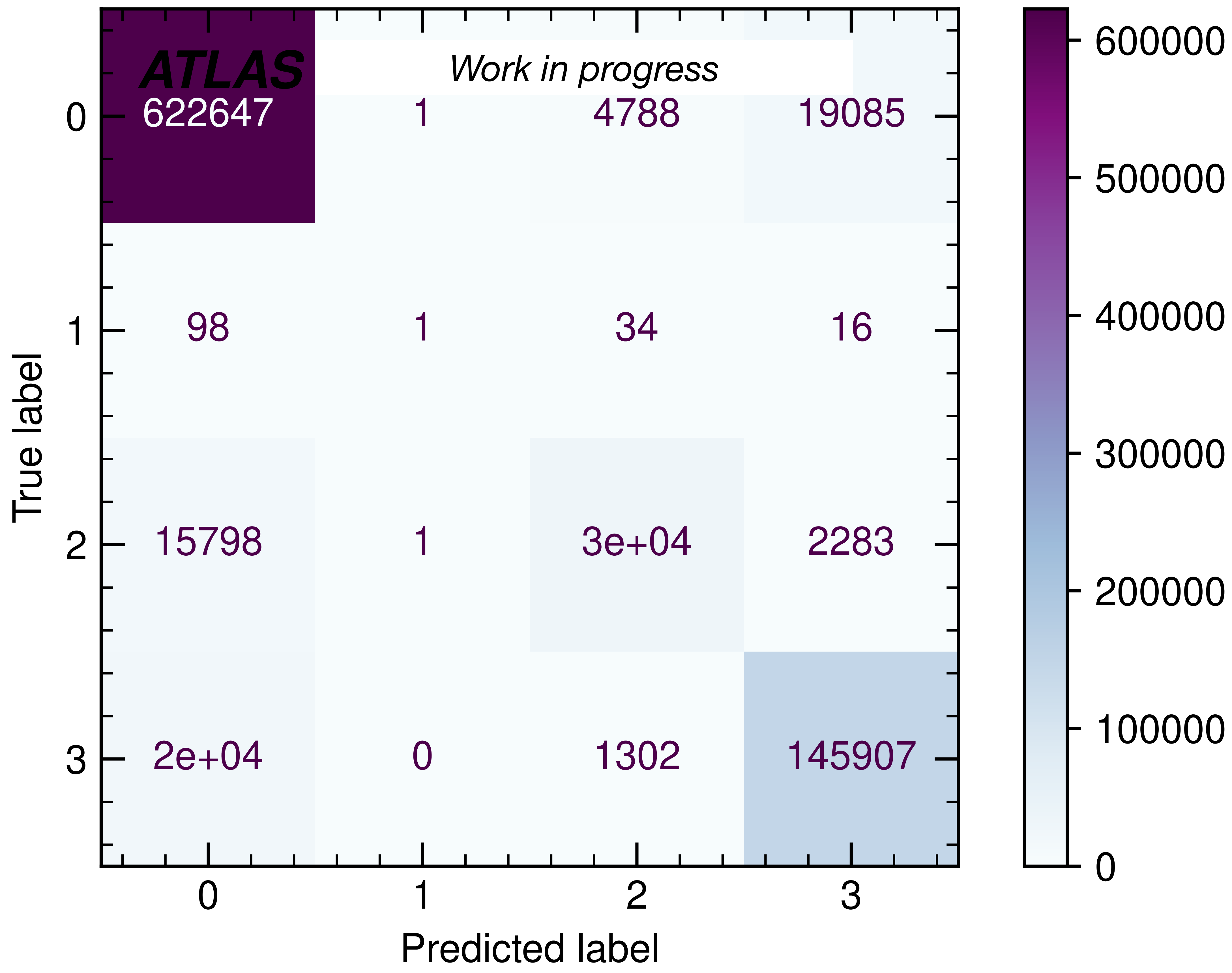


# Input Distribution (Tracks)

# Track Origin Identification: Performance

## Confusion Matrix

- The diagonal elements of the matrix represent correct classification!
  - Pileups and Displaced tracks most accurately classified
  - ~20k “true” displaced tracks classified as pileups and vice versa!
  - ~16k “true” primary tracks classified as pileups



# JetMatrixView

- 40 tracks x 40 tracks confusion matrix
  - Instead of being sorted by trackID's its sorted by truthVertexId of each track
  - For example {TrackId(VertexId)} in a Jet is {2223(1),2224(3),2225(1),2226(2)}

**Track ID Based Sort**

| | **2223** | **2224** | **2225** | **2226** |
|---|---|---|---|---|
| 2223 | 1 | 0 | 1 | 0 |
| 2224 | 0 | 1 | 0 | 0 |
| 2225 | 1 | 0 | 1 | 0 |
| 2226 | 0 | 0 | 0 | 1 |

**VertexID Based Sort**

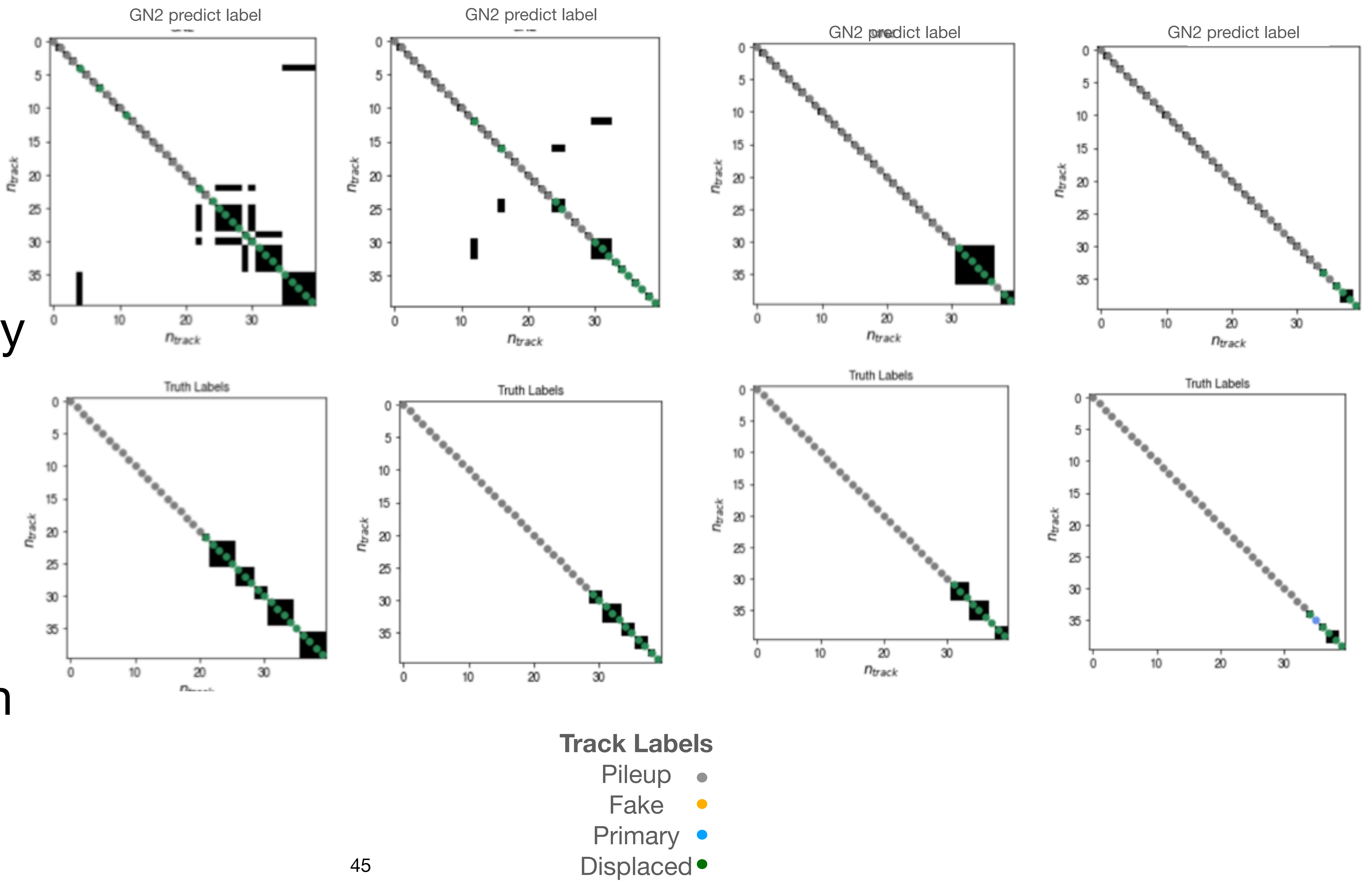
| | **2223** | **2225** | **2226** | **2224** |
|---|---|---|---|---|
| 2223 | 1 | 1 | 0 | 0 |
| 2225 | 1 | 1 | 0 | 0 |
| 2226 | 0 | 0 | 0 | 0 |
| 2224 | 0 | 0 | 0 | 0 |

# Jet View from Classifiers!

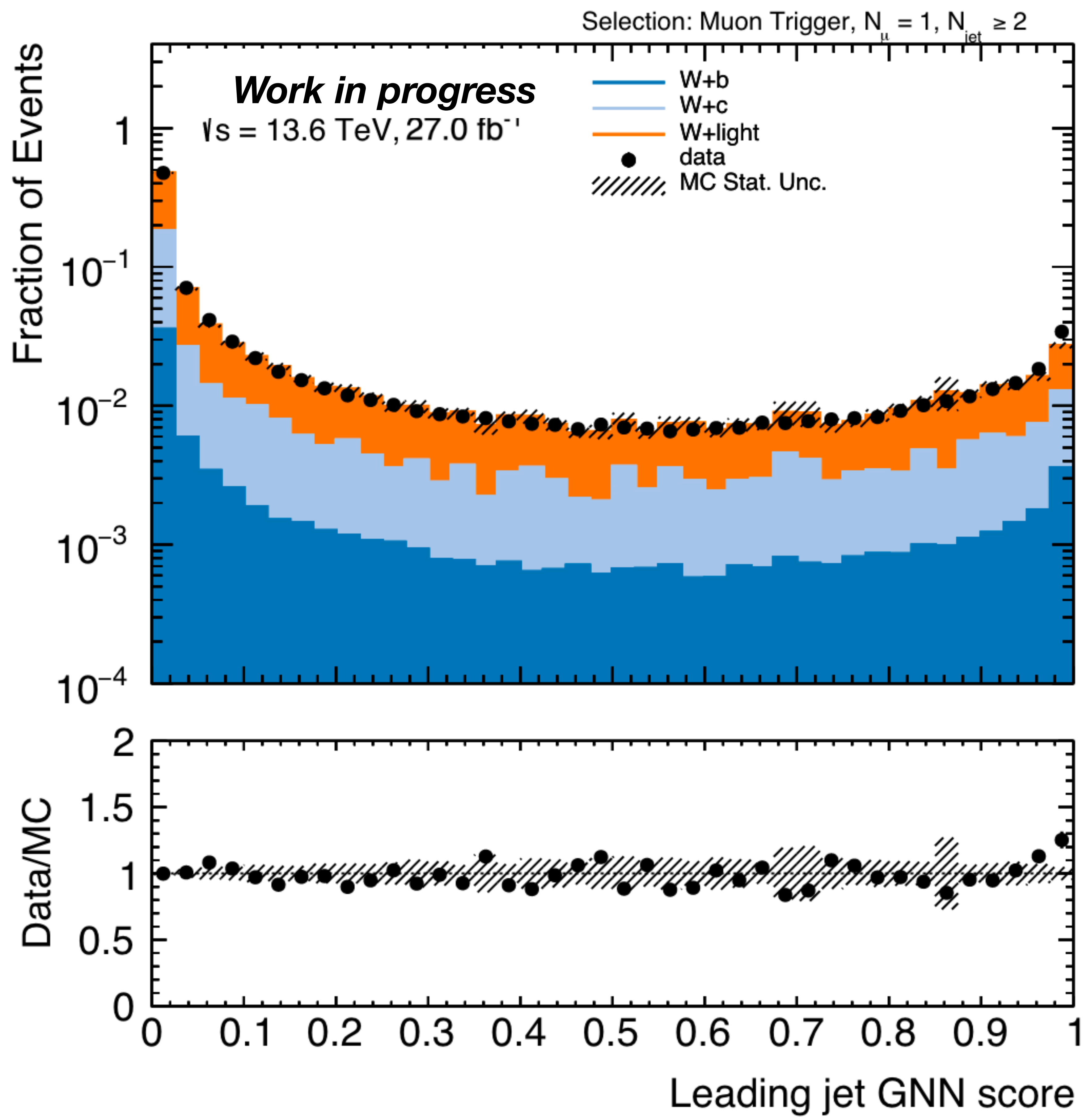
## Use GNN to classify events?

- True labels vs GNN predicted labels visualization for jet, track and vertex prediction
- $n_{trk} \times n_{trk}$ matrix sorted by TruthVertID
    - 1 (Black) if two tracks share the same vertex
    - 0 (White) if two tracks do not share a common vertex

**ATLAS** *Work in progress*



# GNN Validation




- First looks at 2022 data validate GNN performance on real data!