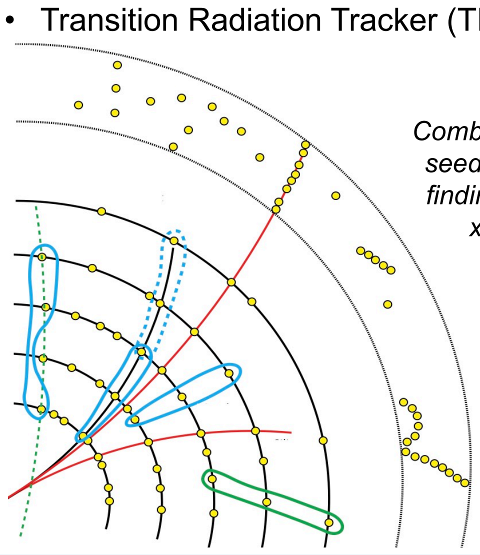
Machine Learning predictor for ‘measurement-to-track’ association for the ATLAS Inner Detector Trigger

Track Finding as a Pattern Recognition Problem

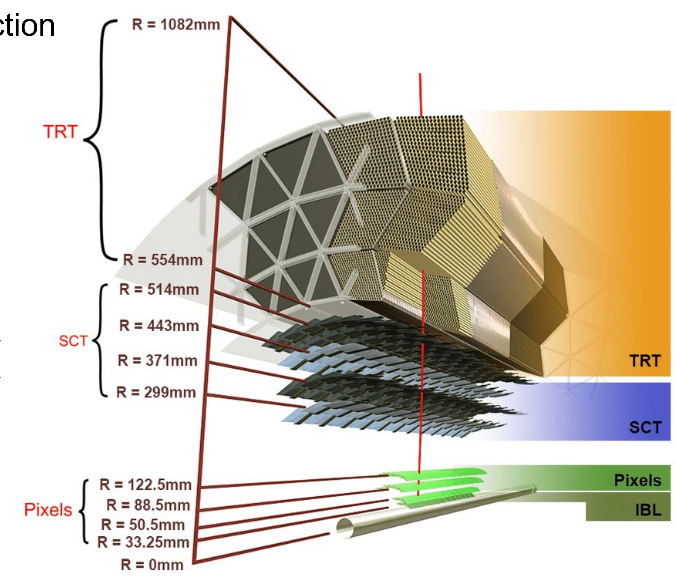
- **Associate individual measurements** into sequences representing tracks, scale: O(10^5) hits per event & several 1000s of tracks
- Track finding algorithm for the LHC Run-2 data taking period is based on **combinatorial track following** using track seeds
- Typically the seed number **scales non-linearly** with the number of hits (~cubical)
- Motivation for novel Machine Learning (ML) approaches in track finding that could lead to large savings in CPU
- **Aims:** create a ML algorithm to **predict if a pair of hits belong to the same track** given input hit features & optimize the High Level Trigger (HLT) Inner Detector (ID) track seeding by **reducing the proportion of fake seeds**

ATLAS Inner Detector

- Dedicated to track & vertex reconstruction
- **Pixel detector & Insertable B-Layer (IBL) are closest to the beamline, have the highest hit occupancy**
- Semiconductor Tracker (SCT)
- Transition Radiation Tracker (TRT)



Combinatorial seeded track finding in the x-y plane



- ID Trigger is part of the HLT & performs **fast online track & vertex finding**
- The **Fast Tracking algorithm** uses seeded track finding using combinatorial track following

Data Exploration

Data:

- Monte Carlo (MC) 13 TeV $t\bar{t}$ events <μ> = 80 Run-2 geometry
- Seeds constructed at the combinatorial stage of ATLAS track seeding are extracted
- **Doublet hit pairs** are formed from **pixel-only layers**
- Inner: (1,2), outer: (2,3)

Input Features:

- Consider r-z plane
- | cot θ |, where θ is the angle of inclination to the doublet pair of hits from the z axis
- w_η, pixel cluster width measured in the η direction

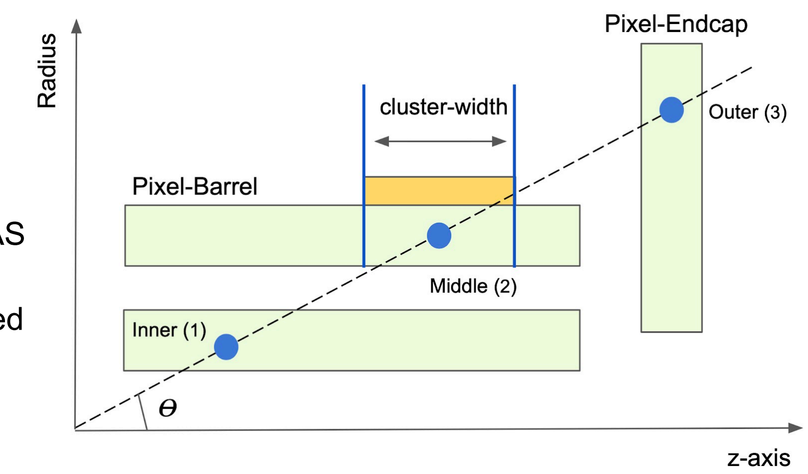


Illustration of a triplet seed in the r-z plane pixel layers of the ID

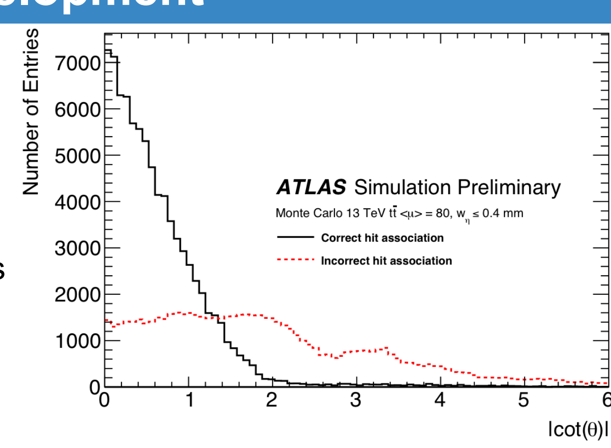
Ground truth labels obtained from MC truth:

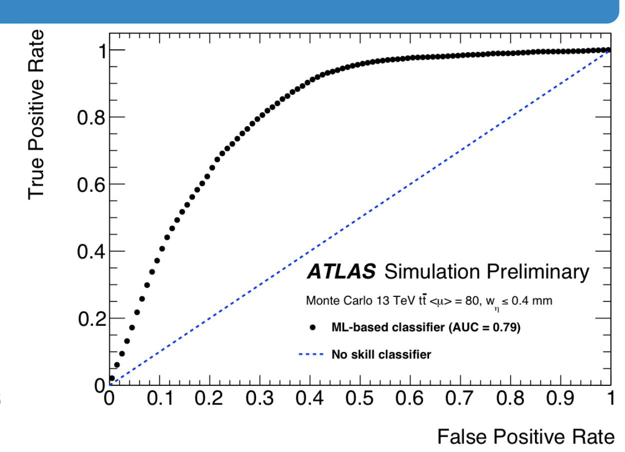
- **Correct hit association (1)**: hit pairs belong to the same track & correspond to a truth particle
- **Incorrect hit association (0)**: hit pairs do not belong to the same track

Classifier Development

The Model:

- Binary classification using a **‘Not-so-Naïve’ Bayes’ classifier** with a generative model
- The likelihood is computed via a Kernel Density Estimate fitted to each doublet class
- **| cot θ |** forms the input feature to be learned for varying **pixel cluster width w_η**
- ROC curve is used to adjust probability threshold cut
- Each classifier is tuned to yield **True Positive Rate = 0.95** to maintain high purity of doublets
- Predictions were made for each 1D distribution & plotted as an ‘acceptance-rejection’ region
- Acceptance region converted to look-up table, ensures fast look-up in HLT ID track seeding
- Classifiers for pixel-barrel doublets & pixel-endcap doublets are trained separately

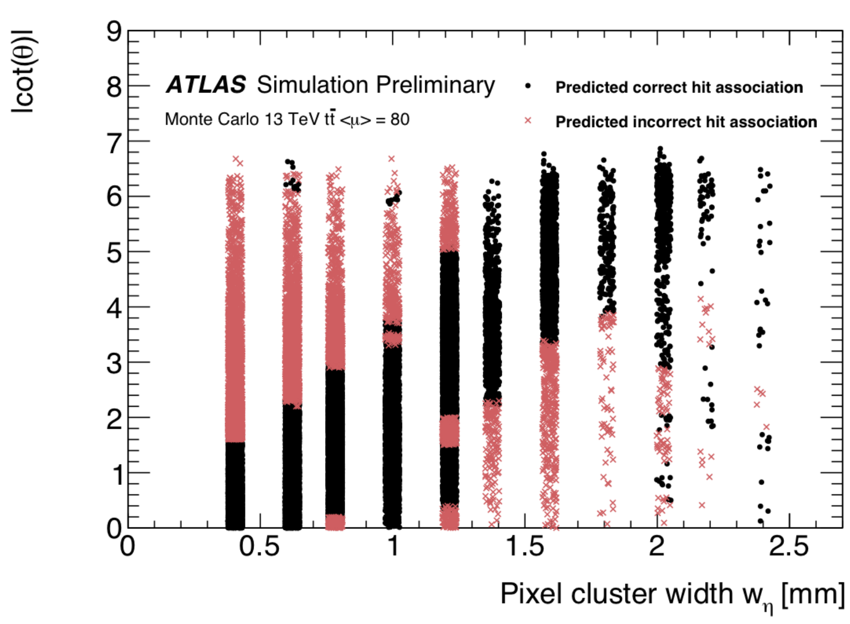




Bayes’ Theorem:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood: $P(x|c)$; Class prior: $P(c)$; Predictor prior probability: $P(x)$; Posterior probability that data x belongs to class c: $P(c|x)$

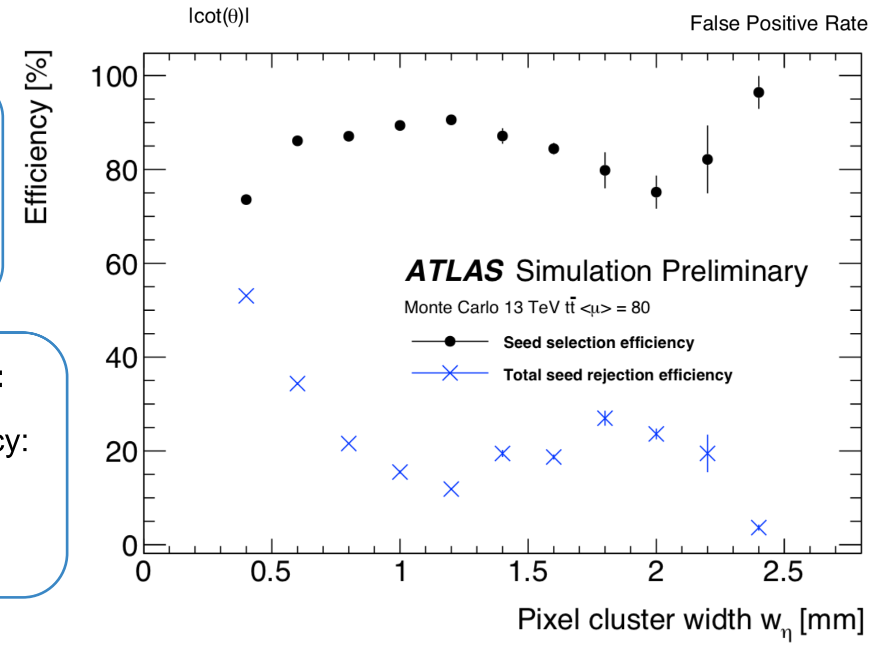


Doublet metrics:

Recall = 95%
Precision = 56%
F1 score = 71%

Triplet seed metrics:

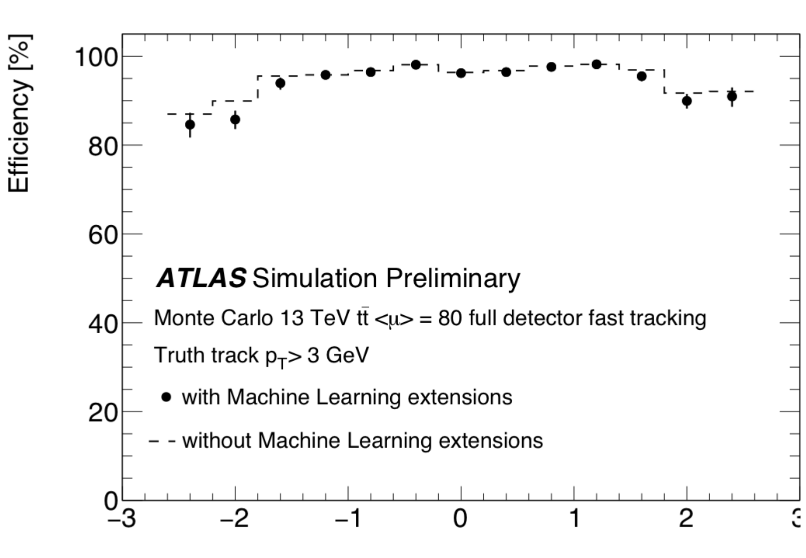
Seed selection efficiency: 74.8 ± 0.1%
Total seed rejection: 41.5 ± 0.1%

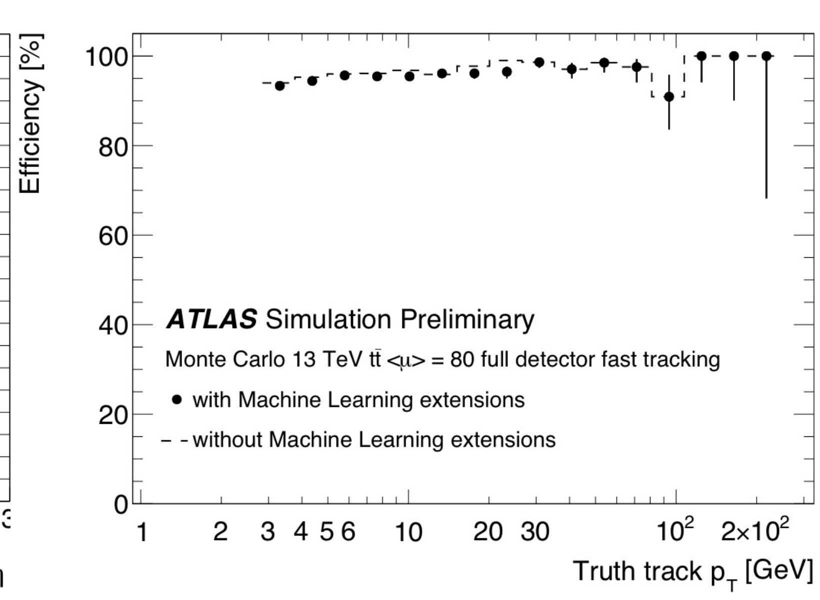


Performance Evaluation

Efficiency vs Track Parameters:

- <μ> = 80: **93.9% average tracking efficiency** (nominal 95%), with **2.3x speed-up factor**
- Efficiency loss is mainly observed at large |η|





Breakdown of speed-up factors observed for different stages in the HLT ID fast tracking algorithm at <μ> = 80 [1]

Total Speed-up Factor	Seed Generation	Seed Processing	Track Fitting
2.3×	1.3×	3.3×	1.5×

Performance of the full detector tracking at varying average pile-up multiplicities with application of ML extensions [1]

$<\mu>$	Efficiency Loss (%)	Total Speed-up Factor
40	0.7	1.6×
60	0.7	2.1×
80	1.1	2.3×

Conclusions:

- ML based classifier for seed selection in the HLT ID fast tracking algorithm has provided significant CPU savings with minimal loss in tracking efficiency
- Reducing the proportion of fake seeds at an earlier stage in the HLT track seeding, ensures the reduction of processing time in later tracking stages & hence saving CPU

Nisha Lad (UCL) on behalf of the ATLAS Collaboration
Dmitry Emeliyanov (STFC RAL), Nikos Konstantinidis (UCL)

References:
[1] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HLTTrackingPublicResults

ATLAS EXPERIMENT