



B-Tagging through G.N.N

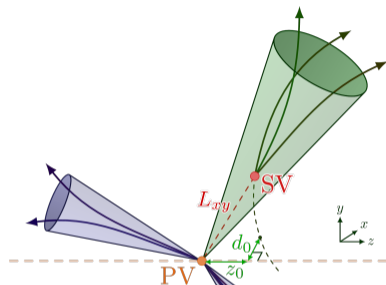
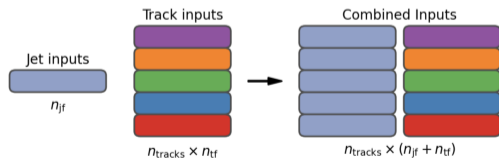
Cauê Evangelista

Sprace

- Previous Chapters
- Today
 - Track Analysis in the CMSSW
 - Overall Structure
 - Available features
 - Documentation
 - B-Tagging in the CMSSW
 - Documentation
 - Some issues
 - Track Initialiser: DeepSets Paper
 - Formalism
 - Application: Sum of digits
 - Track Initialiser: Deep Sets Code
 - Application Code: Sum of digits

Previous Chapters - Tracks

- Just Particle Cloud it's not enough!
- How to simulate Tracks alongside Jets?
- Which features are available to us? ¹



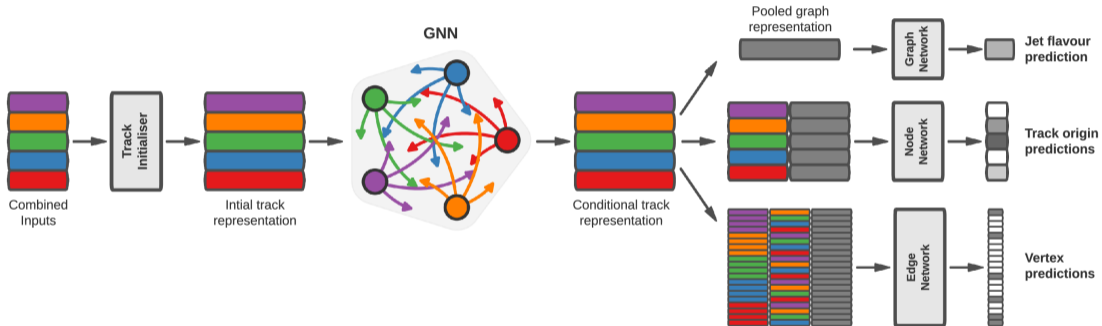
¹Atlas Collaboration. "Graph Neural Network Jet Flavour Tagging with the ATLAS Detector", 2022.

Parameter	Selection
p_T	> 500 MeV
$ d_0 $	< 3.5 mm
$ z_0 \sin \theta $	< 5 mm
Silicon hits	≥ 8
Shared silicon hits	< 2
Silicon holes	< 3
Pixel holes	< 2

Previous Chapters - Tracks Features

Jet Input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	Description
q/p	Track charge divided by momentum (measure of curvature)
$d\eta$	Pseudorapidity of the track, relative to the jet η
$d\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin \theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used in the reconstruction of an electron or muon (only for GN1 Lep)

Previous Chapters - Architecture



Track Analysis in the CMSSW - Overall Structure

□ Three main files: Analyzer, Build File, Configuration File

- Analyzer: is the place to put your analysis code. It creates the classes used on the Configuration's modules later on. It's a .cc file and can be created (alongside all the others needed files) using the [mkedanlzc command](#)
- Build File: .xml file used to compile the classes written in the Analyzer. The 'scram b' command starts the compilation. More information about the structure of a typical BuildFile can be found [here](#)
- Configuration File: .py file that sets information about your analysis as the number of events, source file, what classes are going to be used and in what order (this is important). A detailed example can be found in the [SPRACE CMSSW Tutorial](#)

Track Analysis in the CMSSW - Available features and Documentation

□ Available features

- A lot of methods are stored in the [TrackReco](#) folder in [CMSSW github](#)
- Features accessed: quality criteria, number of valid hits, number of lost hits, point of closest approach(CMS center), PT, Charge, Transverse IP, Longitudinal IP, Curvature
- Hits in each detector region: working on it

□ Documentation

- Detailed information about how to perform a track analysis using an EDAnalyzer can be found in The CMS Offline WorkBook([4.1.2 Writing your own EDAnalyzer](#) and [7.2 Track Analysis](#))

□ Documentation

- Information about B-Tagging can be found in [session 7.9 of the CMS Offline Workbook](#)
- Difference from Track Analysis: the collection of objects we're using
- "In CMSSW 1.5 - 1.6, the "Jet Probability" algorithm gives the best performance and is consequently recommended for most physics analyses"
- The algorithms give a b-tag discriminator.

□ A Issue

- Modification of the BuildFile with the 'proper' line gives rise to a compilation error.
- Open ticket in CMS Talk to discuss it.

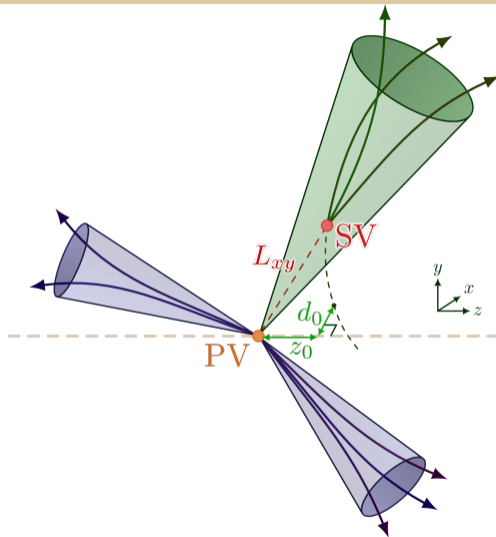
□ Available features

- A lot of methods are stored in the [TrackReco](#) folder in [CMSSW github](#)
- Features accessed: quality criteria, number of valid hits, number of lost hits, point of closest approach(CMS center), PT, Charge, Transverse IP, Longitudinal IP, Curvature
- Hits in each detector region: working on it

□ Documentation

- Detailed information about how to perform a track analysis using an EDAnalyzer can be found in The CMS Offline WorkBook([4.1.2 Writing your own EDAnalyzer](#) and [7.2 Track Analysis](#))

Track Analysis in the CMSSW - Available features and Documentation



□ Formalism

- Theorem 2: A function $f(X)$ operation on a set X having elements from a countable universe, is a valid set function, ie., invariant to the permutation of instances in X , if it can be decomposed in the form $\rho \left(\sum_{x \in X} \phi(x) \right)$, for suitable transformations ρ and ϕ
 - Replacing ρ and ϕ by universal approximators leaves matters unchanged, since, in particular, ρ and ϕ can be used to approximate arbitrary polynomials. Then, it remains to learn these approximators
 - Each instance x_m is transformed (possibly by several layers) into some representation $\phi(x_m)$
 - The representations $\phi(x_m)$ are added up and the output is processed using the ρ network in the same manner as in any deep network (e.g. fully connected layers, nonlinearities, etc.)
- Sum of Digits: 100k sets of at maximum 10 digits. The set label is the sum of the digits in that set. The testing data was another 100k sets, but in this case, the number of digits lies in (5, 100)

²Zaheer, Manzil, et al. "Deep sets." *Advances in neural information processing systems* 30 (2017)

Track Initialiser: DeepSets Code

