Run 1 legacy performance: Tracking, b-tagging and muons

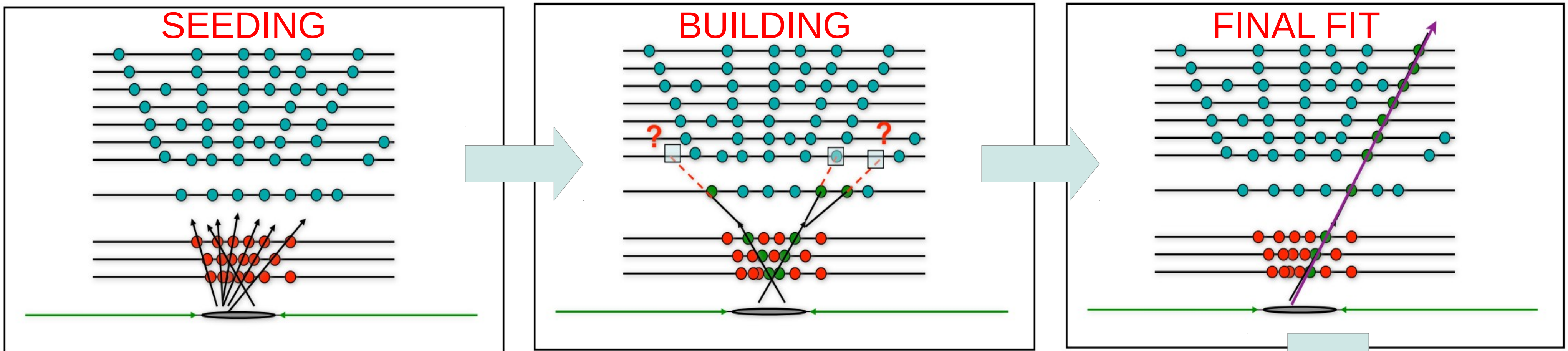
Andrea Rizzi, on behalf of the CMS collaboration
Edward Moyse, on behalf of the Atlas collaboration

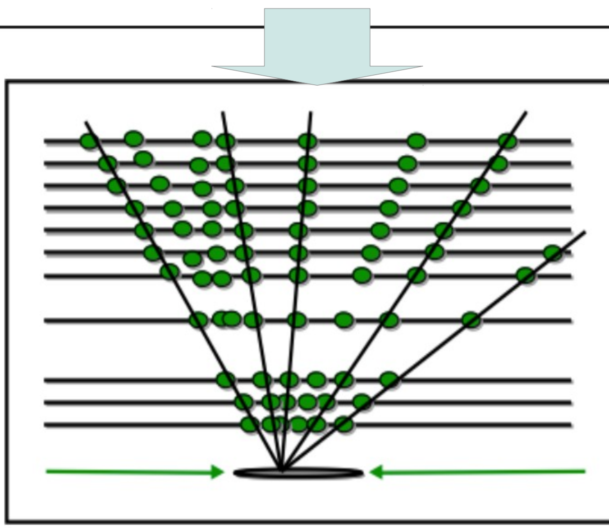
This part: Tracking and b-tagging (a CMS view)
Next part: Tracking and muons (an Atlas view)

Outline

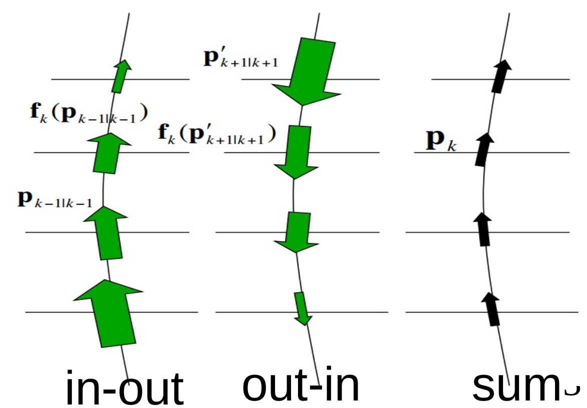
- Tracking in CMS
- High-pt jet tracking and cluster splitting
- B-tagging at LHC
- B-tag efficiency measurements
- Conclusions (Run-1 lessons)

CMS tracking



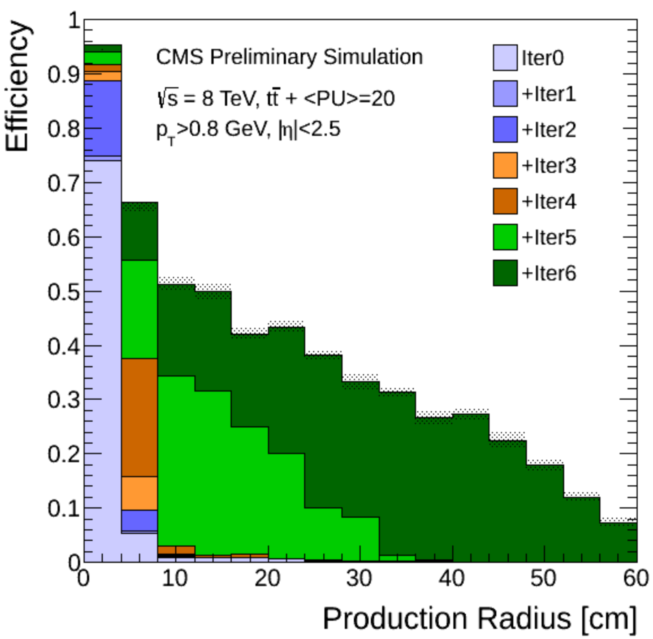


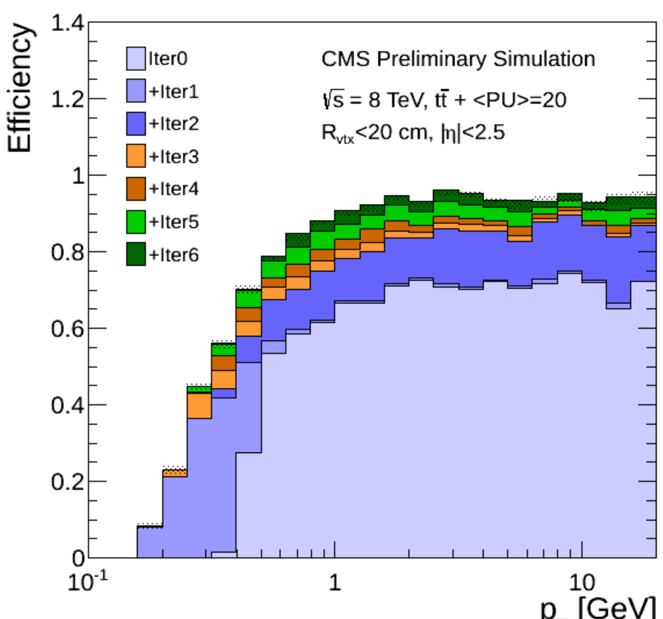
- Track reconstruction in CMS is based on a KF technique: starting from some initial track hypothesis (seed) several measurements (hits) are attached one by one updating at each iteration the track “state”
- The seeding is mostly made with low occupancy, high resolution pixel detector and is constrained to be compatible with track origin in the luminous region
- Only patterns with large number of hits attached and with limited number of “holes” are kept in order to avoid “fake” tracks from random alignments. Tracks sharing too many hits are also cleaned as duplicates
- Final fitting is performed as a smoothing combining the in-out states collection with the out-in one

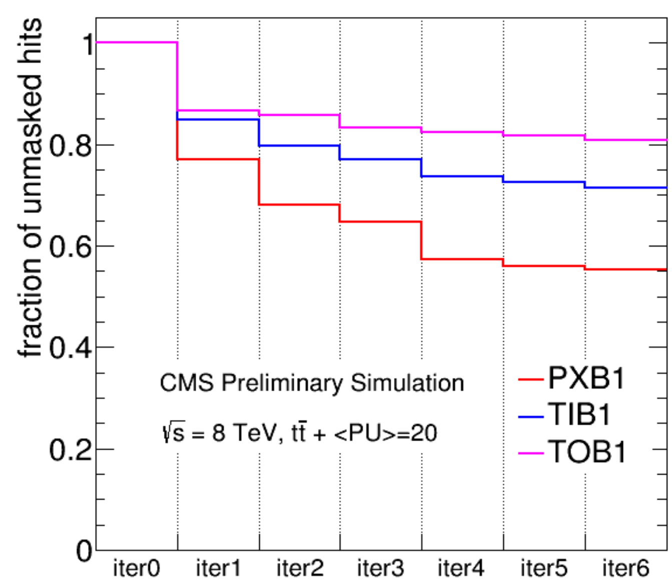


CMS iterative tracking

- In order to keep the reconstruction time under-control while still being able to reach very low pt, a divide&impera approach has been followed
  - First we attempt reconstructing prompt tracks (seeding in pixels) above some pt threshold and with clear pattern (e.g. triplets of pixel hits)
  - Then we go step by step in the more complex scenarios (displaced tracks, low pt tracks, tracks with only two hits in the pixels)
  - At each step we remove assigned hits to ease the next step
- Iterative tracking, including soft particles, is a key element to allow CMS particle-flow approach



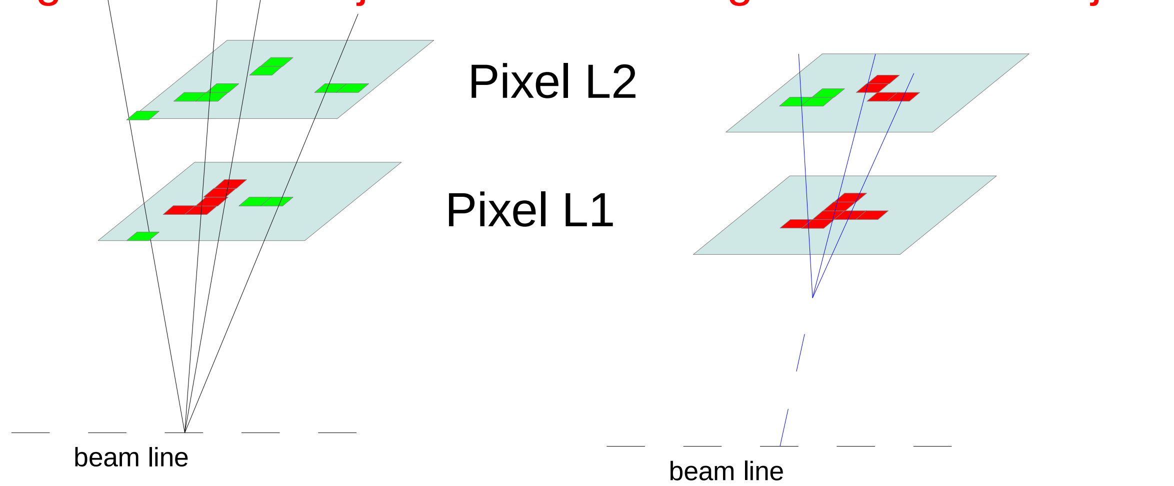


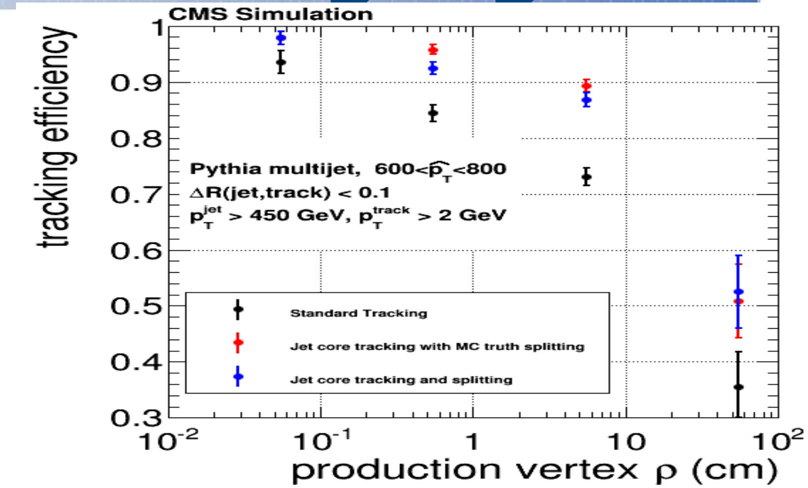


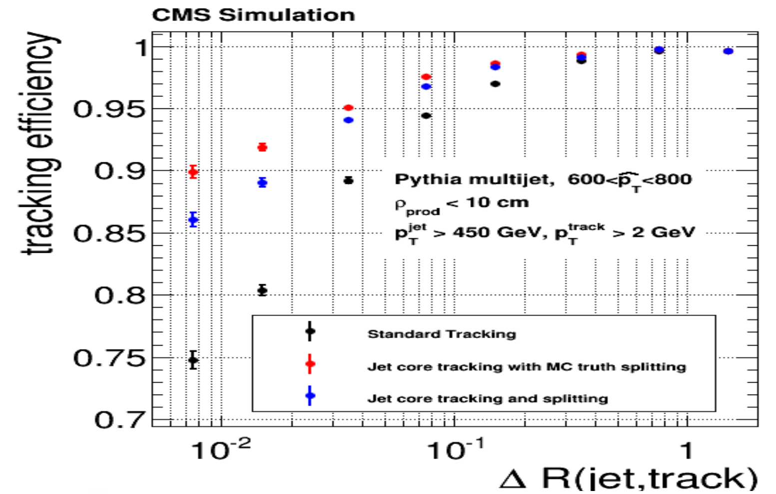
High pt and cluster merging

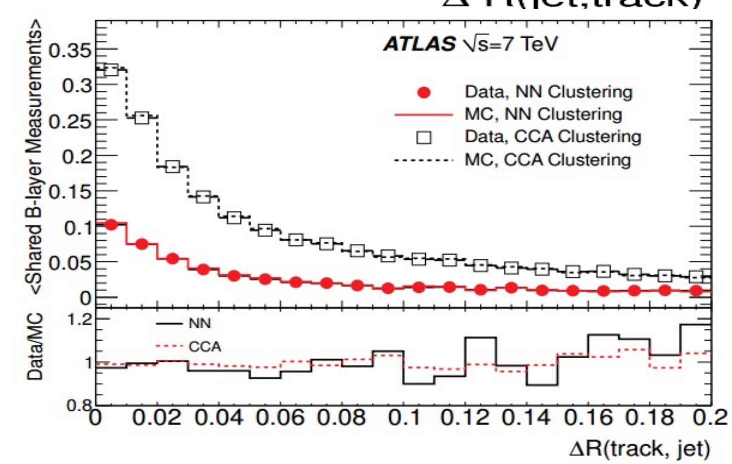
Merged clusters in jets

Merged clusters in b-jets



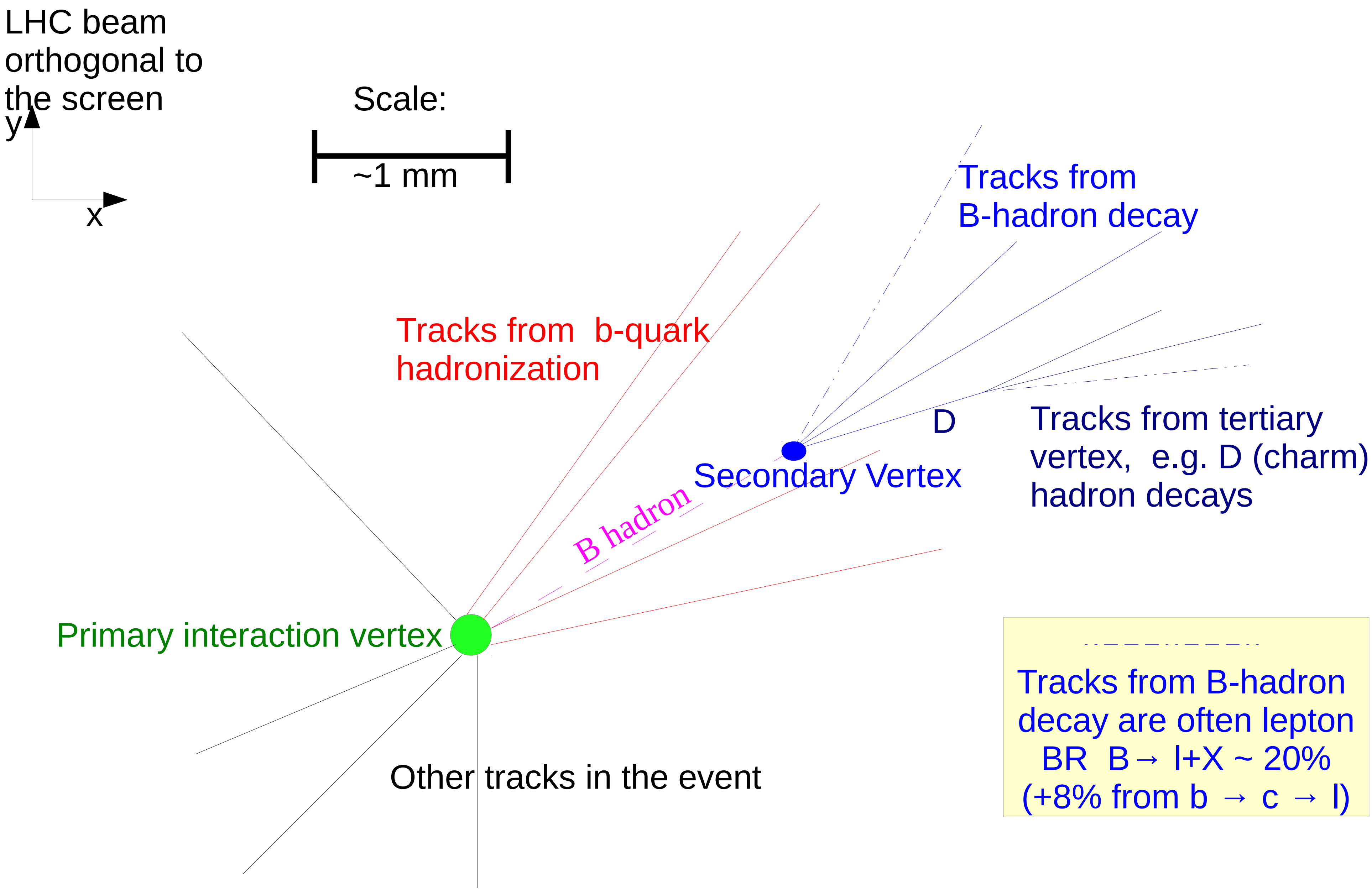






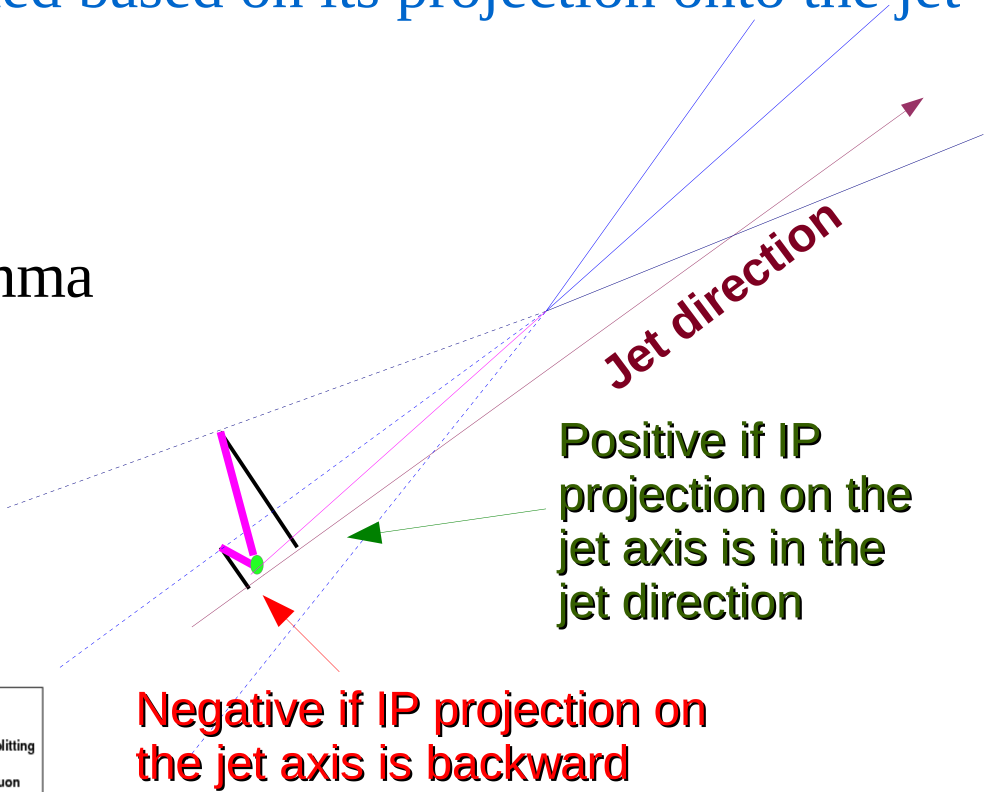
- Hits from different tracks can result in a merged pixel cluster in the core of high pt jets
- The effect is even more pronounced for b-jets because the B decay happens closer to the pixel detector
- In addition the complexity of pattern recognition in the core of a high pt jets increases non-linearly
- Both Atlas and CMS recently developed solutions to mitigate this problem

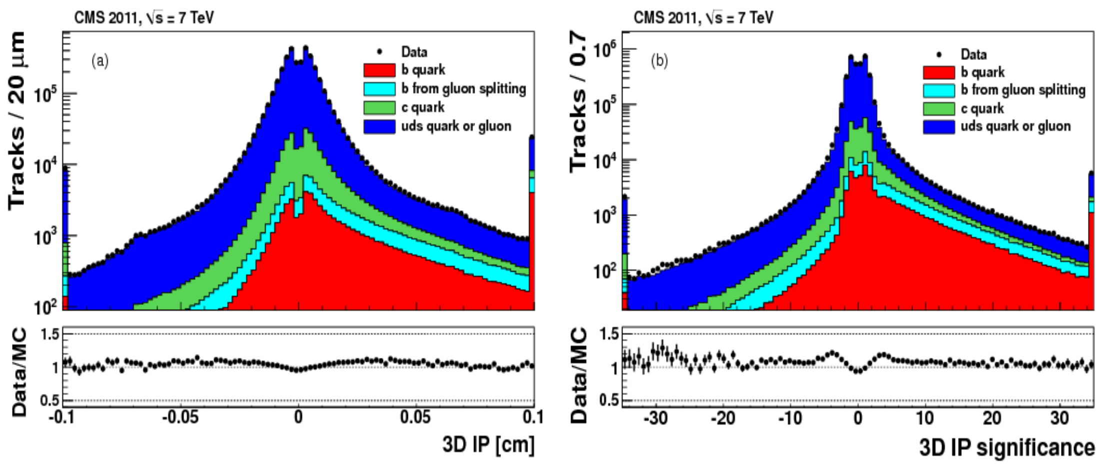
B-jet zoom-in


LHC beam
orthogonal to
the screen
y
x
Scale:
~1 mm
Tracks from
B-hadron decay
Tracks from b-quark
hadronization
D
Tracks from tertiary
vertex, e.g. D (charm)
hadron decays
Secondary Vertex
B hadron
Primary interaction vertex
Other tracks in the event
Tracks from B-hadron
decay are often lepton
BR B→ l+X ~ 20%
(+8% from b → c → l)

Impact parameter information

- A key feature to distinguish b-jets from gluon or light-quark jets is the impact parameter (IP) of the tracks
- For tracks in a jet the IP can be signed based on its projection onto the jet direction
- The IP is boost invariant:
  - The decay angle shrink as 1/gamma
  - The flight path grows as gamma
- Typical B IP is close to resolution limits → use significance

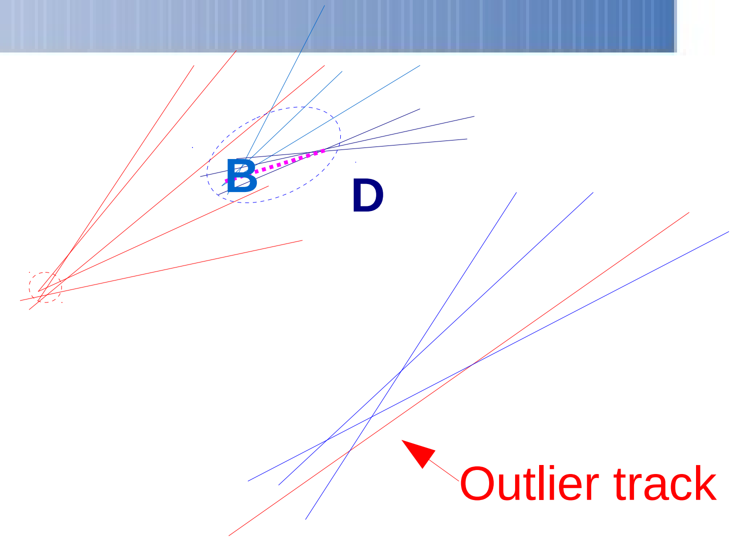


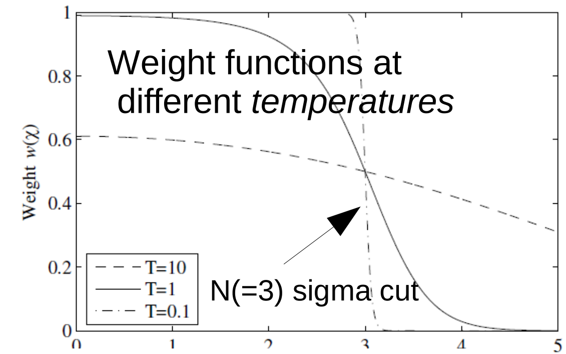


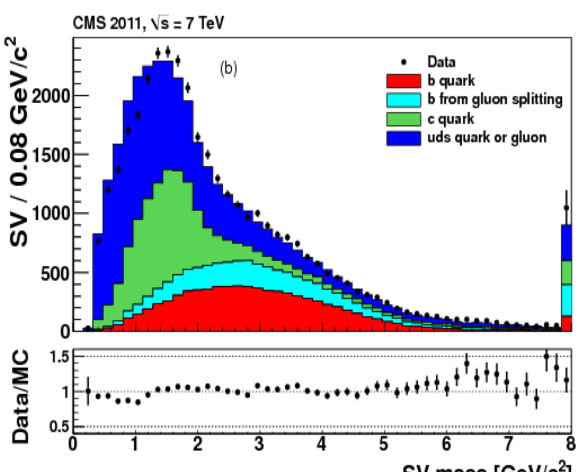
Geant4 full simulation nicely reproducing the IP distribution

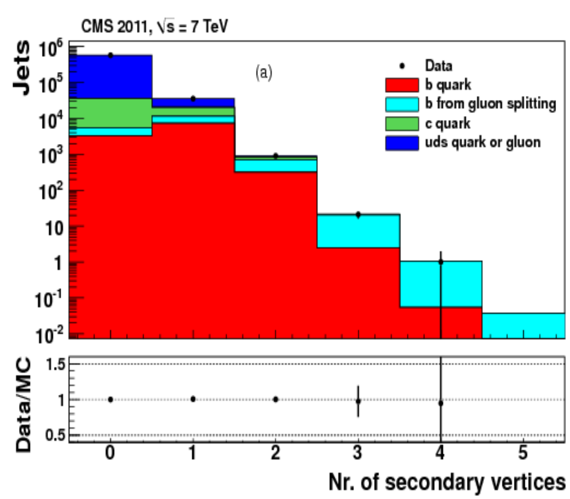
Vertexing

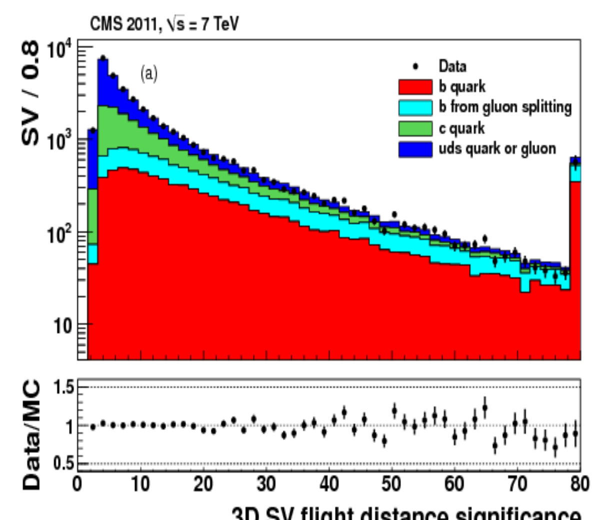
- Tracks from B decay originate from a single point → Secondary Vertex
- A simple vertex finding algorithm is to fit all tracks in a jet with an outlier resistant fitter
  - At each step remove the tracks already used
  - Tracks are assigned depending on a weight, based on their distance to vertex significance
  - An annealing technique is used for better convergence
- While the presence of a secondary vertex is already a strong indication for a B, the vertex properties can be exploited to better discriminate B from C and light







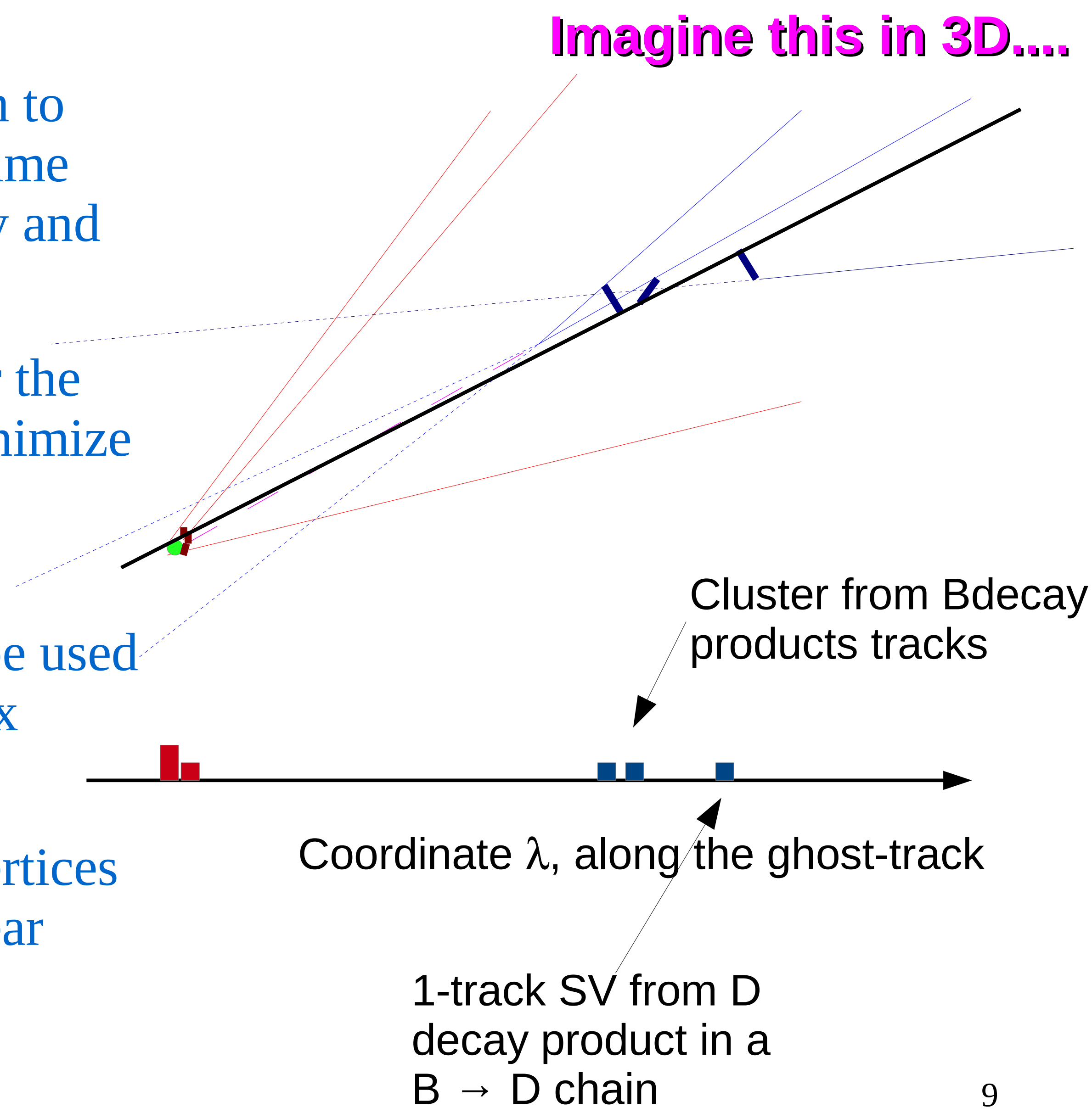




B flight direction & vertex fitting

This is know as “ghost track vertex finder” in CMS and “JetFitter” in Atlas

- An alternative, B specific, approach to vertex finding is to fit at the same time the B flight direction and secondary and tertiary vertices
- All tracks contribute if one look for the direction, starting from PV, that minimize the sum of distances (or chi2) of all tracks in the jet
- Distance along such direction can be used as a simple 1D coordinate for vertex finding
- This in principle allow “1-track” vertices assuming B and D are quasi-collinear



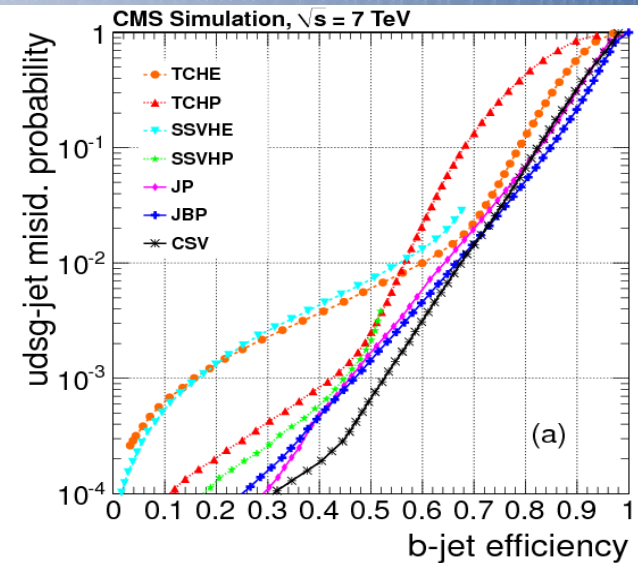
Discriminators

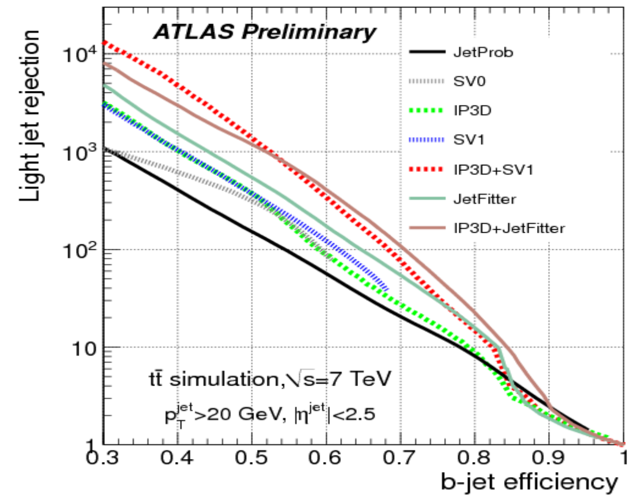
The typical outcome of b-tagging is a continuous variable that can be used to cut or to weight the events in an analysis

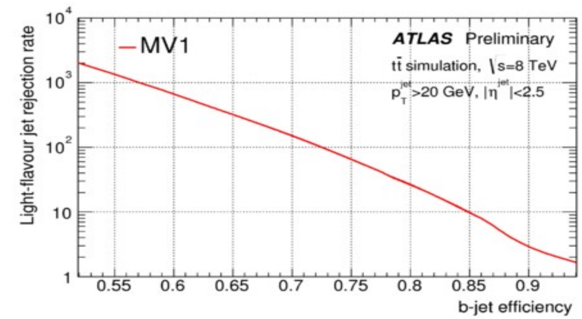
- Simple algorithms
  - Just cut/return the significance of track IP or vertex displacement
  - In presence of lepton, exploit the IP or pTrel of the lepton
- “linear” combined algorithms:
  - Compute jet probability from track IP significance
  - Make likelihood ratios for S vs B from distribution of several variables (e.g. secondary vertex mass, flight distance, etc..) CMS CSV
  - Combine LR discriminant from several sources (IP, SV, lept)
- Multivariate algorithms:
  - Use ANN or other techniques to combine different inputs Atlas MV1
  - Use ANN to combine several of the above outputs

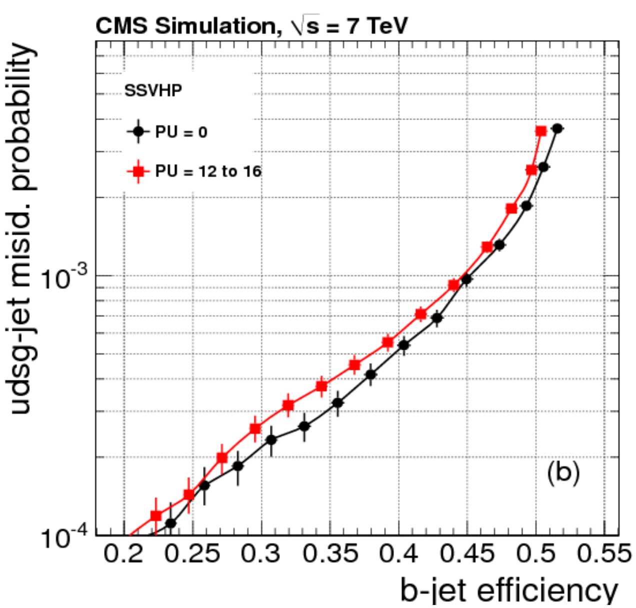
Performance

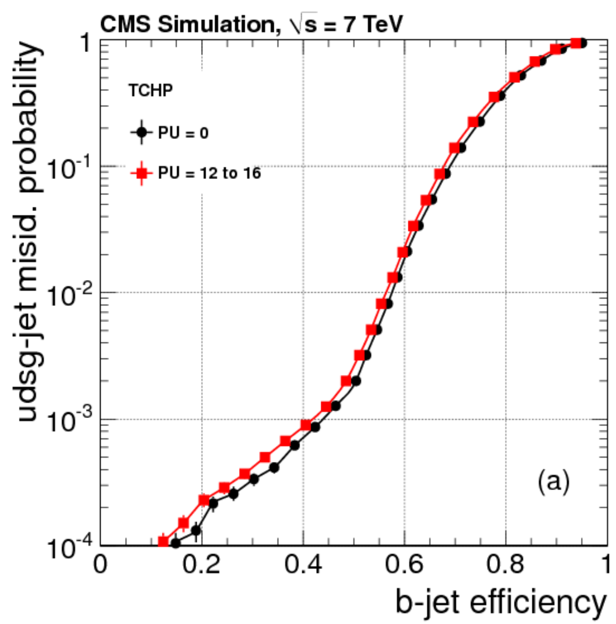
- Similar performance in Atlas and CMS for similar algorithms
- Default algorithms for Atlas (MV1 or IP3D+JetFitter) performs better than CMS CSV
  - CMS should better use its MVA algorithms in Run2!
- Effects of PU not dramatic, but visible, in Run1. PU mitigation is a must for Run2!





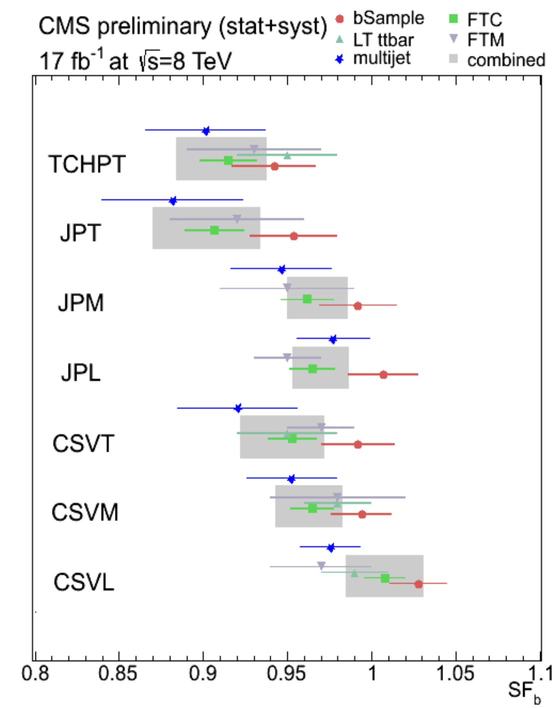


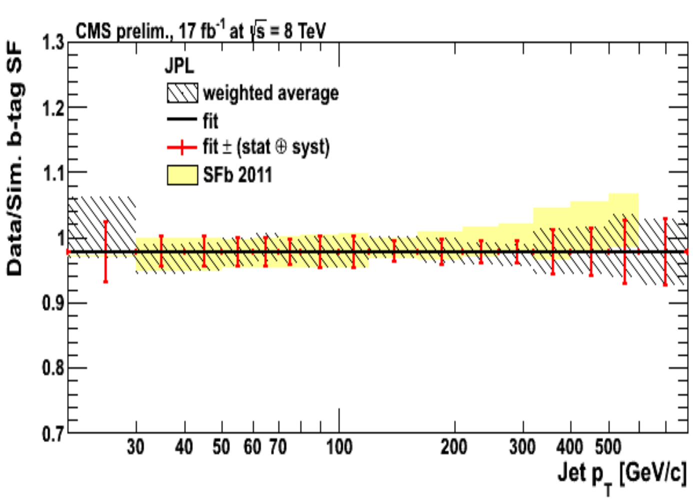


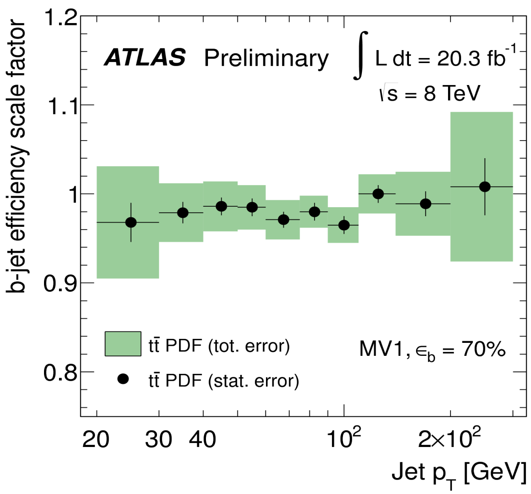


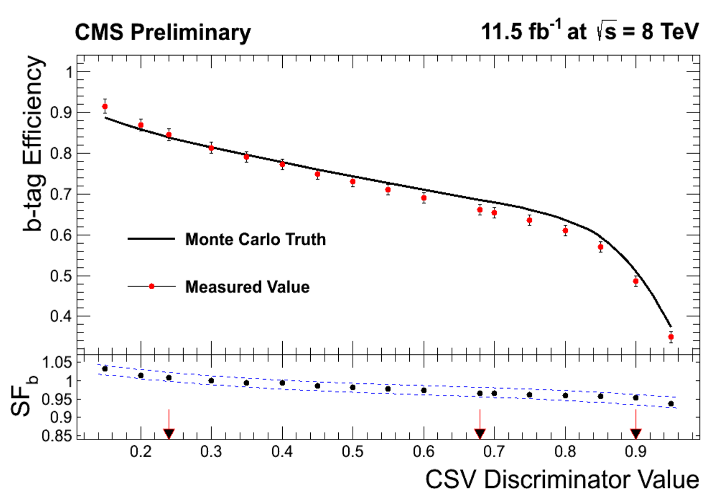
Efficiency measurements

- Historic method for efficiency measurement based on muon ptRel tested at LHC by both experiments
  - Large systematics starting at pt > ~ 100 GeV
  - Dependency on MC for pT-rel distribution prediction
- Algorithm "cross calibration":
  - E.g. use IP algos as a "tag" and CSV as a "probe" → lots of correlations
- Golden method: use ttbar events (dilep and semi-lep)
  - ttbar events, especially di-leptonic can be easily selected in a pure sample without any b-tag requirement
  - Assuming Vtb=1, a pure sample of b is then available. Different stat techniques are developed to fit scale factors from such sample
  - Precision of ~2% achieved both in Atlas and CMS



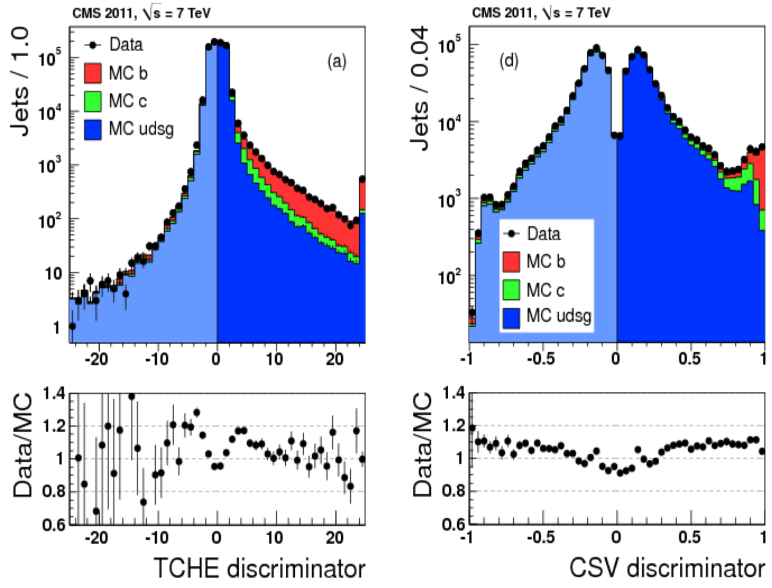


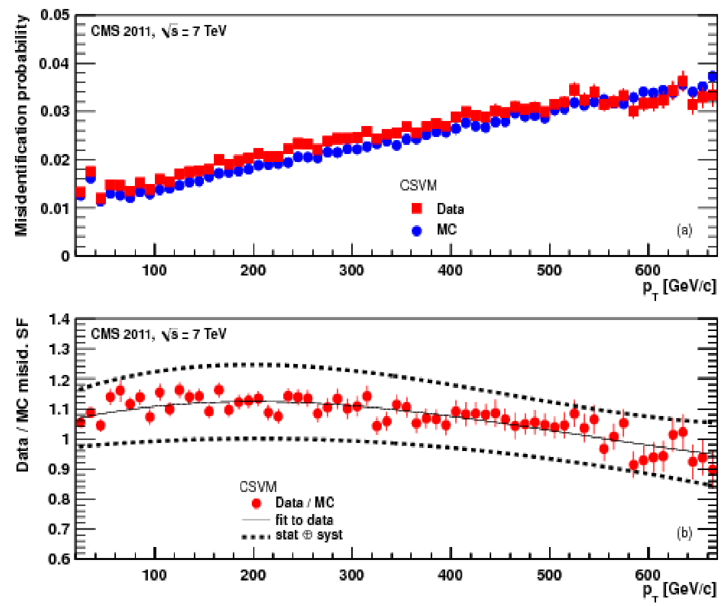




Mistag measurements

- Main method for light mis-tag measurement is the “negative tagging”
  - Give lifetime sign to tracks and vertices decay length based on jet direction
  - While B are mostly positive, the light have rather symmetrical distribution
  - Measure the “negative side” on data to predict the positive
- Several weak points:
  - Light jets have genuine (asymmetric) lifetime (K0, Lambda, Nuc. Int., gamma conversions, etc..) that has to be still corrected from MC
  - Tells very little about the correlations used in MVA based techniqes
- Investigate other techinques?
  - ttbar may also allow to select a pure sample of hadronic W





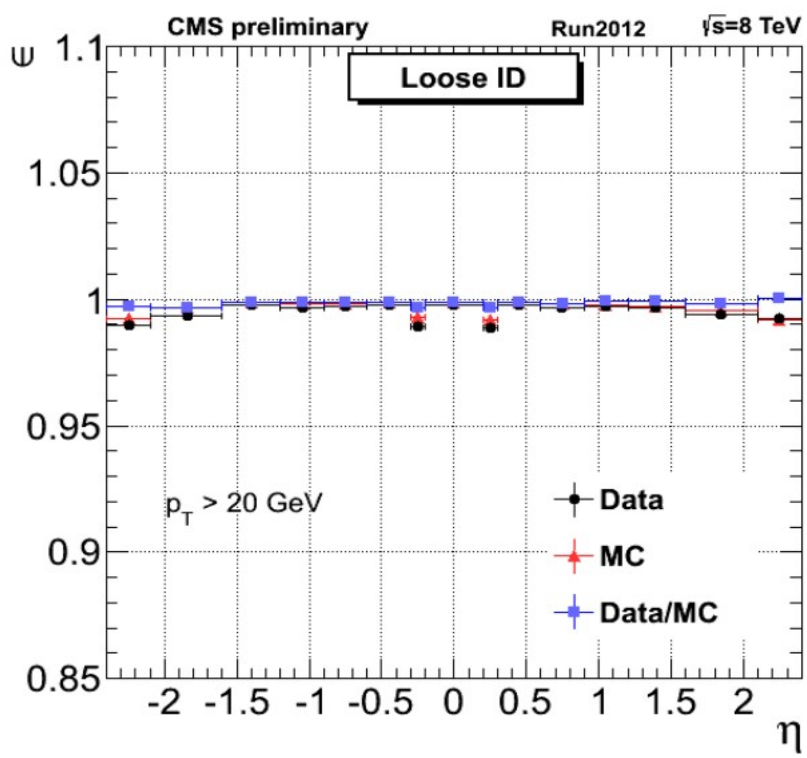
Conclusions

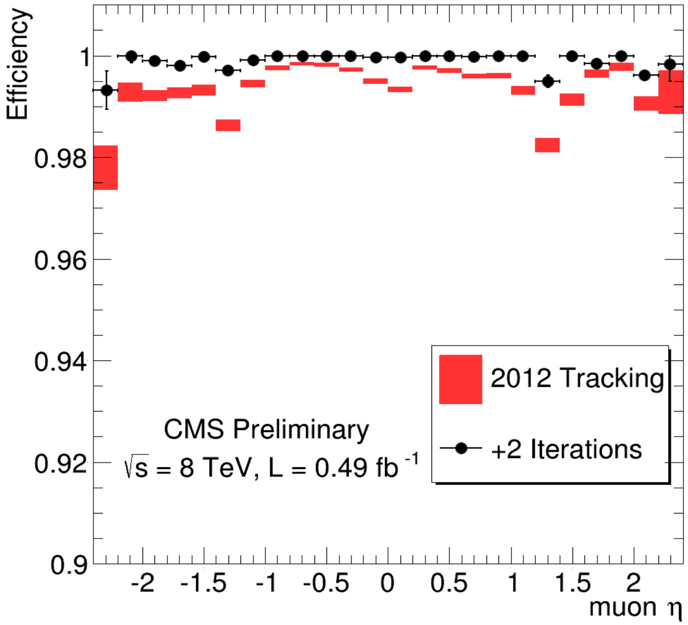
- Tracking and b-tagging performed as expected from simulations in Run1 (data/MC scale factor very close to 1)
- Some problems observed in the “tails” during Run1 (high PU, high pT jets) could be much more common in Run2, so fixes and mitigation campaigns are ongoing
- Performance measurements for b-tagging is at level of few percent and the large top production provides a great source of calibration “b”

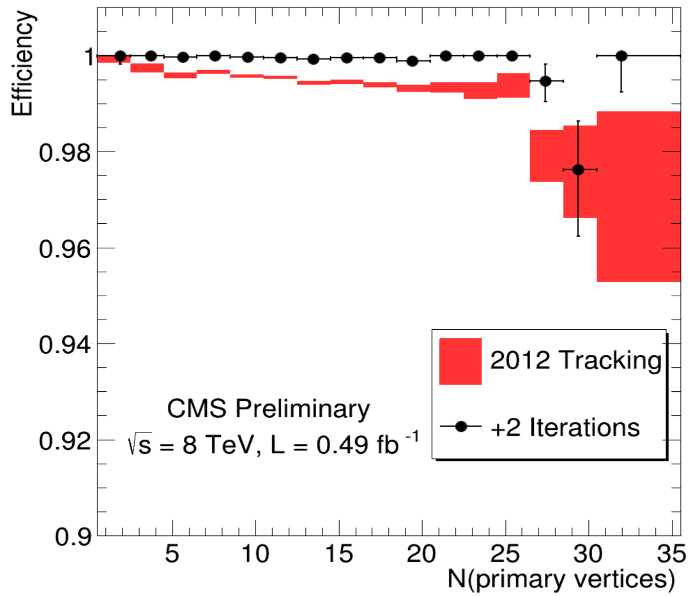
backup

Muon reco

- Great Data/MC agreement: few permille
- Additional iterative steps developed during Run1 to recover performance loss (few % anyhow!) with PU







CMS tracker

- Pixel + Strips
- Barrel (3+4+6 layers)
- Endcap (2+3+9 disks)
- ~70M readout channels
- Covers |eta| <2.5
- Analog readout