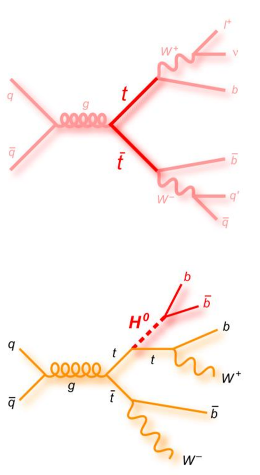
Performance and calibration of b-tagging with the ATLAS experiment at LHC Run 2

Gordon Watts
UW/Seattle
For the ATLAS Collaboration

Abstract: The identification of *b*-flavored jets is key to many physics analyses at the LHC, including measurements involving Higgs bosons or top quarks, and searches for physics beyond the Standard Model. The capacity of ATLAS to efficiently tag b-jets has been enhanced for Run-2 with the addition of the Insertable B Layer (IBL), and improvements in the tracking and b-tagging algorithms. In the algorithm optimization special emphasis has been placed in improving the performance for reconstructing high p_T *b*-jets, addressing the challenges posed by track and vertex reconstruction in such an environment. The efficiency and rejection power of these algorithms have been calibrated on data taken in 2015, in particular by exploiting the copious production of b-jets in top quark decays, complemented by studies in multi-jet events. First results from the 2016 data will also be shown.

Anatomy of a B Decay

Jets containing *b* quarks look a lot like a jet from any other quark!



Typical mass of a Bottom Quark Meson: ~5.4 GeV/c^2

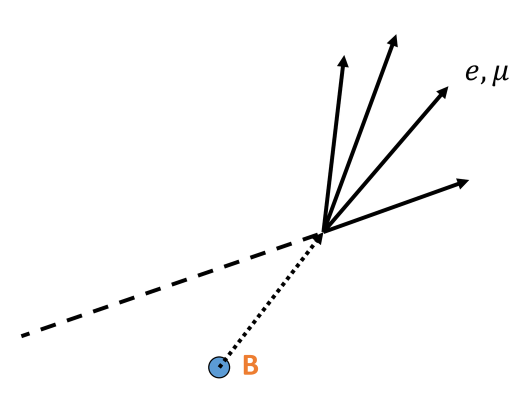
Plays a unique role in physics, for example the Wtb vertex or the Higgs decay ($h \rightarrow b\bar{b}$)

Because B mesons decay via the weak force, decays are delayed compared to other particles. This leads to physics features we can use to identify bottom-quark jets.

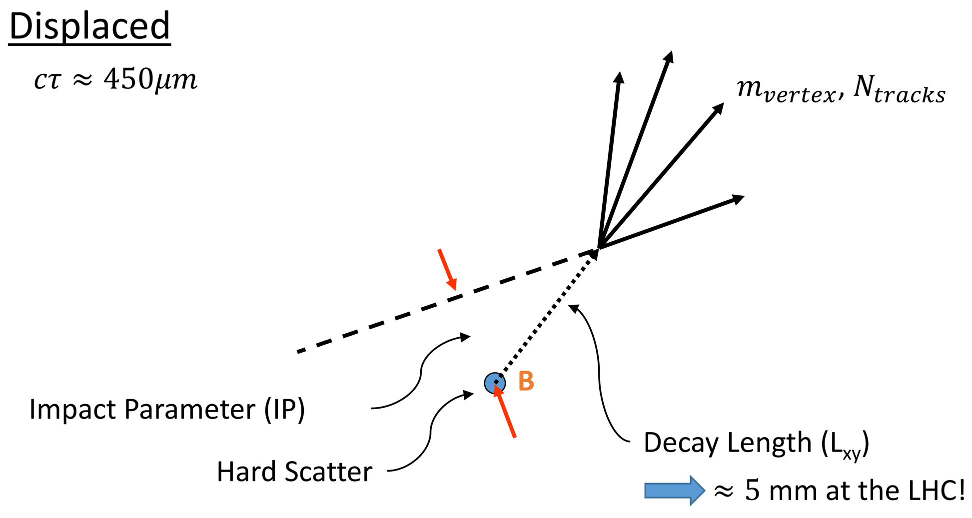
Many standard model backgrounds to these signals contain light jets

- Being able to separate them opens up channels and discoveries not otherwise possible
- Instrumental in the discovery of the top quark at the Tevatron
- Crucial for measuring the Higgs coupling to heavy flavor, almost all $t\bar{t}$ measurements, and many SUSY channels that include heavy flavor.

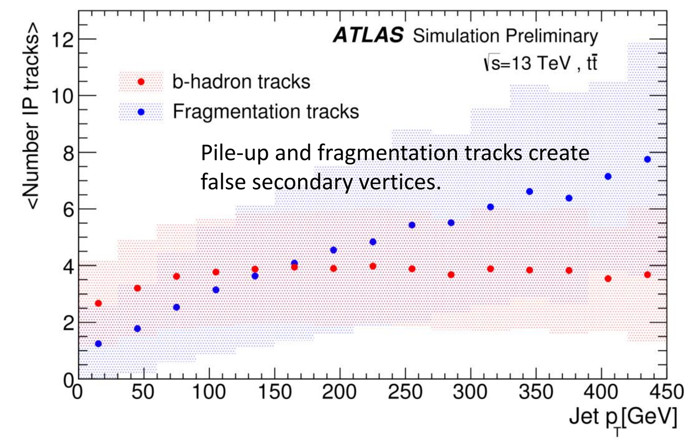
Two types of Decays



- ~20% of decays all *b*-decays
- No need of pixel detectors to identify (e.g. DZERO top discovery)
- Efficiency is low



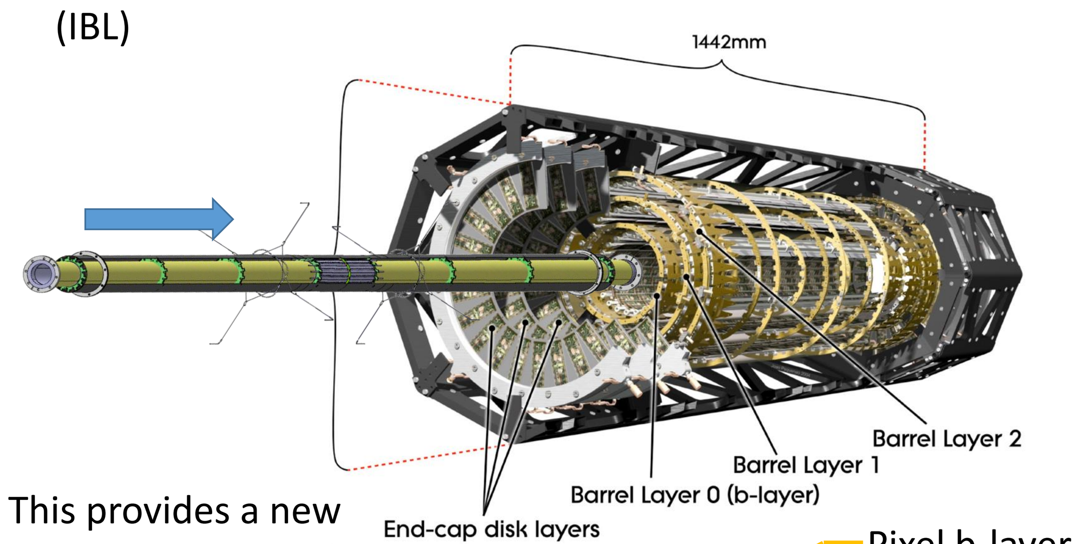
- Look for tracks with large impact parameter
- Fit for a displaced vertex
- Requires high precision tracks



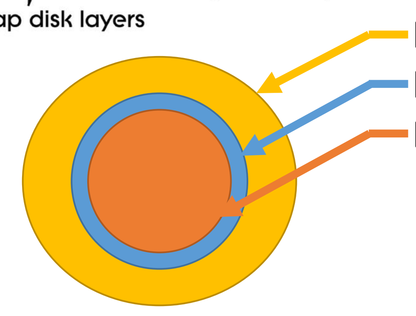
Run 2 Detector Improvements

Taking advantage of small distances requires specialized detectors.

- At first silicon strip and now pixel detectors
- Constant evolution of technology pushing the readout speed and spatial resolution
- Modern pixel detectors in an experiment like ATLAS have O(1 million) active elements.
- The most relevant improvement for *b*-tagging is the insertable b-layer (IBL)



This provides a new measurement much closer to the interaction point, which improves the track d_0 measurement, one of the key variables in *b*-tagging

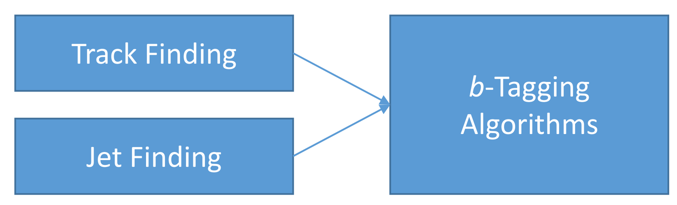


ATLAS-TDR-019

Finding A Bottom Quark Jet

In the most common implementation jets are first found and then tracks near the jet are examined for displaced tracks and vertices.

- ATLAS uses a cone whose size is p_T-dependent around the jet axis to associate tracks.
- Trackless algorithms use jet-like algorithms to cluster tracks before applying the same track based algorithms as in the jet case.



The *b*-Tagging algorithm derives most of its power from track-based kinematic variables

- The impact parameter and secondary vertex finding are the traditional main variables.
- There are algorithms that try to take advantage of the cascade charm meson decay.
- Most modern algorithms involve a neural net or boosted decision tree in some form.

The High p_T Problem

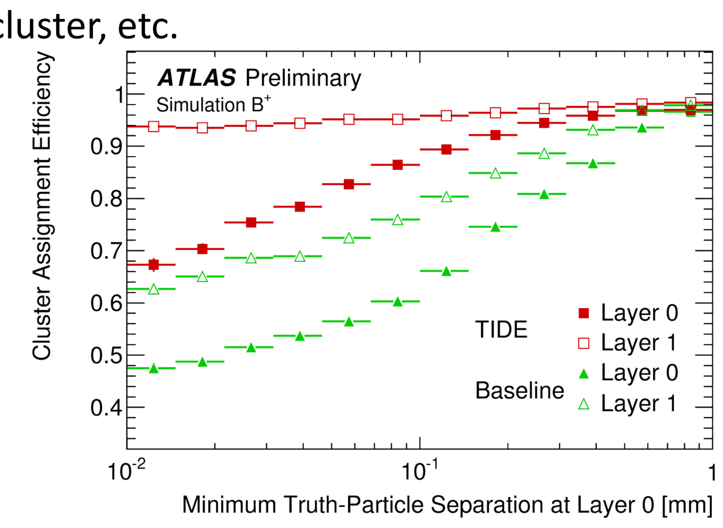
Track Density is the main issue with finding b-jets in high p_T jets.

- ATLAS as improved its hit clustering algorithm in Run 2
- The improvement in hit cluster finding has a direct effect on track finding in dense environments, which improves *b*-tagging in high p_T jets.
- As the LHC and ATLAS probe higher and higher scales, *b*-tagging must work at these higher and higher jet energies. Both the algorithms and calibration are severely impacted by this environment.
- Upgrades were made to the hit-clustering to improve track-hit assignment in high occupancy regions.
- This work concentrated on shared clusters – the cases where two particles went through neighboring or the same pixel hits.
- Hit sharing decisions now incorporate tracking information - # of clusters already on a track, the type of cluster, etc.

A new Neural Net algorithm has been designed to improve the assignment.

New

Old



ATL-PHYS-PUB-2015-006

The MV2 ATLAS *b*-Tagging Algorithm

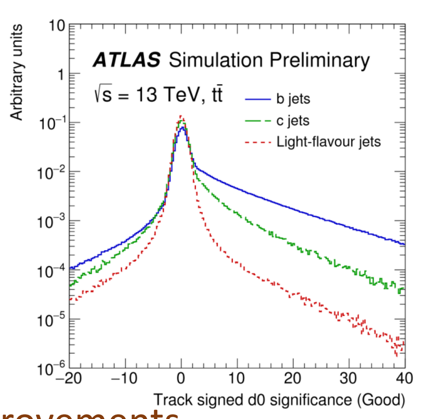
ATL-PHYS-PUB-2015-022
ATL-PHYS-PUB-2016-012

The MV2 algorithm, a boosted decision tree based algorithm, uses the kinematic variables from a jet (p_T, η) and three separate algorithms focused on impact parameter, vertex finding, and cascade $B \rightarrow D$ decays

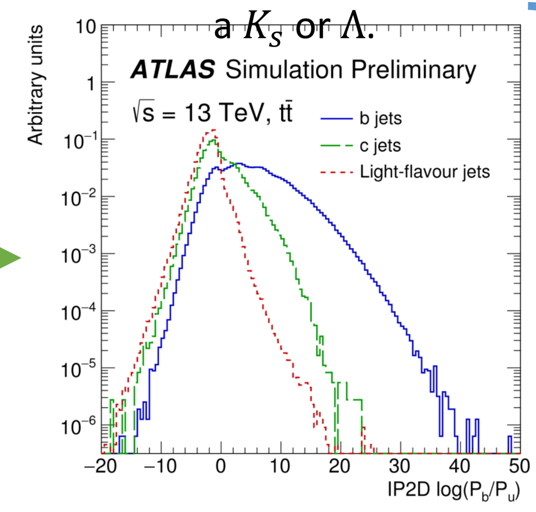
IP2D and IP3D

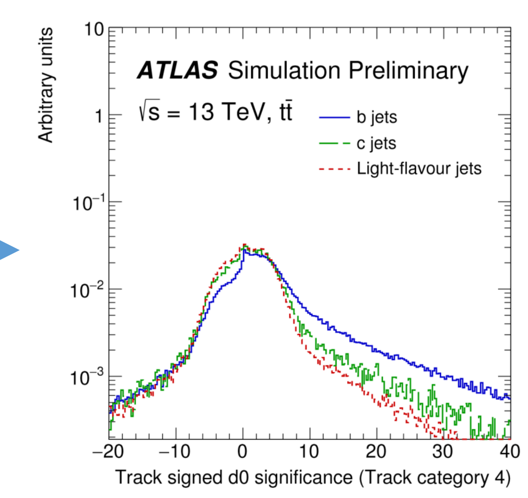
A track from a displaced *b*-decay will likely have a large impact parameter (d_0).

- Because of tracking resolution there will be an associated uncertainty for d_0.
- IP3D and IP2D use the shape of resolution curve to assign a probability to each track.



The impact parameter distributions are PDF's that can be used to determine the probability of a track being from a *b*, *c*, or light quark.





Improvements

Different types of tracks have different resolutions!

- Poor track in the same category-type as good track will degrade a good track-category's probability curve
- Added a number of new track categories – especially ones involving the track having an expected hit in the IBL and that hit being present
- These include categories that acknowledge some high p_T *b*-hadrons will decay *outside* the inner-silicon layer

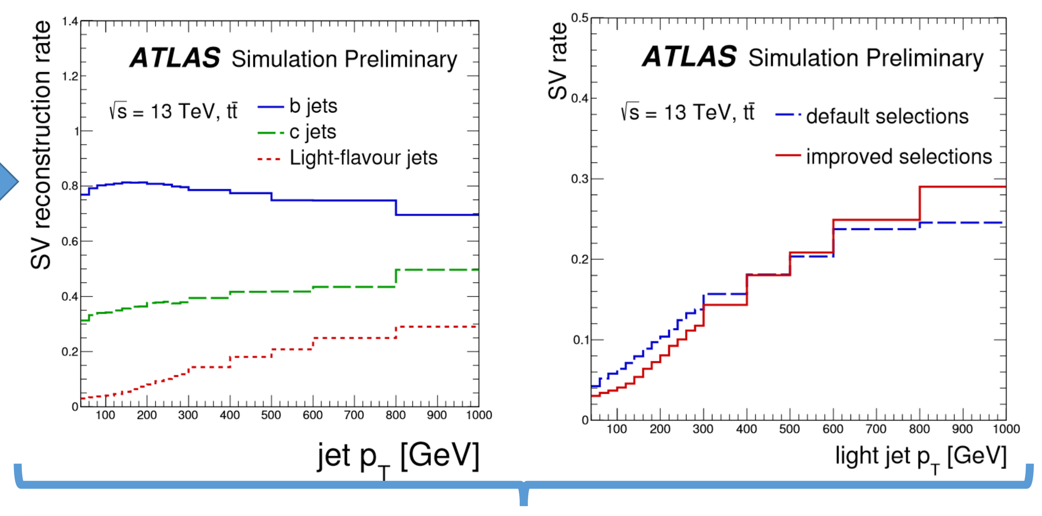
#	Category	Fractional contribution [%] *b*-jets	*c*-jets	light-jets
0	No hits in first two layers; expected hit in IBL and b-layer	1.9	2.0	1.9
1	No hits in first two layers; expected hit in IBL and no expected hit in b-layer	0.1	0.1	0.1
2	No hits in first two layers; no expected hit in IBL and expected hit in b-layer	0.04	0.04	0.04
3	No hits in first two layers; no expected hit in IBL and b-layer	0.03	0.03	0.03
4	No hit in IBL; expected hit in IBL	2.4	2.3	2.1
5	No hit in IBL; no expected hit in IBL	1.0	1.0	0.9
6	No hit in b-layer; expected hit in b-layer	0.5	0.5	0.5
7	No hit in b-layer; no expected hit in b-layer	2.4	2.4	2.2
8	*Shared* hit in both IBL and b-layer	0.01	0.01	0.03
9	At least one *shared* pixel hits	2.0	1.7	1.5
10	Two or more *shared* SCT hits	3.2	3.0	2.7
11	*Split* hits in both IBL and b-layer	1.0	0.87	0.6
12	*Split* pixel hit	1.8	1.4	0.9
13	*Good*	83.6	84.8	86.4

Secondary Vertex Finder (SV)

Iterative Vertex Finding Algorithm. It starts by forming all possible 2-track vertices, and then combining nearby tracks to build a final vertex.

Improvements

- Tracks are ordered in p_T and cut off after 24 to limit fake vertices due to tracks from fragmentation. This is particularly useful in high p_T jets.
- Tracks with too small or too large an impact parameter ($\left|\frac{d_0}{\sigma_{d_0}}\right| < 2, \left|\frac{z_0}{\sigma_{z_0}}\right| > 6$) are removed to reduce effects due to pile up.
- The detector hit pattern must be consistent with the location of each 2-track seed vertex (e.g. remove tracks which have hits before the vertex)
- Remove 2-track vertex seeds consistent with material interactions, mass of greater than 6 GeV/c^2, or consistent with a K_S or Λ.



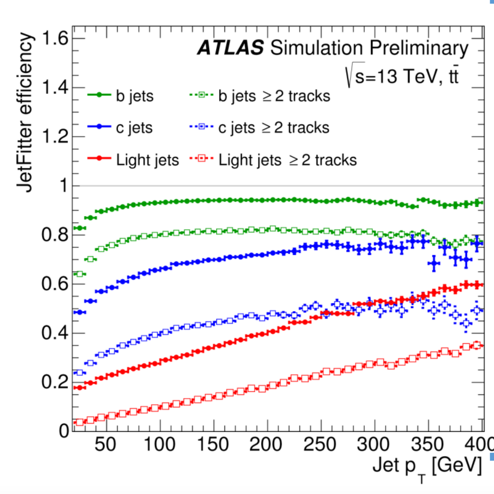
$m(\mathrm{SV})$	Invariant mass of tracks at the secondary vertex assuming pion masses
$f_E(\mathrm{SV})$	Fraction of the charged jet energy in the secondary vertex
$N_{\mathrm{TrkAtVtx}}(\mathrm{SV})$	Number of tracks used in the secondary vertex
$N_{\mathrm{2TrkVtx}}(\mathrm{SV})$	Number of two track vertex candidates
$L_{xy}(\mathrm{SV})$	Transverse distance between the primary and secondary vertices
$L_{xyz}(\mathrm{SV})$	Distance between the primary and secondary vertices
$S_{xyz}(\mathrm{SV})$	Distance between the primary and secondary vertices divided by its uncertainty
$\Delta R(\mathrm{jet},\mathrm{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex

Kinematics	$p_T(jet)$ $\eta(jet)$	Jet transverse momentum Jet pseudo-rapidity

Jet Fitter

Exploit the topology of a cascade $B \rightarrow D$ decay.

- Finds tracks consistent with a travel axis from the IP to the *B/D* decay hypothesis
- Allows it to find single track decays
- Purity is much better for 2-track vertices



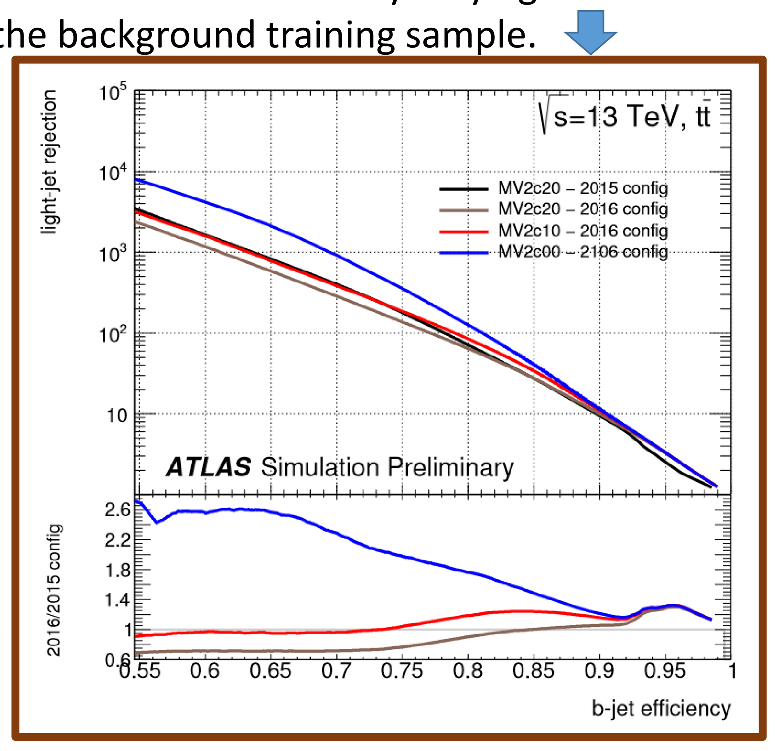
$N_{\mathrm{2TrkVtx}}(\mathrm{JF})$	Number of 2-track vertex candidates (prior to decay chain fit)
$m(\mathrm{JF})$	Invariant mass of tracks from displaced vertices assuming pion masses
$S_{\mathrm{xyz}}(\mathrm{JF})$	Significance of the average distance between the primary and displaced vertices
$f_E(\mathrm{JF})$	Fraction of the charged jet energy in the secondary vertices
$N_{\text{1-trk vertices}}(\mathrm{JF})$	Number of displaced vertices with one track
$N_{\geq\text{2-trk vertices}}(\mathrm{JF})$	Number of displaced vertices with more than one track
$N_{\mathrm{TrkAtVtx}}(\mathrm{JF})$	Number of tracks from displaced vertices with at least two tracks
$\Delta R(\vec{p}_{\mathrm{jet}},\vec{p}_{\mathrm{vtx}})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices

$\log(P_b/P_{\mathrm{light}})$	Likelihood ratio between the *b*- and light jet hypotheses
$\log(P_b/P_c)$	Likelihood ratio between the *b*- and *c*-jet hypotheses
$\log(P_c/P_{\mathrm{light}})$	Likelihood ratio between the *c*- and light jet hypotheses

The MV2 BDT

A boosted decision tree was used to combine all these input variables.

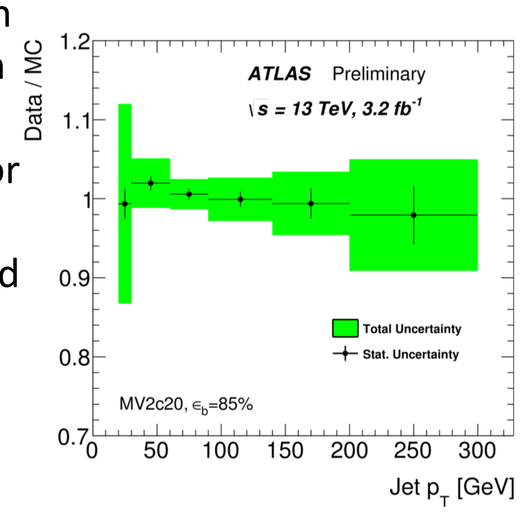
- The input variables are the same, other than specific improvements already noted.
- A great deal of time was spent understanding the training and performance.
- Trained on over 2 million events for background.
- Added a high mass $Z' \rightarrow b\bar{b}$ sample to improve high mass training
- Charm rejection can be altered by varying the amount of charm in the background training sample.

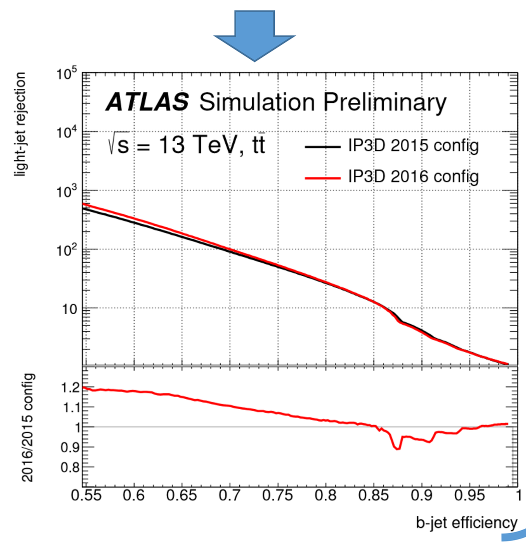


Calibration

Calibrating the *b*-tagging algorithms in Run 2 is well underway. Starting by tuning Run 1 methods for the Run 2 environment.

- Dilepton $t\bar{t}$ results have been released for 2015 data.
- Uses the combinatorial likelihood method which takes into account correlations between jets in a Dilepton event.





ATLAS-CONF-2014-004   FTAG-2016-002