b-tagging in the environment of boosted objects

Dominik Duda

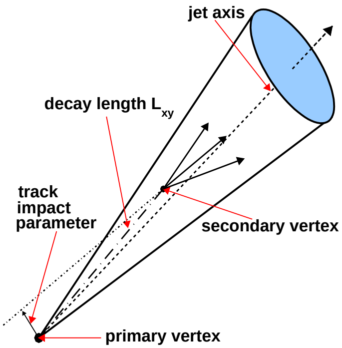
on behalf of the ATLAS collaboration

August 19th, 2014

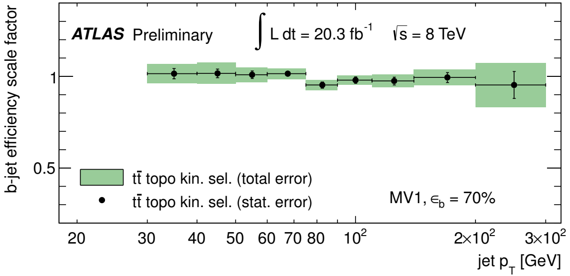
BERGISCHE UNIVERSITÄT WUPPERTAL

BMBF-Forschungsschwerpunkt ATLAS Experiment
Physics on the TeV-scale at the Large Hadron Collider
FSP 101 ATLAS

- ATLAS liftime based *b*-tagging algorithms are based on:
  - Track impact parameters (IP)
  - Secondary vertices (SV)
  - Multivariate techniques (combining both)
- ATLAS default tagger: MV1
  - Based on artificial neural network
  - Uses variables from one IP and two SV based algorithms



- Identification of isolated *b*-jets was intensively studied in the data recorded in Run I by the ATLAS
  - Performance is well understood for jets with a transverse momentum between 20 GeV and 300 GeV
  - *b*-tagging efficiency for jets with a p_T around 100 GeV can be measured with a total systematic uncertainty below 2%.

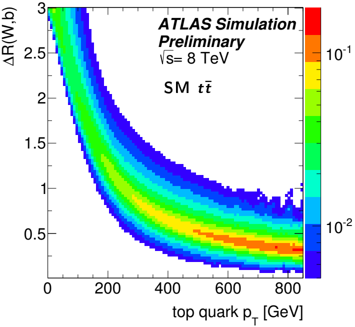


- High luminosity and increased energy $\sqrt{s}$ in LHC Run II will open up new regions of phase space at high p_T
- Will require excellent *b*-tagging performance at higher p_T and in boosted topologies

- Decay products of boosted particles tend to be collimated
- Angular separation of decay products is approximated by

$$\Delta R \approx \frac{2m}{p_T} \qquad (1)$$

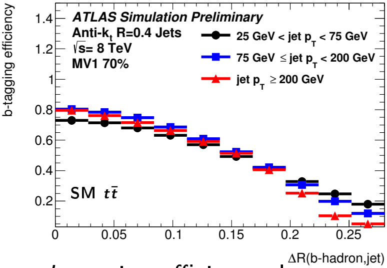
- For $p_T^{\mathrm{top}} > 450\,GeV$ and $p_T^{\mathrm{Higgs}} > 300\,GeV$ decay products tend to have a separation smaller than 0.8 (twice the radius of jets typically used for *b*-tagging in ATLAS)
- Standard reconstruction techniques start failing to resolve decay products individually

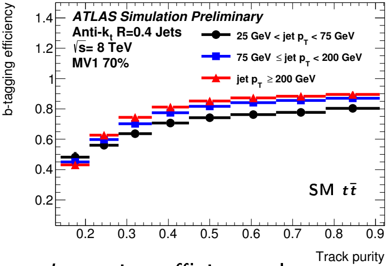


- Ongoing efforts (*b*-tagging in boosted environments) in ATLAS:
  - Flavour tagging with track jets: `https://cds.cern.ch/record/1750681`
  - *b*-tagging in boosted topologies: `https://cds.cern.ch/record/1750682`

Problems related to b-tagging in dense environments

- Degradation of jet direction resolution (Angular separation (b-hadron, jet axis))
  - Several b-tracks do not get matched to the jet
  - Algorithm performance degrades
- Light-flavour contamination
  - Change of jet properties
  - Impacts b-taggers trained upon pre-defined reference distributions
- Use track purity, defined as $N_b^{\text{tracks}} / N_{\text{in Jet}}^{\text{tracks}}$, to quantify

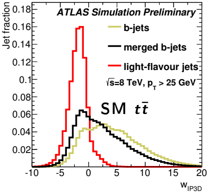


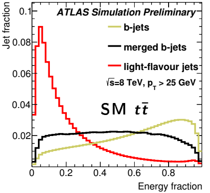


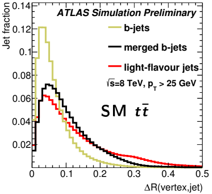
- b-tagging efficiency decreases by up to **a factor of 4.**
- b-tagging efficiency decreases by up to a **factor of 2**

Problems related to b-tagging in dense environments

- Study b-tagging related quantities for merged b-jets (in $t \to bW \to bq\bar{q}$ decays)
  - Merged b-jets: jet contains b-quark and light/c-quark from W-decay
- Several quantities used by MVA tagging tools lose separation power
  - Output of IP based tagger
  - p_T sum of tracks at SV/ p_T sum of all tracks in the jet
  - Angular separation between vertex direction and jet axis
- Quantities dissimilar to the training samples
- Approaches to improve performance
  - Develop dedicated b-tagging algorithms
  - Jets with smaller distance parameter R





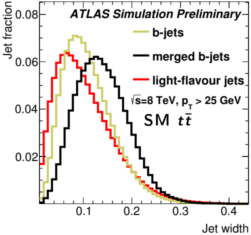


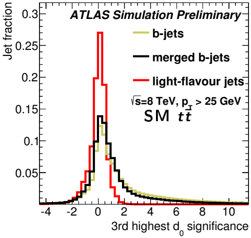
Adjusting b-tagging algorithms to boosted topologies

- b-tagging algorithm dedicated to boosted topologies
  - Introduce additional input quantities that are less overlap dependent
  - Substructure or jet shape related (e.g. jet width)

$$w_{\text{jet}} = \frac{\sum_{i=1}^{N} p_T^{\text{trk}_i} \Delta R(\text{trk}_i, \text{jet})}{\sum_{i=1}^{N} p_T^{\text{trk}_i}}$$

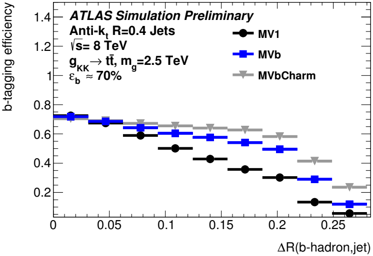
  - Track with the 3rd highest d_0 significance $\equiv d_0/\sigma_{d_0}$
- Emphasise jets from boosted $t\bar{t}$ decays more strongly in the training
  - Enrich statistics of merged jets
- Train b- vs. light-flavour jets (MVb) and b- vs. c-jets (MVbCharm)

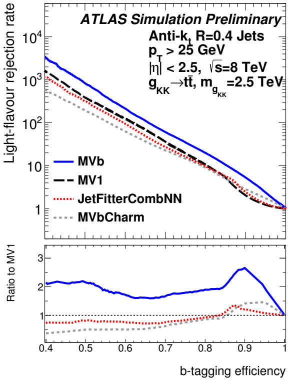




Performance of the new tagger

- Performance comparison to ATLAS *b*-tagging algorithms
- In sample with a high fraction of merged *b*-jets ($g_{KK} \to t\bar{t}$ with $m_{KK} = 2.5\,TeV$)
- *b*-tagging efficiency vs. $\Delta R(b\text{−hadron}, \text{jet})$ (below)
  - MVb tagger is less affected by jet overlap
  - Efficiency loss with increasing shift of jet axis is reduced
  - Improved by **a factor of 1.5**.



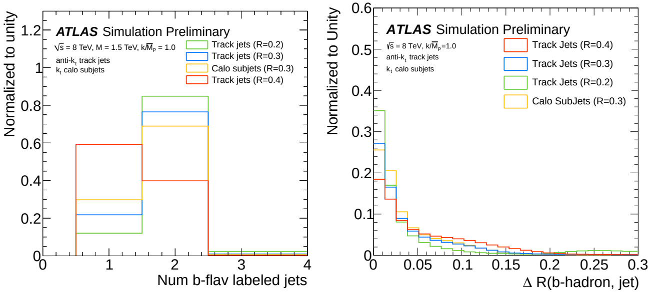


- Light-flavour rejection rate vs. *b*-tagging efficiency
- Performance in jet sample obtained from $g_{KK} \to t\bar{t}$ decays with $m = 2.5\,TeV$ is **improved by 50% to 160%**

Flavour tagging with track-jets in boosted topologies

- *b*-tagging performance is strongly degraded in dense environments due to a worse jet direction resolution
- Possible solution: Using track-jets instead of calorimeter-jets for *b*-tagging in boosted topologies
- Advantages:
  - Optimize track jets for the best *b*-tagging performance (e.g. smaller distance parameters $R = 0.4, 0.3, 0.2$), and calorimeter jets for the best interpretation of the hadronic final state
  - Very flexible - can easily associate track jets via ghost matching to any calorimeter based object (only one data/MC b-tagging calibration needed)
  - Better resolution of jet direction from using small R jets
  - Relatively pileup insensitive
    - Able to easily study jets corresponding to low p_T *b*-hadrons
- Current strategy (work flow):
  - Reconstruct decay of boosted massive particle into an all hadronic final state using 'large'-R jets
  - Run clustering on selected tracks to reconstruct track jets
  - Match track jet and large-R jet with ghost-association procedure
  - *b*-tag track jets (using MV1)

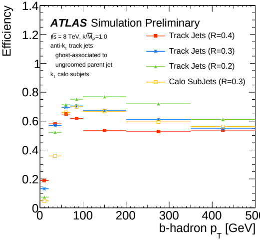
- Example: b-tagging performance in events in which a RS graviton decays via $G_{RS} \to hh \to b\bar{b}b\bar{b}$ with G_{RS} masses between $1.0\,TeV$ and $2.4\,TeV$.
  - Small-R jets more often resolve both decay products
    - $R = 0.4$ track jets: 40% of events have 2 b-jets
    - $R = 0.3$ ($R = 0.2$) track jets: 80% (85%) of events
  - Jets with smaller R have a much better resolution of the jet direction (this is important !!! see slide 5)

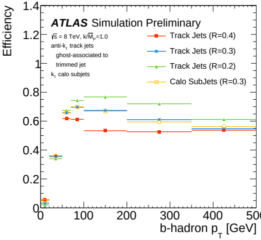


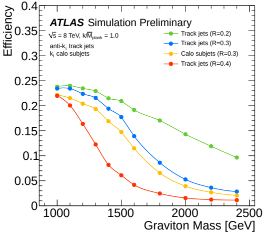
BERGISCHE UNIVERSITÄT WUPPERTAL

D. Duda, BOOST14

- *b*-identification efficiency (efficiency to find a jet around the *b*-hadron and *b*-tag it) for jets ghost-matched to the large-R jets (left) and to its subjets (middle)
  - $R = 0.4$ jets have a worse performance (too large to resolve components individually)
  - $R = 0.2$ track jets have the best performance
  - Small-R track jets can also more easily access low p_T *b*-hadrons
  - Leads to significant enhancement in performance when requiring 4 b-tags(right)







Summary and Conclusion

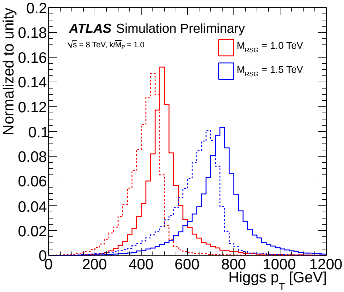
- Important to maximise b-tagging performance in boosted topologies (like in $h \to b\bar{b}$ or $t \to bW \to bq\bar{q}$ decays) in Run II
- Problems in these environments are related to the shift of the jet axis direction and contamination with tracks from other flavour decays
  - Losing tracks in the track-to-jet association
  - Degradation of the basic b-tagging algorithms
  - Change of jet properties wrt. training distributions
- Possible approaches to overcome these problems are:
  - Use track-jets independently optimised for b-tagging.
  - Usage of smaller distance parameters in the jet clustering
  - Developing a dedicated b-tagging algorithm
- Large improvements due to the usage of track jets and a dedicated b-tagging algorithm
- Ongoing work to calibrate both approaches on 2012 data
- Merge both studies into one complimentary approach

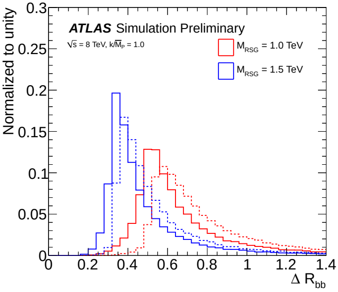
Backup

Track(jet) Selection

- Track selection
  - $p_T > 500 MeV$
  - $d_0 < 1.5 mm$
  - $z_0 \sin\theta < 1.5 mm$
  - Hits in the Inner Detector
    - At least one hit in the Pixel detector
    - At least six hits in the Pixel + SCT
- Trackjet selection
  - At least two tracksb
  - $p_T > 7 GeV$
  - $|\eta| < 2.5$

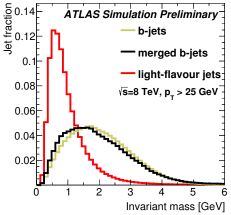
- p_T spectra for the leading and subleading Higgs bosons for two RS graviton masses
- Angular separation of the $b\bar{b}$ pair coming from the leading and subleading Higgs bosons
- For $m_{G_{RS}} = 1\,TeV$ the decay products are almost always in the regime where $R = 0.4$ jets resulting from the two b-hadrons would merge
- Smaller distance parameter in the jet clustering would be beneficial

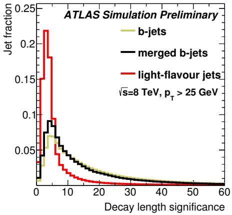




Problems related to b-tagging in dense environments

- Study b-tagging related quantities for merged b-jets (in $t \to bW \to bq\bar{q}$ decays)
  - Merged b-jets: jet contains b-quark and light/c-quark from W-decay
- Several quantities lose separation power wrt light-flavour jets (as shown previously)
- Most quantities are not that strongly affected (like inv. mass or decay length significance)
- In general: Properties of vertices are not that strongly affected as additional tracks are intrinsically rejected by the vertex fit.





- from simulated SM $t\bar{t}$ events

Performance of the new tagger

- Performance comparison to ATLAS tagging algorithms
  - Light-flavour rejection rate vs. b-tagging efficiency
- MVb shows very similar performance to ATLAS default tagger in the SM $t\bar{t}$ sample
- MVb shows much better performance in SM QCD for extremely high p_T jets