Performance Studies of b-Jet Identification in CMS

Leo Yu[1] (For the CMS Collaboration)
lyu@cern.ch

[1]Physics and Astronomy Department, Rutgers University

Jets associated with the production of bottom quarks in pp, pPb, and PbPb collisions are identified by a variety of algorithms developed by CMS. These algorithms exploit the long lifetime and high mass of bottom quarks by using the impact parameters of charged-particle tracks, the properties of reconstructed decay vertices, the presence of a lepton, or combinations of these quantities. In this poster, the performance of these algorithms including their efficiency and purity are discussed by the data measurements and their comparison with expectations based on simulations.

Introduction

Lead-lead collisions at the LHC are expected to reach sufficient energy densities to form quark-gluon plasma (QGP) [1]. Hard-scattered partons in QGP are predicted to lose energy through elastic and inelastic interactions as they travel through the medium (a phenomenon known as jet quenching), and the energy loss is expected to depend on the flavor of the parton [2].

Bottom quarks are characterized by their relatively long lifetime, high mass, and hard fragmentation functions. Identification of b-jets typically exploits observables relating to properties such as charged tracks, secondary vertices, and identified leptons. CMS, with its high precision silicon pixel tracker and its lepton detection facilities, is well suited for "tagging" b-jets.

b-Tagging Algorithms

CMS has developed several b-tagging algorithms that perform similarly in data and simulation. While the discriminator values cannot be compared across tagging algorithms, all tagging algorithms share the property that a higher discriminator value shows a higher likelihood that the jet is a b-jet [3].

The impact parameter of a track with respect to the primary vertex of an event can help distinguish tracks that come from the decay of b hadrons from prompt tracks. The 3D impact parameter significance of a charged track is defined to be the 3D impact parameter divided by its uncertainty. The **Jet Probability** (JP) tagger examines the 3D impact parameter significance of each track in a jet to determine the likelihood that the jet is a b-jet.

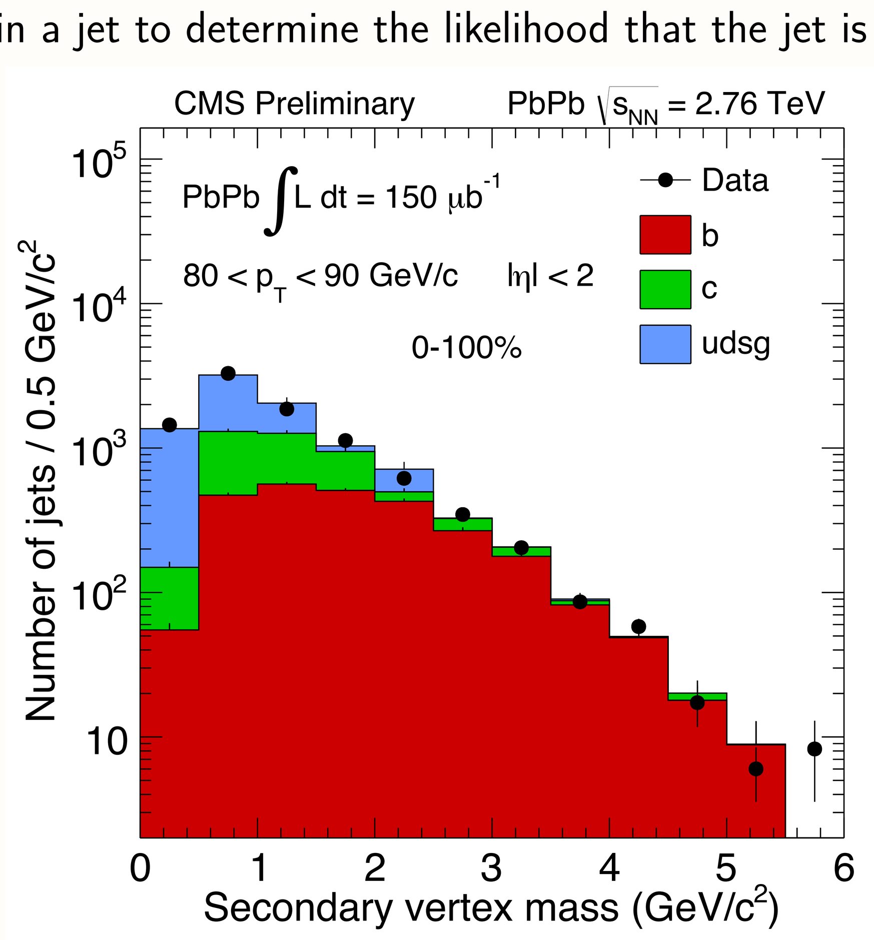

Figure 1: Secondary vertex mass of PbPb data and simulation at 2.76 TeV. The plot is for jets with $80 < p_t < 90$ GeV/c for events in the entire centrality range. The different colors represent the different flavor contributions to each bin in the simulation. The overall normalization of the three flavor categories in simulation is varied by a greatest likelihood fit to approximate the relative flavor contribution in data. [4].

The presence and properties of secondary vertices (SV) provide discriminating power between b-jets and non b-jets. Flight distance is defined to be the distance between the primary vertex and the secondary vertex. The **Simple Secondary Vertex** (SSV) tagger looks at the flight distance significance (flight distance divided by its uncertainty) to determine the likelihood that a jet is a b-jet. The High Efficiency (SSVHE) version of the tagger examines vertices with at least 2 associated tracks. Figure 1 is an example plot of the secondary vertex mass distributions over a specific jet p_t range after a tagger has been applied. The good matching shows that the tagger performs similarly in data and simulation.

SSV is chosen to be the primary tagger for these analyses due to its robustness against a combinatorial background, as it requires a secondary vertex to be present [5]. JP is used a a cross check for SSV.

Performance of b-Taggers

B-tagging **efficiency** is defined to be the fraction of b-jets that are identified by the tagger and is an important quantity for calculating the number of b-jets in a sample. **Misidentification rate** is the fraction of light/c-jets tagged. Figure 2 shows the performance curves of SSV on pp and pPb simulation. The plot shows that a >99% (90%) rejection of light-jets (c-jets) gives a b-tagging efficiency of approximately 50%. From such a plot, a **"working point"** (a maximum threshold on the misidentification rate of light jets) for a tagger is used to select/reject jets.

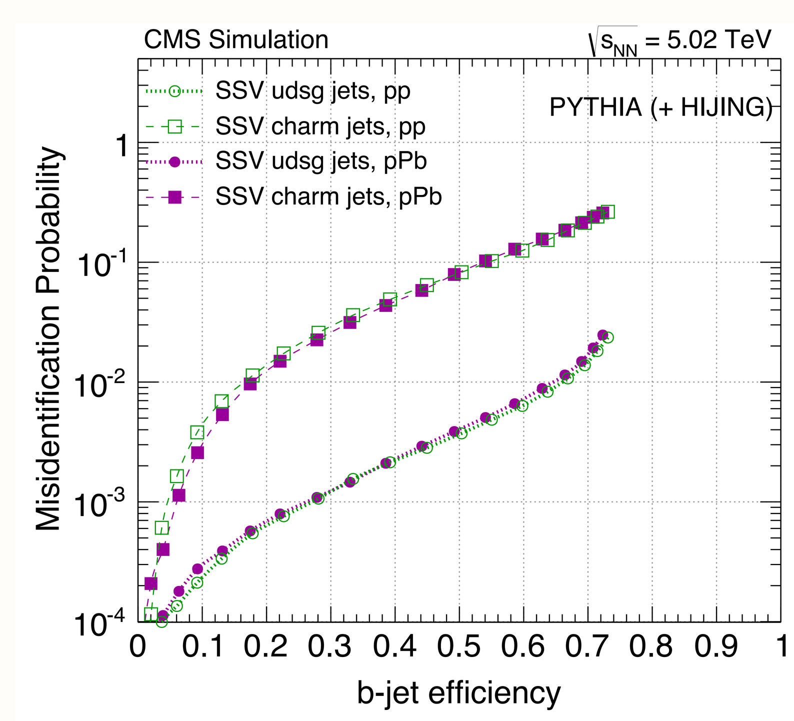

Figure 2: SSV b-tagging efficiency vs. misidentification rate of light/c-jets in pp and pPb simulation at 5 TeV. Performance is similar for pp and pPb due to identical reconstruction methodology [5].

The efficiency and purity of the tagged sample is calculated once a jet selection on the working point is made. B-tag **purity** is defined to be the number of b-jets divided by the total number of jets in the tagged sample; the b-jet fraction is obtained by fitting the simulation SV mass to the data, per jet p_t bin. The left panel in Figure 3 is an example of such a fit from which the relative flavor contributions in data is extracted. The purity as a function of p_t is shown in the right panel of Fig. 3 and is used to calculate the total number of b-jets in a sample.

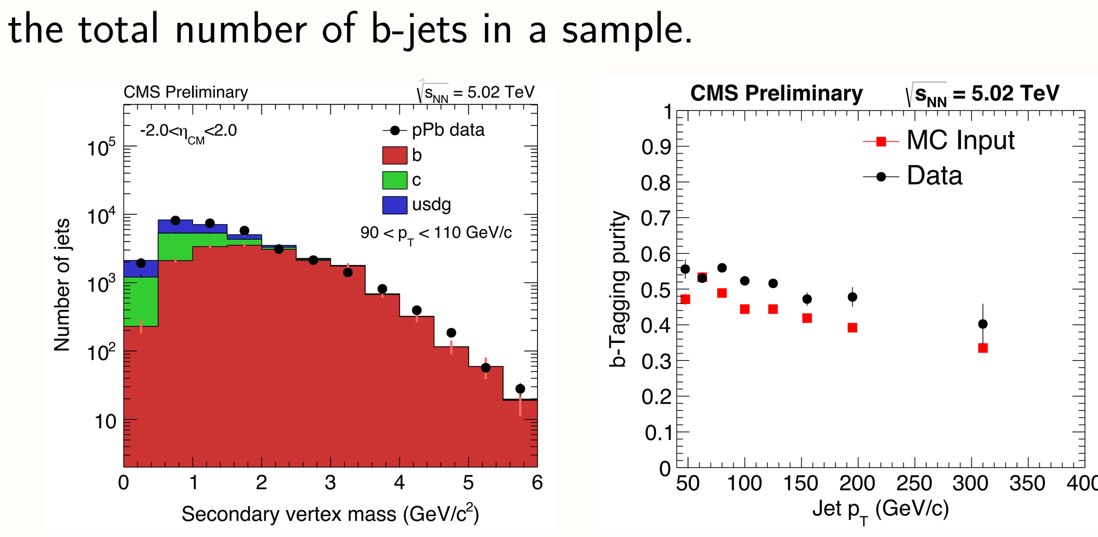

Figure 3: **Left:** template fit of simulation to data for SV mass, for jets with $90 < p_t < 110$ GeV/c in pPb. The different colors are the different flavor contributions in simulation. The overall normalization of the three flavor categories in simulation is varied by a greatest likelihood fit. **Right:** b-tag purity as a function of jet p_t for pPb [5].

Muon-triggered Jets

Muons are much more likely to come from b-jets than from lighter flavored jets due to the bottom quark's large semi-leptonic branching ratio; thus a sample requiring the presence of a jet with a muon in every event will be enriched in b-jets. Additionally, the relative p_t of the muon with respect to its jet's axis (p_t^{Rel}) tends to be higher in b-jets compared to lighter-flavored jets due the bottom quark's high mass, which makes it a good discriminator for b-jets.

Figure 4 is an example fit of Pythia simulation to pp data showing very good performance of a b-tagger on muon-triggered events. The fraction of tagged light/c-jets is very small here, which shows that high purity samples can be created from muon-triggered events [6].

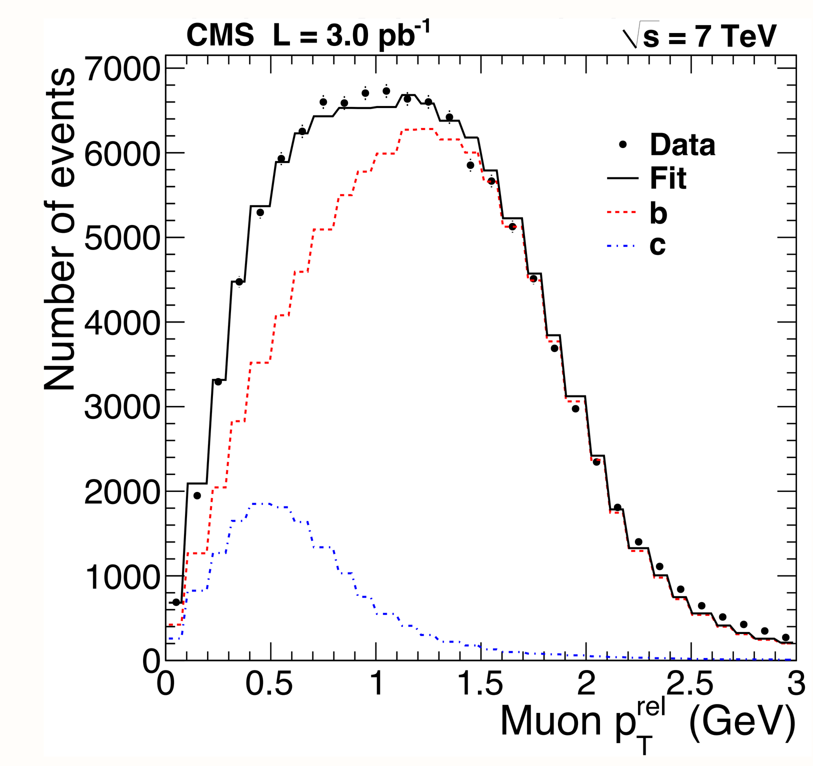

Figure 4: Template fits of the muon p_{tRel} simulation spectra to pp data (7 TeV) after a tagger (SSV) selection has been applied. The different colors show the contributions of different flavored jets in the simulation [6].

Results and Outlook

The suppression of b-jets in PbPb collisions is shown in the R_{AA} plot on the left panel of Figure 5. However, a more detailed understanding of the parton mass and flavor dependence of jet quenching will require a reduction in the systematic uncertainties of the analysis methods, including those related to b-tagging. There may be slight enhancement of b-jets in the right panel of Fig. 5, which would be consistent with cold nuclear matter predictions [7], though more details about parton energy-loss models will require higher precision measurements.

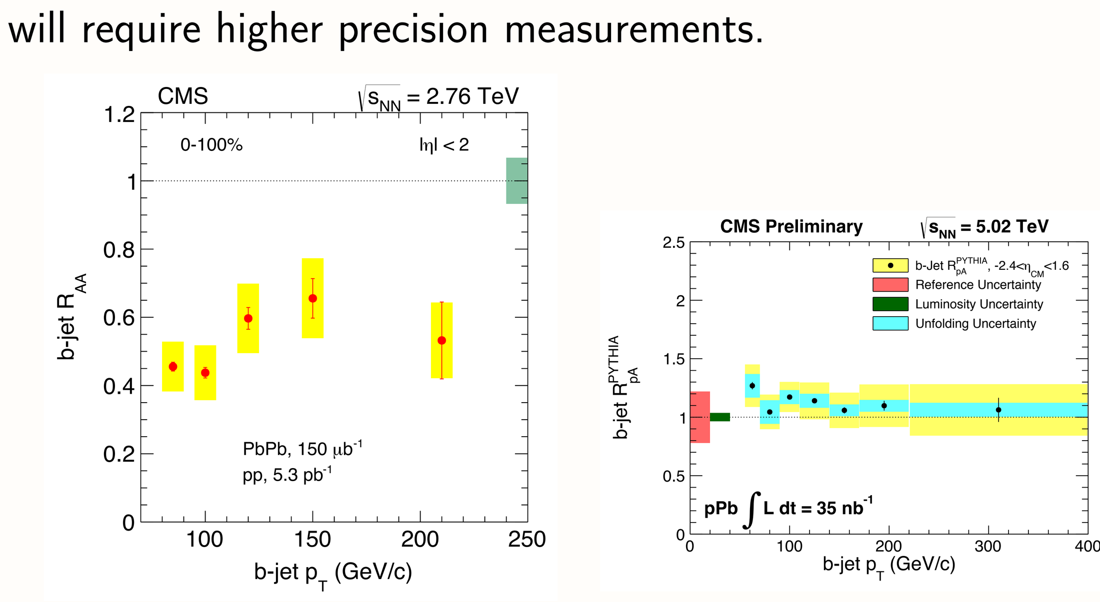

Figure 5: b-jet nuclear modification factor for PbPb and pPb collisions as a function of jet p_t. **Left:** R_{AA} between PbPb data and pp data at 2.76 TeV [4]. **Right:** R_{pA} between pPb data and pp simulation at 5.02 TeV (no pp collision data exists at 5.02 TeV) [5].

Future use of muon-triggered data, which require a lower minimum jet p_t, will allow for more precise measurements of b-jets in the low jet p_t region.

References

[1] E. V. Shuryak, "Theory of hadron plasma," Sov. Phys. JETP **47**, 212 (1978).

[2] Y. L. Dokshitzer and D. E. Kharzeev, "Heavy quark colorimetry of QCD matter," Phys. Lett. B **519**, (2001) 199 [arXiv:hep-ph/0106202].

[3] S. Chatrchyan *et al.* [CMS Collaboration], "Identification of b-quark jets with the CMS experiment," JINST **8**, (2013) P04013 [arXiv:1211.4462 [hep-ex]].

[4] S. Chatrchyan *et al.* [CMS Collaboration], "Evidence of b-jet quenching in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," CMS-HIN-12-003, (2013) [arXiv:1312.4198 [nucl-ex]].

[5] The CMS Collaboration, "Nuclear Modification Factor R_{pA} of b jets in pPb collisions," CMS-HIN-14-007, (2014).

[6] The CMS Collaboration, "Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV," JHEP **04**, (2012) 084 [arXiv:1202.4617 [hep-ex]].

[7] C. Salgado *et al.*, "Proton-Nucleus Collisions at the LHC: Scientific Opportunities and Requirements," J. Phys. G **39**, (2012) 015010 [arxiv:1105.3919].

XXIV QUARK MATTER
DARMSTADT 2014