

FLAVORS IN THE SOUP:
AN OVERVIEW OF HEAVY-FLAVORED JET ENERGY LOSS AT CMS

A Dissertation
Submitted to the Faculty
of
Purdue University
by
Kurt E. Jung

In Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

May 2016
Purdue University
West Lafayette, Indiana

“In the beginning, the universe was created. This has made a lot of people very angry and has been widely regarded as a bad move.”

- Douglas Adams

ACKNOWLEDGMENTS

A typical graduate student relies on countless people to complete a dissertation, especially so when he or she is part of a large collaboration like CMS. I would like to gratefully acknowledge the hard work of the hundreds of people that were involved in optimal detector performance, along with the LHC operators that were relied upon to deliver excellent performance of the accelerator.

On a more personal note, I want to thank my advisor Wei Xie for his unyielding support during the many hundreds of impromptu discussions we had. His intuition and guidance helped me develop as a scientist and a person. In addition, my parents Tom and Maureen were always there to provide both emotional and financial assistance as well as the encouragement I needed to complete this dissertation. I also want to thank Clare McCarty for introducing me to physics and sending me on this journey. Finally, I would like to thank my wife Jennifer who has been incredibly supportive during this process, especially for those months when I was at CERN and she had to hold down the fort all by herself! I could not have completed this without any one of them and I owe all of them a huge debt of gratitude.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
ABSTRACT	xiii
1 INTRODUCTION	1
1.1 The Quark-Gluon Plasma	1
1.2 Cold Nuclear Matter Effects	4
1.3 Jet Physics at Hadron Colliders	4
1.4 Flavor-Tagged Jets	5
2 THEORETICAL BASIS	7
2.1 The Standard Model	7
2.2 Quantum Chromodynamics	10
2.3 The Factorization Theorem	12
2.4 Jets	15
2.4.1 Heavy-Flavored Jets	16
2.4.2 Jets in Smaller Systems	17
3 THE CMS DETECTOR	19
3.1 Overview	19
3.2 Inner Tracking System	20
3.3 Calorimetry	21
3.4 Triggering	26
3.5 Data Availability	31
4 CENTRALITY	35
4.1 Theoretical Interest	35
4.2 Experimental Methodology	36
5 EVENT SIMULATION	39
5.1 PYTHIA Event Generator	39
5.2 Heavy-Ion Generators	41
6 JETS	47
6.1 Track reconstruction	47
6.2 Jet Reconstruction	50
6.3 Underlying-Event Subtraction	53
6.4 Jet Corrections	54

	Page
6.5 Flavored Jet Identification	62
6.5.1 b-Jet Tagging	63
6.5.2 c-Jet Tagging	74
7 MODIFICATION OF B-TAGGED JETS	81
7.1 Results in PbPb (2.76 TeV)	81
7.2 Results in pPb (5.02 TeV)	94
8 MODIFICATION OF C-TAGGED JETS	111
8.1 Results in pPb (5.02 TeV) and pp (2.76 TeV)	111
8.2 Systematic Uncertainties	116
9 SUMMARY	129
REFERENCES	131
A COORDINATE SYSTEM	137
B CONSISTENCY OF HEAVY-FLAVOR TAGGER INPUT BETWEEN SIM- ULATION AND DATA	139
B.1 Secondary Vertex	139
B.2 Single-Track Displacement	141
C ATTEMPTS TO INCLUDE NLO GENERATOR COMPARISONS . . .	147
VITA	151

LIST OF TABLES

Table	Page
2.1 Relative strengths of the coupling constants of the four fundamental forces at the scale of the Z boson ($q^2 = 91 \text{ GeV}$)	9
3.1 Prescale factors for jet triggers used during the 2013 run period	29
6.1 Iterative track reconstruction steps for pPb and pp reconstruction . . .	48
6.2 Modified iterative track reconstruction steps for PbPb regional tracking	65
7.1 Major contributors to the PbPb b-jet spectrum systematic uncertainty at 2.76 TeV for three jet p_T bins of inclusive centrality.	94
7.2 Major contributors to the pp b-jet spectrum systematic uncertainty for three jet p_T bins at 2.76 TeV.	95
7.3 Major contributors to the pPb b-jet spectrum systematic uncertainty at 5.02 TeV for three jet p_T bins.	103
8.1 Relative abundances and MC tagging efficiency of each c-hadron species	122
8.2 Decay multiplicities of charmed hadrons from data and PYTHIA	123

LIST OF FIGURES

Figure	Page
1.1 A subset of the R_{AA} results from the CMS collaboration.	3
2.1 The current pieces of the Standard Model (c/o Wikimedia Commons) .	8
2.2 Two examples of Feynman diagrams possible in QCD that contribute to effects like asymptotic freedom.	12
2.3 PDF distributions from the NNPDF collaboration at two energy scales: 10 GeV^2 (left) and 10^4 GeV^2 (right) [26]	14
2.4 Production mechanisms of b-jets at the lowest two production orders. Shown are the flavor creation (left), flavor excitation (center) and the gluon splitting (right) processes.	16
3.1 Cartoon of CMS outlining the various subsystems of the detector (CERN, 2010.)	19
3.2 Schematic of the Inner Tracking System layout [45]	20
3.3 CMS track impact parameter resolution performance in real data ($\sqrt{s} = 900 \text{ GeV}$) for the plane transverse to the beam (left) and longitudinally, along the beam (right) [46].	22
3.4 CMS track impact parameter resolution performance for $t\bar{t}$ simulations for the plane transverse to the beam (left) and longitudinally, along the beam (right), as a function of charged particle p_T . The solid (open) symbols correspond to the half-width at 68% (90%) confidence levels [47].	22
3.5 Quadrant view of the detail of the positions of all CMS subsystems [49].	25
3.6 Example trigger efficiency as a function of fully-reconstructed jet p_T from pPb data. Shown are the jet triggers with thresholds of 40 GeV (left) and 100 GeV (right) [12].	28
3.7 Closure plot of the single jet spectra against the spectrum obtained via trigger combination	32
3.8 Schematic of hierarchal network setup of the LHC Computing Grid including ATLAS and ALICE computing centers (Worldwide LHC Computing Grid, 2010).	33
4.1 Cartoon showing an example of the collision impact parameter (b) . . .	36

Figure	Page
4.2 Correlation of pixel multiplicity with forward and backward HCAL energy in PbPb collisions, showing a clear, tight correlation.	37
6.1 Simulated jet clustering behavior using the sequential recombination algorithm with $p = 1$ (left), and $p = -1$ (right).	52
6.2 Four-step procedure for underlying-event subtraction for jets at CMS.	53
6.3 Correction factors from simulation for anti- k_T jets for various η intervals as a function of jet p_T	55
6.4 Jet resolution (top panel) and response (bottom panel) for jets reconstructed with the anti- k_T algorithm using size parameters between 0.3 and 0.5 units in η/ϕ space.	55
6.5 The Relative jet response is shown for pPb (asymmetric) data, both before and after applying the C_{asym} correction factor as compared to simulation (left), along with the ratio of data to MC both before and after the correction factor is applied (right).	57
6.6 The γ +jet residual response function for MC and pPb data (left) and the ratio of data/MC (right) is shown.	59
6.7 Unfolding matrix created from simulated pPb collisions (left) and pp collisions (right)	61
6.8 The jet p_T spectrum both before and after unfolding using various methods is shown (left), along with a ratio of various unfolded spectra to the simulated “truth” values (right)	62
6.9 Distributions of the SSV discriminator. The black filled points are from data while the filled histograms denote contributions from various jet flavors using a PYTHIA simulation.	64
6.10 Distributions of the SSV tagging performance in pPb collisions. Shown is the tagging purity (left) and efficiency (right) from data and simulation.	65
6.11 Distributions of the SSV tagging performance in PbPb collisions. Shown is the tagging purity (left) and efficiency (right) from data and simulation.	66
6.12 Distributions of the JP discriminator values for pPb collisions (left) and for PbPb collisions (right) shown for different jet flavors. The black filled points are from data while the filled histograms are from a PYTHIA simulation.	69

Figure	Page
6.13 Distributions of the b-jet efficiency vs the light jet rejection for both the JP and SSV taggers. Shown is the performance in a pp simulation (left) and a PbPb simulation (right).	70
6.14 Distributions of the b-jet efficiency vs the charm jet rejection for both the JP and SSV taggers. The performance is shown for pp simulation (left) and PbPb simulation (right).	71
6.15 Template fit of the SV mass distributions for data (filled black points), and in MC from the three flavor components of light jets (filled blue histogram), charm jets (filled green histogram) and bottom jets (red) in pPb collisions (left) and PbPb collisions (right).	73
6.16 Template fits of the JP discriminator both before (left) and after (right) applying the SSV tagger. Contributions from b-tagged jets (red), charm jets (green), and light jets (blue) are shown against a distribution in pPb data (closed black points)	73
6.17 Distributions of the charm-jet tagging efficiency vs the b-tagging efficiency (left) and the light-jet tagging efficiency (right), comparing the standard SSV 2+ track algorithm and the high-purity 3+ track version.	76
6.18 Cartoon showing an example of the usefulness of the corrected secondary vertex mass calculation in a typical heavy-flavored decay.	76
6.19 The c-jet purity is shown as a function of the high-purity SSV selection for three selections of jet p_T . The working point is chosen at 1.68 such that the c-jet purity is maximized for a wide range in jet p_T	78
6.20 Template fits of secondary vertex mass (left) and corrected secondary vertex mass (right) are shown for pPb collisions	79
7.1 Cross-section for b-tagged jets in PbPb (scaled by T_{AA}) for various centrality classes, along with the cross section in pp at 2.76 TeV. Results are also compared to a PYTHIA prediction.	82
7.2 The b-Jet R_{AA} is plotted as a function of jet p_T , and is compared to a theoretical prediction based on pQCD with various gluon-to-medium coupling strengths.	83
7.3 The b-Jet R_{AA} is plotted as a function of centrality for two jet p_T bins where a clear trend can be seen, indicating that b-jets are more suppressed in more central events	84
7.4 The b-Jet and inclusive jet R_{AA} as a function of jet p_T are shown. The consistency between the two results is indicative that flavor-dependent effects are very small at high energies.	86

Figure	Page
7.5 The gaussian fits to the systematic variation of the b-jet purity in PbPb at 2.76 TeV derived from MC statistical uncertainty in jet p_T bins from 80-100 GeV/c (left) and 120-150 GeV/c (right)	88
7.6 The contribution to the total systematic uncertainty for PbPb (top) and pp (bottom) is shown as a function of jet p_T	90
7.7 The contribution to the jet reconstruction systematic uncertainties for PbPb (left) and pp (right) is shown as a function of jet p_T	92
7.8 The contribution to the total systematic uncertainty from jet reconstruction for the R_{AA} measurement after partial cancellation of pp and PbPb uncertainties	93
7.9 The R_{pA}^{PYTHIA} as a function of jet p_T for b-jets in pPb collisions at 5.02 TeV. A pQCD prediction from Huang et. al. [77] is also shown.	96
7.10 The inclusive-jet R_{pPb}^* and b-jet R_{pA}^{PYTHIA} as a function of jet p_T . We observe agreement between the two results.	98
7.11 The b-jet fraction in pPb collisions at 5.02 TeV is shown and is compared to a prediction from a simulation of PYTHIA jets embedded in a HIJING background. The b-jet fraction from 2.76 TeV pp data is also shown	99
7.12 The $b\bar{b}$ cross-section from multiple scattering orders is shown using various theoretical models as a function of collision energy. See Ref. [80] for more details. Note the horizontal axis scale is mislabeled by a factor of 10 so LHC energies correspond to 2.76×10^4 and 5×10^4 GeV on the figure.	100
7.13 The b-jet cross section for various pseudorapidity selections is shown (left), along with the PYTHIA-based nuclear modification factor R_{pA}^{PYTHIA} (right) for the same four pseudorapidity bins. Note the absence of a $1.5 < \eta_{CM} < 2.5$ bin is due to the fact that jets are only calibrated to $ \eta_{lab} < 2$ units.	102
7.14 The contribution to the b-tagging systematic uncertainties for pPb collisions at 5.02 TeV is shown as a function of jet p_T	104
7.15 The contribution to the jet reconstruction systematic uncertainties for pPb collisions at 5.02 TeV is shown as a function of jet p_T	105
7.16 The b-jet cross-section (left) and the b-jet fraction (right) from 7 TeV data is compared to predictions from the PYTHIA D6T tune and to another next-to-leading order generator known as MC@NLO.	106
7.17 The b-jet cross-section in 2.76 TeV data is compared to the PYTHIA Z2 tune prediction.	107

Figure	Page
7.18 The ratio of predicted pp b-jet cross-sections between the PYTHIA D6T and Z2 tunes is shown for both 2.76 and 7 TeV	108
8.1 The c-jet fraction for pPb (left) and pp (right) is shown and compared to the PYTHIA prediction at center-of-mass energies of 5.02 TeV (left) and 2.76 TeV (right).	112
8.2 The c-jet spectra for pPb collisions at 5.02 TeV (left, scaled by T_{pA}) and pp collisions at 2.76 TeV (right) are shown and compared to PYTHIA predictions at both center-of-mass energies. The lower panels of both plots show the Data/PYTHIA ratio.	114
8.3 Global heavy flavored energy loss picture	115
8.4 SSVHP working point variation	117
8.5 toyMC variation	118
8.6 The distribution of the number of generator-level charmed hadrons (left) and the reconstructed vertex multiplicity (right) in a jet cone of $R = 0.3$ for jets with $p_T > 80$ GeV/ c	120
8.7 The distribution of the number of generator-level charmed hadrons (left) and the reconstructed vertex multiplicity (right) in a jet cone of $R = 0.3$ for jets with $p_T > 80$ GeV/ c	121
8.8 Distribution of tagging efficiencies obtained by variation of the charged decay multiplicities	124
8.9 The total c-tagging systematic uncertainties are shown for pPb at 5.02 TeV (left) and pp at 2.76 TeV (right) using various colored histograms. The quadratic sum is shown as a solid black histogram.	125
8.10 The total jet reconstruction systematic uncertainties are shown for pPb at 5.02 TeV (left) and pp at 2.76 TeV (right) using various colored histograms. The quadratic sum is shown as a solid black histogram.	126
A.1 Schematic of the x, y, z coordinate system, along with visual representations of the θ and ϕ angles.	137
B.1 Distributions of number of reconstructed secondary vertices and number of associated tracks per secondary vertex in PbPb	140
B.2 Secondary vertex mass distributions in PbPb	140
B.3 Secondary vertex p_T distributions in PbPb	141
B.4 Simple secondary vertex discriminator distributions in PbPb	141

Figure	Page
B.5 Distributions of number of reconstructed secondary vertices and number of associated tracks per secondary vertex in pPb	142
B.6 Flight distance and flight distance significance distributions	142
B.7 Simple secondary vertex discriminator distributions	143
B.8 2-D IP distributions in PbPb	143
B.9 3-D IP distributions in PbPb	144
B.10 Distribution of the number of selected jet-associated tracks for the impact parameter-based taggers and their p_T distribution in PbPb	144
B.11 The distributions of distance from tracks to the axis of the associated jet and track decay length in PbPb	145
B.12 2-D IP distributions in pPb	145
B.13 3-D IP distributions in pPb	146
B.14 The distributions of distance from tracks to the axis of the associated jet and track decay length in pPb.	146
C.1 b-jet cross section comparison to NLO generators.	147
C.2 b-jet cross section of data, Pythia and POWHEG at 7 TeV	148
C.3 b-jet cross section of data, Pythia and POWHEG and 2.76 TeV.	149

ABSTRACT

Kurt E. Jung PhD, Purdue University, May 2016. Flavors in the Soup: An Overview of Heavy-Flavored Jet Energy Loss at CMS. Major Professor: Wei Xie.

The energy loss of jets in heavy-ion collisions is expected to depend on the flavor of the fragmenting parton. Thus, measurements of jet quenching as a function of flavor place powerful constraints on the thermodynamical and transport properties of the hot and dense medium. Measurements of the nuclear modification factors of the heavy flavor tagged jets from charm and bottom quarks in both PbPb and pPb collisions can quantify such energy loss effects. Specifically, pPb measurements provide crucial insights into the behavior of the cold nuclear matter effect, which is required to fully understand the hot and dense medium effects on jets in PbPb collisions. This dissertation presents the energy modification of b-jets in PbPb at $\sqrt{s_{NN}} = 2.76$ TeV and pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, along with the first ever measurements of charm jets in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and in pp collisions at $\sqrt{s} = 2.76$ TeV. Measurements of b-jet and c-jet spectra are compared to pp data at $\sqrt{s} = 2.76$ TeV and to PYTHIA predictions at both 2.76 and 5.02 TeV. We observe a centrality-dependent suppression for b-jets in PbPb and a result that is consistent with PYTHIA for both charm and bottom jets in pPb collisions.

1. INTRODUCTION

1.1 The Quark-Gluon Plasma

In the mid-2000s, the experiments at the Relativistic Heavy Ion Collider (RHIC) announced the discovery of a new state of matter known as the Quark-Gluon Plasma (QGP) [1–4], which was later confirmed at very high energies by experiments at the Large Hadron Collider (LHC) [5–7]. The extremely hot and dense QGP medium is a strongly-coupled collection of subatomic partons known as quarks and gluons that are liberated from their parent hadrons due to an effective reduction of the strong nuclear force through a process known as color screening. It is expected that the universe was composed primarily of this QGP a few milliseconds after its creation, when quarks and leptons had formed but the universe was still too hot to allow quarks to bind into larger particles, like protons and neutrons. Free quarks are never otherwise observed in nature due to the inherent properties of quantum chromodynamics (QCD), which govern the behavior of the strong nuclear force. At low energies, the strong force has a potential that increases linearly as a function of distance unlike, for example, the gravitational potential which decreases quadratically as a function of distance (see e.g. [8]). As two quarks separate, the potential between them increases such that at some point it becomes more energetically favorable to create a new quark pair from the quantum vacuum between the two original quarks, such that quarks are always in contact with other quarks. This inability to separate quark pairs means that quarks are never truly deconfined, at least outside of the QGP. In the QGP, however, the energy and sheer density of quarks and gluons is so large that it leads to an effect known as color screening, which decreases the effective potential of the collective strong force between all participating partons. Color screening (and anti-screening) are purely quantum mechanical effects, where the vacuum itself surrounding the color

charge is polarized, leading to an change in the effective force felt by surrounding particles. In fact, some theoretical models [9] rely solely on parton proximity estimates to define the onset of QGP creation. In any case, this color screening effect increases as a function of temperature (through a decrease of the colored Debye mass – see e.g. Ref. [10]) such that in heavy-ion collisions the color screening effect is strong enough to allow the quarks and gluons to travel throughout the medium, relatively unbound by the strong force.

This deconfinement leads to a number of interesting properties, and indeed the primary objective of many of the heavy-ion experiments at hadron colliders today is to quantify various properties of the QGP. One such property of interest is parton energy loss. By identifying the amount of energy quarks and gluons (collectively known as partons) lose as they traverse the medium, the effective coupling strength of the partons in the medium can be qualitatively observed. Essentially what is done is a calculation of the change of a particle’s momentum transverse to the particle beam direction (p_T) in heavy-ions relative to that same situation in proton-proton collisions. This observable is known as the nuclear modification factor (R_{AA}) and is defined in Eq. 1.1.

$$R_{AA} = \frac{1}{T_{AA}} \frac{dN^{PbPb}/dp_T}{d\sigma^{pp}/dp_T} \quad (1.1)$$

The scale factor T_{AA} in Eq. (1.1) is equal to the average number of nucleon-nucleon collisions divided by the total proton-proton inelastic cross-section. This factor accounts for the increase in particle production in heavy-ions relative to proton-proton collisions just due to multiple scattering and other simple geometric effects unrelated to the QGP. The scale factors have been tested by calculating the R_{AA} of electroweak probes (e.g. photons and the W and Z bosons), which do not interact via the strong nuclear force and are expected to be unmodified by the QGP medium.

Figure 1.1 shows a sample of the R_{AA} measurements from the CMS experiment using inclusive charged particles, B mesons, inclusive jets and b-jets.

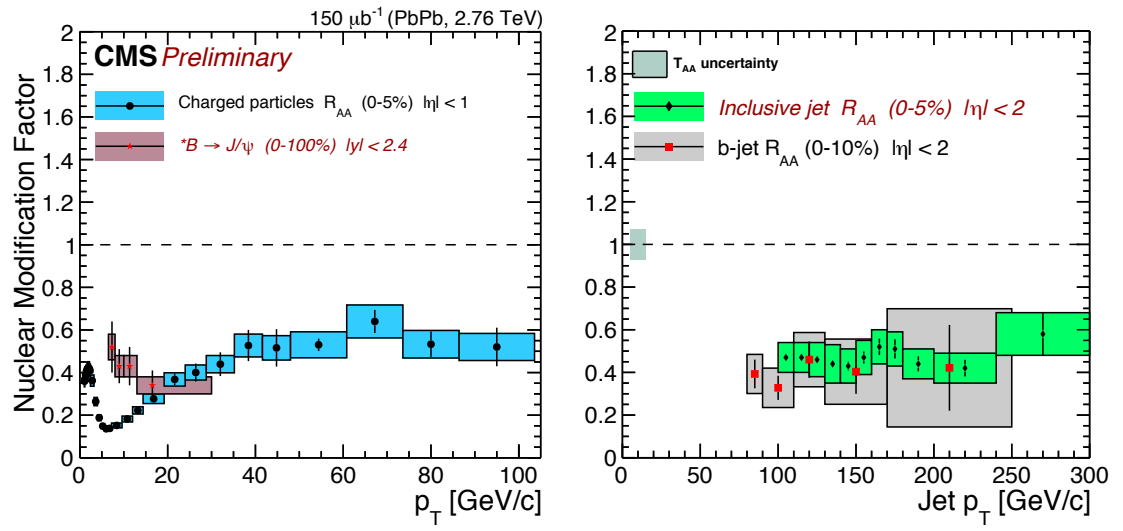


Figure 1.1. A subset of the R_{AA} results from the CMS collaboration.

1.2 Cold Nuclear Matter Effects

While R_{AA} is a useful tool to measure the energy loss seen in lead-lead (PbPb) collisions, it does not tell the whole story. The mere presence of a nucleus can influence an observed particle spectrum, resulting in behavior like that described in the so-called Cronin effect [11]. The Cronin effect describes a process where one or more extra scatterings off the remaining nuclear fragments in a nuclear collision tend to preferentially kick partons toward higher momentum, resulting in a harder p_T spectrum. To quantify this sort of behavior, studies have also been performed in proton-lead (pPb) collisions. As the density of a pPb collision is significantly lower than that of PbPb collisions, the expectation is that QGP is not created in this collision system, even at LHC energies. This has been verified by a number of R_{pA} measurements at CMS and ATLAS, e.g., Refs. [12,13], where R_{pA} is the proton-nucleus collision system analogue of R_{AA} . As QGP is not expected to exist in these collisions despite the presence of a heavy nucleus, these collision systems allow for tests of initial-state (pre-collision) nuclear effects, like the Cronin effect.

1.3 Jet Physics at Hadron Colliders

This dissertation is primarily concerned with measurements of energy loss of particle jets, which is a particularly interesting observable for studying QGP properties. These jets typically originate from a single hard scattering event between two subatomic particles, where a quark-antiquark pair is produced with considerable transverse momentum, that is, with momentum perpendicular to the original particle beam direction. As the quarks separate, the strong nuclear force potential energy increases, which stimulates the production of particles from the vacuum as governed by the color fields between the quarks. This process is known as “fragmentation”. What results from this process is a collimated collection of particles, which can be easily identified using a hadronic or electromagnetic calorimeter, which are commonly used tools in modern particle detectors.

While jets are somewhat trivial to identify in an experimental apparatus due to the lack of background at typical jet energies, the rigorous definition of “total jet energy” is actually somewhat ambiguous. Jets fragment both the transverse direction and the direction parallel to the beam, so the total area in which particles are associated to the jet is not well-defined, even jet-by-jet. In practice, a defined cone size is specified for each analysis, where larger cones are preferred to ensure the majority of the jet energy is captured. It is crucial to ensure that theoretical predictions use the same cone size, as a larger cone will, of course, capture more energy. If a cone size is too large, however, cones will also capture large contributions from soft, low-momentum particle production that is unrelated to the jet. This soft production can artificially enhance the energy scale of the jets used in an analysis, especially in PbPb events.

Most importantly, however, the nature of fragmentation means that for sufficiently large jet cone sizes, the majority of the initial quark energy can be captured, so jets provide near direct access to initial quark kinematics. In addition, the fact that jets are produced at high- p_T , in great quantities (at least at the LHC), and have virtually no background mean that they are excellent tools to study quark-medium interaction.

1.4 Flavor-Tagged Jets

Once jets are identified, additional discrimination can be applied to separate jets based on the flavor of their parent quark. Of particular interest are jets that stem from beauty (b) quarks and charm (c) quarks. These two quarks have masses much larger than the lighter up, down and strange quarks [14], and consequently are produced much earlier in the collision, though in much smaller number. Because the heavy quarks are produced so early, they experience the full evolution of the QGP and carry with them information about the QGP shortly after it was formed. By comparing results of inclusive jets, c-tagged and b-tagged jets, a mass-dependent picture of energy loss might be observed. In addition, flavor tagging can be done in multiple collision systems for a collision species-dependent result.

This dissertation will measure the nuclear modification factors of b-quark jets in the PbPb, pPb and pp collision systems, quantifying both the energy loss and initial-state effects of heavy-flavored jets in these systems. In addition, this dissertation presents the first ever measurement of c-quark jets in a heavy ion system and measures the initial state effects of these jets in the pPb and pp systems. Collectively, these measurements are the first of their kind at the LHC and provide a nearly complete picture of quark mass or quark flavor-dependent energy loss.

2. THEORETICAL BASIS

2.1 The Standard Model

The Standard Model of particle physics represents decades of study into the fundamental nature of the subatomic world and describes the interactions and behavior of three of the four known fundamental forces [15, 16]. The successful prediction of many previously “missing pieces” including the electroweak bosons (1983 by the UA1 and UA2 Collaborations [17, 18]), the top quark (1995, by the CDF and D0 Collaborations [19, 20]), and the Higgs boson (2013, CMS and ATLAS Collaborations [21, 22]) prove the model’s validity and predictive power. Though overwhelmingly successful at describing present-day high-energy physics results, the Model is necessarily incomplete as it remains silent on new discoveries like dark matter [23] and neutrino oscillation [24], among others.

At its heart, the Standard Model uses the property of gauge invariance to describe all fundamental interactions of particles. Each piece of the model is point-like and is either a particle with half-integer spin “fermion” or a particle with integer-spin “boson”. All six known quarks and six known leptons are fermions, which are further classified into three generations consisting of two quarks and two leptons each. While the leptons participate in both the electromagnetic and weak nuclear forces, the quarks additionally participate in the strong nuclear force through “colored” interactions. While this color is unrelated to the everyday definition, it allows for the explanation of color confinement, where individual observable particles are always observed as color-neutral. Explicitly, the three colors can be thought of as red, green, and blue, where all composite particles must be color neutral, or white. This color neutrality can manifest itself in one of two ways. The first color neutral combination is found in baryons, like protons and neutrons, where each of the three constituent

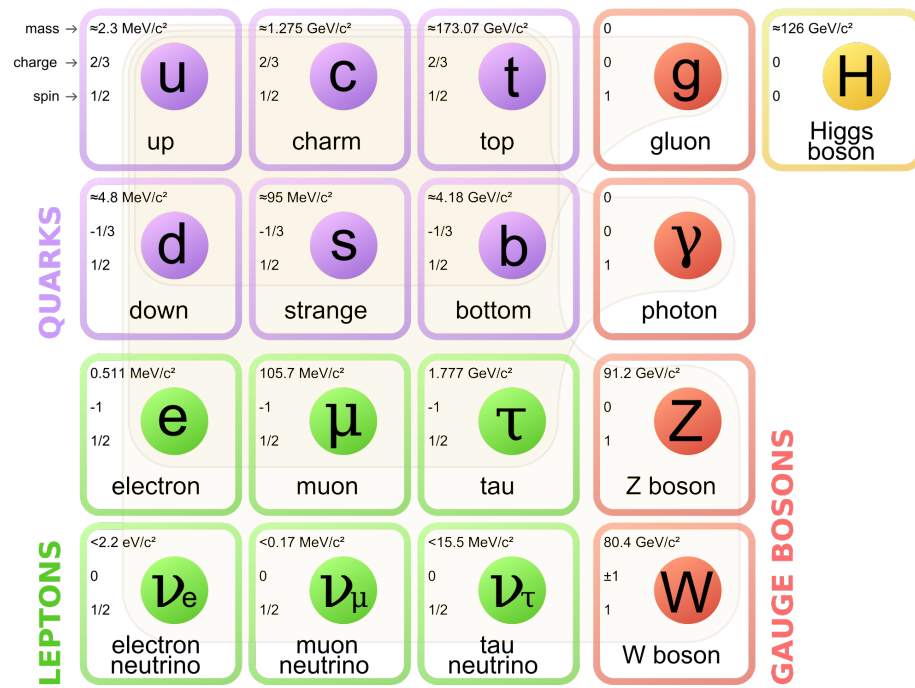


Figure 2.1. The current pieces of the Standard Model (c/o Wikimedia Commons)

Table 2.1.
Relative strengths of the coupling constants of the four fundamental forces
at the scale of the Z boson ($q^2 = 91 \text{ GeV}$)

Force	Relative Coupling
Strong	$\alpha_s \approx 0.1$
Electromagnetic	$\alpha \approx 1/127$
Weak	$\alpha_W \approx 10^{-5}$
Gravity	$\alpha_G \approx 10^{-40}$

quarks has a unique color (RGB = white), while the second neutral combination is found in mesons, like the pion or the J/ψ , where the two constituent quarks are color opposites of one another ($R\bar{R} = \text{white}$) .

Finally, the bosons are sometimes referred to as force carriers or mediators, since each one is responsible for propagating one of the fundamental forces between particles. Gluon exchange propagates the strong nuclear force, while the W and Z bosons mediate the weak nuclear force and photons mediate the electromagnetic force.

One final aspect of the model which requires an introduction is the idea of coupling constants. These dimensionless constants are different for each fundamental force and they change as a function of q^2 , or the momentum transfer in the interaction. Since they are dimensionless and universal, they allow for a way to compare the relative strength of the various forces. The values relative to the strong coupling constant are shown in Table 2.1. Since the electromagnetic and weak coupling constants are much less than 1, calculations of these interactions can rely on perturbative methods, where higher-order interactions contribute less and less to calculations of particle interaction probabilities. This is not the case, however, with strong-force interactions, which cannot be accurately calculated with perturbative methods (at least not without some clever tricks). Instead, full calculations must be performed either with strong assumptions regarding behavior, or on a lattice, which requires

large amounts of computational time. While these calculations can provide valuable insight, these difficulties underscore the importance of high-quality experimental studies of fundamental interactions, which can guide theoretical development.

2.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) describes the behavior of the strong force and the particles that couple to it. The Lagrangian of strong interactions is a member of an $SU(3)$ gauge group in color, contrary to Quantum Electrodynamics (QED) which has underpinnings based on an $SU(2) \times U(1)$ gauge group. The $SU(3)$ symmetry of the QCD Lagrangian naturally leads to the presence of three colors and eight color force carriers. These force carriers are gluons that form a color octet, with eight unique color-anticolor combinations. In addition to its symmetries, the QCD Lagrangian is also unique in that the gauge field couples to itself. In other words, the force carrier (gluon) is also colored and can couple to itself, unlike the QED force carrier (photon), which does not have electrical charge. Mathematically, this self-interaction allows for an ever-expanding number of Feynman diagrams which can be drawn with more and more gluon vertices, each with a similar contribution to a process's total probability. At fixed order in perturbation theory, however, the momentum integrals involved in summing the diagrams do not converge. This is known as an ultraviolet divergence and can only be solved through a regularization procedure, where a cutoff value is imposed on the momentum transfer. This regularization effectively limits the applicability of the theory unless one assumes that some yet undiscovered theory takes over at higher orders and the UV divergence is an artifact of improper extrapolation. With this assumption, some calculations can be performed. When renormalizing, it is important to account for the length (or energy) dependence of the α_s strong coupling constant. The value of α_s is given by:

$$\alpha_s(|q^2|) = \frac{12\pi}{(11n - 2f) \ln(|q^2|/\Lambda^2)} \quad (|q^2| \gg \Lambda^2) \quad (2.1)$$

where n is the number of colors, f is the number of quark flavors, q^2 denotes the momentum transfer in the interaction, and Λ (often denoted Λ_{QCD}) is an experimentally determined constant, which essentially defines the non-perturbative scale of QCD.

Experimental constraints on Λ place the value somewhere around 200 MeV [25], which (maybe not coincidentally) is qualitatively similar to measurements of the QGP temperature at RHIC, hinting that there may be some correspondence between the transfer into the perturbative regime and the onset of parton deconfinement through creation of the QGP.

The nature of the asymptotic freedom of the QGP (as discussed in Section 1.1) is a unique consequence of the QCD Lagrangian and can be investigated by understanding the evolution of α_s as a function of collision energy or effective length scale. Fundamentally, the variation in the coupling strength as a function of distance are due to the nature of the virtual particles that carry the force. Using electrodynamics (QED) as a comparison again, we find that in the vicinity of a charge, the quantum vacuum itself becomes polarized, such that the virtual antiparticles are attracted to the charge, while virtual particles are repelled. This is known as charge screening and reduces the effective coupling of the theory at any finite distance. In QCD, however, the ability of gluons to self-interact adds yet another complication to the picture. Rather than reduce the field strength, the additional interactions actually reduce the net effect of surrounding virtual gluons and increase the field strength at finite distances by diminishing the screening effect. A few examples of diagrams that contribute to this “antiscreening” are shown in Fig. 2.2. Therefore, the question of how the strong force behaves at short distances comes down to whether the screening or antiscreening components dominate the interactions of the field. This can be solved by again referring to Eq. (2.1). The term $(11n - 2f)$ in the denominator governs the direction of the evolution of α_s , where n is the number of colors (3) and f is the number of flavors (6). From this, it is straightforward to see that if the number of quark flavors is smaller than 16, the coupling is positive and decreases as a function of q^2 . Since only six quarks have been found, this indicates that asymptotic



Figure 2.2. Two examples of Feynman diagrams possible in QCD that contribute to effects like asymptotic freedom.

freedom exists for large energies (or very small length scales). The reduced coupling strength at large energies also means that at high-energies, QCD actually is relatively perturbative. Fortunately, these scales are accessible using the LHC, so perturbative methods can generally be used for predictions of the very highly energetic collisions that take place there.

2.3 The Factorization Theorem

Typical high energy collisions are very complex and consist of interactions over many particles at varying energy scales. For collisions at very high energies like in those that produce jets, the effective length scale becomes small, such that the collision dynamics can be factorized. A typical cross-section for a $2 \rightarrow 2$ scattering process of partons a and b to final state hadrons a' and b' consists of only a few components:

$$\sigma(p_a, p_b; Q) = \sum_{a,b} \int dx_a dx_b f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \otimes \sigma_{ab \rightarrow a'b'} \otimes D_a(z_{a'}, Q^2) D_b(z_{b'}, Q^2) \quad (2.2)$$

From left to right, these components include the parton distribution functions of the quarks and gluons within the proton $f(x, \mu_F^2)$, the two-body scattering cross-section $\sigma_{ab \rightarrow a'b'}$, and the fragmentation functions of the quark that define its transition into final-state hadrons that make up the identified jet $D(z, Q^2)$, where z is the momentum fraction of the initial quark carried by the final state hadron.

The parton distribution functions (PDFs) not only include the three primary valence quarks, but also include contributions from the sea quarks or virtual gluons that make up the total proton or neutron. These sea quarks are essentially manifestations of the total nucleon binding energy that become apparent when probed with another high-energy particle. As seen in Figure 2.3, over all particles that carry a low momentum fraction of the nucleus (known as Björken- x , or just “ x ”), the gluon contribution dominates, while at very high momentum fractions, valence quark contributions tend to dominate. For typical jet production at the LHC, where the total exchanged momentum Q^2 is on the order of 100-1000 GeV², the x range sampled is from roughly 10^{-3} to 10^{-1} . The PDFs in this kinematic regime from the NNPDF2.3 global fit is shown in Figure 2.3 [26].

The fragmentation functions describe the probability of a parton a (b) to produce a hadron a' (b') and are essentially the final-state analogue to the initial-state PDFs. Here the parameter $D_a(z, Q^2)$ refers to the probability of forming a hadron from parton a with a fraction z of the parton momentum at some momentum scale Q . These functions can change dramatically from parton to parton and from hadron to hadron. Properties of these functions are of particular interest in the heavy flavor sector, where it is expected that the heavy flavored mesons should carry a large fraction of momentum from their parent partons, and, as such, the heavy flavor fragmentation functions should be significantly harder than their light flavor counterparts [27].

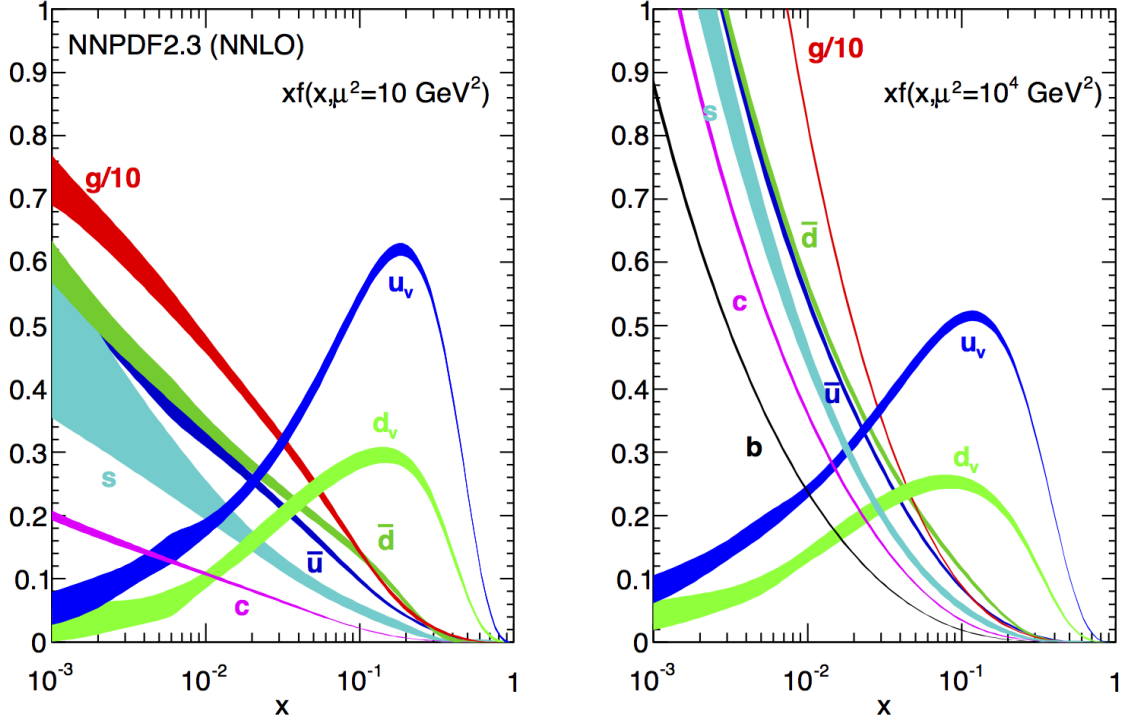


Figure 2.3. PDF distributions from the NNPDF collaboration at two energy scales: 10 GeV^2 (left) and 10^4 GeV^2 (right) [26]

2.4 Jets

As discussed in the introduction, jets are an excellent tool to study the dynamics and properties of high energy collisions. The length scales at which jets are produced mean that the factorization theorem applies, and when produced in conjunction with the QGP in heavy-ion collisions, mean that the jets must traverse the medium to be observed. Therefore, observations of jets can generally be assumed to directly connect to their parent quarks, such that any quark energy loss from the QGP will translate to a reduction of measured jet energy.

At a basic level, in the vacuum, jet production is a simple QCD hard-scattering process. As the jet evolves, however, the final-state hadroproduction is less clear. Jets radiate gluons and split into quark/antiquark pairs, forming a large particle shower during a process known as fragmentation. This evolution is governed by a set of equations known as the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [28] down to very low momentum particles (≈ 1 GeV). Below this threshold, the calculation of jet energy at leading order becomes infrared and collinear unsafe. In other words, the amplitude of the gluon production probability diverges as energy and the angle between the outgoing gluon and initial parton tend toward zero. It is important to ensure that jets, both theoretically and experimentally, are safe from these divergent effects, typically referred to (rather straightforwardly) as being “collinear and infrared safe”. In practice, particularly in heavy-ion experiments, very low momentum particles inside the jet cone are actually subtracted when calculating the total jet energy in an attempt to remove the effects of soft background particles that are not produced by jet fragmentation. This is discussed further in Chapter 6.

In the presence of the QGP medium, however, the situation changes. Through a process known as jet “quenching” it is expected that jets lose significant amounts of energy as they traverse the medium, primarily through induced gluon radiation [29] and collisional energy loss [30]. This feature of jets was one of the original mechanisms proposed to test for the presence of the QGP and has been extensively observed at

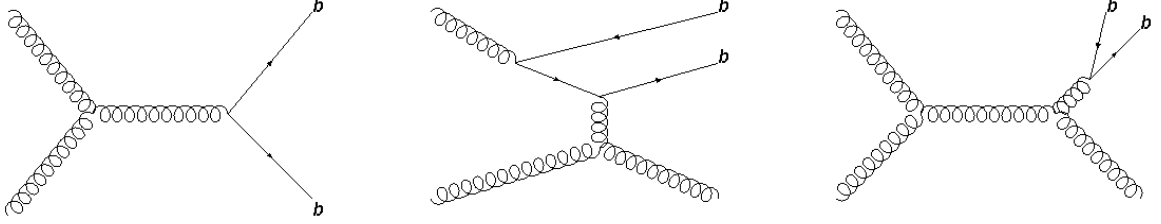


Figure 2.4. Production mechanisms of b-jets at the lowest two production orders. Shown are the flavor creation (left), flavor excitation (center) and the gluon splitting (right) processes.

both RHIC [31,32] and the LHC [5,33]. As the amount of energy loss depends on QGP parameters like temperature, thickness, and medium-to-parton coupling strength, much can be learned about the QGP just from these measurements.

2.4.1 Heavy-Flavored Jets

While the QGP colored interaction strength might be obtained using inclusive jets, further studies can be done by studying the energy loss of heavy-flavored jets. These heavy-flavored jets are seeded by a beauty or charm quark, which ought to impact the amount of jet quenching for a few different reasons. First, gluon-jet production (where a gluon is the highest momentum parton in the jet cone) in the heavy flavor sector is a much larger fraction of the total jet production than for inclusive-jets. These gluon jets are expected to quench more strongly than light flavored jets due to the larger color factor for gluon emission from a gluon than from a quark [34].

Furthermore, the heavy flavor production mechanisms are very different than that of light jets. While the classic Drell-Yan-like ($gg \rightarrow b\bar{b}$) production [35] dominates inclusive jet production, b-jets in particular have three primary production types, outlined in Fig. 2.4. These three diagrams show the flavor creation process (FCR, Fig. 2.4 [left]), which is the classic leading-order production mechanism, but what makes b-jets interesting is the fact that the next-to-leading order diagrams actually

dominate the production at high energies [36]. The flavor excitation (FEX, Fig. 2.4 [center]) and gluon splitting (GSP, Fig. 2.4 [right]) production mechanisms can dramatically influence the jet energy that is reconstructed in collider experiments, as the $b\bar{b}$ pair is tightly collimated and generally contained within the same jet cone. These merged cone jets typically force conclusions from inclusive b-jet measurements to have a number of caveats, but these measurements can be directly compared to a theoretical calculation using a perturbative expansion to (at least) next-to-leading order.

Due to these expected differences in jet energy loss as a function of jet flavor, most theoretical models predict different quenching behavior for light and heavy jets. While most models predict a reshuffling of the relative strengths of the radiative and collisional energy loss components, some models introduce new mechanisms entirely. One model by Ralf Rapp, et. al., describes an energy loss model where heavy quarks lose energy via continual association and disassociation into mesons within the medium [37,38]. Finally, models based on results from a correspondence between conformal field theories and other exotic particle models using anti-deSitter spaces (known as AdS/CFT correspondence) predict behavior that is very different between heavy and light flavored quarks (though future work is needed to fully understand the effects of this theoretical framework) [39]. No matter the model, however, there is a universal expectation that there should be some partonic mass dependence of jet energy loss, if only at low momentum, where the beauty or charm quark masses are no longer a negligible contribution to the total jet energy. Measurements of jet energy loss or jet energy modification in the heavy flavor sector will be the primary focus of this dissertation.

2.4.2 Jets in Smaller Systems

In addition to studies performed in PbPb collisions, the 2013 run period of the LHC featured a proton-lead (pPb) collision system. In such systems, initial state

effects can be effectively factorized from final state quenching effects such that a true measure of medium quenching strength can be obtained, unaffected by initial-state nuclear effects. This initial and final state factorization becomes especially important in light of surprising results from CMS (and others) that show the nuclear modification factor in proton-lead collisions “ R_{pA} ” of inclusive charged particles is quite large [40]. This indication of large initial-state effects suggests that final state quenching measurements are influenced by the mere presence of a nucleus in the collision system and should be taken into account when drawing conclusions regarding energy loss in heavy-ion collisions.

In addition to energy loss, the natural asymmetry in the proton-lead collision system can be used to an experimental advantage. Measurements of dijets in pPb have indicated that the dijet production as a function of the polar angle relative to the beamline (hereinafter known as pseudorapidity, see Appendix A) correlate strongly with the momentum fraction carried by the incident parton with respect to the parent nucleus. This nuclear analog of the parton distribution function, or nPDF, can be constrained in this way and has been studied extensively by CMS [41]. There are also further attempts to measure the nPDFs via flavor-tagged jets. For b-jets especially, the production mechanisms predominantly constrain the measured nPDFs to gluons, which can prove extremely useful to the understanding of these nPDFs. Typically, most modern nPDF calculations use a set of fits to all available experimental data, where results from different energies can be related to one another using assumed extrapolations of Q^2 via the DGLAP evolution equations, and the Björken- x dependence derived from experiment. As the relationships between all existing data are complicated and the calculation of these nPDFs remain a major source of uncertainty for analyses dependent on NLO effects, additional clean measurements in this area would certainly be welcome. This area has been explored theoretically [42], and only recently has data come available using flavor tagged jets in an asymmetric system [43]. These results will be expanded on in Section 7.

3. THE CMS DETECTOR

3.1 Overview

The Compact Muon Solenoid (CMS) detector is a member of the latest generation of particle detectors and includes full hermetic coverage over 2π units in azimuthal angle (ϕ) and ± 2.4 units in pseudorapidity (η), though additional forward calorimetry extends the coverage to ± 5 units in η . Note that additional technical details of the

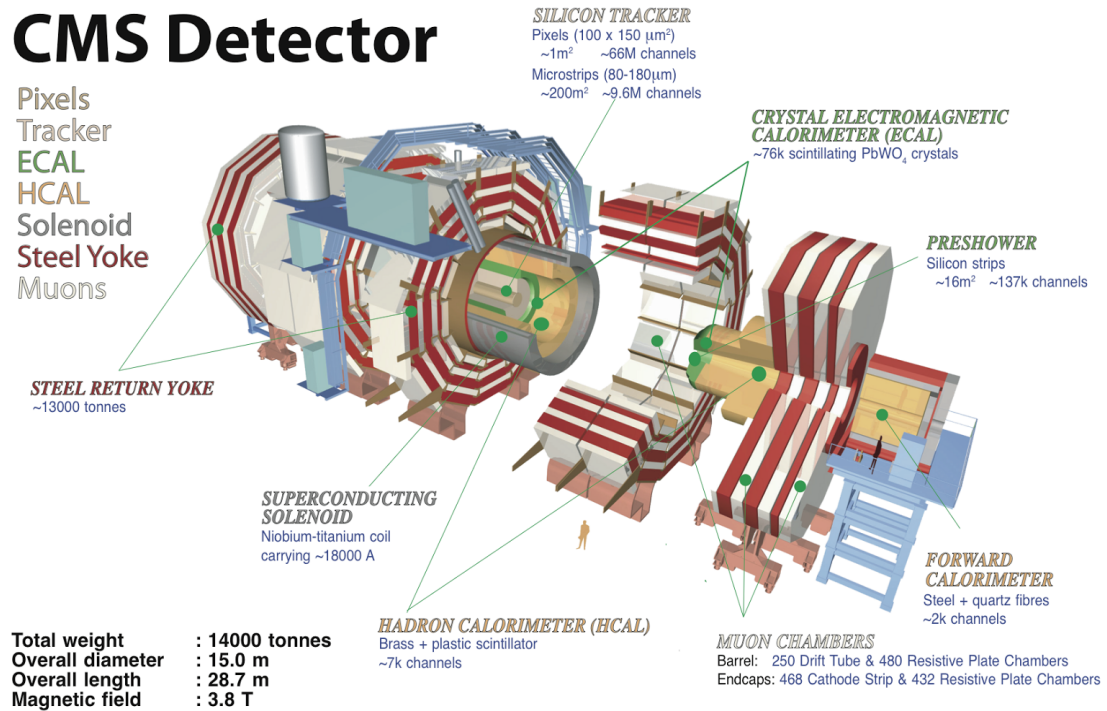


Figure 3.1. Cartoon of CMS outlining the various subsystems of the detector (CERN, 2010.)

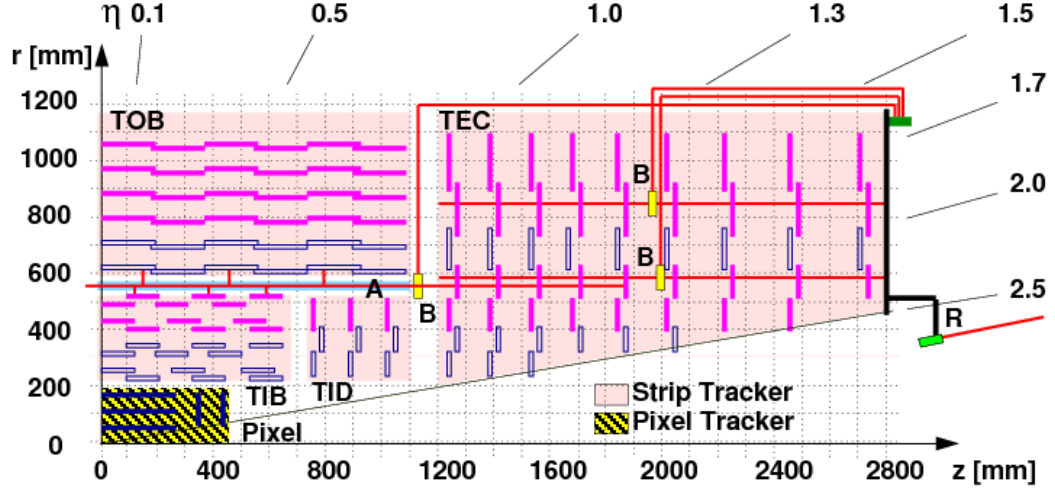


Figure 3.2. Schematic of the Inner Tracking System layout [45]

CMS detector not covered in this chapter can be found in the CMS Technical Design Report [44].

At its heart, the CMS detector has a superconducting solenoid magnet with an inner diameter of 5.9 m capable of generating a magnetic field of up to 3.8 T. The size of the magnet (and therefore the detector) is essentially governed by the desired momentum resolution power at high-momentum. To obtain the required performance for unambiguous muon identification from narrow resonance decays, the magnet size and bending power allows for a particle momentum resolution of $\Delta p/p \approx 10\%$ at $p = 1 \text{ TeV}/c$. Surrounding and within the magnet are a number of subsystems which work together to accurately measure a number of particle properties. This chapter will describe the CMS subsystems starting from those closest to the beamline and working outwards toward the outermost subsystems.

3.2 Inner Tracking System

The inner tracking system is a combination of two subsystems, namely, the pixel tracker and the silicon strip tracker. The layout of these detectors can be seen in Fig.

3.2. The pixel tracker (shaded in yellow) is comprised of three layers of hybrid pixel detector in the barrel region with radii of 4.4, 7.3, and 10.2 cm, with a pixel size of $100 \times 150 \mu\text{m}^2$. Complementary to the pixel tracker, the silicon strip tracker (shaded in pink) is much larger, with ten offset layers of silicon sensors in the barrel region placed between 20 and 110 cm from the interaction point. The tracker endcaps are comprised of two pixel and nine strip layers such that efficient tracking is achieved out to $|\eta| < 2.4$ units. All together, 66 million pixels and 9.6 million strips comprising an area of around 200 m^2 are used in the inner tracking system.

The inner tracker is essential to the measurement of flavor-tagged jets. In essence, these measurements rely on tagging jets based on their displacement from the initial parton-parton interaction point (known as the primary collision vertex). These displacement lengths for heavy quarks are on the order of $100 \mu\text{m}$, so it is imperative that the CMS tracking system is able to resolve individual vertices at this order of magnitude. One measure of track resolution is known as the track impact parameter. The impact parameter is simply the minimal distance between the track helix and the reconstructed primary vertex. The individual track impact parameters for all high-purity tracks is shown in Fig. 3.3. From this, we see that even for small values of transverse momentum (p_{T}), the transverse track resolution is roughly $500 \mu\text{m}$. The track resolution improves at high- p_{T} , as is shown in the CMS simulations shown in Fig. 3.4. For jets, therefore, it is expected that the resolution is within the kinematic constraints of b- and c-quark decay. Additional discussion of the performance of the CMS tracking system can be found in Refs. [46, 47].

3.3 Calorimetry

The extensive calorimetry installed as part of the CMS detector is essential to the study of jets. CMS has three primary calorimeter systems, namely the electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL) and the muonic calorimeter (muCAL). In general, calorimeters are designed in one of two ways. In one configu-

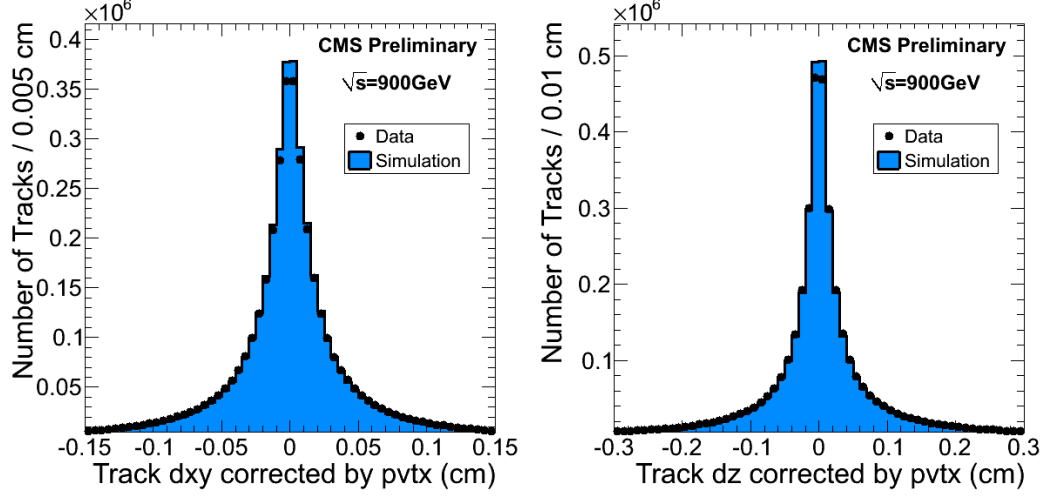


Figure 3.3. CMS track impact parameter resolution performance in real data ($\sqrt{s} = 900$ GeV) for the plane transverse to the beam (left) and longitudinally, along the beam (right) [46].

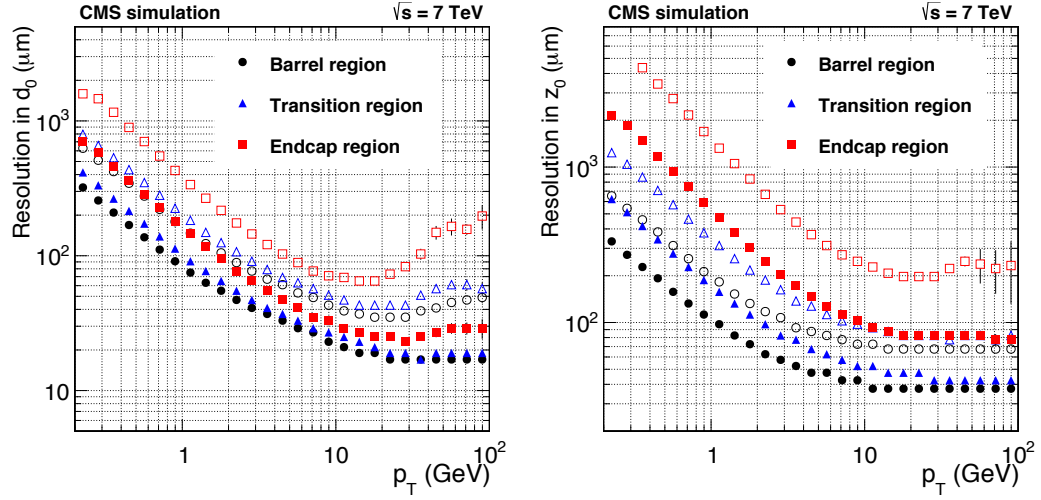


Figure 3.4. CMS track impact parameter resolution performance for $t\bar{t}$ simulations for the plane transverse to the beam (left) and longitudinally, along the beam (right), as a function of charged particle p_T . The solid (open) symbols correspond to the half-width at 68% (90%) confidence levels [47].

ration, alternating layers of active detector material and passive material, such that the particle energy is measured in each active detector layer and each passive layer induces some sort of energy loss. The second configuration is monolithic, where a single block of material is used both to induce particle energy loss and to measure the lost energy. While the second configuration is generally faster to readout, it is also usually more expensive. In either configuration, the goal is to capture all energy of the particle in the calorimeter in order to obtain an accurate measurement of the initial particle energy. The energy loss of the particle as it punches through the detector layers is governed by the Bethe-Bloch equation, shown in Eq. (3.1), where z is the projectile charge, β is its relativistic velocity, and Z is the atomic number and n is the electron number density of the target:

$$-\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi}{m_e c^2} \cdot \frac{n z^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \cdot \left[\ln \left(\frac{2m_e c^2 \beta^2}{(10 \text{ eV}) Z (1 - \beta^2)} \right) - \beta^2 \right] \quad (3.1)$$

By dividing the calorimetry into subcomponents, the active and passive material can be optimized for various particles. For example, muons have a much smaller interaction cross-section than do electrons, so much more material is needed to fully stop muons than is needed to fully stop electrons or photons. In addition, the granularity of each subcomponent can be optimized for cost based on the expected radiation shower profile that each type of particle creates as it interacts with the calorimetry.

The ECAL is a homogeneous and hermetic detector consisting of more than 68,000 lead tungstate (PbWO_4) crystals with an inner radius of 129 cm. One of the primary design considerations for calorimetry at the LHC is the speed at which the various subdetectors can be read out and cycled through in preparation for the next interaction. Lead tungstate was chosen because of its speed, as 80% of scintillator light is emitted within 25 ns, or 1 proton bunch crossing at nominal energy. In addition, the crystals have a short radiation length ($\chi_0 = 0.89 \text{ cm}$) and are radiation hard and therefore very well suited for good performance in a high-intensity collision environment. The primary drawback is the low yield, so photomultiplier tubes need to be

directly adjacent to the crystals, and therefore need to be able to operate in a high magnetic field. This is accomplished through the use of silicon avalanche photodiodes and vacuum phototriodes. The detector has a full 2π coverage in azimuth, while the barrel and endcap together provide a pseudorapidity coverage that extends to $|\eta| < 3.0$.

The HCAL is heavily influenced by the magnet system, as it is located primarily within the magnet coil, though there are sub-components located outside of the magnet. The advantage to segmenting the HCAL into two subcomponents inside and outside of the magnet is that the magnet is essentially used as an attenuator and extends the effective thickness of the detector to more than 10 radiation lengths. By catching the “tails” of the hadronic radiation showers in this way, the effective resolution of the calorimeter is significantly improved. The entire system consists of the hadron barrel, outer, endcap, and forward components, where the forward HCAL extends the pseudorapidity coverage to ± 5 units in η . For the barrel, outer and endcap calorimetry, the active medium is plastic scintillator tiles, interspaced with brass passive absorber material. The plastic scintillator tiles are readout via wavelength-shifting fibers and fiber optic cables such that the readout system can be placed outside of the center barrel. It is general procedure to benchmark the performance of a hadronic calorimeter using either jet energy resolution or via a missing transverse energy resolution study. Through simulation of the HCAL response, and confirmed via 7 TeV data, the overall resolution is $\sigma \approx 1.0\sqrt{\Sigma E_T}$ [48].

As seen in Fig. 3.5 at roughly $z = 3.5$ m and $R = 2$ m, the interspacing between the ECAL and HCAL is positioned such that no single non-instrumented section points directly to the interaction point. For reference, note that each dotted line at fixed pseudorapidity intervals point back to the interaction point. This allows for full hermetic coverage while still allowing access to the inner detectors’ critical systems, like data readout and cooling.

The muon calorimetry system is the largest subsystem of CMS and dominates the physical appearance of the detector. Due to the massive size and the varying radiation

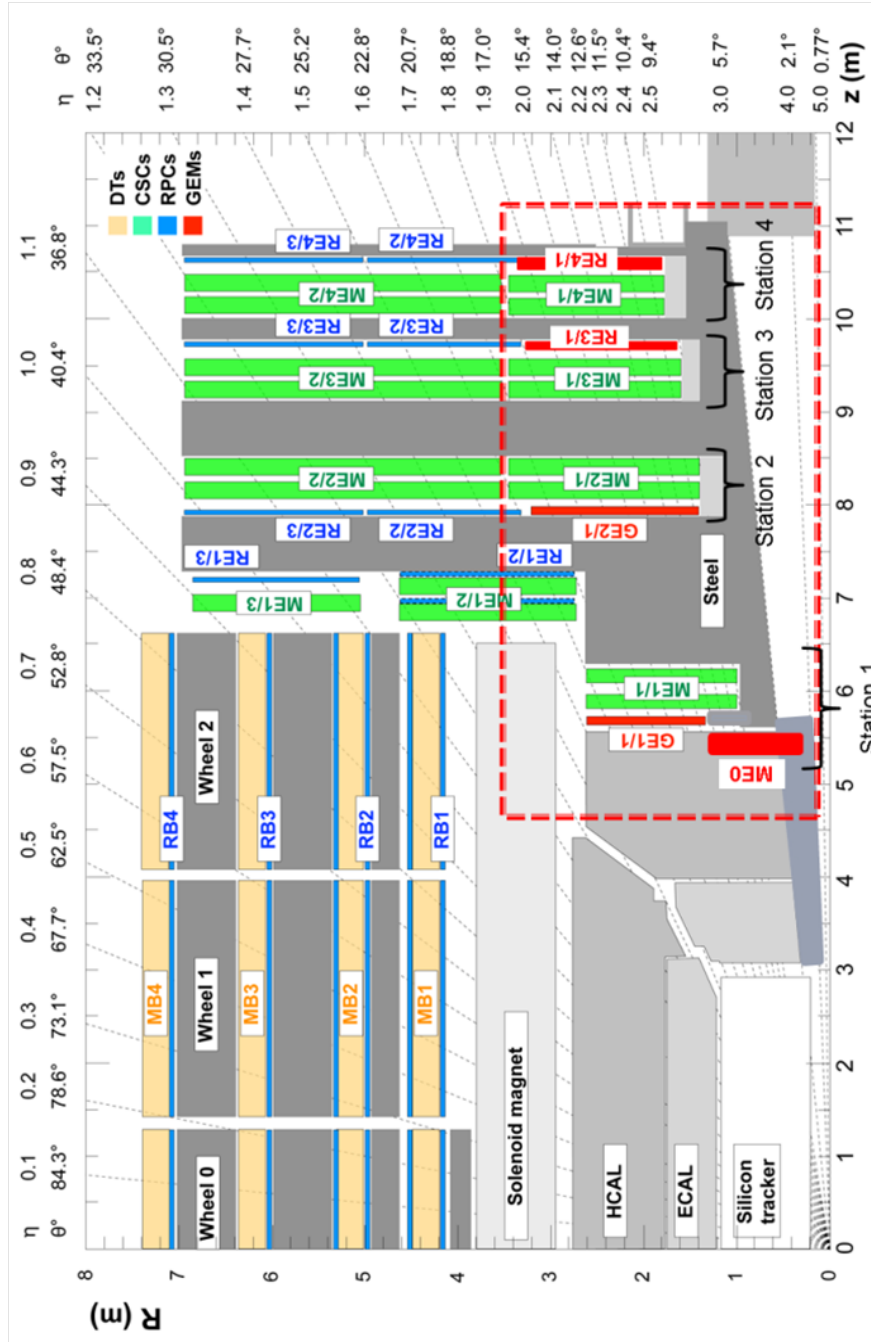


Figure 3.5. Quadrant view of the detail of the positions of all CMS subsystems [49].

requirements as a function of pseudorapidity, three types of gaseous detectors are used. In the barrel region, the muon rate and residual magnetic field are relatively low, so drift tube chambers are used. Conversely, the magnetic field and neutron flux are higher in the endcap region ($1.2 < |\eta| < 2.4$) so cathode strip chambers are used. As drift tubes and cathode strip chambers are relatively slow, additional instrumentation in the form of resistive plate chambers is added in both the barrel and endcap regions. Resistive plate chambers are essentially parallel-plate capacitors filled with gas such that a particle interaction with the gas triggers an electromagnetic avalanche, dropping the voltage across the plates. This avalanche leads to a poor spatial resolution but is fast enough to unambiguously identify the bunch crossing of muon production such that event triggering is possible even with a high bunch-crossing rate. While triggering will be discussed in detail in Section 3.4, the intent is to flag events that are interesting in some way in order to prioritize these events for write-out to a disk. As extensive muon calorimetry is a focus point of CMS, it is beneficial to flag all muonic events such that all events are kept on disk for analysis.

In conjunction with the tracking system, these calorimetry systems provide excellent jet resolution, on the order of 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV.

3.4 Triggering

Triggering is essential in a modern collider experiment, where the time between each bunch crossing is on the order of tens of nanoseconds. Since the maximum write speed to disk per event is on the order of 100 Hz, a reduction factor of approximately 10^6 is required purely from a computing power standpoint. Furthermore, it would be prohibitively expensive both from a computational and storage point of view to process every single proton bunch crossing event for each analysis, especially considering the great majority of bunch crossings are not interesting for analyses.

Trigger Setup at CMS

CMS uses a two-level trigger system, using both hardware and software-based algorithms. The first level (L1 triggers) is entirely hardware-based, while the second-level (HLT triggers) is software-based. L1 triggers are very simple and are typically implemented using simple threshold algorithms written to field-programmable gate arrays (FPGAs). These FPGAs allow for a fully customizable hardware circuit that can be dynamically altered via firmware. These provide all the advantages of customizeability, while still maintaining the effective speed of a single-use integrated circuit. The high- p_T jet triggers are a good example of a typical use of an L1 trigger, which triggers on a cluster of calorimeter towers above a certain energy threshold. This logic can easily be encoded in a programmable integrated circuit and satisfy the speed requirements at the lowest level.

The second trigger level is more complicated, but the reduced rate afforded by the L1 trigger decisions means that each algorithm has a longer time to process the data. L1 trigger developers target an output rate of roughly 100 kHz, which is directly sent to the HLT trigger algorithms. This reduced rate allows for the ability to perform tasks like jet reconstruction as part of the trigger algorithm. While full event reconstruction at 100 kHz is not possible, even with large processor farms, there are tricks to maximize event throughput. First, the CMS trigger system uses a buffer such that L1 trigger acceptance rate fluctuations can be compensated for. During run periods with more interesting events, the buffer will fill, and will gradually clear during periods with less interesting events. Clever algorithms will reject an event as soon as possible such that the buffer throughput is also maximized. Finally, advanced techniques like partial event reconstruction and multiple L1 seeding are used to ensure the timing requirements are fulfilled.

It is important to note that even the HLT triggers that perform partial or full event reconstruction are not fully efficient right at the trigger threshold. Especially in the heavy-ion environment, which has a large background from soft (low- p_T) particles,

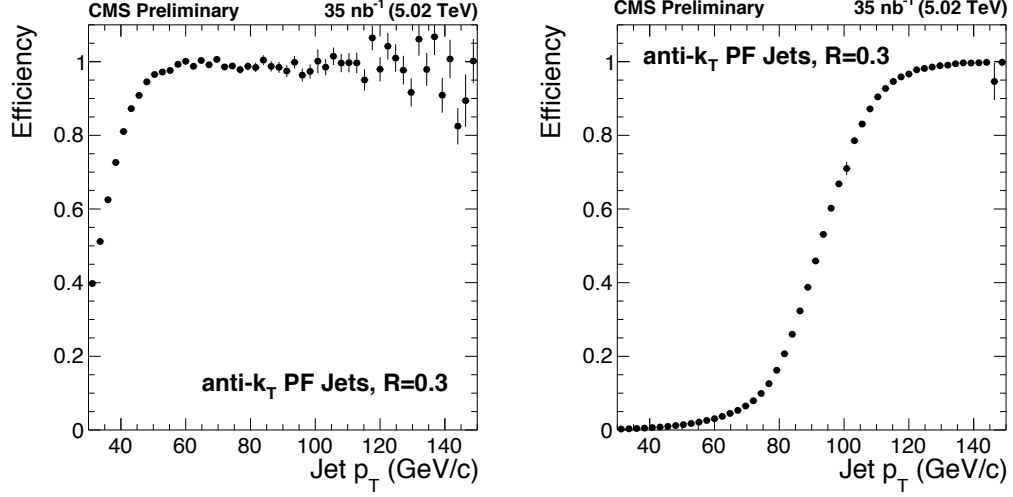


Figure 3.6. Example trigger efficiency as a function of fully-reconstructed jet p_T from pPb data. Shown are the jet triggers with thresholds of 40 GeV (left) and 100 GeV (right) [12].

additional calibrations and background subtractions (in the case of jets) are always performed offline such that there is a smearing effect between the online reconstructed jet p_T and the offline, fully-calibrated jet p_T . This smearing leads to a soft turn-on curve for the trigger efficiency as a function of offline jet p_T . An example of this is seen in Fig. 3.6, which shows the efficiency of two jet triggers with thresholds of 40 GeV (left) and 100 GeV (right).

Trigger Combination

Though every effort is made to filter events using the trigger system, there is often a need to introduce an additional event rejection factor when the throughput from the trigger system exceeds the maximum speed at which the system can write data to tape. This is typically implemented for triggers of physics processes that have a production rate higher than the disk write speed at CMS. For this, we introduce the concept of trigger prescales which artificially reduce the trigger acceptance rate by

Table 3.1.
Prescale factors for jet triggers used during the 2013 run period

Single-Jet Trigger Threshold	Average Prescale	Maximum Prescale
20 GeV	1795	8900
40 GeV	49	121
60 GeV	7	31
80 GeV	1.5	3
100 GeV	1	1

simply throwing away a fraction of the accepted events. The prescales are calibrated such that the triggers that accept events less frequently are prescaled less, while events that accept many events are prescaled more heavily. This ensures that a balance of events over a wide range of physics topics is written to disk. The prescale values themselves refer to the additional rejection factor applied to the events that pass the trigger condition. For example, a prescale value of five means that an additional factor of five events are not written to disk, even though the events pass the trigger condition. The prescale implementation at CMS is a simple running counter, which is in contrast to other experiments where a random selection is performed. In addition, the prescaled values are allowed to vary throughout the run, to adjust for variations in the instantaneous luminosity. Table 3.1 shows the maximum and average prescale factors for the five single-jet triggers used in the 2013 pp and pPb run period.

While this system is useful to ensure that disk space is well-utilized from interesting physics events, it presents a problem for an analysis that wishes to measure an observable over a wide range in p_T . In this case, the triggers must be combined together to take full advantage of the statistics on disk, but this combination is made fairly non-trivial due to the combination of the trigger efficiency curves and the trigger prescale factors. Based on the offline jet- p_T alone, it is impossible to know *a priori* whether or not an event should have passed a certain trigger. To circumvent

this problem, we have developed a novel trigger combination algorithm such that the statistical precision is maximized over a wide range in p_T . The sample is split into a number of exclusive categories that each correspond to a trigger used in the combination. These categories are split using the trigger object, that is, the online partially-calibrated jet that the trigger uses to make its decision. The maximum “trigger p_T ” from all trigger objects used removes the ambiguity when dealing with the smearing from post-reconstructed objects. For an explicit example, we can use the procedure developed for the b-jet analysis in pPb collisions. As five jet triggers with thresholds of 20, 40, 60, 80, and 100 GeV are used in the analysis, the five event categories are

1. $20 \text{ GeV} \leq \text{MAX}(\text{trigger } p_T) < 40 \text{ GeV}$
2. $40 \text{ GeV} \leq \text{MAX}(\text{trigger } p_T) < 60 \text{ GeV}$
3. $60 \text{ GeV} \leq \text{MAX}(\text{trigger } p_T) < 80 \text{ GeV}$
4. $80 \text{ GeV} \leq \text{MAX}(\text{trigger } p_T) < 100 \text{ GeV}$
5. $\text{MAX}(\text{trigger } p_T) \geq 100 \text{ GeV},$

where again, $\text{MAX}(\text{trigger } p_T)$ refers to the maximum p_T from an object that fired any of the five triggers. It is also important to note that in the CMS trigger configuration if an event is prescaled away, the trigger objects are not stored. In other words, if a trigger fires but is prescaled, the $\text{MAX}(\text{trigger } p_T)$ value is zero. Once events are separated into these five mutually exclusive categories, it is a simple matter of restoring the unbiased spectrum by weighting each category with the prescale factor of the respective trigger. In other words, the weight for each category is

1. $w_{20} = \text{Prescale}_{\text{Jet}20}$
2. $w_{40} = \text{Prescale}_{\text{Jet}40}$
3. $w_{60} = \text{Prescale}_{\text{Jet}60}$

4. $w_{80} = \text{Prescale}_{Jet80}$

5. $w_{100} = \text{Prescale}_{Jet100},$

noting that the prescale factor that is used is the prescale factor of the trigger on an event-by-event basis and not the average prescale factor for the entire run. Figure 3.7 shows the closure of each (weighted) single trigger spectra against the combined spectrum. From the figure, it is clear that once each trigger is fully efficient, the combined spectrum matches very well with that trigger, as expected. This matching ensures that the recovered spectrum is truly unbiased above the point at which the lowest trigger is fully efficient (in this case, ≈ 30 GeV/c). In addition, comparison of the statistical precision shows that the combined spectrum is significantly more accurate at high- p_T than any prescaled trigger is, confirming that the statistics saved on disk are used efficiently.

3.5 Data Availability

Once the data is stored on tape, it can be readily transferred around the world via a global network of CMS computing centers as part of the LHC Computing Grid. The CMS component of the grid is hierarchal, with four “tiers” of computing sites, as outlined in Fig. 3.8. One Tier-0 center, on-site at CERN, is responsible for handling the data acquisition directly from the detector and for the primary reconstruction of the event content. Once the reconstruction is complete, a copy of the datasets is shipped to at least one of the six CMS Tier-1 facilities, which is typically designated based on geographical considerations. The Tier-0 and Tier-1 sites assume custodial responsibility for datasets that they manage. From there, datasets can be shipped to one (or multiple) of the 25 regional CMS Tier-2 facilities, which support iterative analysis of the data by allowing any authorized CMS user to request computing time to analyze local datasets. In this way, long queues for computing resources and network bandwidth are minimized since there is no need to transfer large datasets to a user’s home computing facility.

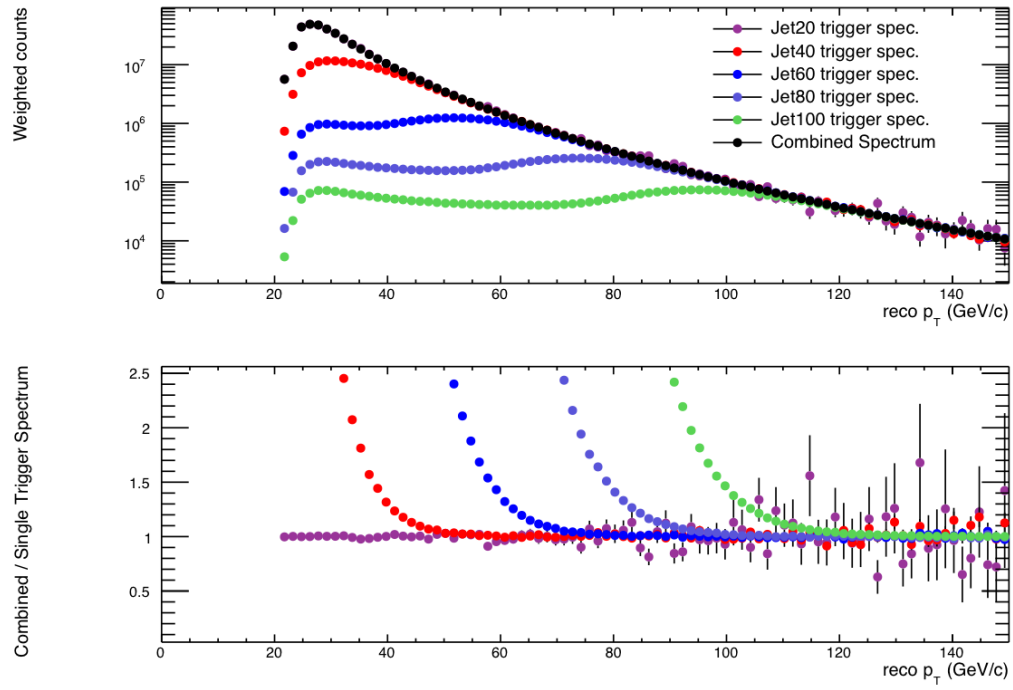


Figure 3.7. Closure plot of the single jet spectra against the spectrum obtained via trigger combination

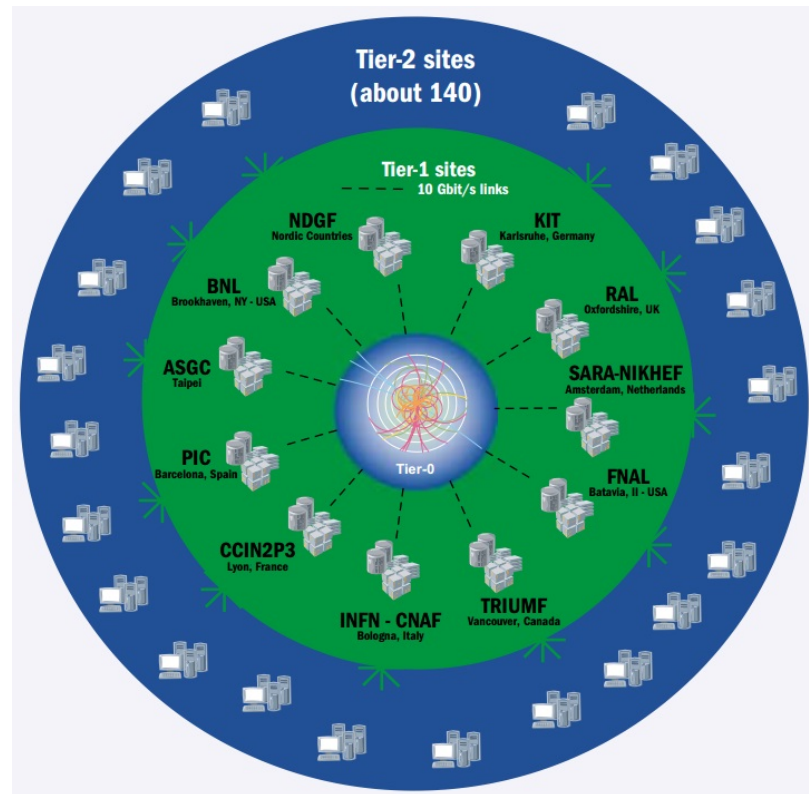


Figure 3.8. Schematic of hierarchal network setup of the LHC Computing Grid including ATLAS and ALICE computing centers (Worldwide LHC Computing Grid, 2010).

4. CENTRALITY

4.1 Theoretical Interest

In proton-proton collisions, the specific location of the collision point relative to the proton center is relatively meaningless experimentally. With large nuclei, however, this concept becomes very important. For lead nuclei, the number of binary nucleon-nucleon collisions (N_{coll}) can range from a few up to almost two thousand depending on the degree of overlap the two colliding lead nuclei share. The physics involved will be very different between a glancing blow, where only vacuum-state QCD aspects are needed to describe the collision, and a central head-on collision, where the final state observables are often modified significantly by the presence of an extended medium. In addition, it is often interesting to observe the QGP medium effects as a function of centrality. Generally, there is a direct correlation between increasing centrality and more dramatic QGP effects, however this correlation is very smooth. This smoothness is indicative that the crossover point between ordinary hadronic matter and deconfined QGP medium is not well-defined, and acts much like a second-order phase transition. Searches for a hadronic matter critical point continue to this day [50], but as of writing, there has been no such observation.

Fortunately, a model exists that allows for a prediction of cross-section behavior as a function of collision centrality. The Glauber model [51] assumes a simple linear scaling based on the number of total binary collisions from two colliding nuclei. By further assuming a Woods-Saxon density profile for all heavy nuclei [52], the expected cross-section for a hard-scattering of heavy nuclei can be found by extrapolating the cross-section of a single binary collision at the same center-of-mass energy. This simple model performs remarkably well and is used both to estimate centrality from

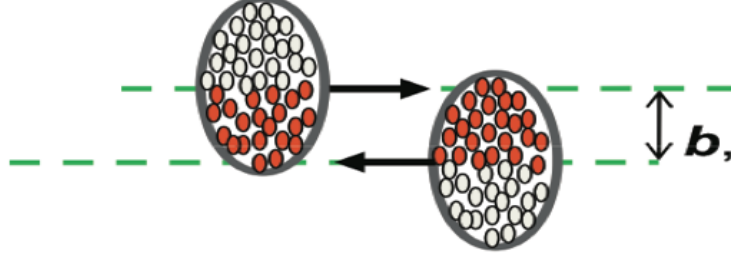


Figure 4.1. Cartoon showing an example of the collision impact parameter (b)

final-state information and in order to directly compare proton-proton results with those observed in heavy-ion collision environments.

More formally, the model simply states that the hard scattering cross-section of any two nuclei A and B can be approximated by:

$$\sigma_{AB}^{\text{hard}} = \int d^2b \sigma_{NN}^{\text{hard}} T_{AB}(b), \quad (4.1)$$

where $T_{AB}(b)$ is the nuclear density function as a function of impact parameter (b). The impact parameter is the distance between the centers of the two colliding nuclei and is schematically demonstrated in Fig. 4.1. Finally, the nuclear density is approximated by the Woods-Saxon distribution and normalized such that

$$\int d^2b T_{AB}(b) = AB \quad (4.2)$$

where A and B are the number of nucleons in each nuclei.

4.2 Experimental Methodology

It is impossible to monitor the beam position well enough to determine centrality before the collision, so experimentalists are constrained to using final-state observables to approximate the nuclear overlap of a given collision. Fortunately, the Glauber model provides an estimate of centrality directly from the number of binary collisions.

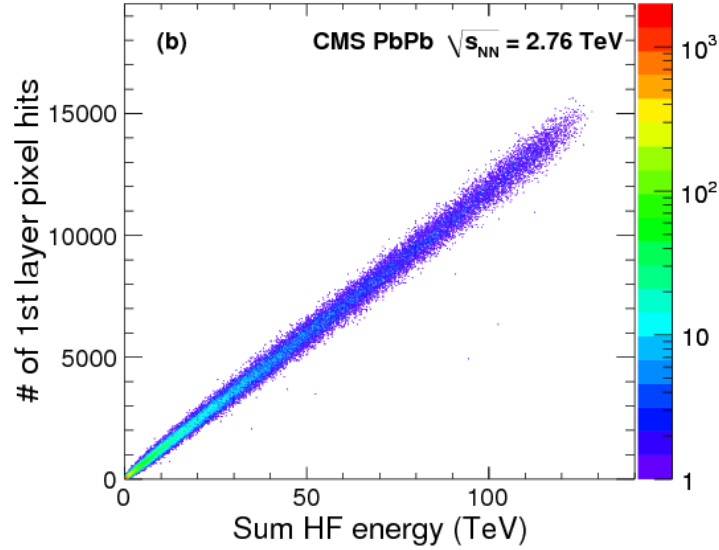


Figure 4.2. Correlation of pixel multiplicity with forward and backward HCAL energy in PbPb collisions, showing a clear, tight correlation.

Since it is expected that there is a direct correlation between the number of binary collisions and the number of final state particles produced transverse to the beam direction, we can simply use the multiplicity of the collision, or number of observed final-state particles, as a proxy for the collision centrality.

There is a small technical catch. Jet analyses typically involve the observation of hard fragmenting processes, which produce many particles that are uncorrelated with the degree of nuclear overlap. To avoid biasing the centrality classification by jet production, an alternate definition of centrality is used. It is expected that the enhanced particle production should scale in similar ways over all values of pseudorapidity, such that energy collected in very forward or backward detectors can also provide a reasonable correlation to collision centrality. To test this, the correlation of pixel multiplicity to energy collected in the forward and backward hadronic calorimeters ($3 < |\eta| < 5$) is shown in Fig. 4.2 for a collection of PbPb data at $\sqrt{s_{NN}} = 2.76$ TeV. This very tight correlation indicates that forward and backward HCAL energy can

be used to classify centrality. In this way, by only measuring jets in the central barrel ($|\eta| < 2.4$) and using the HCAL endcaps to measure the centrality, we ensure that jet production does not bias the centrality definition. Events are divided into percentage classes, such that 100% centrality corresponds to the most peripheral events and 0% centrality corresponds to the most central events.

5. EVENT SIMULATION

Experimentalists and theorists share a symbiotic relationship, where each cannot exist without the other. While theorists drive the frontiers of the field, attempting to explain effects with new models, these attempts are fairly fruitless without experimental results and guidance. In a fairly similar way, Monte Carlo (MC) event generation is a microcosm of this effect. Event generators are code packages that attempt to simulate one (or many) aspects of particle collisions, usually guided by a theoretical model or framework. Without theorist input to event generators, experimentalists have no baseline with which to search for new physics, but experimental data also provides additional constraint to event generators which can be retuned (or sometimes completely overhauled) to reflect new experimental discoveries.

5.1 PYTHIA Event Generator

Modern day event generators tend to be fairly specific, attempting to replicate a few (or even one) particular subprocess with great accuracy. This stems from the fact that the strong force coupling constant is large, such that perturbative methods are generally only valid in special cases. Because of this, a full descriptive theory of an evolution of a collision from the initial hard scattering through the particle breakup (“hadronization”) is not possible at this time. The PYTHIA event generator, however, attempts to bridge pQCD calculations and phenomenological models in an attempt to fully describe the behavior of proton-proton collisions. PYTHIA version 6.424 is used in all analyses described in this dissertation. A comprehensive manual describing the generator can be found in Ref. [53].

The PYTHIA generator makes use of the factorization theorem (Eq. (2.2)) and attempts to fully replicate the known Standard Model processes with the highest

possible accuracy. That being said, this feat is highly non-trivial especially considering that a fundamental (read: nonperturbative) model of color interactions is not currently possible without time-consuming numerical integration methods on a lattice. Therefore, PYTHIA attempts to prioritize leading-order processes and is optimized for $2 \rightarrow 1$ and $2 \rightarrow 2$ body scattering processes. Though a general treatment of interactions with a three-body final state is not implemented, specific processes are tacked-on to the particle fragmentation framework when necessary. It is important to note here that flavored jet production has a large fraction from $2 \rightarrow 3$ body interactions and, as such, PYTHIA is not expected to faithfully reproduce all aspects of heavy-flavored jets. Nevertheless, the generator does perform remarkably well, and uncertainties stemming from the uncertainty of PYTHIA calculations are generally small.

Jet physics are reproduced in PYTHIA primarily via the implementation of initial and final state radiation in conjunction with the method used for jet hadronization. Parton radiation is modeled as a series of $1 \rightarrow 2$ body scatterings, where momentum, energy, color, charge, etc. are conserved. The radiation probabilities are governed by the DGLAP evolution equations until the infrared cut-off value is reached. In PYTHIA version 6.424 this cutoff is 1 GeV.

Hadronization is performed via an implementation of the Lund string fragmentation model. After two quarks undergo a hard scattering, the distance between them rapidly increases, such that the potential energy becomes very large. Due to this large potential, a second $q\bar{q}$ pair may be created from vacuum fluctuations. This new pair may hadronize with the original pair, or may pair off with other extended quark pairs in the vicinity. In addition, both pairs of quarks now may create additional pairs from the vacuum which, in turn, may create more pairs, such that a large multiplicity of particles are created in a parton shower. The parton shower continues iteratively until all quarks are fully hadronized. Once hadronization is complete, a decay table is used to fully decay unstable particles with proper resonances, lifetimes, etc. Further details of Lund string fragmentation can be found in Ref. [54].

5.2 Heavy-Ion Generators

For all PYTHIA's successes, the relative simplicity of the model does not make it a great candidate for the simulation of a heavy-ion collision. Rather than directly simulate each parton scattering subprocess, modern Heavy-Ion event generators take a different approach and attempt to simulate large collections of nucleons with statistical methods instead of simulating the details of each individual nucleon-nucleon interaction. Because the QGP is thought to act much like a strongly interacting liquid with collective behavior, this approach is a more natural starting point for the simulation of the QGP medium.

HIJING

The HIJING generator [55] was one of the first attempts to fully simulate jet production in a heavy-ion event. It relies strongly on the same theoretical framework as PYTHIA for jet production, while including a model for jet quenching via radiative energy loss based on the expected parton-medium interaction. Most relevant to these analyses, however, HIJING implements an implementation of the Woods-Saxon nuclear density distributions and an eikonal formalism to determine the impact-parameter dependence of inelastic process production. This allows for an estimation of centrality-dependent effects, such that a more central event produces a denser medium and a more peripheral event acts more like a p+p event. Finally, nuclear effects in the hard scattering are accounted for by the inclusion of an impact-parameter-based parton distribution function.

HIJING also invokes the concept of Minijets, which are very soft (≈ 2 GeV) jets produced from hard scattering events, but instead of being clearly resolveable in a detector, tend to drive the particle correlation behavior seen in the medium. In the model, minijet production scales with the number of inelastic scattering events in a given A+A crossing.

AMPT

The AMPT generator [56] is an extension to the HIJING generator such that the output parton distribution from a modified version of HIJING is used as a transient input to the AMPT program. Instead of proceeding with the production of minijets to deal with partonic energy loss, an alternate formulation is used. AMPT forces hard partons to suffer additional rescatterings rather than model energy loss as a series of simple straight-line induced radiation. A parton phase space distribution is calculated from the initial HIJING parton distribution and given a formation time. Once they are formed, the minijet parton cascade is performed using the Zhang’s Parton Cascade (ZPC) model [57], which consists primarily of elastic rescattering. The ZPC model includes gluon-gluon elastic scatterings at leading order, regulated by a color screening mass that is currently obtained from an input parameter.

Once the parton cascade is complete, the Lund string fragmentation model is evoked in order to produce the resulting hadrons from the parent partons. There is also a “string melting” option of AMPT that invokes a different quark coalescence model, but this is not used in these analyses.

At this point, chemical freezeout is reached, where only elastic scatterings remain. To finish the collision evolution, the ART hadronic transport model is used, albeit modified slightly to include additional interactions that only occur at RHIC and LHC energies. The final output of the generator occurs when a cutoff time is reached - generally 1.2 fm/c after the last elastic scattering occurs.

HYDJET

The HYDJET generator [58] gets its name from the combination of tools it uses to generate simulated heavy-ion collision events. The premise is that it is a combination of hydrodynamics, PYTHIA, and PYQUEN (which itself is PYTHIA incorporating jet quenching effects). The focus of the generator is to simulate jet quenching effects as accurately as possible, while still simulating a realistic underlying event, including

collectivity and soft hadroproduction. A standard simulated event goes through the following processes: First, the total inelastic and hard scattering cross-sections are calculated by combining the PYTHIA prediction from the center-of-mass energy and the expected number of nucleon-nucleon interactions from the Glauber model. Next, the hard scattering is generated, based on expectations of the jet production cross-section from PYTHIA, where the jet partons propagate forward using the quenched PYQUEN subroutines. These subroutines modify the initial parton states by including radiative and collisional energy loss according to the BDMS gluon radiation spectrum [59] and assuming an isotropic Boltzmann distribution of medium partons at the QGP temperature. Nuclear shadowing effects may or may not be included at this point, depending on user parameters, and the jets are hadronized using the PYTHIA Lund string fragmentation.

The soft part of the HYDJET event can be considered a fully thermalized collection of soft hadrons and is generated first by calculating the chemical and thermal freeze-out hypersurfaces obtained from relativistic hydrodynamics. Next, the size of the medium is calculated based on the collision's initial state properties and expected effective temperature. Finally, partons are distributed randomly around the medium assuming a Poisson multiplicity around the mean expectation value. Finally, the boosted parton collective flow and four-momentum is calculated and left to evolve naturally. Particles decay to final-state hadrons via decay tables and include natural two- and three-body resonances.

EPOS

EPOS generates events using a somewhat alternative theoretical framework than the other generators mentioned in this section [60]. The generator models multiple interactions not as simple overlaps of single nucleon-nucleon interactions, but in a style governed by Regge-Gribov theory and pQCD. The model is expressed in terms of virtual composite particles known as pomerons. In exchange diagrams, pomerons

consist of multiple gluons (or even multiple gluon ladders) and, consequently, a collection of quantum numbers. The advantage, however, is that pomeron exchange allows for a p+p and Pb+Pb collision to be treated with the same theoretical framework which reduces the need for effective or phenomenological-driven theories. As pomerons are the starting point for the generation of particles, the multiplicity of an event is, on average, proportional to the cross-section of the number of cut pomerons, where cut pomerons refer to string breaking phenomena.

The underlying event evolves like HYDJET, where relativistic thermodynamics is used, however, the initial conditions given to the underlying event are governed by the number of pomerons that are cut during the collision. The evolution of a cut pomeron into a jet or soft particle is obtained via one of three scenarios unfolding. First, a string can form in the cut pomeron tube with too little energy to escape. In this case, the particles evolve with the hydrodynamic simulation. Second, a string can form outside the cut pomeron tube, such that it simply escapes the event as an unmodified jet. Third, the string can form inside the cut pomeron tube but with enough energy such that it escapes. This situation reflects the production of a jet that has been modified by the medium.

Once the strings are produced, they hadronize via a combination of vacuum string fragmentation (à la the Lund model) and through collective hadronization procedures. Hard particles typically escape the medium too quickly to hadronize inside the “core” of the heavy-ion event, but soft partons hadronize as well as dynamically interact until they reach an effective freeze-out cutoff, governed by a given energy density.

Generator Deployment in CMS

Once of the primary drawbacks of the ideology used in most heavy-ion collisions is that while they simulate the medium and effective collective behavior very well, their statistical nature makes it impossible for any of the heavy-ion specific generators to produce a given scattering process *a priori*. For example, in a PYTHIA event, specific

subprocesses can be invoked, such that every event generates a b-jet, for example. This specificity is impossible in heavy-ion generators. However, it is very undesirable to produce millions of heavy ion events, hoping a few events that correspond to a physics process of interest are created. This is made especially so considering that a detector response simulation of a very central event at top LHC energies can take around ten hours.

To get around this impracticality, a process known as “embedding” is used, where PYTHIA subprocesses are manually inserted into a minimum-bias style background from a heavy-ion generator by matching the generator vertices. The products of both generators are then simulated and reconstructed as if they were a single event. Unfortunately, that means that while production speed is dramatically improved, all jet-to-medium interaction is lost, since embedding does not allow for interactions between the PYTHIA and background events. This is, therefore, really only useful for hard processes like jet production, where we can invoke the factorization theorem and assume that jet production is not affected by the presence of the medium. Let me be clear that while jet energy *after production* is certainly modified once it interacts with the medium, the initial production mechanisms are likely not modified. So while embedded events may accurately reflect the jet production cross-section in heavy-ion events, they can only be used as a baseline for studies of jet quenching, since this phenomenon will not be present in embedded simulations.

The two interaction vertices are matched in the PYTHIA and background event such that, for all intents and purposes, information from both generators is treated as a single interaction event. Once generated, the event is propagated through a full detector simulation using the Geant4 package [61], which contains a complete physical description of the detector, including materials properties, such that particle-detector interactions and responses are simulated accurately.

Once data has been taken, the simulation is generally reweighted to account for discrepancies between the data and MC. Distributions of centrality and vertex po-

sition are always reweighted to data to accurately portray any biases from event selection, though additional distributions may be reweighted as necessary.

As is stated in the introduction to this chapter, these models are necessarily incomplete, or else the experimentalist's job would be unnecessary! Therefore it is always of utmost importance to consider the limitations of any generator and accurately reflect any shortcomings or misleading results provided by any generator. They are not catch-alls and must be used intelligently!

6. JETS

The primary physics object of interest in the analyses discussed in this dissertation is the jet. Jets are composite structures typically produced in hard-scattering events, where two bound quarks are forced apart very rapidly in a collision and strong-force string breaking induces a particle shower. This process is known as “fragmentation” and forms the basis of jet production. Experimentally, fragmentation manifests itself through the observation of clusters of tracks and calorimeter hits in a detector. In this sense, it is important to discuss individual track reconstruction first before moving on to full jet reconstruction techniques.

6.1 Track reconstruction

Track reconstruction forms the basis of jet reconstruction, as at an experimental level, jets are simply somewhat arbitrary cones containing a collection of tracks (or calorimeter energy depositions). Reconstructing tracks in a very dense environment like a PbPb collision at the LHC is a challenging task, but the very fine granularity of the inner tracking system and a number of iterative algorithms assist in properly reconstructing the tens of thousands of detector hits in a typical central PbPb collision.

Except in the case of special muon reconstruction, CMS uses an “inside-out” tracking ideology, meaning that tracks are first found by creating seeds from the inner tracking system. These seeds are created by finding charge deposition on pixels that are radially aligned, indicating a coincidence from a single track. When three charge depositions (“hits”) align this way, it is referred to as a pixel triplet. These pixel triplets form the basis for the majority of high-quality tracks used in physics analyses. Once these are identified, additional pixel doublets (where only two distinct

Table 6.1.
Iterative track reconstruction steps for pPb and pp reconstruction

Step	Seeds	Max. Transv. Displ. (cm)	Min. trk p_T (GeV/c)
0	Pixel Triplets	0.02	0.6
1	Low- p_T Pixel Triplets	0.02	0.2
2	Pixel Pairs	0.015	0.6
3	Detached Pixel Triplets	1.5	0.3
4	Mixed Triplets (Pixel + SiStrip)	1.5	0.4
5	SiStrip Pairs	2.0	0.7
6	SiStrip Endcap Pairs	6.0	0.6

hits are radially aligned) are found in the next iteration. This process of iteratively identifying obviously aligned track seeds with successively looser constraints in the pixel detector continues until all tracker hits are used or until no more hits pass the required quality selections. The iterative procedure for prioritizing high-quality seeds is shown in Table 6.1.

At this point, the collection of seeds is used to build tracks. The magnetic field inside the solenoid is nearly homogeneous so typical solutions to a charged particle moving at high velocity through a uniform magnetic field are helices. Assuming a helical shape, the track reconstruction algorithm propagates a track outward by estimating a track momentum and attempts to match the track extrapolation with any energy deposition in the silicon strip tracker. Using Kalman filter techniques [62], the algorithm finds the best fit for each track seed until the seed list or the silicon strip hits are exhausted or until there are no more unassociated hits that are compatible with the remaining track seeds.

Once the inner tracking has completed, there is a list of criteria each track must satisfy for inclusion in an analysis. Generally, tracks must satisfy the “highPurity” requirements, described in detail in Ref. [46]. Although the specifics of the highPu-

rity selection has evolved over time, this selection essentially requires tracks to be associated to a minimum number of hits on the full inner tracking system, along with strict requirements on the normalized detector hit discrepancy from the best fit helix (χ^2) and compatibility with a pixel vertex. The χ^2 requirement mitigates any poorly reconstructed tracks, where a combinatorial alignment of a few hits might convince the track reconstruction code that a real particle propagated through the detector. The pixel track to vertex requirement ensures that tracks from cosmic particles or beampipe interactions do not dramatically contaminate the sample. Pixel vertices are a spatial association of at least two pixel-only tracks such that they are compatible with either the primary interaction point between the two original colliding protons or lead nuclei, known as the “primary vertex” or compatible with a long-lived particle decay, known as a “secondary vertex”.

Finally, a minimum p_T selection of 0.4 GeV/ c is imposed to both mitigate the extremely high multiplicity of low momentum particles and to ensure that particles do not bend in the magnetic field so much that they actually form full circles and repeatedly interact with the tracking system. These “looper” particles are especially problematic for reconstruction and great care must be taken to remove contamination from these effects if one desires to reconstruct particles at very low momentum. In principle, however, these particles contain very little information about the physics of the initial hard process so they are often simply rejected. The radius of curvature can be effectively calculated by:

$$R(m) = \frac{p_T}{cB} \quad (6.1)$$

where p_T is in eV/ c , B is the magnetic field in Tesla and c is the speed of light. At CMS, the looper p_T threshold corresponds to about 250 MeV/ c , since the inner tracker has a radius of 20 cm.

Finally, the calorimeter information is incorporated into the tracking through the use of the “Particle Flow” algorithm. This CMS-specific reconstruction technique allows for the selection of the optimal combination of tracker and calorimeter information to build single particle candidates. In this way, the algorithm can determine

whether a particle is one of three primary types. First the algorithm looks for charged hadrons, where energy is deposited in both the ECAL and HCAL and this energy is associated to a track. Next, the algorithm looks for photons, where energy is deposited in the ECAL, but not in the HCAL, and the location of the energy deposition is not associated to a track. Finally, neutral hadrons are found by looking for energy in the HCAL that is not associated to either a track or to energy in the ECAL. Again, starting inside-out, tracks are extrapolated to the ECAL and HCAL, and matching energy depositions are associated to each track. The algorithm does not necessarily assign the entirety of the energy in a particular calorimeter hit to the track, rather it builds an expectation of the energy based on the track momentum and particle type. This allows for the reconstruction of overlapping neutral and charged hadrons in the calorimeter.

In addition to the Particle Flow jets, jets are also reconstructed using only calorimeter information, both as a cross-check of the particle flow algorithm and for the study of systematic uncertainties obtained by using the combined information of the tracker and calorimeters.

6.2 Jet Reconstruction

Once the individual tracks are reconstructed and the particle flow algorithm is run, jets can be clustered. There are many choices for jet clustering algorithms, driven both by physics interest and computing constraints. No matter the choice, however, a good algorithm should satisfy at least a few basic constraints. First, the algorithm must be infrared and collinear safe. In other words, the introduction of very soft infrared gluon emission and the presence of collinear splitting (where a high- p_T particle splits into lower- p_T daughter particles that are Lorentz-boosted to be collinear to the original particle momentum) during jet fragmentation should not affect the jet reconstruction. Consider the performance of a notoriously infrared-unsafe algorithm, like a seeded fixed-cone algorithm. This algorithm starts with a

high- p_T seed and defines a fixed cone around the seed. Then, the algorithm checks to see if the vector sum of particles in the jet cone matches the seed vector. If so, the jet is declared stable. Else, the cone is redefined around the particle vector sum and iterates again. When jets overlap, however, the algorithm must play tricks (e.g. adding additional seeds at the midpoint) to distinguish components of the two jets. It turns out, however, that the presence of very soft emission in this region between the jets can dramatically affect the jet reconstruction, so much so that additional (fake) jets can be reconstructed. Additional discussions of infrared safety can be found in Ref. [63], while additional discussion of jet reconstruction techniques can be found in Ref. [64].

Instead of seeded algorithms, computational techniques, e.g., those implemented by the FastJet package [65], are now powerful enough to efficiently use sequential recombination algorithms that inherently are both collinear and infrared safe. One such example is the anti- k_T algorithm [66], which itself is a specific case of a general recombination algorithm. The general recombination procedure essentially defines a quantity for each calorimeter hit or track

$$d_i = k_{T,i}^{2p} \tag{6.2}$$

and a similar quantity for each pair of calorimeter hits or tracks.

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{6.3}$$

In this algorithm, i and j are the particle indices, k_T is the particle transverse momentum, Δ is the distance between two particles, R is the requested radius of each jet cone and is a user input to the algorithm, and p is an input parameter that defines the clustering behavior of the algorithm. The anti- k_T algorithm chooses what appears to be an almost pathological choice of $p = -1$. It does, however, lead to some interesting consequences and ends up performing very well. First, the algorithm loops over all particles d_i and particle pairs d_{ij} . If $\min(d_{ij}) < \min(d_i)$, the particle pair is merged into a single effective particle and all d_{ij} values are recalculated. If instead

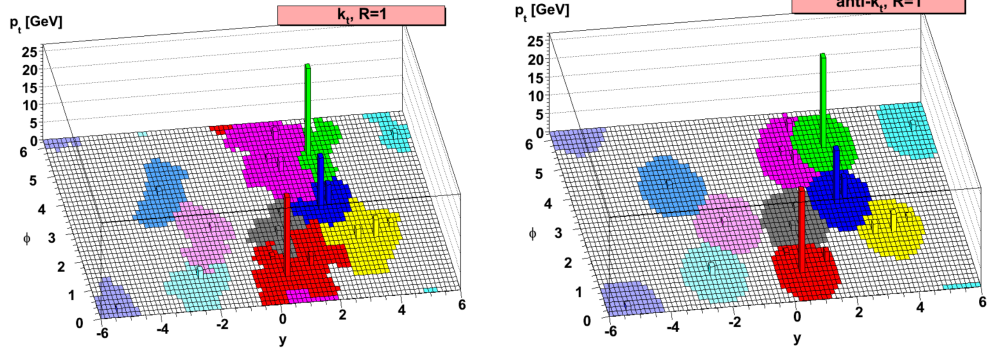
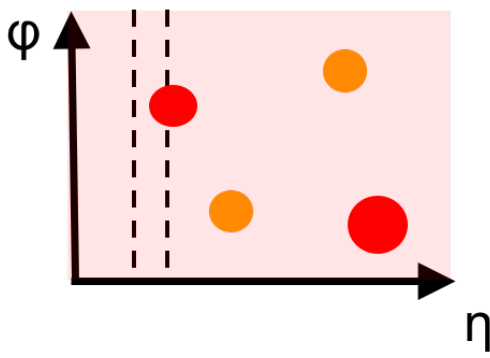


Figure 6.1. Simulated jet clustering behavior using the sequential recombination algorithm with $p = 1$ (left), and $p = -1$ (right).

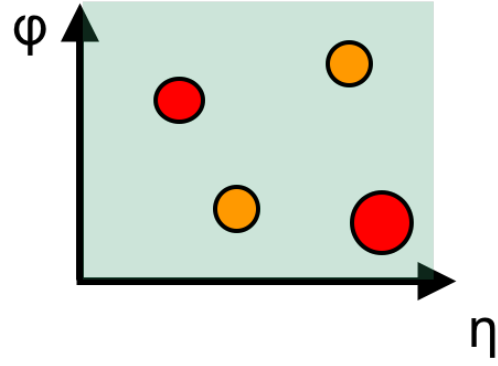
$\min(d_i) < \min(d_{ij})$, the particle is designated as a jet and removed from the collection. One can see the collinear and infrared safety in the algorithm by observing that:

- When the k_T of a particular candidate is very small, it is almost guaranteed to be merged with an adjacent particle, so it has no bearing on the jet, ensuring infrared safety.
- When the distance between two particles of any momentum approaches zero, they are also almost guaranteed to be merged, ensuring collinear safety.

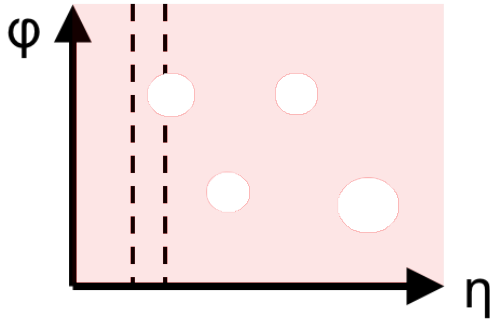
What the choice of $p = -1$ achieves is the prioritization of the hard components when clustering. This leads to jet cones that are nearly circular and have sizes that correspond well to the input R parameter. Other choices ($p = 0$, $p = 1$) lead to many deformed cones, which introduces a bias based on the jet fragmentation behavior. This can be observed in a simulation using the HERWIG generator [67], shown in Fig. 6.1, where the choice of $p = 1$ is shown on the left, and the choice of $p = -1$ is shown on the right for a cone size of $R = 1.0$.



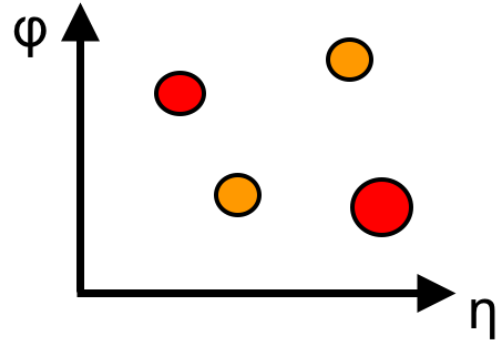
(a) Step 1, Background energy calculated in strips of η



(b) Step 2, anti- k_T algorithm run on background subtracted towers



(c) Step 3, Background energy recalculated in strips of η , excluding jet areas



(d) Step 4, anti- k_T algorithm rerun on background subtracted towers

Figure 6.2. Four-step procedure for underlying-event subtraction for jets at CMS.

6.3 Underlying-Event Subtraction

One major difference between jets in a heavy-ion environment and those in a proton-proton environment is the presence of a large soft particle background in the heavy-ion environment that is mostly uncorrelated to the hard scattering event that

produced the jets. Unfortunately the removal of the underlying event is especially non-trivial. It is simply not possible to answer experimentally whether a given soft particle originated as part of some thermal radiation inside the jet cone or whether it is truly from the jet fragmentation process. To complicate matters further, the soft background can fluctuate significantly event-by-event, so a statistical subtraction is not feasible. What is done at CMS for all PbPb and pPb jet analyses is shown in Fig. 6.2. The idea is to first cluster the total energy in towers using the HCAL granularity (since it is the largest) and calculate the average energy in strips of pseudorapidity. This average energy is removed from all towers, and the anti- k_T algorithm is run. Then, the area that contains reconstructed jets is removed from the towers and the background is recalculated, again in strips of pseudorapidity. Finally, the jets are rebuilt again after subtraction using the anti- k_T algorithm. These final jets are used for analysis.

6.4 Jet Corrections

Once jets are reconstructed, they must be corrected to account for detector inefficiencies and the non-linear response of the calorimeters. The jet energy is corrected in a series of steps, starting with the largest correction factors and moving toward smaller corrections, such that each successive correction is applied on top of the previous correction. The total correction is defined as:

$$C = C_{MC}(p_T^{raw}, \eta) \times C_{asym}(\eta) \times C_{abs}(p_T''), \quad (6.4)$$

where the three C factors are, in order, from MC jet energy response, the jet response asymmetry as a function of η , and the absolute jet energy scale derived from photon or Z boson balancing.

The largest factor is due to the jet energy response correction, as derived by Monte Carlo simulation. After simulation and reconstruction of an event, generated jets are matched to their post-reconstruction counterparts. After matching, Gaussian fits are made in narrow ranges of jet p_T of the ratio of reconstructed jet energy to the

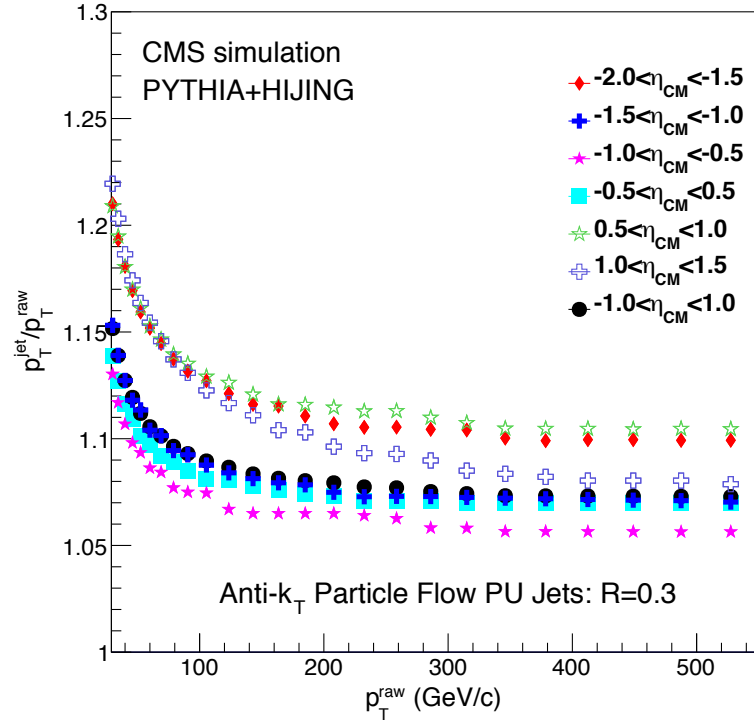


Figure 6.3. Correction factors from simulation for anti- k_T jets for various η intervals as a function of jet p_T .

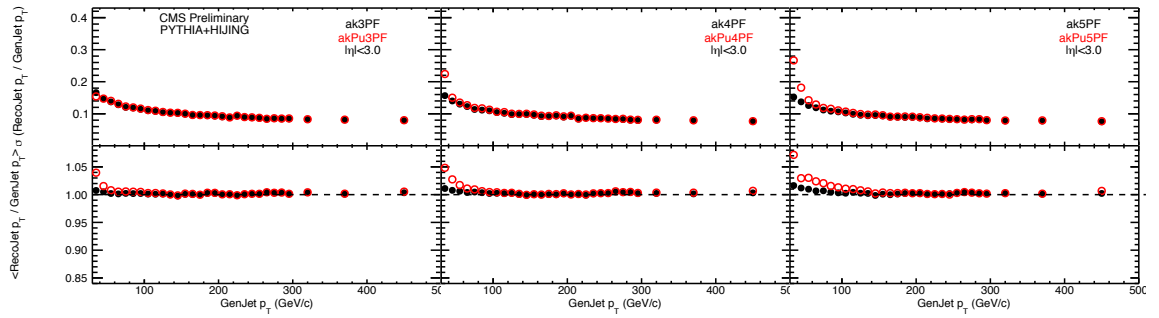


Figure 6.4. Jet resolution (top panel) and response (bottom panel) for jets reconstructed with the anti- k_T algorithm using size parameters between 0.3 and 0.5 units in η/ϕ space.

generated jet energy, where the gaussian mean is taken to be the average deviation of jet energy from generator truth, and the Gaussian width is the effective resolution of jets in that narrow range of jet p_T . The gaussian means are plotted as a function of jet p_T and are applied as the first correction factor $C_{MC}(p_T^{raw}, \eta)$ to the jets in data. As shown in Fig. 6.3, the average correction factor ranges between 5% and 20%, depending on both jet p_T and jet η . Once this correction factor is applied, a closure test is performed as a function of generator-level jet p_T . This closure test is shown in Fig. 6.4, where the top panel shows the jet resolution, defined as the gaussian width (σ) of the corrected jet p_T divided by the MC truth p_T . The bottom panel shows the jet energy scale closure, which is the average of corrected jet p_T divided by the MC truth p_T . When applying the correction factor to a MC simulation, we observe closure to within 5% for all jet p_T .

The second correction factor aims to correct for the jet response as a function of jet η . Essentially, the barrel region of the detector ($|\eta| < 1.3$) is assumed to be well-calibrated after the MC response correction is applied, and the jet response as a function of pseudorapidity is calculated relative to the barrel region. Dijet events are used to calibrate the region outside the barrel, where one “reference” jet is required to be in the barrel region, and the response is measured against a “probe” jet, which can be found anywhere in the detector. If both jets are in the barrel region, the reference and probe jets are randomly assigned. The full correction is the ratio of two response functions:

$$C_{asym}(\eta) = R_{rel}^{MC} / R_{rel}^{data}, \quad (6.5)$$

where the response functions in MC and data are defined relative to the dijet balance function

$$R_{rel} = \frac{2 + \langle B \rangle}{2 - \langle B \rangle}, \quad (6.6)$$

and the dijet balance function B is defined as

$$B = \frac{2(p_T^{probe} - p_T^{ref})}{p_T^{probe} + p_T^{ref}}. \quad (6.7)$$

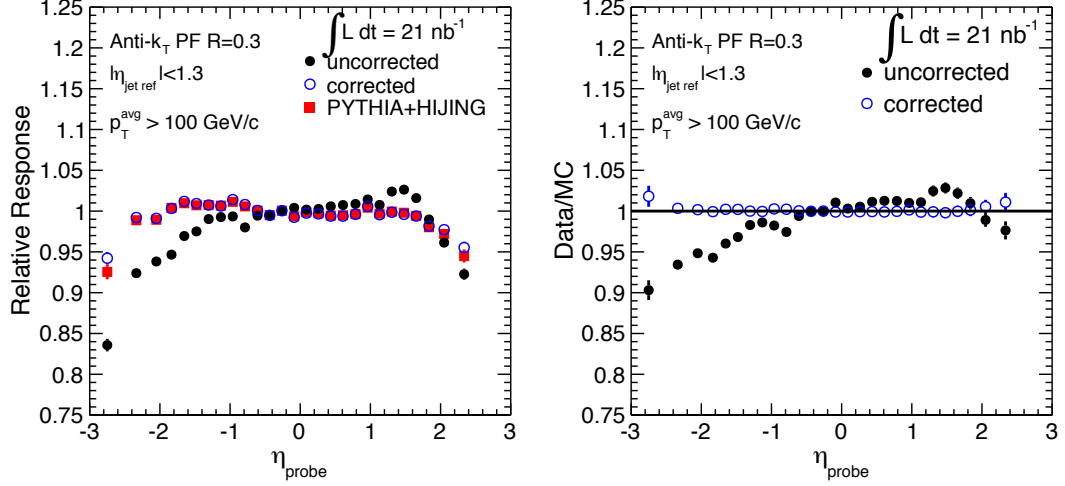


Figure 6.5. The Relative jet response is shown for pPb (asymmetric) data, both before and after applying the C_{asym} correction factor as compared to simulation (left), along with the ratio of data to MC both before and after the correction factor is applied (right).

Since the method assumes that dijets carry roughly the same amount of energy in order to calibrate the response, events with three or more jet must be accounted for. These three-jet events are suppressed by requiring $p_T^{\text{third}}/p_T^{\text{avg}} < 0.2$, and by requiring the two largest jets in an event to be relatively back-to-back in azimuth, where $|\Delta\phi| > 2.5$. The performance of this correction in pPb data can be seen in Fig. 6.5, for dijet events with $p_T^{\text{avg}} > 100$ GeV. This correction is not applied to PbPb data, as we observe significant jet suppression in PbPb collisions that breaks the back-to-back jet symmetry. As we no longer have any control over the expected jet energy in the barrel, this correction is not viable in the heavy-ion collision systems. We account for this by assigning an additional jet reconstruction systematic uncertainty to the PbPb jet analyses.

At this point, jets are calibrated relative to simulation, and to each other, but an overall scale factor may be missing from the jet energy. The final required correction factor applies an absolute energy scale correction factor, based on the assumption that

photons or Z bosons are well-calibrated, since their reconstruction is independent of the HCAL reconstruction. Similar constraints are placed on photon+jet events as are on the dijet events used in the previous step. Events are required to have a photon in the barrel region of the detector, with $p_T^\gamma > 40$ GeV. These photons are required to be isolated, with the full constraints detailed in Ref. [68], and again, the photon-jet pair is required to be mostly back-to-back in azimuth, with $|\Delta\phi| > 2.5$, and photon+2 jet events are suppressed using a selection of $p_T^{\text{second}}/p_T^\gamma < 0.2$. Though these tight selections limit the statistical precision of this energy correction, the remaining energy scale difference derived from this correction is negligible for pPb and pp data, as shown in Fig. 6.6. Note the value of 0.9 in the left panel of Fig. 6.6 does not indicate poor jet energy scale closure, rather it is a simple kinematic effect obtained by comparing photons with jets of a fixed cone size. As the ratio of the two response functions in MC and data is essentially unity, this confirms the good closure of the reconstructed jets and the “true” jet energy. This correction is also not applied using PbPb data, since again, jet quenching effects ruin the correlation of jet and photon energy. However, as quenching is expected to factorize from the detector response, the PbPb residual response can be calculated using pp data at the same center-of-mass energy.

Unfolding

Once the jets are fully calibrated and corrected, a final procedure must be applied to the jets to directly compare them to any theoretical prediction. While the calibrations and corrections derived in the previous section correct the average jet energy, an intrinsic jet-by-jet uncertainty remains due to the detector resolution. This leads to jet p_T smearing behavior that will affect any non-flat distribution of jets, e.g. the jet cross-section as a function of jet p_T . As a jet may have a resolution of up to a $\pm 20\%$ difference from the “true” jet p_T , this gaussian smearing convoluted with a steeply-falling spectrum means that jets are preferentially smeared toward larger

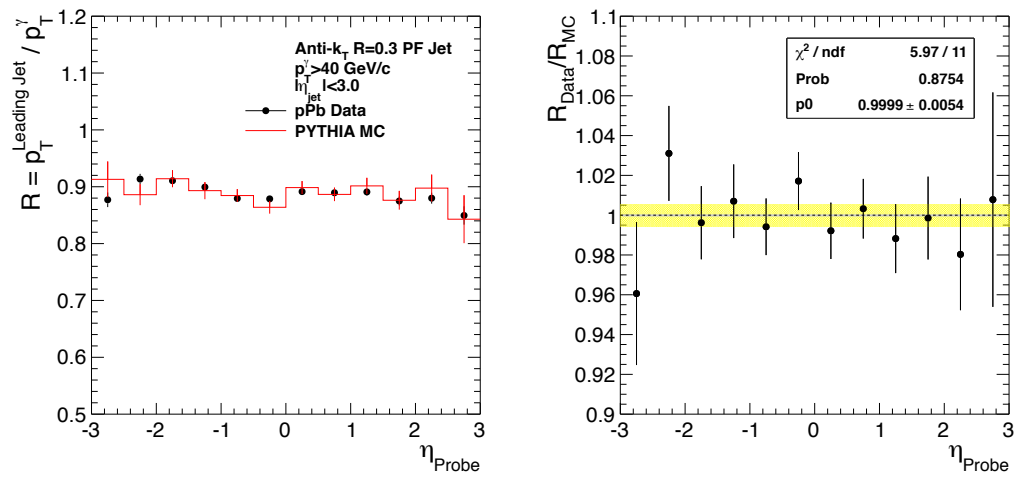


Figure 6.6. The γ +jet residual response function for MC and pPb data (left) and the ratio of data/MC (right) is shown.

momentum. In other words, this resolution smearing actually artificially flattens the jet p_T spectrum. In addition, the non-linearity of the jet resolution (as seen in the top panel of Fig. 6.4) affects the shape of the distribution. Lower momentum jets tend to have a broader resolution, so the reconstructed jet p_T distribution becomes more steep than the generator-level jet p_T distribution. The combination of these two competing effects is very difficult to identify jet-by-jet and powerful statistical tools are required to deconvolute the resolution smearing from the spectrum. The ratio of the reconstructed to the generator “truth” p_T spectra can be seen in the filled red points on the right panel of Fig. 6.8.

To correct for this smearing effect, the jets are “unfolded”. At its heart, the unfolding procedure is a matrix inversion:

$$\frac{dN}{dp_T^{\text{reco}}} = \mathbf{A}_{r,t} \frac{dN}{dp_T^{\text{truth}}}, \quad (6.8)$$

where $\mathbf{A}_{r,t}$ denotes the two-dimensional correlation matrix between reconstructed and MC “truth” jet p_T . While it is tempting to simply apply the inverted matrix $\mathbf{A}_{r,t}^{-1}$ in bins of p_T to the reconstructed jet spectrum, the transformation matrix itself has an uncertainty which can dramatically affect the final spectrum. Furthermore, small fluctuations in the reconstructed spectrum can be transformed into large nonphysical fluctuations after a simple matrix inversion, so more robust tools are needed. An example unfolding matrix can be seen in Fig. 6.7 in pPb (left) and pp (right).

The analyses discussed in this dissertation use two primary methods of unfolding. First is a method based on Bayes’ Theorem as developed by D’Agostini [69], and second is a method based on the principles of Tikhonov regularization known as “Singular Value Decomposition” or SVD unfolding [70]. Both methods are accessed via the RooUnfold implementation [71] and attempt to minimize the large fluctuations from the matrix inversion, though in somewhat different ways. Bayesian unfolding uses an iterative (though non-convergent) method to successively minimize the uncertainty on the output spectrum, while SVD unfolding provides a procedure to choose an optimized regularization parameter, which the algorithm uses as an input. Both methods are attempted in these analyses and the resulting closure in simulation can

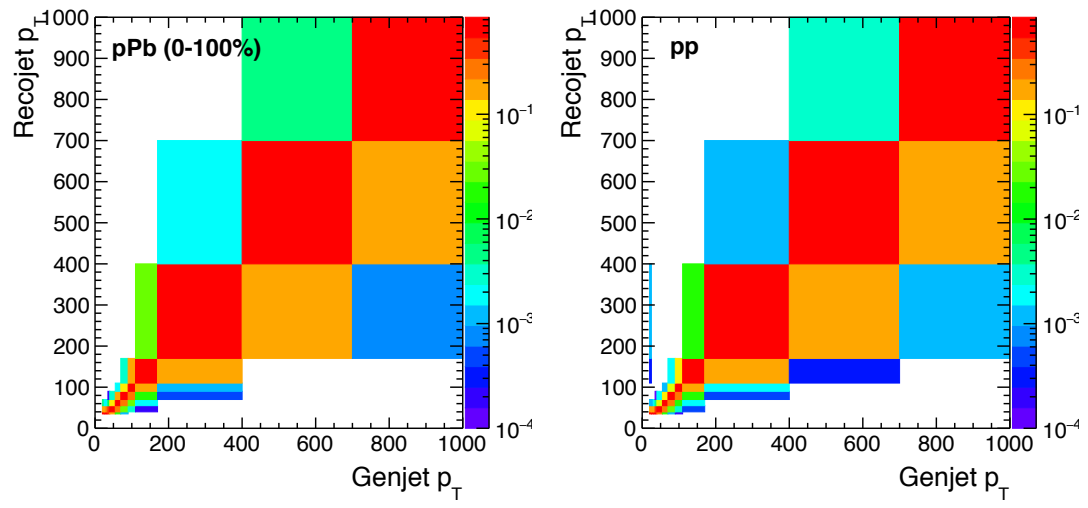


Figure 6.7. Unfolding matrix created from simulated pPb collisions (left) and pp collisions (right)

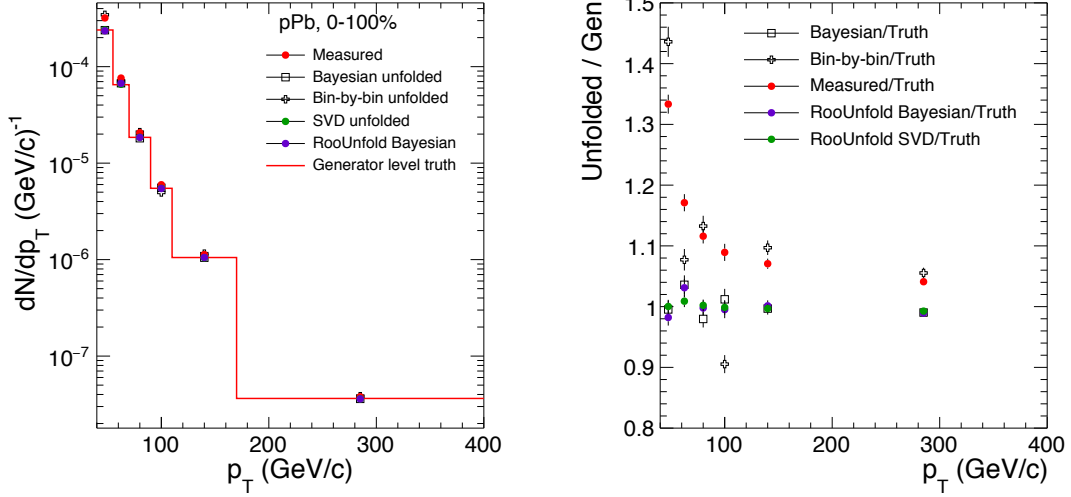


Figure 6.8. The jet p_T spectrum both before and after unfolding using various methods is shown (left), along with a ratio of various unfolded spectra to the simulated “truth” values (right)

be seen in Fig. 6.8 for pPb, where Bayesian unfolding is shown as closed purple points and the SVD unfolding is shown as the closed green points. In addition to these useful algorithms, the closure from the simple bin-by-bin matrix inversion is shown, which has very poor closure. This indicates that these advanced unfolding algorithms are needed to properly remove the resolution uncertainty from any spectrum result.

Due to the non-convergence of the iterative Bayesian unfolding procedure, the SVD unfolding algorithm is preferred by CMS. In addition, with the proper choice of regularization parameter, the SVD unfolding performance is slightly better than that of the Bayesian unfolding.

6.5 Flavored Jet Identification

The main focus of this dissertation is to show results of b-tagged and c-tagged jets. This section describes the procedure for flavor tagging jets at CMS. While jet-by-jet

flavor tagging is not yet implemented, statistical modeling allows for the extraction of the heavy-flavored jet fraction, both of b and c-jets, such that a cross-section can be obtained. From this, R_{AA} and R_{pA} values can be extracted.

6.5.1 b-Jet Tagging

Identifying b-jets in collider experiments relies primarily on the long lifetime of the b-quark. Due to the inability for the b-quark to decay into the top quark because of the huge mass difference between the quarks, the b-quark is forced to decay via a weak, quark generation-changing process, namely $b \rightarrow c$ or $b \rightarrow u$, via the exchange of a W-boson. This process is suppressed by the CKM quark mixing matrix, requiring a charge and parity violating decay, so the typical lifetime of the b-quark is fairly long (relative to particles that decay via strong force exchanges) and is on the order of 10^{-12} s [14]. Assuming a Lorentz factor of ≈ 10 -50 at the LHC, the b-quark decay can be displaced from the primary vertex by at least 300 μm and up to roughly 1 cm, which is easily identifiable at CMS.

The Simple Secondary Vertex Tagger

The primary algorithm used to tag b-jets in these analyses is known as the “simple secondary vertex” (SSV) tagger and takes advantage of the fact that the displaced b-quark will decay into a resolvable secondary vertex. This algorithm is described in detail in Ref. [72], but the idea is that the algorithm calculates a discriminator value for each jet based on the secondary vertex divided by its uncertainty, hereafter called the secondary vertex significance. Tight selections are placed on the secondary vertices such that other compatible particle decays (like that of K_s^0) are rejected. The secondary vertex (SV) selection requires:

- Fewer than 65% of tracks are shared between the primary and secondary vertex
- A SV transverse (2D) flight distance between 0.01 and 2.5 cm

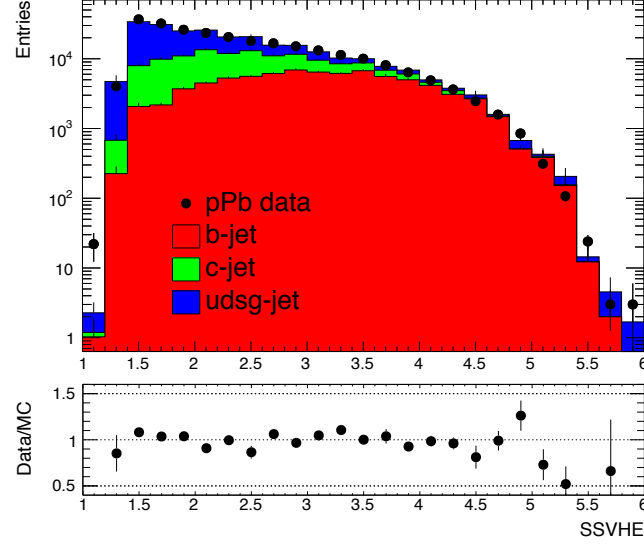


Figure 6.9. Distributions of the SSV discriminator. The black filled points are from data while the filled histograms denote contributions from various jet flavors using a PYTHIA simulation.

- The SV ΔR to the jet is less than 0.5 units in η, ϕ space
- An explicit veto on SV masses $\pm 0.05 \text{ GeV}/c^2$ around the mass of the K_s^0 meson
- A transverse flight distance significance > 3.0
- A SV Mass $< 6.0 \text{ GeV}/c^2$

These selections provide good b-jet SV identification performance while avoiding an overwhelming contamination of the sample of vertices that are from light-flavor decays. In addition, b-jets are allowed to have more than one reconstructed secondary vertex to allow for the case where both the b and subsequent charm decays are reconstructed. Figure 6.9 shows the fraction of various jet flavors as a function of three-dimensional vertex displacement significance. It is clear that at large values of displacement, the sample is dominated by heavy-flavored jets.

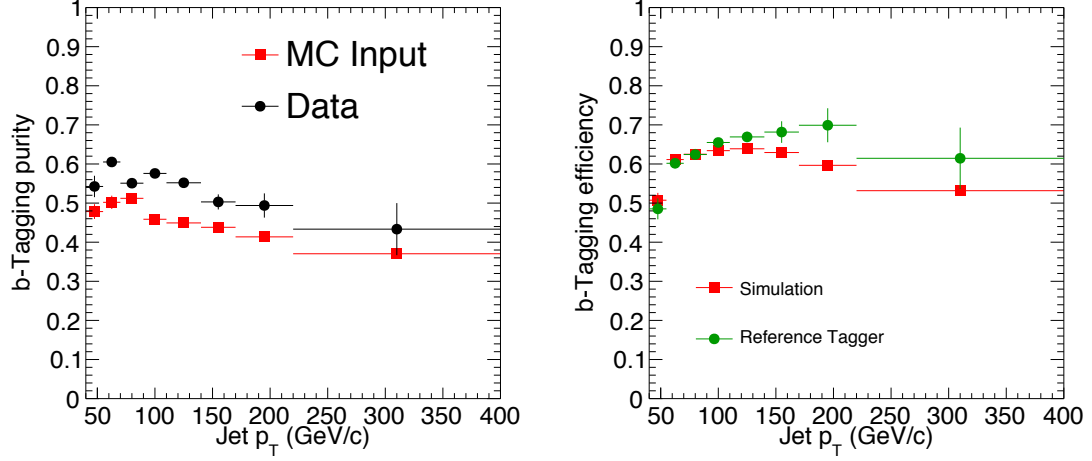


Figure 6.10. Distributions of the SSV tagging performance in pPb collisions. Shown is the tagging purity (left) and efficiency (right) from data and simulation.

Table 6.2.

Modified iterative track reconstruction steps for PbPb regional tracking

Step	Seeds	Max. Transv. Displ. (cm)	Min trk p_T (GeV/c)
0	Pixel Triplets	0.02	1.2
1	Low- p_T Pixel Triplets	0.02	0.4
2	Pixel Pairs	0.015	1.2
3	Detached Pixel Triplets	1.0	1.2
4	Mixed Triplets (Pixel + SiStrip)	0.5	1.0

A working point is chosen on the tagger to ensure that a large fraction of light-flavor jets are rejected while retaining as many heavy flavor jets as possible. An SSV working point selection of 2.0σ provides a 40-50% b-jet tagging purity and a 60% tagging efficiency in pPb and pp (Fig. 6.10) and a 35% tagging purity and 45% efficiency in PbPb collisions (Fig. 6.11).

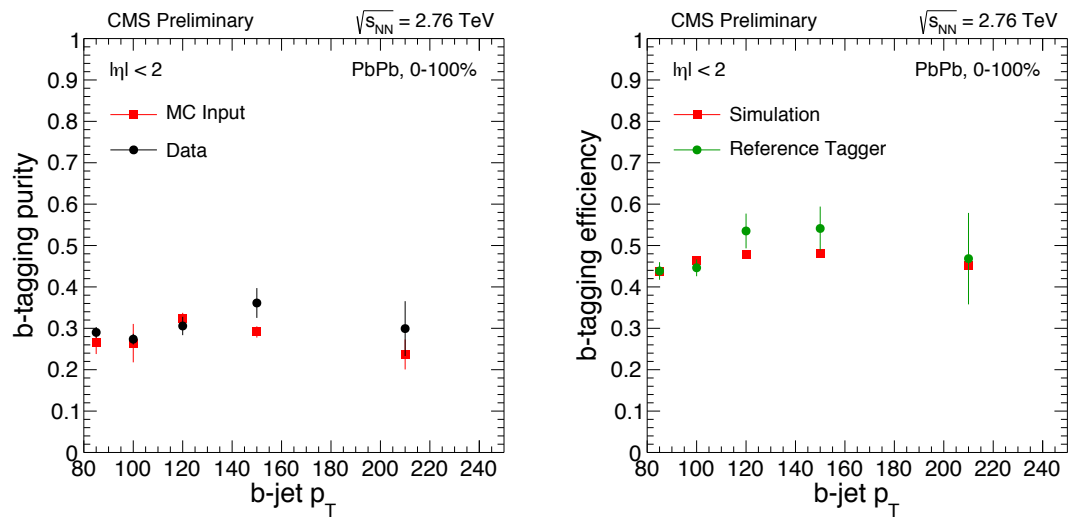


Figure 6.11. Distributions of the SSV tagging performance in PbPb collisions. Shown is the tagging purity (left) and efficiency (right) from data and simulation.

The primary reason that the tagging efficiency is lower in PbPb is due to the difference in displaced track reconstruction. In PbPb, the enormous multiplicity in central PbPb collisions requires massive computing resources to successfully track all particles. Due to the tight memory requirements during online reconstruction, the primary reconstruction is limited to only tracks that are compatible with the primary vertex. As one might imagine, this reconstruction severely limits the ability to reconstruct displaced vertices, so additional reconstruction is needed. A procedure known as “regional tracking” was implemented, where additional reconstruction steps were added to regions corresponding with the cones of high- p_T jets. Using jets with $p_T > 60$ GeV as seed areas, a slightly modified version of the iterative tracking algorithms from Section 6.1 are run, such that displaced tracks and vertices are reconstructed with reasonable efficiency within the jet cones. The modified parameters of the iterative tracking are shown in Table 6.2, where steps 3 and 4 are only run within cones of high- p_T jets. While the limited size of the regional tracking reduces the efficiency of b-jet tagging, the performance remains sufficient such that b-jets can be identified, even in events with very large track multiplicities.

The Jet Probability Tagger

In addition to the SSV tagger, a somewhat orthogonal tagging method is used as a cross-check. This second method is known as the “Jet Probability” (JP) tagger and relies on the displacement of individual tracks from the primary vertex and is independent of any secondary vertex reconstruction. The algorithm works by assigning each jet a discriminator based on a combination of all track displacement values, defined in Eq. (6.9), where each P_i value represents the compatibility of each jet track with the primary vertex and the full product provides a numerical value for how likely all jet tracks are associated to the primary vertex:

$$P_{jet} = \Pi \cdot \sum_{i=0}^{N-1} \frac{(-\ln \Pi)^i}{i!}, \text{ where } \Pi = \prod_{i=1}^N \max(P_i, 0.005) \quad (6.9)$$

The implementation of a minimum compatibility value of 0.5% is a protection against poorly reconstructed tracks severely affecting the jet probability as a whole. The JP tagger then builds a discriminator value for each jet based on $-\ln P_{jet}$. Examples of these distributions are shown in Fig. 6.12 for pPb (left) and PbPb (right) collisions. Note that performance in pPb collisions is virtually identical to that in pp collisions due to the identical reconstruction procedure.

Strict requirements are placed on the reconstructed tracks to ensure the track reconstruction properly constrains the track impact parameter and to disfavor tracks from light-flavored displaced decays. These requirements are the same in all collision species and are imposed on top of the standard tracking selection:

- Two-dimensional track impact parameter (d_{xy}) < 0.2 cm
- Longitudinal track impact parameter (d_z) < 17 cm
- Track $\chi^2 < 5$
- Track $p_T > 1$ GeV/c

In addition, track-to-jet compatibility selections are required:

- Shortest distance between track and jet axis < 0.07 cm
- Decay length < 5 cm
- ΔR to jet < 0.3 units in η, ϕ space

The JP tagger has two distinct advantages over the standard SSV tagger. First, the JP tagger is independent of secondary vertex reconstruction, so the SV reconstruction efficiency can be evaluated and cross-checked through the use of this tagger. In addition, the JP tagger can be calibrated to a data sample using the negative values of the tagger. While each jet track is required to have a minimum track-to-vertex probability of 0.005 during the jet probability calculation, it does not preclude the track from having a negative impact parameter, where the track is displaced on the

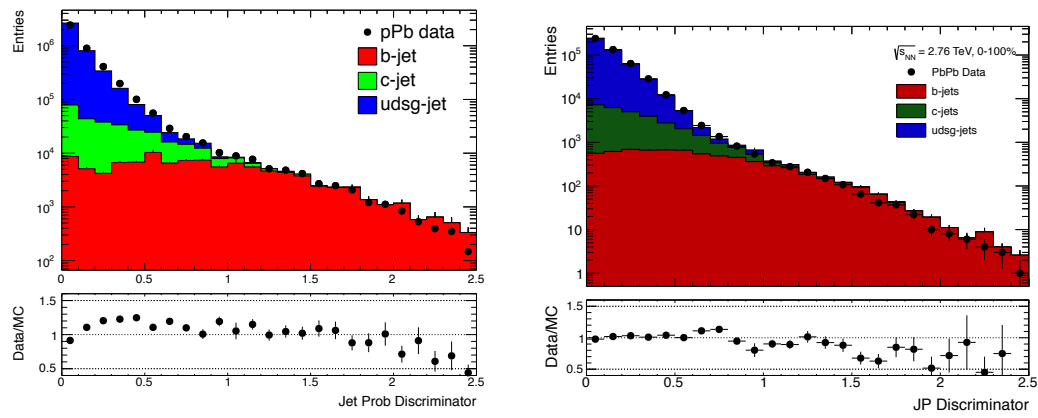


Figure 6.12. Distributions of the JP discriminator values for pPb collisions (left) and for PbPb collisions (right) shown for different jet flavors. The black filled points are from data while the filled histograms are from a PYTHIA simulation.

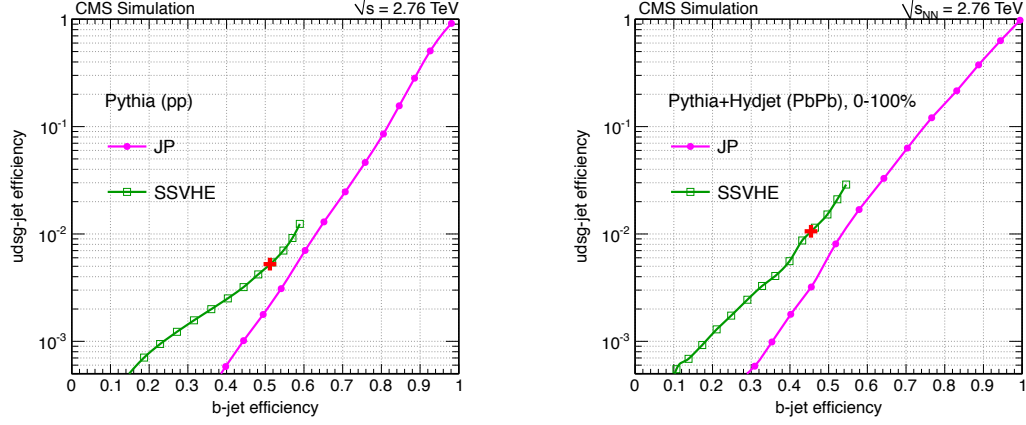


Figure 6.13. Distributions of the b-jet efficiency vs the light jet rejection for both the JP and SSV taggers. Shown is the performance in a pp simulation (left) and a PbPb simulation (right).

opposite side of the primary vertex from the jet. This of course is non-physical and stems from poorly reconstructed or combinatorial track association from tracks that simply happen to be within the jet cone. Because these track-to-jet associations are non-physical, the distribution of negative track impact parameters should be flat. If this distribution is not flat, a weighting factor is applied to each track in order to flatten the distribution, ensuring that the track impact parameters from positively displaced tracks are purely from physical sources, and not due to a miscalibration.

The performance of the two b-taggers (SSV and JP) can be summarized by plotting the b-tagging efficiency against the light and charm jet rejection power from simulation and is shown in Figs. 6.13 and 6.14. As the discriminator selections are loosened, more b-jets are captured, but additional contamination from light and charm jets is also collected. The goal is to find a relatively flat area of the discriminator curve such that b-tagging efficiency is maximized while minimizing the light-jet contamination. While no obvious candidates exist, the red cross shown in the figure

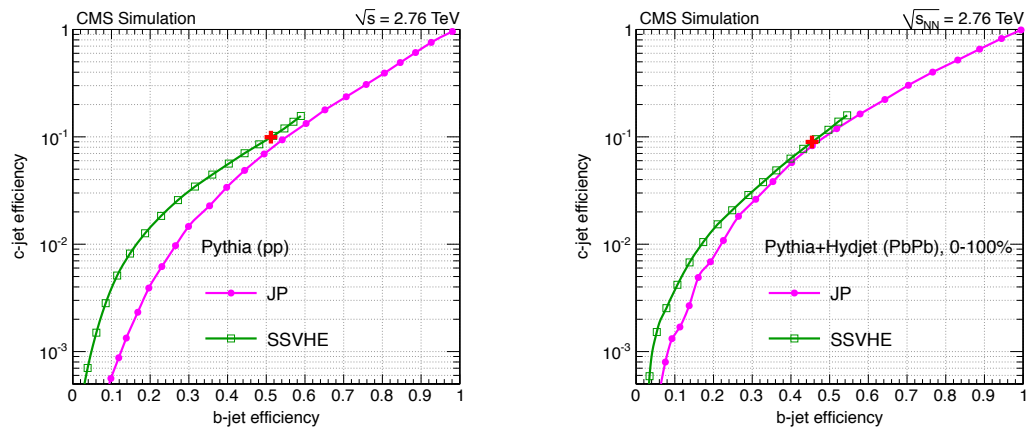


Figure 6.14. Distributions of the b-jet efficiency vs the charm jet rejection for both the JP and SSV taggers. The performance is shown for pp simulation (left) and PbPb simulation (right).

denotes the “flattest” part of the SSV performance curve and is the working point of the tagger used in the analyses.

It is also important to note that based on Figs. 6.13 and 6.14, the SSV tagger seems to perform worse than the JP tagger. It is natural, then, to ask why the SSV tagger is used instead of the JP tagger. While the JP tagger does provide superior b-jet identification performance, its ability to be calibrated independently from simulation is invaluable and essential to cross-check the primary tagger’s performance. In other words, if the JP tagger is used as the primary b-tagging method, additional independent cross-checks are very difficult and systematic uncertainties increase dramatically because the SSV tagger cannot be independently calibrated using data. An additional increase in tagging efficiency of roughly 10 percentage points is negligible compared to the decrease in systematic uncertainty that the JP tagger can provide, especially considering that the tagging efficiency and purity is corrected in the analysis.

After tagging, the next step in the analysis procedure is to extract the b-jet cross-section. This is accomplished by first applying the working point selection on the SSV tagger ($SSV > 2.0$) and then creating template fits to the secondary vertex mass, as seen in Fig. 6.15. These template fits are created using an unbinned maximum likelihood fitting method, as implemented by the ROOFIT package [73]. Each jet p_T bin is fit independently such that a total spectrum can be built bin-by-bin, where the shape of each flavor contribution is estimated in simulation, and the flavor yields are allowed to float up and down independently such that the overall shape in data is matched as closely as possible. By selecting only jets that have a SSV discriminator value larger than 2.0, the b-jet contribution dominates the sample (especially at large SV mass values) and drives the fit. Once the fits are completed, it is a simple matter to read off the purity value in each jet p_T bin by dividing the integral of the b-jet component of the fit by the total data integral.

To calculate the tagging efficiency, additional fits are used. Template fits are made of distributions of the JP tagger discriminator, both before and after the $SSV > 2.0$

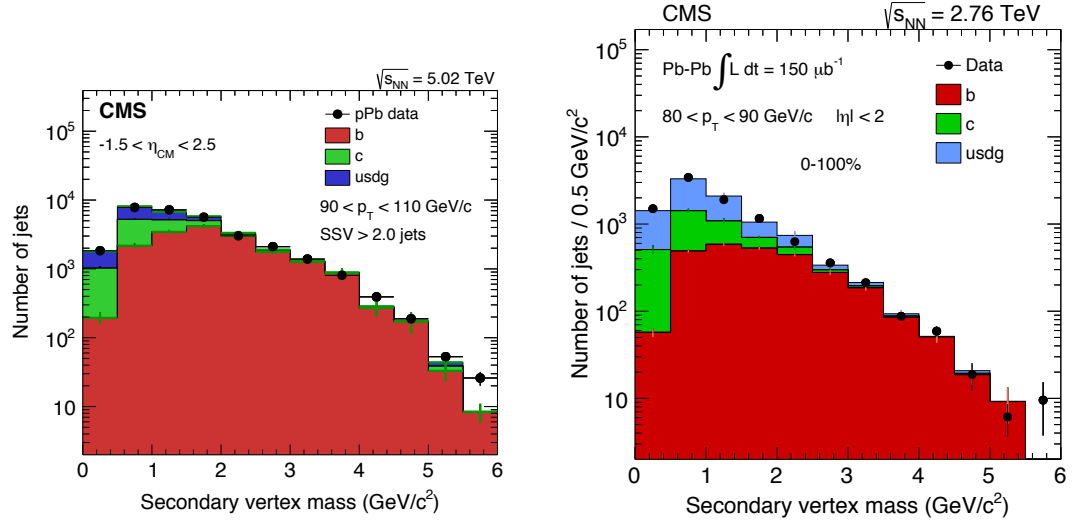


Figure 6.15. Template fit of the SV mass distributions for data (filled black points), and in MC from the three flavor components of light jets (filled blue histogram), charm jets (filled green histogram) and bottom jets (red) in pPb collisions (left) and PbPb collisions (right).

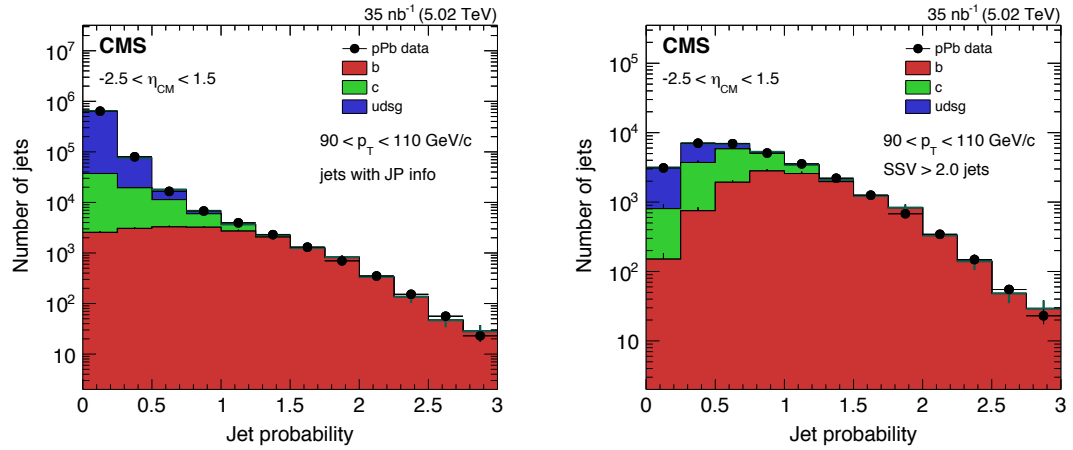


Figure 6.16. Template fits of the JP discriminator both before (left) and after (right) applying the SSV tagger. Contributions from b-tagged jets (red), charm jets (green), and light jets (blue) are shown against a distribution in pPb data (closed black points)

tagging selection. Again, the contributions from the three flavor categories are fit to the data and the efficiency value is extracted using Eq. (6.10), where f_b is the purity value as derived from the template fit to the secondary vertex mass and C_b is the fraction of jets that have a JP discriminator value.

$$\epsilon = \frac{C_b f_b^{\text{tagged}} N_{\text{jets}}^{\text{tagged}}}{f_b^{\text{untagged}} N_{\text{jets}}^{\text{untagged}}} \quad (6.10)$$

In principle, a jet could have no tracks associated with it (only calorimeter hits) and a JP discriminator would not be calculated, though this is extremely unlikely. In practice, the C_b fraction is greater than 99.9%. Example distributions of template fits of the JP tagger discriminator in pPb collisions before and after SSV tagging are shown in Fig. 6.16. Similarly to the secondary vertex mass fits, all flavors contribute significantly to the fit, but the b-jet contribution dominates the sample above a JP discriminator value of roughly 1.5, allowing a robust fit to the contributions from all three flavors.

Once the tagging purity and efficiency are calculated for each jet p_T bin, the total number of b-jets can be found by the simple ratio $N_b = N_b^{\text{tagged}} f_b / \epsilon$, where N_b^{tagged} is the number of b-tagged jets, f_b is the tagging purity and ϵ is the tagging efficiency.

6.5.2 c-Jet Tagging

In addition to verifying the b-tagging performance in heavy-ion collisions, I have developed the first successful charm jet-tagging algorithm for use in a heavy-ion environment. As previously mentioned, charm-flavored objects have a much shorter lifetime than beauty-flavored objects, such that the vertex resolution required to identify these objects is only possible with the most advanced tracking systems, like those implemented at the primary LHC experiments. In addition, the kinematics of charm jets is very similar to light jets, and as the charm-jet fraction is only on the order of 10% of an inclusive-jet sample, the separation of these components is generally very challenging. Nevertheless, there are subtle differences between light and charm jets that can be exploited.

This new c-tagging algorithm uses a subtle variation of the existing b-tagging framework in order to minimize contributions from light-jets and separate b-jets from charm-tagged jets as much as possible. There are two primary components. First, a modified version of the SSV tagger is used to tag the displaced vertices from heavy-flavor decays. This SSV “high-purity” algorithm requires the same selections as the original “high-efficiency” version with the additional constraint that the vertex must have at least three associated tracks, rather than just two, as was the case for b-tagging. The addition of the third track seems fairly trivial at first glance, but the overwhelming majority of two-track displaced vertices are not heavy-flavored. Even at the same tagger working point, this simple additional requirement improves the light jet rejection of the sample from a factor of 100 to a factor of 300. This additional rejection is crucial to accurately extract the relative contribution of charm-tagged jets to a sample, since the charm jet and light jet kinematics are so similar. The difference in performance between the high-efficiency (2+ track) SSV and the high-purity (3+ track) SSV algorithms can be seen in Fig. 6.17. The left panel shows the performance of c-tagging relative to b-tagging, while the right panel shows the performance improvement of c-tagging relative to light-jet rejection. The figures indicate that the additional requirement of a third track in the vertex leaves the b-tagging and c-tagging performance essentially unmodified relative to each other, while the high-purity selection improves the light jet rejection.

The second difference between the existing b-tagging and the new charm-tagging algorithms is the use of a modified version of the secondary vertex mass. While b-tagging relies on the basic secondary vertex mass, that is, the relativistic difference of jet energy (from the calorimeter) and jet momentum (from the vector sum of particle momenta), charm-tagging relies on the so-called corrected secondary vertex mass. The definition of this quantity is:

$$M_{corr} = \sqrt{M^2 + p^2 \sin^2 \theta} + p \sin \theta \quad (6.11)$$

where M and p denote the total invariant mass and momentum of the constituent particles associated to the vertex and θ is the angle between the vertex displacement

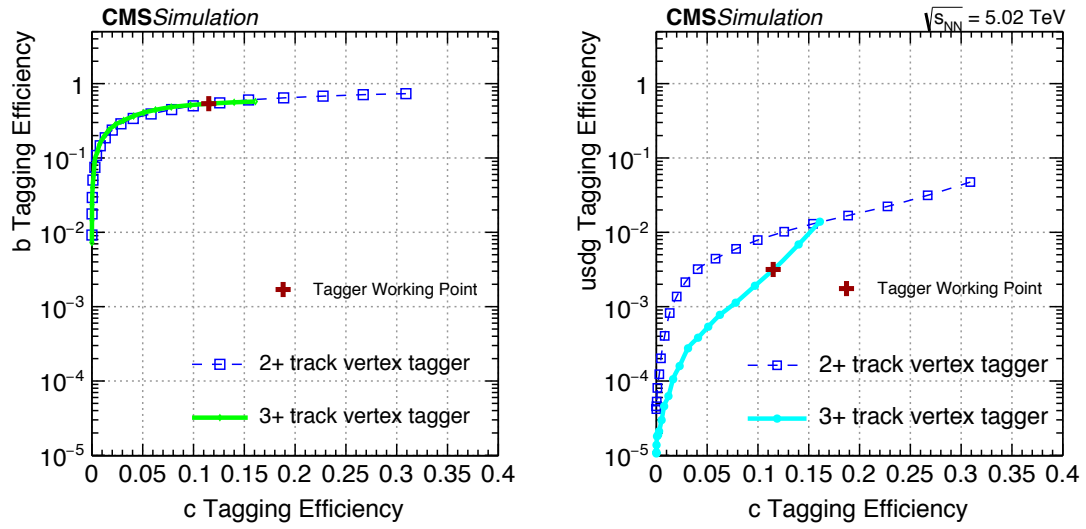


Figure 6.17. Distributions of the charm-jet tagging efficiency vs the b-tagging efficiency (left) and the light-jet tagging efficiency (right), comparing the standard SSV 2+ track algorithm and the high-purity 3+ track version.

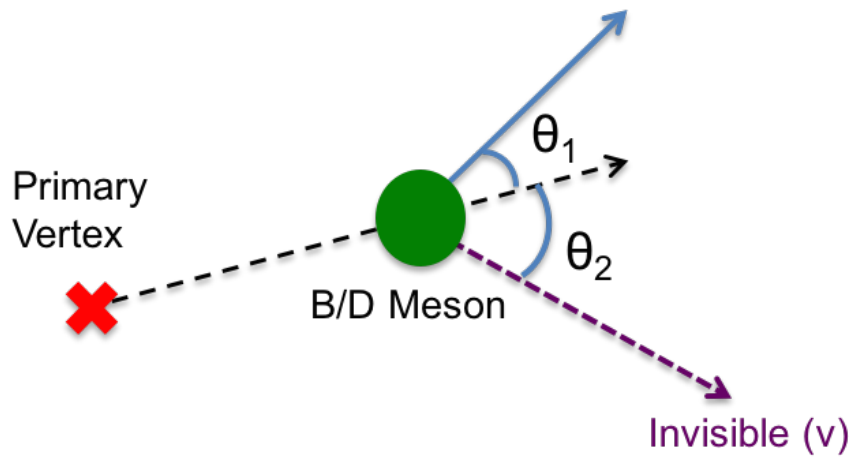


Figure 6.18. Cartoon showing an example of the usefulness of the corrected secondary vertex mass calculation in a typical heavy-flavored decay.

vector and the vector sum of the momenta of the particles associated to the vertex. This quantity was first developed as a tool for identifying b-jets by experiments at the LEP and SLAC colliders [74] and is also used by the LHCb Collaboration [75]. Figure 6.18 describes the motivation behind this correction. The idea is that heavy-flavored objects tend to decay into invisible or other easily missed particles due to a combination of the tracking performance of significantly displaced tracks and the branching ratios of heavy-flavored objects into non-reconstructible objects like neutrinos. In addition, while neutral mesons like π^0 particles can be reconstructed in the calorimeter, the position uncertainty of a neutral particle means that it is impossible for it to be associated to a particular vertex and is therefore not used when calculating the classic secondary vertex mass. The correction also allows for these non-associated particles to be accounted for.

The assumption behind the procedure is that of momentum conservation, where any decay is expected to conserve its four-momentum. When it does not, we invoke Occam’s razor and assume that a particle was not reconstructed, rather than assume a breaking of momentum conservation. Due to the presence of 3+ body decays, it is impossible to directly calculate the total missing energy, rather Eq. (6.11) simply adds the minimum possible missing energy that restores the conservation of four-momentum in the decay process.

Once the c-tagging is applied, the extraction of the c-jet cross-section continues in an analogous way to that used in b-jet tagging. Template fits are again created for each jet p_T bin, though instead of using the secondary vertex mass, the corrected secondary vertex mass is used. In addition, instead of applying a tagging selection of $SSV > 2.0$, a selection is made on the high-purity $SSV > 1.68$, which is chosen in order to maximize the c-jet purity, as seen in Fig. 6.19. The application of the high-purity SSV tagger further reduces the contribution of light-jets to the sample, and the use of the corrected SV mass increases the kinematic separation of the charm and bottom template shapes. Figure 6.20 shows a comparison of the template fits of secondary vertex mass (left) and corrected secondary vertex mass (right). The difference is fairly

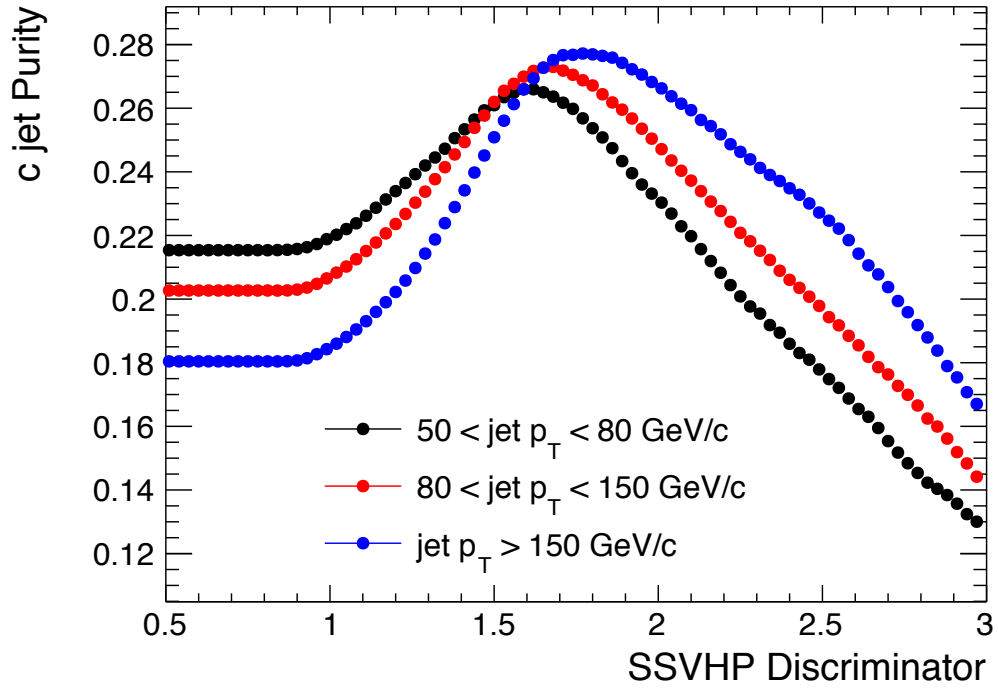


Figure 6.19. The c-jet purity is shown as a function of the high-purity SSV selection for three selections of jet p_T . The working point is chosen at 1.68 such that the c-jet purity is maximized for a wide range in jet p_T .

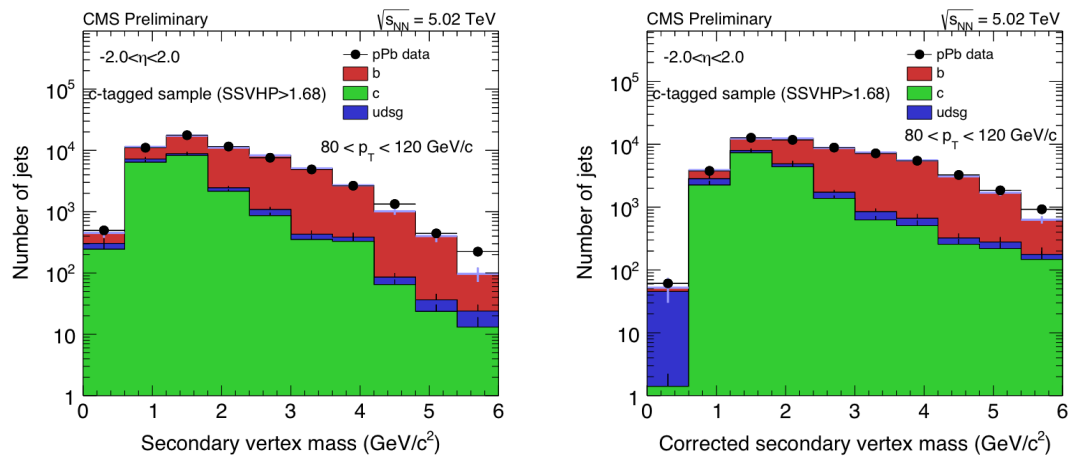


Figure 6.20. Template fits of secondary vertex mass (left) and corrected secondary vertex mass (right) are shown for pPb collisions

subtle, but the b-jet contribution to the corrected SV mass template is shifted toward higher values of corrected SV mass. This additional difference in template shape between b-jets and c-jets is enough to properly constrain the fit with reasonable precision. From these, the c-jet purity can be calculated by simply dividing the integral of the c-jet contribution to the template by the integral of the data (after the SSV high-purity selection). After the purity is calculated, the efficiency calculation is obtained in exactly the same fashion as in the b-jet analysis, discussed in the previous section.

7. MODIFICATION OF B-TAGGED JETS

This chapter describes the b-tagged jet results in all three collision systems of interest: PbPb, pPb, and pp, at 2.76 TeV and 5.02 TeV. Additional details can be found in the published papers for the PbPb and pp collision systems at 2.76 TeV [76] and for the pPb collision system at 5.02 TeV [43].

7.1 Results in PbPb (2.76 TeV)

Identification of b-tagged jets in PbPb is prioritized in order to measure the heavy-flavored jet energy loss in the hot and dense QGP medium. Two types of results are considered: the first, as a function of jet p_T , and the second as a function of collision centrality. Figure 7.1 shows the b-jet spectrum as a function of p_T for various centrality classes from PbPb and pp data at 2.76 TeV. We observe a significant b-jet energy suppression in PbPb with respect to proton-proton collisions after scaling the PbPb cross-section by the effective Glauber nuclear overlap factor. An extra scaling factor T_{AA} is applied to the PbPb spectrum that accounts for the increase in jet production just from the increase in the number of nucleon-nucleon collisions per bunch crossing relative to proton-proton collisions. This factor is defined as:

$$T_{AA}(\text{cent}) = \langle N_{coll} \rangle(\text{cent}) / \sigma_{MB}^{pp}, \quad (7.1)$$

where σ_{MB} denotes the effective minimum-bias cross section of a single scattering event and $\langle N_{coll} \rangle(\text{cent})$ is the average number of binary collisions as a function of centrality.

From the spectra in Fig. 7.1, it is immediately clear that there is significant suppression of b-tagged jets at all measured values of jet p_T , indicating that suppression effects are large even for jets that punch through the medium at very high momenta.

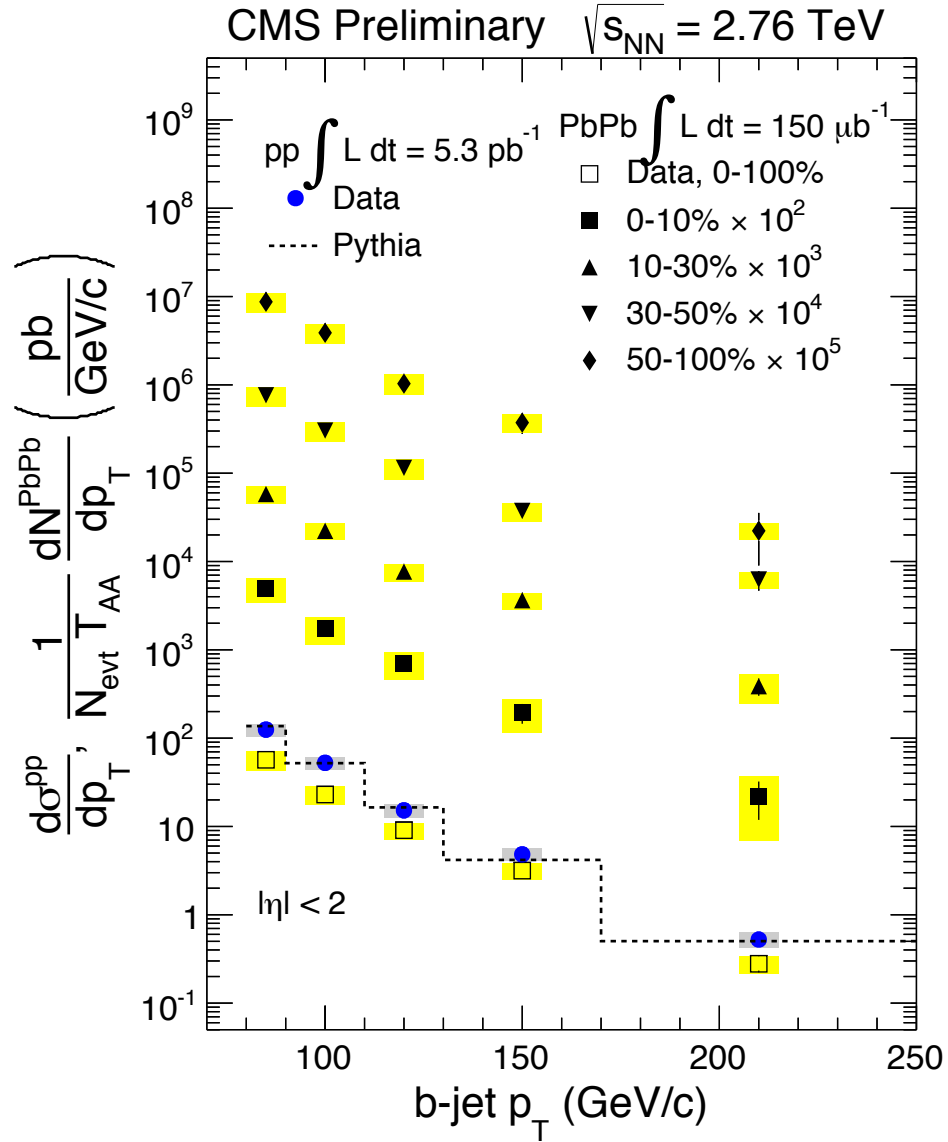


Figure 7.1. Cross-section for b-tagged jets in PbPb (scaled by T_{AA}) for various centrality classes, along with the cross section in pp at 2.76 TeV. Results are also compared to a PYTHIA prediction.

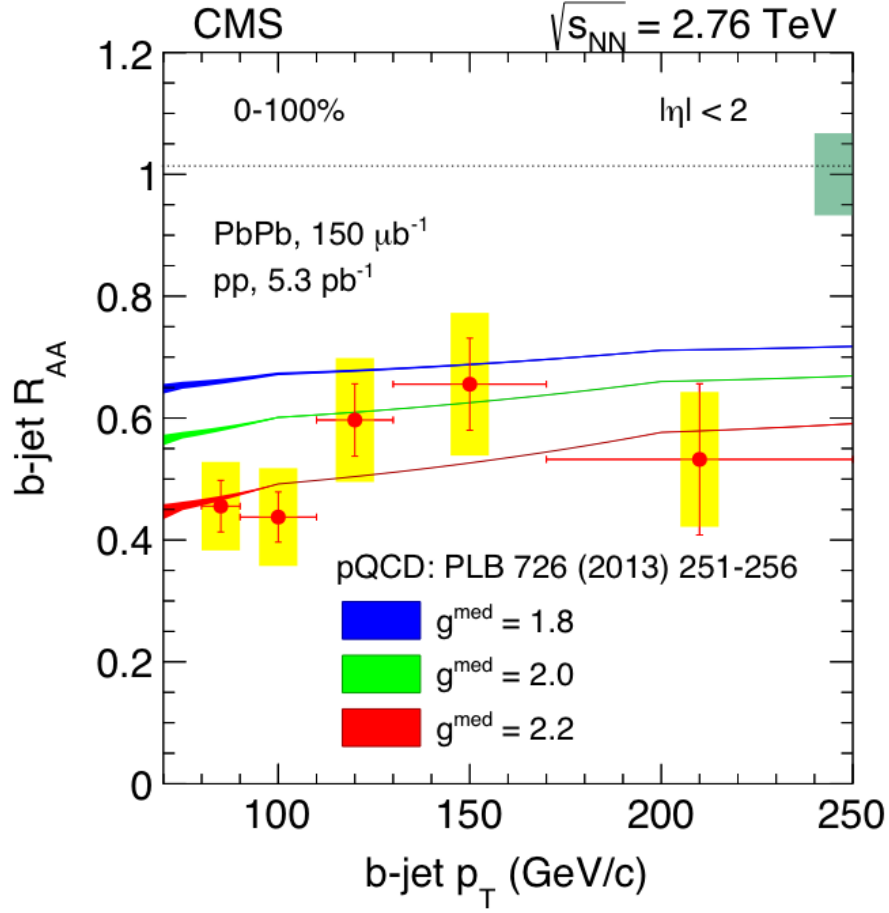


Figure 7.2. The b-Jet R_{AA} is plotted as a function of jet p_T , and is compared to a theoretical prediction based on pQCD with various gluon-to-medium coupling strengths.

The suppression factor can be numerically calculated by finding the b-jet R_{AA} , which is the ratio of the b-jet cross-section in PbPb to that in pp, scaled by T_{AA} , as discussed in the introduction. The b-jet R_{AA} can be seen in Fig. 7.2, where the value hovers around 0.5 for all jet p_T , for the inclusive centrality bin.

The amount of suppression is large, but is consistent with a pQCD prediction that includes both radiative and collisional energy loss [77]. Furthermore, the gluon-to-

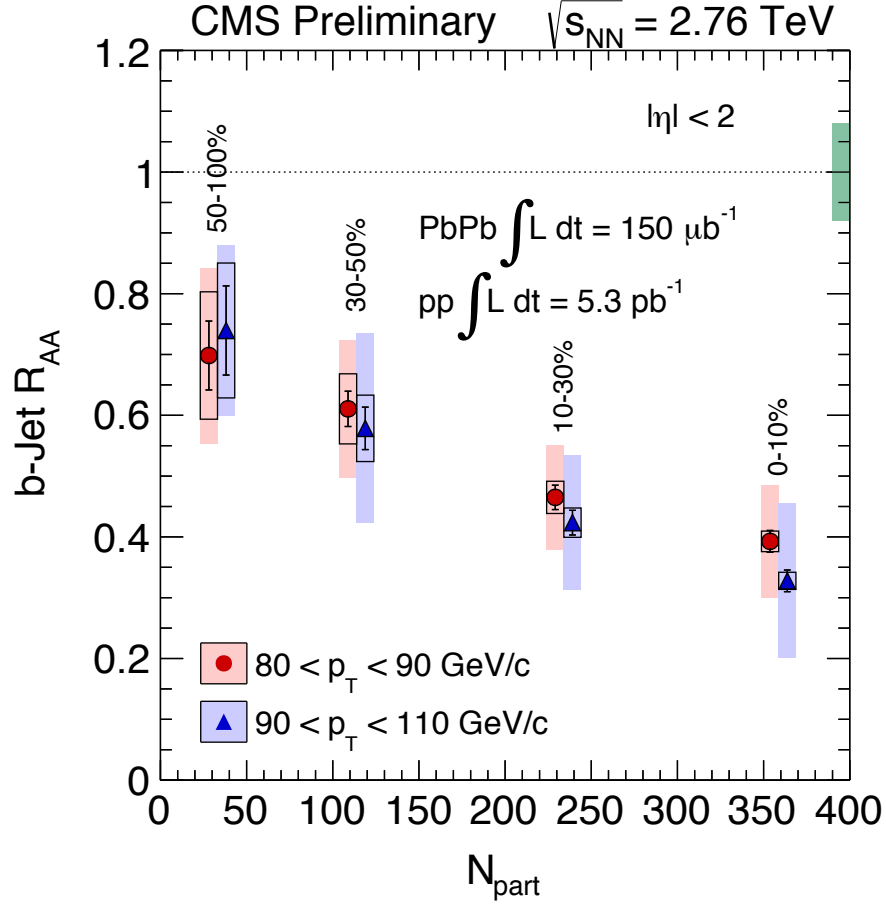


Figure 7.3. The b-Jet R_{AA} is plotted as a function of centrality for two jet p_T bins where a clear trend can be seen, indicating that b-jets are more suppressed in more central events

medium coupling strength is taken as a free parameter in the model, and the data seems to favor a coupling strength on the stronger side of the pQCD expectation.

Additional conclusions can be drawn by observing the result as a function of centrality. Figure 7.3 shows the relative b-jet suppression as a function of centrality, where the horizontal axis moves from peripheral to central events. The N_{part} used in the figure's horizontal axis refers to the number of nucleon participants, as estimated

via the Glauber model. The number of participants ranges from 2 to $2A$ ($= 416$ for PbPb) such that more central events correspond to a higher number of participants. The figure indicates that a centrality ordering is present, where jets produced in more central events are clearly more suppressed than those produced in peripheral collisions. This result confirms the expectation of a QGP medium-induced effect that influences jet production, even at high- p_T .

It is important to make the distinction that jets are not being lost in the medium, rather their energy is just modified. In other words, the total number of jets is not significantly changed, and therefore the average jet yield is roughly the same (after T_{AA} scaling). In this assumption and by assuming that the jet momentum spectrum falls as a function of $p_T^{-5.5}$, a b-jet R_{AA} value of 0.5 indicates that the jets lose around 12-15% of their energy, shifting the entire jet spectrum backwards along the horizontal axis. Due to the steepness of the jet cross-section as a function of jet p_T , a relatively modest energy loss in terms of percentage can lead to a large effective suppression value when considering jet R_{AA} .

Finally, one can consider the flavor dependence of the observed energy loss by comparing the b-jet R_{AA} to that of inclusive jets. Figure 7.4 shows the comparison of the inclusive-jet R_{AA} to the b-jet R_{AA} . Within the experimental uncertainties, we observe no difference between the light and heavy jets, indicating that mass-dependent effects are negligible at very high energies. This observation tends to confirm predictions from pQCD models that suggest that energy loss ought to be similar for jet mass values much greater than the bare b-quark mass ($\approx 5 \text{ GeV}/c^2$) [77, 78].

Systematic Uncertainties

There are a number of systematic uncertainties that factor into this measurement, though they fall into three general categories:

- b-Tagging stability

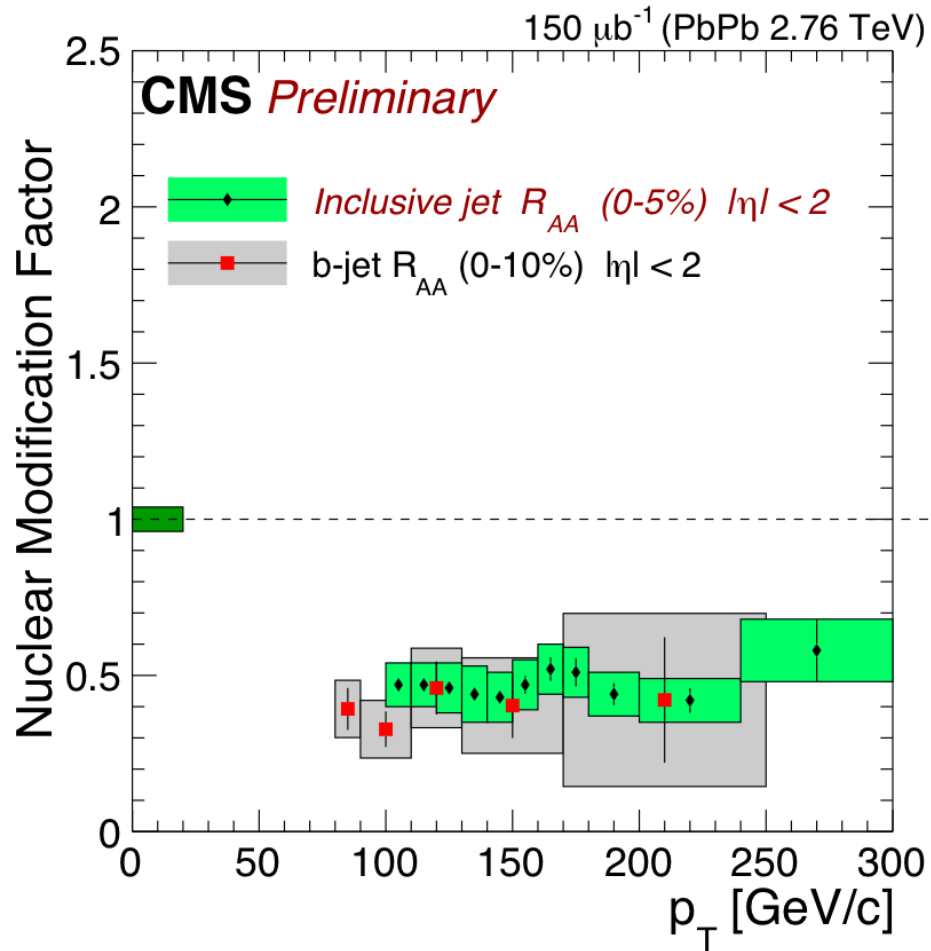


Figure 7.4. The b-Jet and inclusive jet R_{AA} as a function of jet p_T are shown. The consistency between the two results is indicative that flavor-dependent effects are very small at high energies.

- Jet Reconstruction
- Scaling Uncertainties

The b-tagging stability uncertainties deal primarily with the constraint that by template fitting, we require the Monte Carlo to precisely estimate the distributions of jet kinematic quantities. This of course is impossible, so we vary the shapes of the different flavor contributions in various ways to test the stability of the template fits. The first way this is tested is by comparing the b-tagging efficiency from simulation and the tagging efficiency calculated from using the JP tagger, as described in Eq. (6.10). The difference between these two efficiency calculations accounts for about 50% of the total systematic uncertainty at high- p_T . Next, the working point of the SSV tagger is varied such that the tagging purity shifts up and down by about 10%. This accounts for nearly the maximum variation of tagging purity from a secondary vertex-based tagging method such that all uncertainty from the tagger working point is accounted for. Another cross-check is made by fixing the charm jet normalization to the light jet normalization and only allowing the b-jet and charm+light jet contributions to float independently during the fitting procedure. This uncertainty accounts for the sensitivity of the results to the non-b-tagged jet template shapes and is a relatively small part of the total uncertainty for all jet p_T values.

An additional source of uncertainty stems from the limited Monte Carlo statistics used in the analysis. Because computing time and disk resources are limited, it is obviously not possible to produce infinitely precise simulations. To account for this, the contribution to the template fits from each flavor is varied within their statistical uncertainties. The template fits are run hundreds of times and gaussian fits are made to the distribution of the b-jet purity extracted from each template fit. The width of the gaussian fit for each jet p_T bin is assigned as a systematic uncertainty. An example of these gaussian fits is shown in Fig. 7.5 for two jet p_T bins.

The last uncertainty stemming from the b-tagging procedure has to do with the uncertainty of the b-jet production types. As discussed in the Theoretical Basis

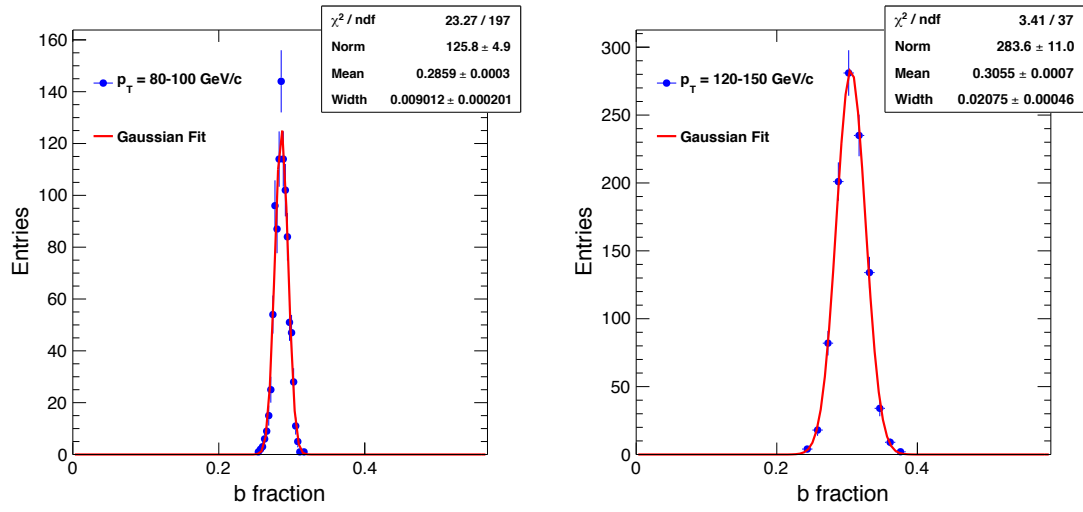


Figure 7.5. The gaussian fits to the systematic variation of the b-jet purity in PbPb at 2.76 TeV derived from MC statistical uncertainty in jet p_T bins from 80-100 GeV/c (left) and 120-150 GeV/c (right)

section of this dissertation (Chapter 2), the evaluation of b-jet energy can be contaminated by a next-to-leading order b-jet production method known as gluon splitting. In this type of production, the b and $\bar{b}$ quarks are highly collimated and are likely to be reconstructed as a single jet. While this first b-tagging result does not distinguish between production types, there has been an attempt to assess the contribution of this production process to the results. In order to mitigate the effect of counting the heavy quark suppression twice, the gluon splitting fraction from the PYTHIA simulation has been varied up and down by 50%. The overall effect of this variation is the smallest uncertainty in the analysis, so the results are not expected to be sensitive to the b-jet production method.

The final uncertainty contribution is only applied in pp collisions and stems from the uncertainty due to pile-up removal. Pile-up is defined as multiple hard-scattering events that occur in a single bunch crossing and can dramatically affect a cross-section measurement if not properly accounted for. By selecting on only good collisions that have at least one vertex with a tight requirement on the longitudinal beam position ($|v_z| < 15$ cm), this effect can be mitigated. The filter does have a minor uncertainty in the calculation of its efficiency, accounting for a very small contribution to the b-tagging uncertainty at high p_T . This uncertainty is not applied in PbPb because the removal of pile-up in a heavy-ion environment is such a monumental task, that the Pb ion beams in the LHC were collided with a lower instantaneous luminosity such that pile up events were very rare.

The total b-tagging uncertainty is shown for both PbPb and pp data in Figure 7.6, where the contribution from all previously discussed sources are added in quadrature. The total systematic uncertainty ranges from about 18% at low p_T to about 15% at high p_T in PbPb, whereas the uncertainty is roughly 12% at low p_T and almost 20% at high p_T in pp.

After b-tagging, additional uncertainties are applied due to jet reconstruction uncertainty, along with scaling uncertainties from the luminosity and Glauber nuclear overlap function uncertainties. The jet reconstruction uncertainty can be factored

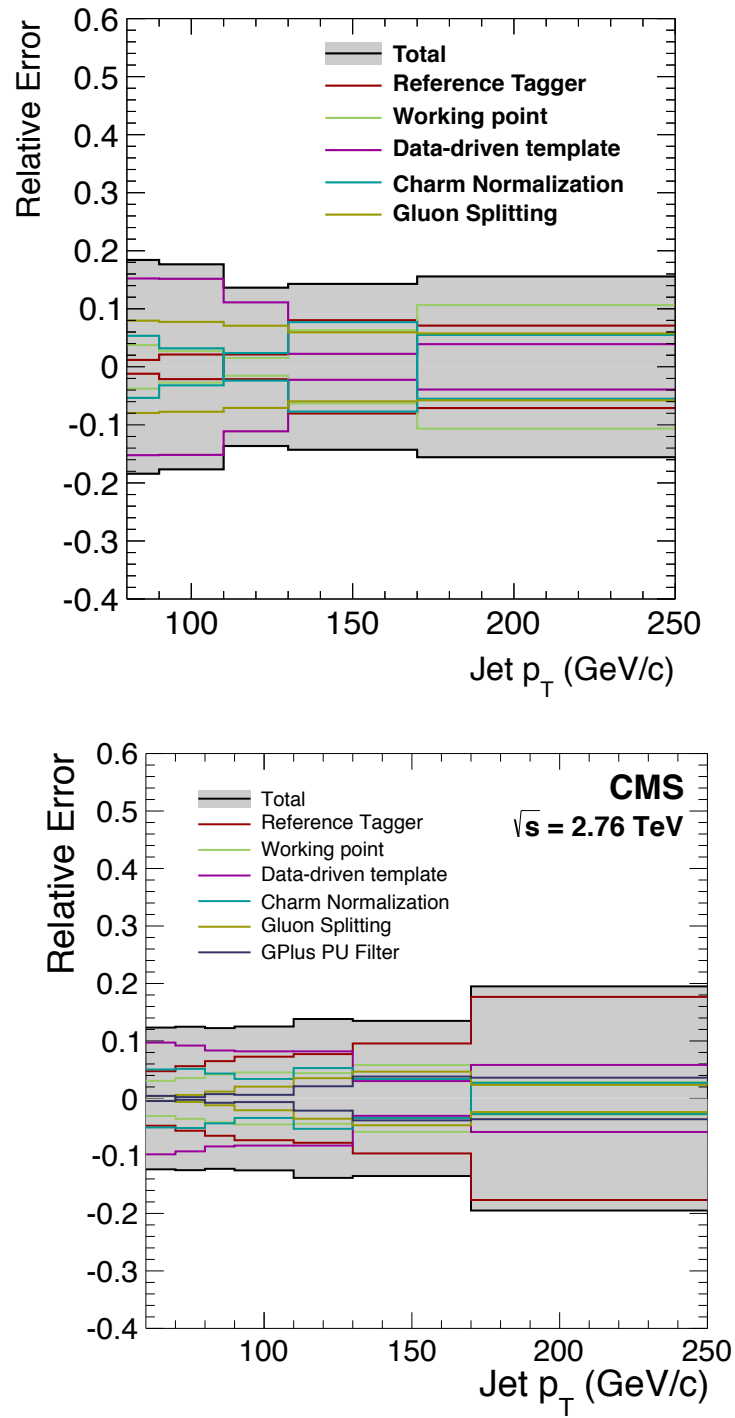


Figure 7.6. The contribution to the total systematic uncertainty for PbPb (top) and pp (bottom) is shown as a function of jet p_T .

into two pieces based on uncertainties in the jet energy resolution and the jet energy scale, where the resolution accounts for the jet energy smearing based on the detector response to the jet and the jet energy scale accounts for an overall normalization relative to a generator-level jet, both as a byproduct of the nonlinear HCAL energy response. These uncertainties are tested by an unfolding procedure, discussed in Section 6.4.

While the unfolding removes the estimated jet uncertainties from the reconstruction procedure, these uncertainties are removed only relative to the accuracy of the Monte Carlo generation. To account for discrepancies in the MC simulation of the jet momentum spectrum shape, the yield of the generator-level spectrum is shifted by 50% at 80 and 250 GeV/c in opposite directions. While the jet yield in MC certainly changes, the template shapes are not affected dramatically, such that the result in data is virtually unchanged. The total result variation from this test is around 1-2%, depending on collision centrality. In addition, the jet resolution is also parameterized and allowed to vary randomly within its uncertainty (approximately 15%, depending on jet p_T). This resolution smearing is correlated between the pp and PbPb results such that the R_{AA} uncertainty is smaller than either of the single-jet spectra uncertainties.

Finally, the unfolding itself is assigned an uncertainty by testing the stability of the Bayesian unfolding procedure by varying the number of iterations. A value for this uncertainty is obtained by finding the ratio of the spectrum at the third, fifth, and sixth iterations to the fourth iteration, where the maximum discrepancy is used as the total uncertainty. This uncertainty, as well as uncertainties in the jet energy scale, resolution and underlying event estimation are shown in Fig. 7.7, where PbPb is shown on the left and pp is shown on the right. The partial cancellation of these systematics for the final R_{AA} uncertainties is shown in Fig. 7.8.

All these uncertainties are added in quadrature to obtain the final systematic uncertainty values used in the analysis. Tables 7.1 and 7.2 show the contributions from

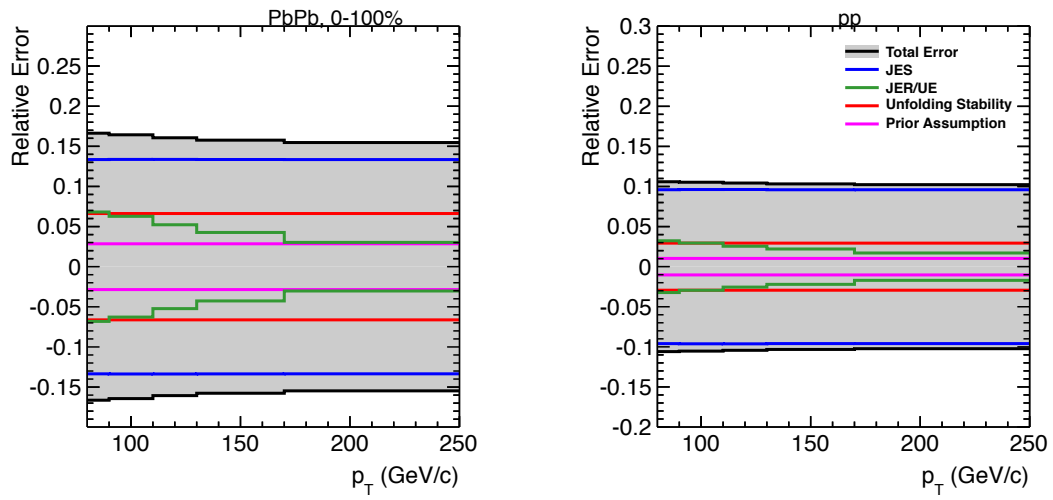


Figure 7.7. The contribution to the jet reconstruction systematic uncertainties for PbPb (left) and pp (right) is shown as a function of jet p_T .

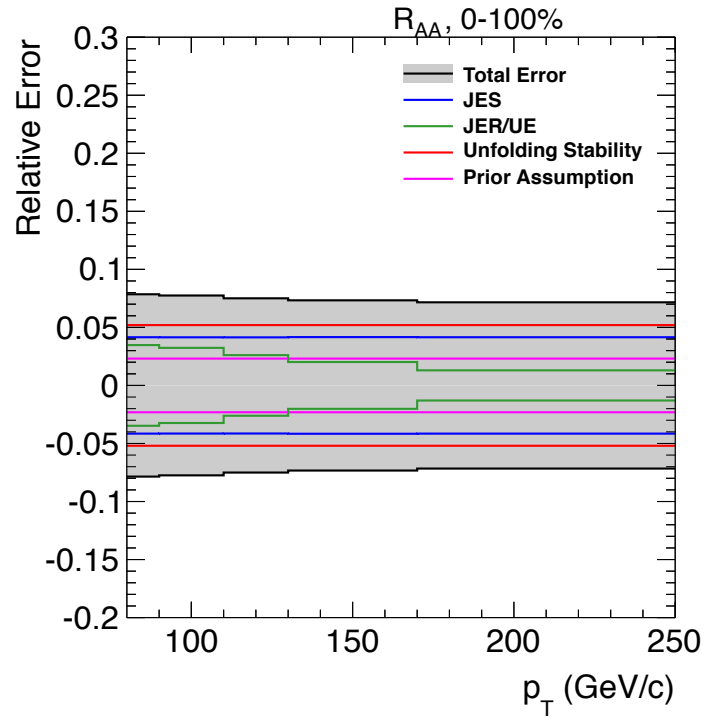


Figure 7.8. The contribution to the total systematic uncertainty from jet reconstruction for the R_{AA} measurement after partial cancellation of pp and PbPb uncertainties

Table 7.1.
Major contributors to the PbPb b-jet spectrum systematic uncertainty at 2.76 TeV for three jet p_T bins of inclusive centrality.

Source	80 - 90 GeV/c	110 - 130 GeV/c	170 - 250 GeV/c
Jet Energy Scale	13%	13%	13%
Jet Energy Resolution	7%	5%	3%
Unfolding Stability	7%	7%	7%
b-Tag Efficiency	1%	3%	7%
Tagger Work. Pt. Variation	4%	2%	11%
Data-Driven Template	15%	12%	4%
Fix Light/Charm ratio	6%	3%	6%
Gluon Splitting	8%	7%	6%
Total	24%	21%	22%

all sources for a selection of jet p_T bins for PbPb and pp collisions, respectively, rounded to the nearest percentage.

7.2 Results in pPb (5.02 TeV)

To complement the clear jet suppression effects seen in PbPb collisions, an additional study of the b-jet cross-section was performed using pPb collisions. Unfortunately, the center of mass energy is different between the pPb and PbPb datasets, so direct comparisons cannot be made, but generalizations regarding the relative size of various effects can be stated. In addition, when published, the pPb data had no complementary proton-proton reference data at the same center-of-mass energy, so all comparisons that rely on a pp reference are from a PYTHIA simulation. Although uncertainties in the simulation are accounted for, this is not as robust as a true nuclear modification measurement. To clarify this point, all jet energy modification results

Source	80 - 90 GeV/c	110 - 130 GeV/c	170 - 250 GeV/c
Jet Energy Scale	9%	9%	9%
Jet Energy Resolution	3%	3%	2%
Unfolding Stability	3%	3%	3%
b-Tag Efficiency	5%	3%	7%
Tagger Work. Pt. Variation	7%	8%	18%
Data-Driven Template	8%	8%	6%
Fix Light/Charm ratio	4%	5%	3%
Gluon Splitting	<1%	4%	3%
Total	16%	17%	23%

Table 7.2.

Major contributors to the pp b-jet spectrum systematic uncertainty for three jet p_T bins at 2.76 TeV.

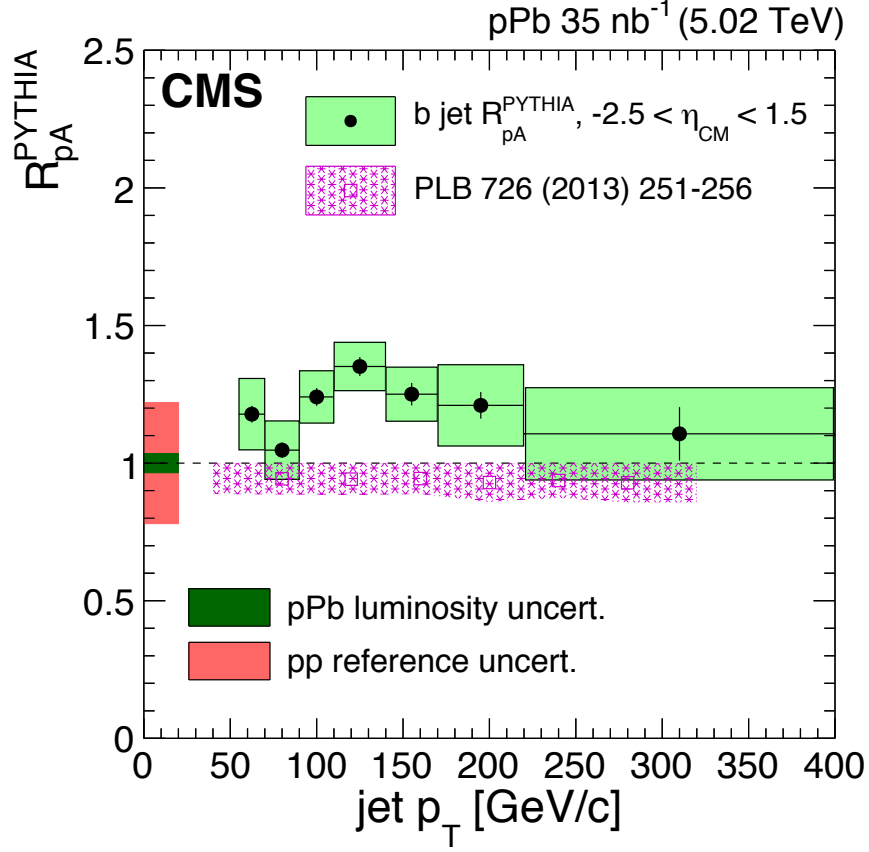


Figure 7.9. The R_{pA}^{PYTHIA} as a function of jet p_T for b-jets in pPb collisions at 5.02 TeV. A pQCD prediction from Huang et. al. [77] is also shown.

are denoted R_{pA}^{PYTHIA} , instead of simply R_{pA} , as would be the case with a true pp reference from data.

One of the primary motivations for the collection of a pPb dataset was to investigate the initial-state jet effects just due to the presence of a nucleus in the collision. It is not expected that QGP is created in pPb collisions¹, so any jet modification seen is expected to correspond to the presence of initial-state scattering. These ini-

¹It should be noted that this point is still highly controversial, thanks to the success of hydrodynamic models in describing pPb data, as well as a number of puzzling experimental results indicating collective behavior. See Ref. [79] for a comprehensive review.

tial state effects can lead to a R_{pA} value larger than unity due to multiple scattering and effects from the nuclear parton distribution functions. While an R_{pA} value significantly below unity might indicate that final state suppression effects are present, the more likely scenario is that some additional initial state effects contribute to jet suppression, and the “true” suppression values observed in PbPb collisions must take such effects into account.

With that in mind, a few interesting conclusions can be drawn from Fig. 7.9, which shows the R_{pA}^{PYTHIA} as a function of jet p_T . First, the R_{pA}^{PYTHIA} value is above unity, indicating that initial-state jet energy suppression effects are relatively minor. Fitting a constant to the pPb data gives a value of 1.22 ± 0.15 (stat.+syst.) ± 0.27 (syst. PYTHIA), indicating that the results are consistent with PYTHIA within the uncertainties. If we momentarily assume the PYTHIA cross-section corresponds well to data, then an order of magnitude estimate can be made regarding the size of the initial-state effects. Assuming a jet cross-section that scales by jet $p_T^{-5.5}$, a 20% effect in R_{pA}^{PYTHIA} corresponds to roughly a 5% enhancement in jet energy, again assuming the jet yields are unchanged and the spectra is merely shifted along the jet momentum axis. This small jet energy enhancement can be attributed to multiple scattering, where the additional nucleon-nucleon interactions in the pre-collision environment give an additional soft kick to the partons that are about to undergo a hard scattering. Though this result is not precise enough to confirm the presence of such an effect, the constraint placed on the size of the initial-state scattering effects leads to the second interesting conclusion. Because we observe a R_{pA}^{PYTHIA} value larger than one, we can conclude that the great majority of the b-jet suppression effects seen in PbPb collisions (as shown in the previous section) can be attributed to jet interaction with the QGP. The mere presence of a nucleus actually drives an increase in jet energy, so the fact that we see a strong suppression in b-jets in PbPb means that the QGP medium must be responsible for the effects.

In addition to the data result, a model prediction is also shown in Fig. 7.9. This model uses a pQCD framework, including both radiative and collisional loss and is

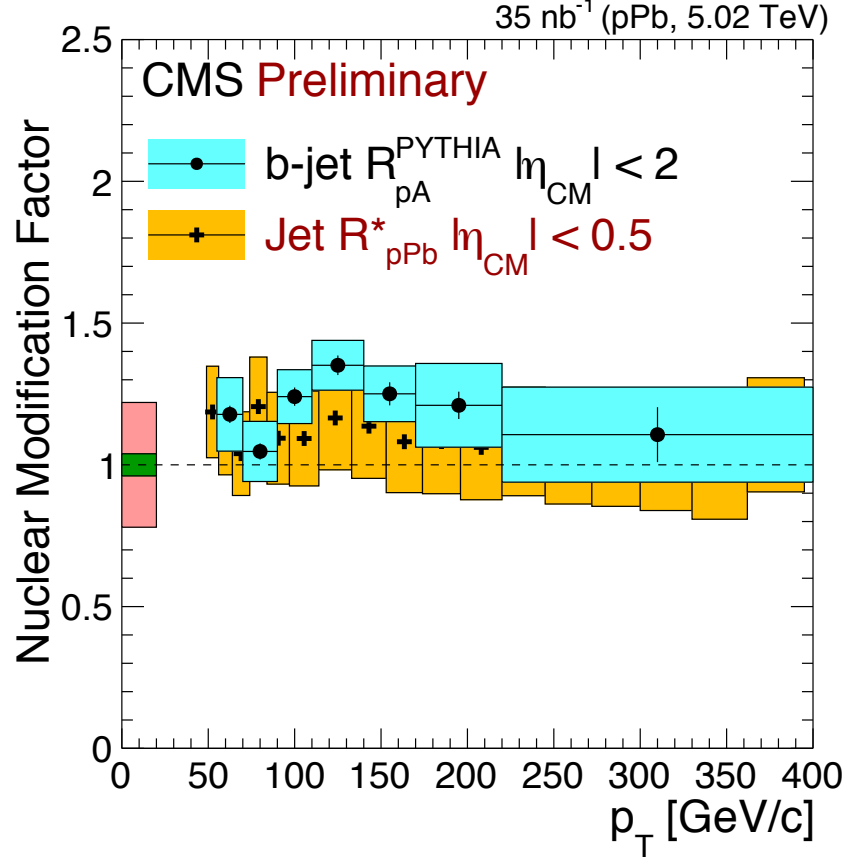


Figure 7.10. The inclusive-jet R_{pPb}^* and b-jet R_{pA}^{PYTHIA} as a function of jet p_T . We observe agreement between the two results.

described in detail in Ref. [77]. Though at a first glance, the model seems incompatible with the data, the prediction does not include any initial state multiple scattering and effectively only accounts for final state suppression effects. Since these effects are quite small in both data and in the model, we conclude that the model is compatible with the observed effects of the final state suppression mechanisms in pPb collisions.

Finally, we can again compare the effects of the b-jet mass by overlaying the the result of the b-jet R_{pA}^{PYTHIA} and the inclusive jet R_{pPb}^* [12], which is measured against an interpolated pp reference. The inclusive-jet analysis assumes a jet yield scaling based on the variable $x_T \equiv \frac{2p_T}{\sqrt{s}}$ to build an interpolated jet spectrum at 5 TeV based

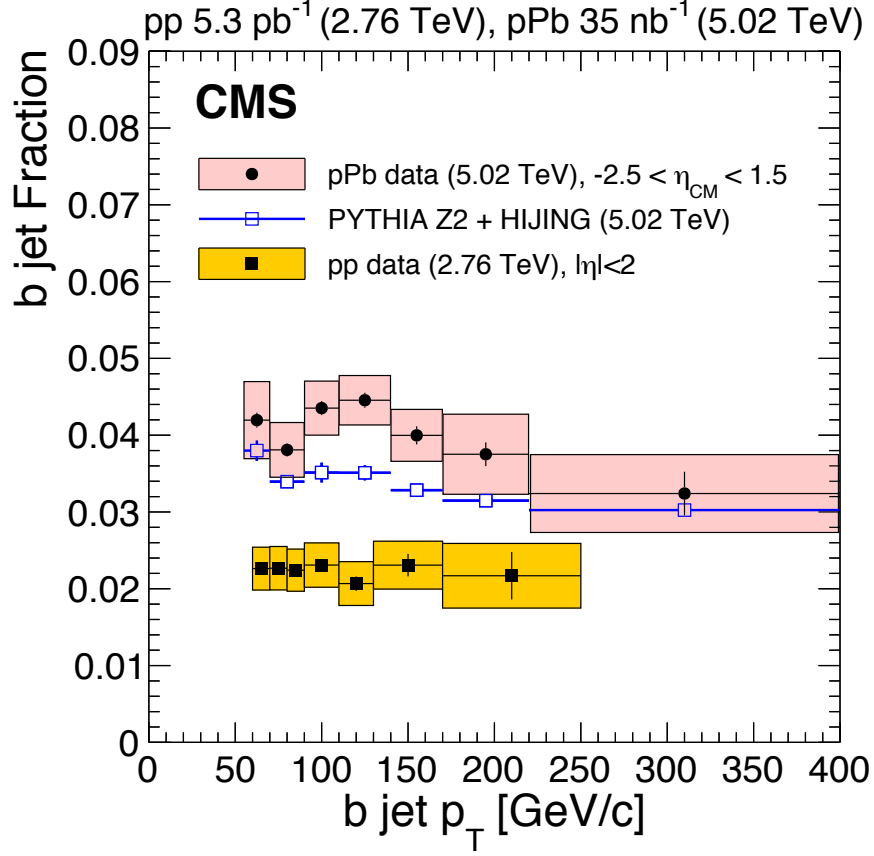


Figure 7.11. The b-jet fraction in pPb collisions at 5.02 TeV is shown and is compared to a prediction from a simulation of PYTHIA jets embedded in a HIJING background. The b-jet fraction from 2.76 TeV pp data is also shown

on measurements of the jet yield at 2.76 and 7 TeV. Figure 7.10 shows that we observe relative consistency between the two results, even without the uncertainty on the pp reference. This indicates (as in the PbPb analysis) that the mass-dependent effects of jet energy modification are still more or less negligible for transverse momentum above ≈ 50 GeV, for jets within $-2.5 < \eta_{CM} < 1.5$.

The b-jet fraction is shown in Fig. 7.11 and is compared to a prediction from a PYTHIA + HIJING simulation, where jets from PYTHIA are embedded into a pPb

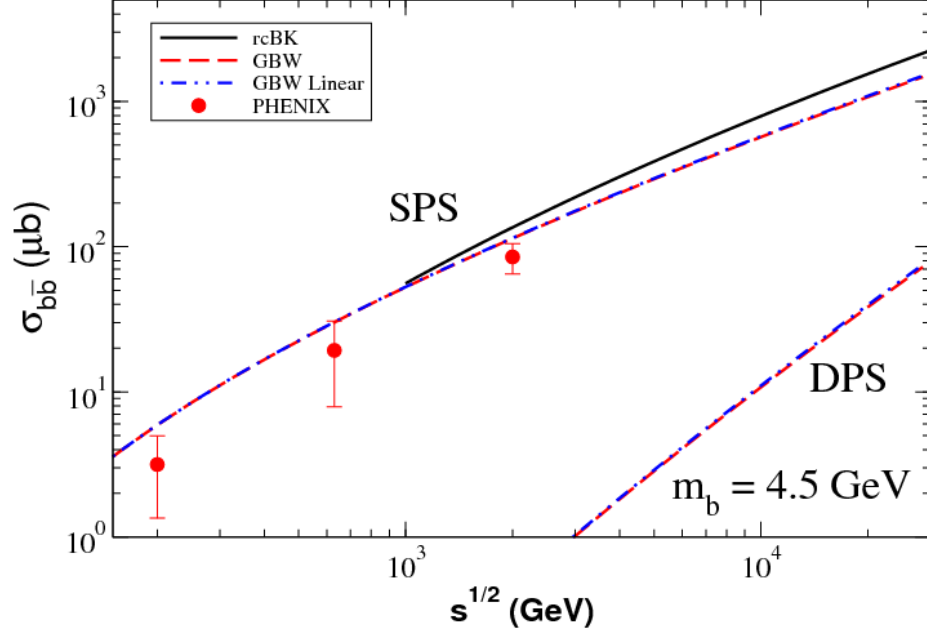


Figure 7.12. The $b\bar{b}$ cross-section from multiple scattering orders is shown using various theoretical models as a function of collision energy. See Ref. [80] for more details. Note the horizontal axis scale is mislabeled by a factor of 10 so LHC energies correspond to 2.76×10^4 and 5×10^4 GeV on the figure.

background as simulated by HIJING. Though PYTHIA is not expected to perform well in the heavy flavor sector due to the large next-to-leading order (NLO) contributions to b-jet production, and PYTHIA's limitation of two-body scattering processes only, the consistency between the data and simulation is relatively good. The PYTHIA generator does attempt to account for NLO processes via successive two-body scattering processes in quick succession, but this phenomenological workaround is not a perfect substitute for true NLO calculations.

We can also compare the b-jet fraction as a function of collision energy, assuming the pPb underlying event does not significantly change the jet flavor production. A comparison of the b-jet fraction in 2.76 TeV pp and 5.02 TeV pPb collisions is shown in Fig. 7.11. We see a clear increase in the b-jet fraction as a function of collision

energy, though recall that the pPb yield is enhanced by roughly 20% just from the initial state nuclear effects. This increase is qualitatively consistent with theoretical models, which predict a large increase in the gluon cross-section at higher center-of-mass energies, leading to an enhancement of heavy flavor production as a function of collision energy [80]. An example of this can be seen in Fig. 7.12, where a clear increase of $b\bar{b}$ cross section can be seen as the collision energy increases.

In addition to the jet p_T dependent results, this analysis also presents measurements as a function of jet η . Previous measurements in asymmetric colliding systems (like pPb) have shown indications that next-to-leading order jet production effects from nuclear parton distribution function (nPDF) distributions influence the pseudorapidity distributions of di-jet production [13,41]. In other words, the pseudorapidity distributions of jets correlate with the initial parton momentum fraction or Björken- x , so the η dependence of jet production can be used as a probe of the nPDFs. While PYTHIA simulations predict fairly weak correlations of single jets to Björken- x (as opposed to di-jets in previous measurements), we can investigate these effects for the presence of large unanticipated nPDF modifications in the heavy-flavor sector. By selecting on b-tagged jets only, we measure primarily gluon nPDFs, as gluon-gluon fusion is the dominant mechanism responsible for the production of b-quark pairs. This is in contrast to previous measurements, where any observed modifications are from a combination of the nPDFs of multiple partons.

Figure 7.13 shows the results of the b-jet cross-section (left) and the nuclear modification factor R_{pA}^{PYTHIA} (right) as a function of jet p_T in four pseudorapidity selections. We observe R_{pA}^{PYTHIA} values consistent with unity as a function of pseudorapidity, which is indicative that very large nPDF effects do not exist in the gluon sector, as anticipated. While the systematic uncertainties in this measurement are fairly large, they provide a baseline for future studies, including measurements of back-to-back b-tagged jets. Di-b-jet measurements in an asymmetric system can begin to place real constraints on the gluon nPDFs at large Björken- x . In addition, these types of measurements constrain the b-jet production mechanism to purely the leading-order

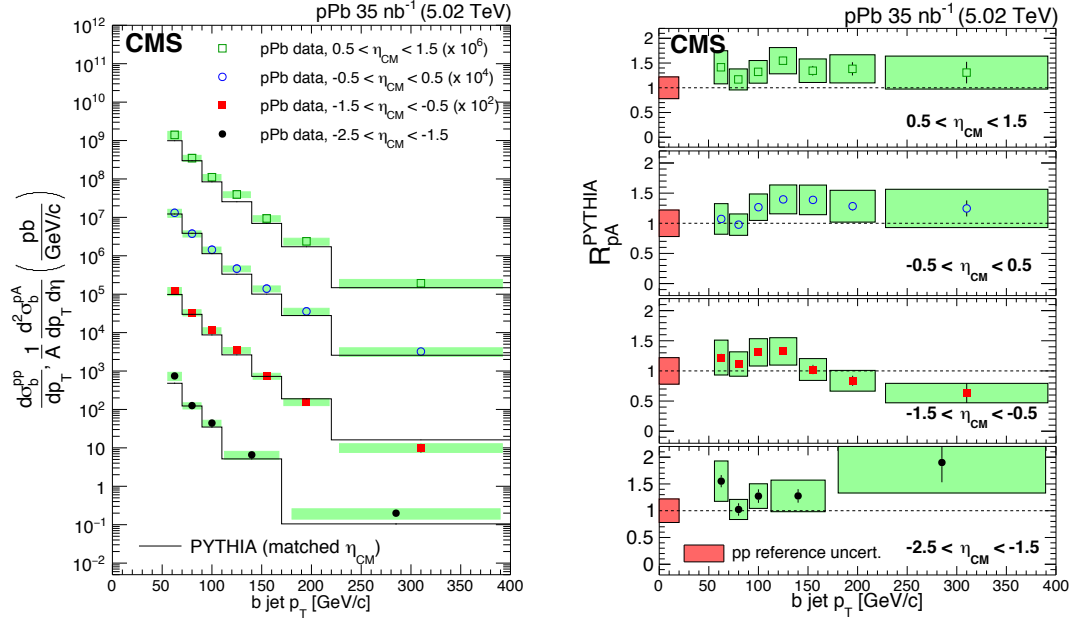


Figure 7.13. The b-jet cross section for various pseudorapidity selections is shown (left), along with the PYTHIA-based nuclear modification factor R_{pA}^{PYTHIA} (right) for the same four pseudorapidity bins. Note the absence of a $1.5 < \eta_{CM} < 2.5$ bin is due to the fact that jets are only calibrated to $|\eta_{lab}| < 2$ units.

Table 7.3.
Major contributors to the pPb b-jet spectrum systematic uncertainty at 5.02 TeV for three jet p_T bins.

Source	40 - 55 GeV/c	95 - 110 GeV/c	220 - 400 GeV/c
Jet Energy Resolution	5%	3%	2%
Unfolding Stability	4%	4%	4%
b-Tag Efficiency	4%	2%	14%
Tagger Work. Pt. Variation	1%	5%	7%
Data-Driven Template	14%	2%	<1%
Fix Light/Charm ratio	7%	<1%	< 1%
Gluon Splitting	4%	6%	3%
Total	18%	10%	17%

flavor creation process, further deconvoluting existing measurements of b-jet energy loss.

Systematic Uncertainties

The systematic uncertainties on this measurement are largely the same as in the PbPb collision system, however, in this case, a portion of the systematics between pPb and pp no longer cancel. As we use a simulation as a stand-in for pp data, any correlation of detector resolution effects is broken. Instead, we simply apply the full pPb uncertainties to the R_{pA}^{PYTHIA} measurement, excluding the jet energy scale, which is removed thanks to unfolding and the in-situ data-driven scale calibration from photon+jet events (see Section 6.4 for more details on this calibration). To account for the uncertainty of the PYTHIA b-jet cross-section, an additional overall normalization uncertainty is used.

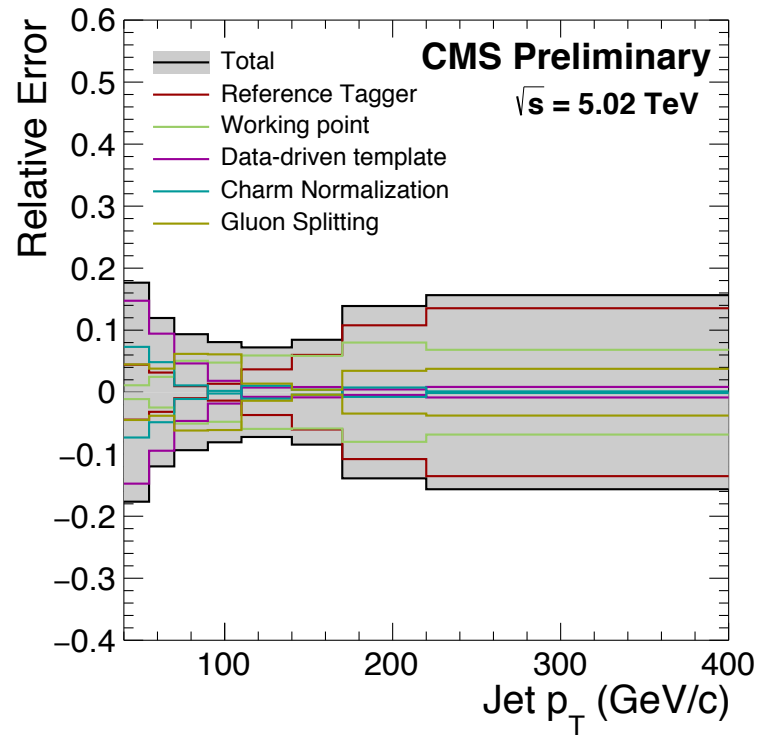


Figure 7.14. The contribution to the b-tagging systematic uncertainties for pPb collisions at 5.02 TeV is shown as a function of jet p_T .

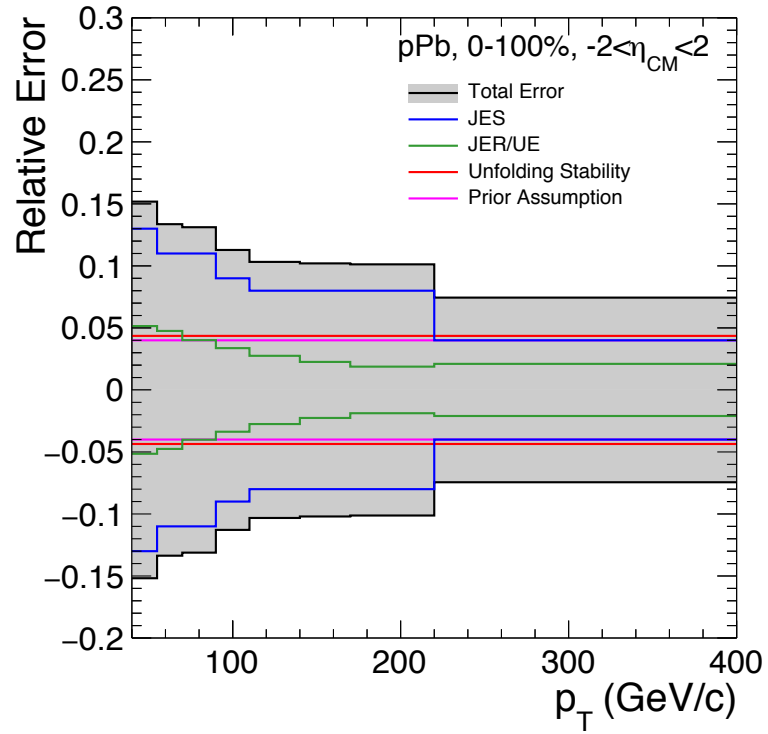


Figure 7.15. The contribution to the jet reconstruction systematic uncertainties for pPb collisions at 5.02 TeV is shown as a function of jet p_T .

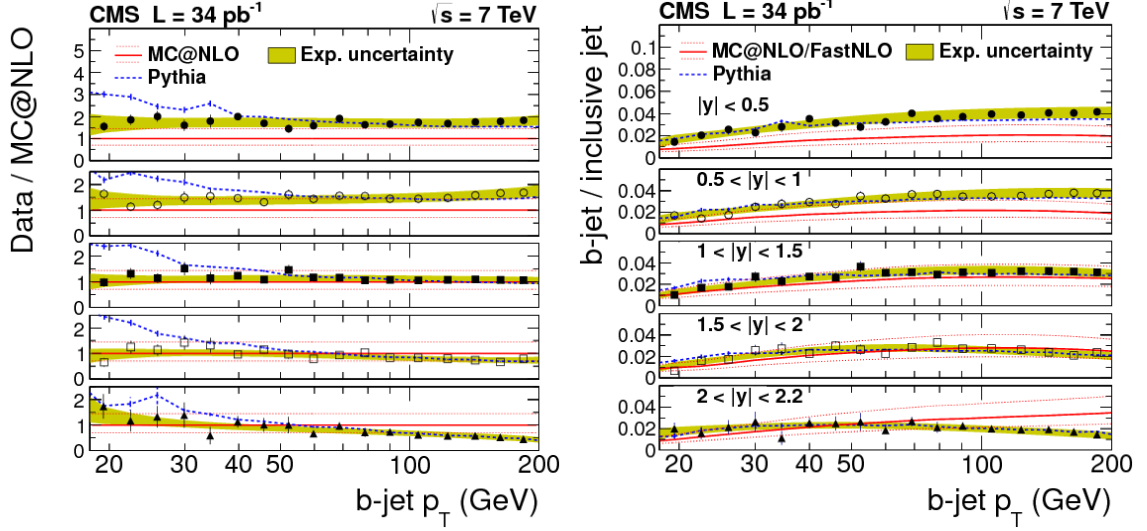


Figure 7.16. The b-jet cross-section (left) and the b-jet fraction (right) from 7 TeV data is compared to predictions from the PYTHIA D6T tune and to another next-to-leading order generator known as MC@NLO.

The same b-tagging and jet reconstruction uncertainty evaluation procedures are applied in pPb data at 5.02 TeV as are applied to PbPb and pp data at 2.76 TeV, described in Section 7.1. The systematic uncertainties from b-tagging are shown in Fig. 7.14, and uncertainties from jet reconstruction are shown in Fig. 7.15. We observe that the largest contributions to the uncertainties at low jet p_T is the use of the data-driven template, where anti-b-tagged events are used for the light jet templates. This contribution makes sense as at low- p_T , b-jet displacement is smaller, and light and b-jets appear kinematically more similar. At high- p_T , the largest contributions to the uncertainty come from the data-driven efficiency cross check (from the JP tagger). In total, the uncertainty ranges from roughly 10% to 20%. The various contributions to the total uncertainty are outlined in Table 7.3.

Due to the use of PYTHIA as a reference for this measurement, an additional scaling uncertainty is calculated based on the coherence of the generator to b-jet results at 2.76 and 7 TeV. At 7 TeV, we find that the PYTHIA D6T tune reproduces

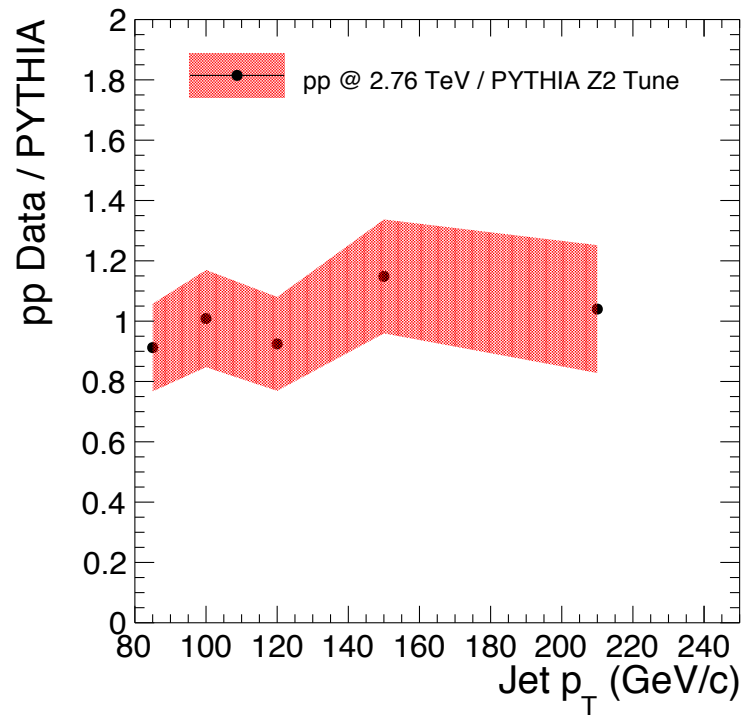


Figure 7.17. The b-jet cross-section in 2.76 TeV data is compared to the PYTHIA Z2 tune prediction.

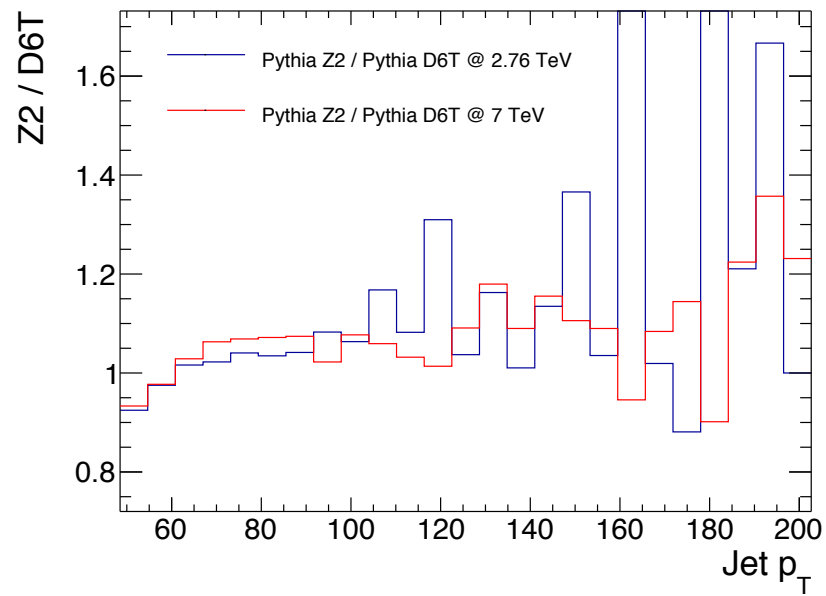


Figure 7.18. The ratio of predicted pp b-jet cross-sections between the PYTHIA D6T and Z2 tunes is shown for both 2.76 and 7 TeV

the calculated cross-sections for jet $p_T > 50$ GeV in all pseudorapidity classes and corresponds extremely well to the b-jet fraction, as shown in Fig. 7.16 [81]. The comparison between the PYTHIA Z2 tune and 2.76 TeV data is shown in Fig. 7.17. We observe relative consistency between the Z2 tune and the 2.76 TeV data, to within about 20%. These two effects together show that we can trust the PYTHIA generator to reproduce the 5.02 TeV b-jet spectrum to within 20%, however as two different PYTHIA tunes were used, an additional uncertainty must be factored in based on the difference between the two tunes. The ratio of the predicted b-jet cross-section between the Z2 and the D6T tunes is shown in Fig. 7.18, where we observe a maximum of 8% difference between the two tunes, for both energies, across all jet p_T . The two uncertainties added in quadrature lead to an overall scaling uncertainty from PYTHIA of 22%. This is assumed to be the maximum deviation from 5 TeV data and is shown in the pPb result plots as a red band around unity.

8. MODIFICATION OF C-TAGGED JETS

This chapter describes the c-tagged jet results in the pPb and pp collision systems at 5.02 TeV and 2.76 TeV, respectively. As the statistics are not great enough to perform c-tagging accurately with the 2011 PbPb data ($150/\mu\text{b}$), c-tagging is not attempted in this system. New statistics collected by CMS in 2015 ought to be large enough to make a measurement of c-tagged jets, but this has yet to be attempted. Additional details can be found in the preliminary paper, describing the methodology and all c-tagging results [82].

8.1 Results in pPb (5.02 TeV) and pp (2.76 TeV)

Due to the nature of the heavy-ion reconstruction and the difficulty in extracting a reliable measurement with limited statistics, the priority for c-jet tagging in a heavy ion environment is using the pPb collision system. Like the b-jet analysis, the measurement of the charm-tagged jets in pPb provides constraints on the initial-state heavy-flavored jet production due to the presence of a nucleus in the collision system.

More importantly, this is the first ever measurement of charm-tagged jets in a heavy ion system, and one of the first in the high-energy field. While b-jet taggers can simply require large displacement values to remove the great majority of light and charm jets from a sample, charm jets are much more difficult to extract. Charm jets have a very small displacement, such that only the most modern tracking systems are able to even resolve the vertex of a typical charm-jet decay. Fortunately, as described in Section 3.2, the resolution of the CMS inner tracking system can resolve these vertices and robustly identify charm-jet decays. In addition to the small decay length, kinematically charm jets behave similarly to light jets so tagging algorithms

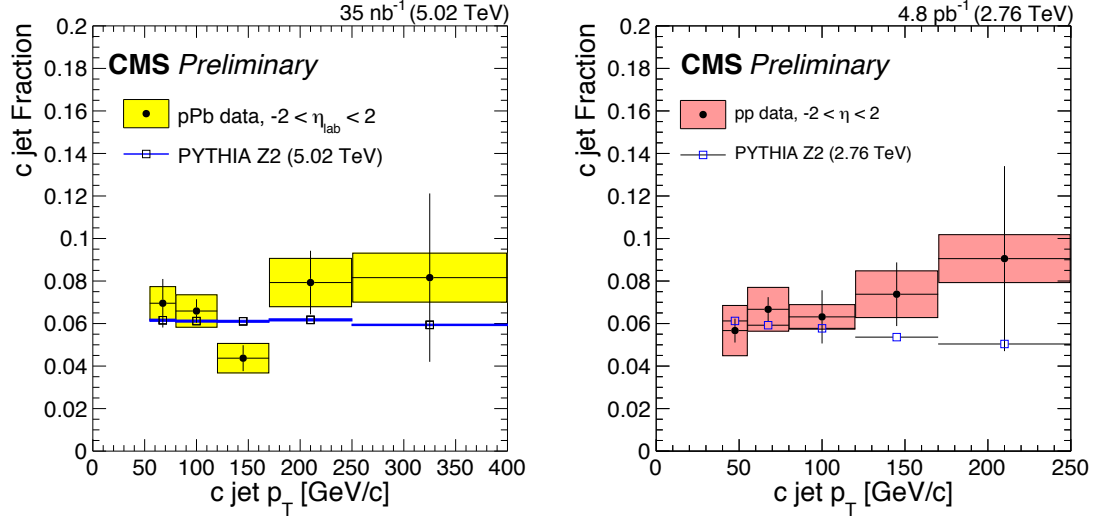


Figure 8.1. The c-jet fraction for pPb (left) and pp (right) is shown and compared to the PYTHIA prediction at center-of-mass energies of 5.02 TeV (left) and 2.76 TeV (right).

must be carefully tuned to avoid removing large fractions of charm jets when filtering out light jets.

Section 6.5.2 describes how these hurdles are overcome. The analysis uses the corrected secondary vertex mass (Eq. (6.11)) to separate the charm from beauty jets, and uses the high-purity version of the SSV tagger to separate the light from charm jets. These two selections create enough separation between the three flavors of jets such that the template fits converge with reasonably small uncertainties.

Figure 8.1 shows the fully corrected c-tagged fraction measurement in both pPb at 5.02 TeV (left) and pp at 2.76 TeV. Though it is difficult to make any dramatic conclusions regarding the fraction as a function of center-of-mass energy due to the large uncertainties, the lack of a large increase in c-jet fraction from 2.76 to 5 TeV is somewhat surprising. Recall that Fig. 7.11 shows a fairly large increase in the b-jet fraction - almost a factor of two, however it seems that this effect does not occur in the charm-tagged jet sector.

Another important note is the strained overlap in Fig. 8.1 between PYTHIA and the pp data at 2.76 TeV. We observe nearly opposite trending behavior, though again the uncertainties are quite large, and each data point is consistent with PYTHIA with fewer than two standard deviations. This is also important to remember when observing the difference between pPb and the PYTHIA predictions at 5.02 TeV, as it is clear that as PYTHIA may not describe the 2.76 TeV data very accurately, it is therefore possible that the PYTHIA stand-in for 5.02 TeV pp data may not match the distribution in data. Like the b-jet analysis, we can only say that the PYTHIA simulation represents the expected c-jet spectrum at 5.02 TeV to within 22%. Unlike the b-jet spectrum, however, the total lack of c-jet measurements in data means that the expected PYTHIA coherence to data is virtually untested. While we do not anticipate that the performance of PYTHIA should dramatically differ between charm and bottom jets, a 22% uncertainty (as in the b-jet analysis) on the c-jet cross-section from PYTHIA is merely our best guess at present.

The full c-jet spectrum is shown in Fig. 8.2 for 5.02 TeV pPb (left) and 2.76 pp (right) data, where we observe reasonable consistency between the PYTHIA simulation and data after scaling the pPb data by the Glauber nuclear overlap factor T_{pA} . Note that while “Data/PYTHIA” is directly analogous to R_{pA}^{PYTHIA} , the naming convention was changed between the b-jet and c-jet publications for political reasons. In any case, the ratio is again fit to a constant value, where we observe that the fraction is 1.00 ± 0.19 (stat.+syst.) in pPb data, and 1.15 ± 0.27 (stat.+syst.) in pp data, not including uncertainties from the PYTHIA simulation. We can conclude, therefore, that no statistically significant jet energy modification exists in the pPb collision system relative to the PYTHIA simulation. Furthermore, to within the PYTHIA uncertainties of 22% (derived from the b-jet analysis), we can conclude that jet energy is not largely modified relative to data.

Finally, we can plot the b-jet and c-jet nuclear modification factors together on the same plot to get some semblance of a global picture of energy loss. This is shown in Fig. 8.3. While unfortunately the c-jet R_{AA} is still missing, it’s becoming very

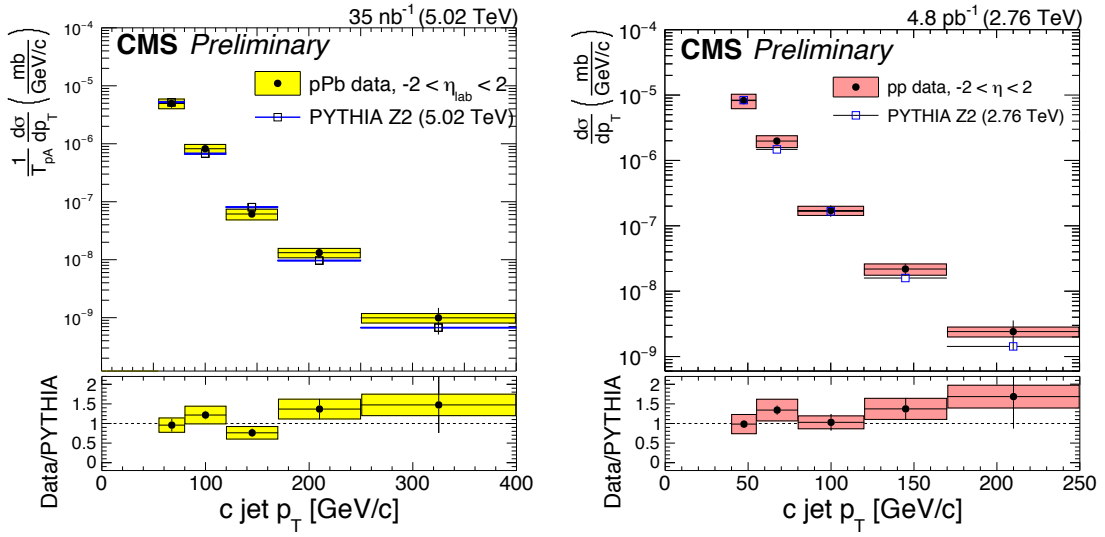


Figure 8.2. The c-jet spectra for pPb collisions at 5.02 TeV (left, scaled by T_{pA}) and pp collisions at 2.76 TeV (right) are shown and compared to PYTHIA predictions at both center-of-mass energies. The lower panels of both plots show the Data/PYTHIA ratio.

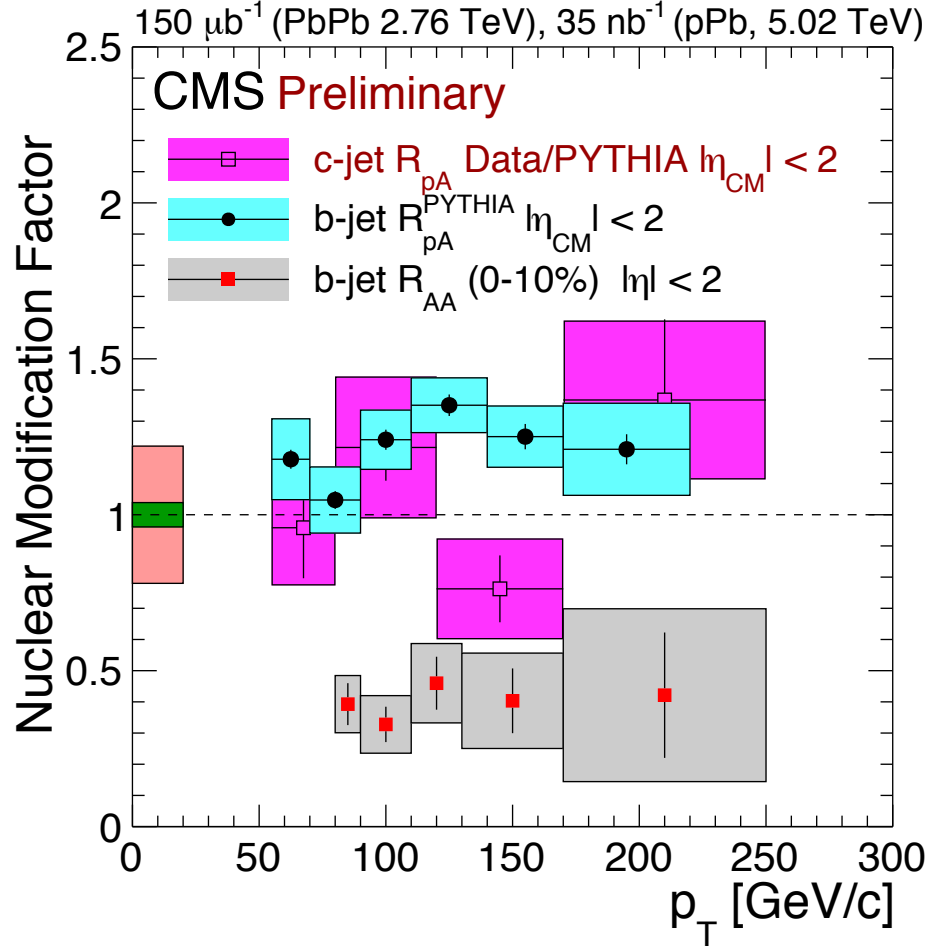


Figure 8.3. The b-jet R_{AA} , $R_{\text{pA}}^{\text{PYTHIA}}$ and c-jet data/Pythia ratios are shown for 2.76 TeV PbPb and 5.02 TeV pPb collisions. We observe that the mass dependent effects at high- p_T are small and within our systematic uncertainties.

clear that at high- p_T , jet energy loss is very similar for heavy, intermediate and light flavored jets and mass-dependent effects are virtually negligible.

8.2 Systematic Uncertainties

While the sources of systematic uncertainties remained largely the same between the b-jet and c-jet analyses, the methodology used to calculate the errors have changed. For c-tagging, the uncertainties that are used are from:

- Tagging efficiency calculation from JP tagger fits, instead of MC
- Working point variation
- MC template statistical uncertainties
- D meson decay
- Gluon splitting

The first item remains the same between the b-tagging and c-tagging analyses. Fits of the JP tagger distribution are created both before and after selecting on the SSV high-purity working point. Again, via Eq. (6.10), a value for c-tagging efficiency can be obtained. The difference between the simulation and the JP-derived tagging efficiencies is taken as the first systematic uncertainty.

To account for uncertainties in the tagger itself, the working point of the c-tagger is varied. We observe an enhancement of the c-jet purity as a function of the tagger discriminator for SSV high purity values between 1.2 and 2.4, shown in Fig. 6.19. To account for the total variation as a function of the tagger, c-jet yield is calculated for every working point between 1.2 and 2.4 in steps of 0.2 units. The average value across all working points is calculated for each jet p_T bin and discrepancies from the average value are taken as a source of systematic uncertainty. An example of the variation of the c-jet yield as a function of tagger working point is shown in Fig. 8.4 in pPb collisions.

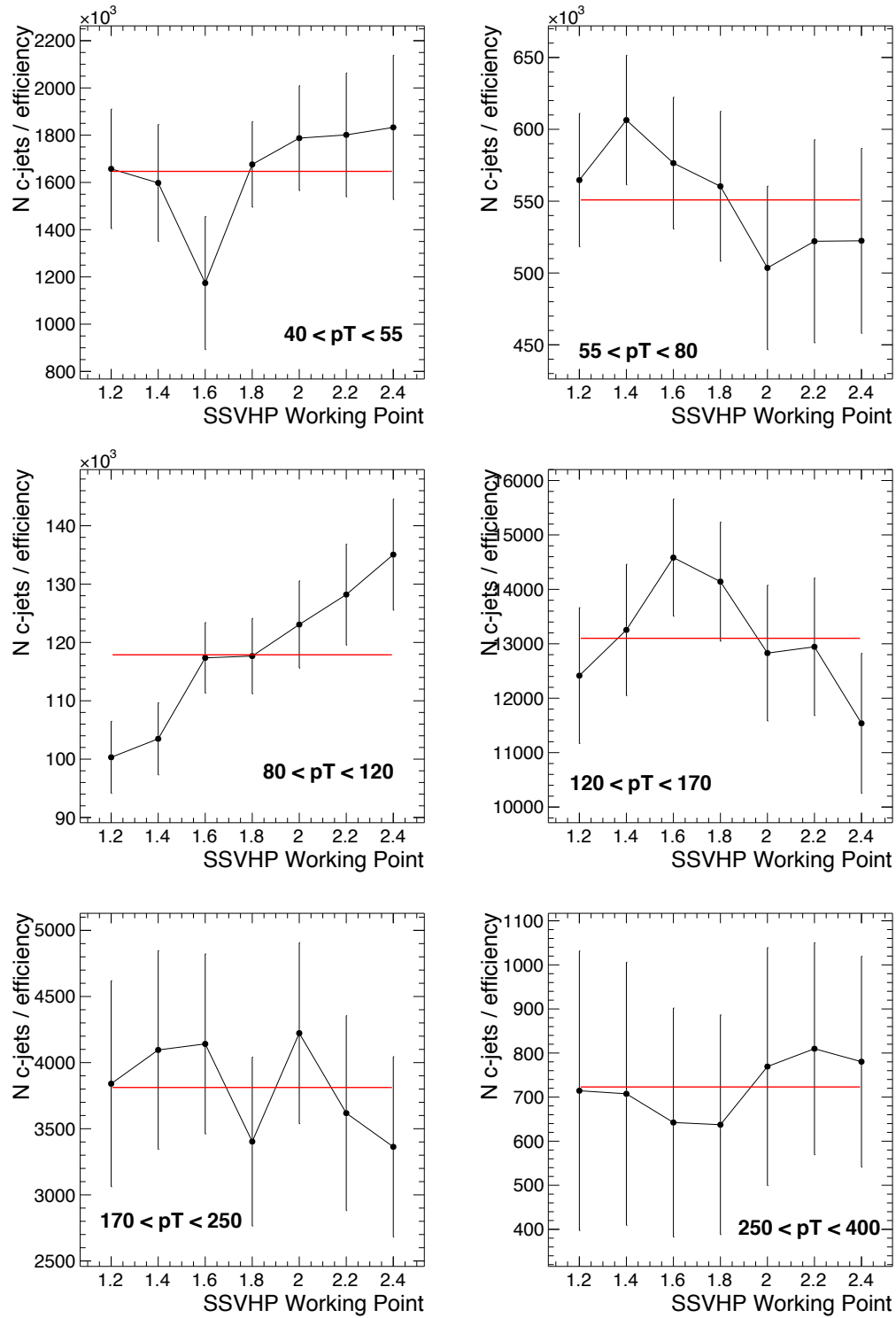


Figure 8.4. The efficiency-corrected number of charm-tagged jets is shown for pPb data at 5.02 TeV as a function of SSVHP working point for the six p_T bins used in the analysis. The solid red line denotes the average value.

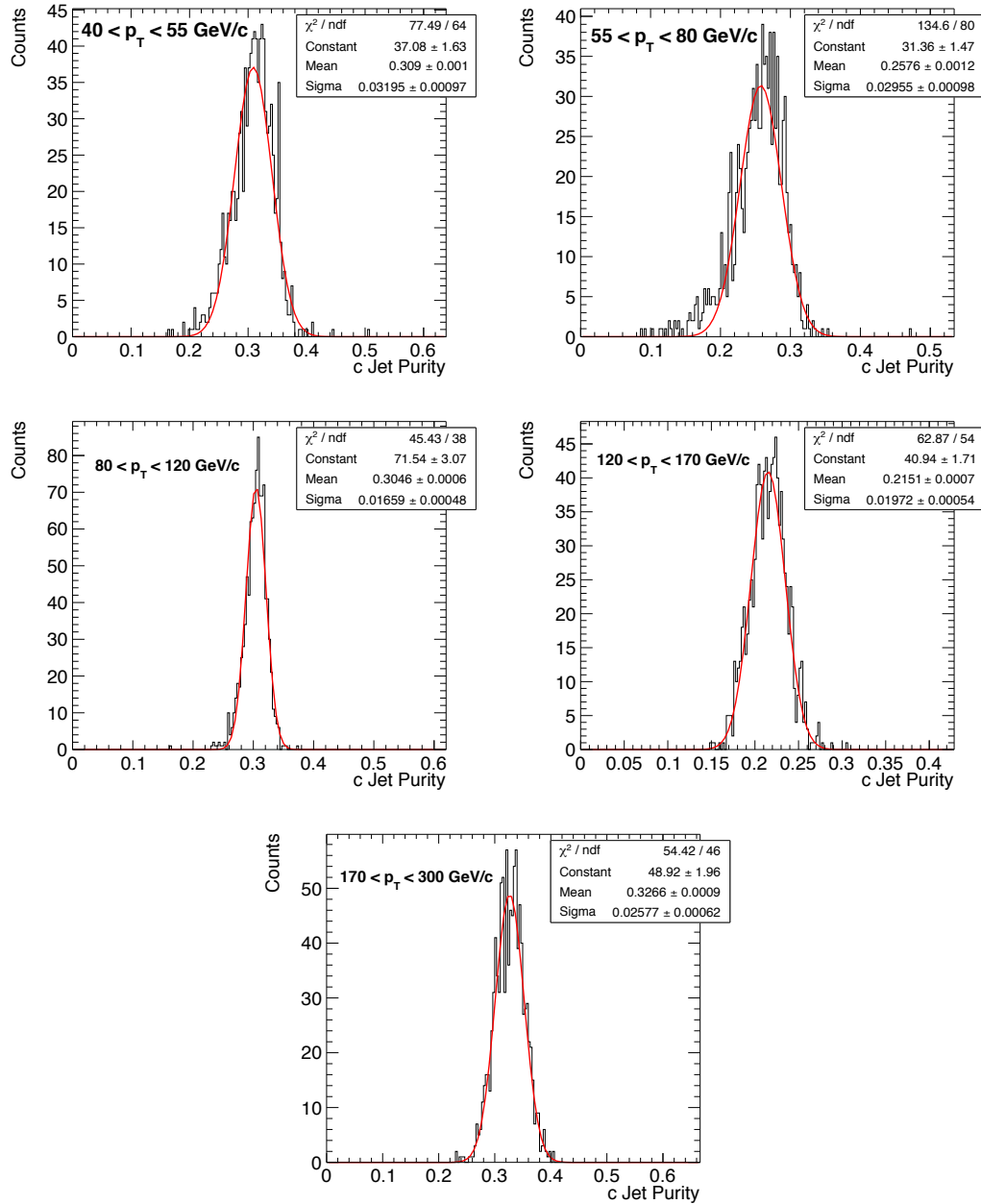


Figure 8.5. Distributions of the charm jet purity obtained from varying the flavor templates from the pPb PYTHIA+HIJING simulation at 5.02 TeV within their statistical uncertainties. These are then fit to a gaussian distribution, where the sigma value of the gaussian is used as the uncertainty. The five figures each show one of the five p_T bins used, as denoted in the figure.

The statistical precision of the PYTHIA simulation is relatively weak in this analysis due to analysis time constraints. While the convergence to the template fits is relatively robust, statistical uncertainties on the templates themselves must also be accounted for. To calculate these uncertainties, a toy simulation is used, where the template fitting procedure is performed thousands of times, each time varying the simulated template shape within its uncertainty. The extracted purity values from each of these fitting procedures is plotted and then fit to a gaussian for each jet p_T bin. The width of the gaussian is used as the systematic uncertainty from this source. Examples for each jet p_T bin in pPb collisions is shown in Fig. 8.5.

Due to the charm jet requirement of a 3+ body vertex, an additional source of uncertainty is implicitly included, which has to do with the ability of PYTHIA to accurately represent the D meson decay fragmentation functions and branching ratios. A test is done to ensure that the fraction of charmed mesons along with their decay multiplicities are consistent with the world-averaged values from data. To quantify the total uncertainty, we investigated the jet-to-meson matching efficiency and reweighted and varied the charm-to-meson branching ratios within their data uncertainties.

Jet flavor matching in CMS is defined by the presence of a parton within 0.4 units (slightly larger than the cone size) in η , ϕ space from the reconstructed jet axis. Using the flavor matching as a guideline, this systematic error calculation matched D mesons to jets using the same association window of 0.4 units in η , ϕ space. To ensure feed-down effects are accounted for, only charm hadrons without further charm decay products are used. The distribution of the number of charm hadrons in a jet cone is shown in the left panel of Fig. 8.6, where we observe that the average number of associated charm hadrons increases slightly when applying the tagger. This is likely due to an enhancement of the gluon splitting fraction, where at least one of the produced charmed mesons decays via a 3+ body process. There is also a small fraction of tagged jets without a charm quark in the jet cone, which is a boundary effect where the fragmentation process happens to push the meson outside of the

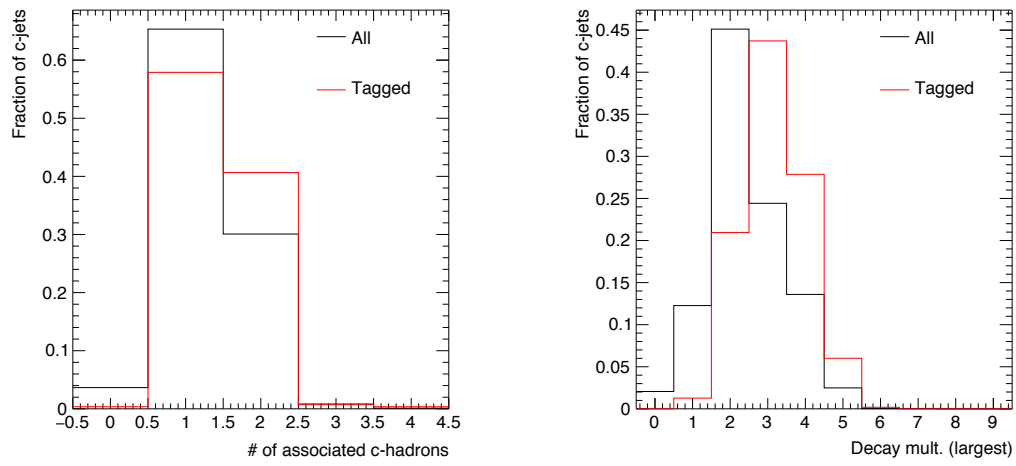


Figure 8.6. The distribution of the number of generator-level charmed hadrons (left) and the reconstructed vertex multiplicity (right) in a jet cone of $R = 0.3$ for jets with $p_T > 80 \text{ GeV}/c$.

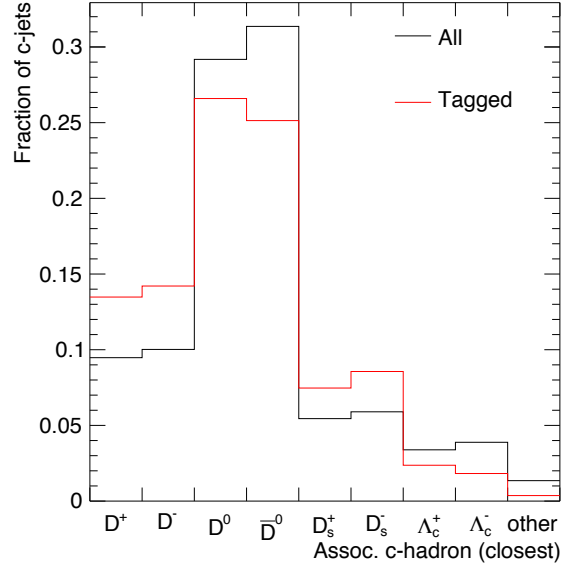


Figure 8.7. The distribution of the number of generator-level charmed hadrons (left) and the reconstructed vertex multiplicity (right) in a jet cone of $R = 0.3$ for jets with $p_T > 80 \text{ GeV}/c$.

association window. The right panel of Fig. 8.6 shows the decay multiplicity of the jet secondary vertex, with and without applying the SSV high purity selection. As expected, the generator-level vertex multiplicity increases dramatically when the selection is applied. Note that jets still may have less than three generator-level decay products due to combinatorial effects that affect the post-reconstruction track-to-vertex association performance.

In addition to the number of mesons in the jet cone, the meson type is also measured. To avoid ambiguity, the meson closest to the jet axis is used as the identifying particle. Figure 8.7 shows the distribution of meson type within the c-jet cone both before and after selecting on the SSV high purity tagger. It is clear from the plot that the tagging efficiency depends on the hadron type. To fully understand these effects, we investigate the natural abundance of the charmed meson types, along with their

Table 8.1.

The relative abundance of each charged hadron species from published data and PYTHIA, as well as their MC tagging efficiency.

Species	Data (%)	PYTHIA (%)	MC tagging efficiency
D^+	23.2 ± 1.1	19.5	18.5
D^0	54.9 ± 2.6	60.8	11.1
D_s^+	10.1 ± 2.7	11.3	18.3
Λ_c	7.6 ± 2.1	7.2	7.4
other	4.2	1.1	3.5

Table 8.2.

The mean charged hadron decay multiplicities from data and from PYTHIA for D^0 , D^+ and D_s^+ .

Species	Data	PYTHIA
D^+	1.96 ± 0.08	2.13
D^0	2.25 ± 0.08	2.25
D_s^+	2.41 ± 0.38	2.46

fragmentation functions. Table 8.1 shows the contribution of each hadron type to the total charmed hadron abundance from data and pythia, and also shows the tagging efficiency as a function of hadron type. To test the effects from the misrepresentation of the charmed hadron fraction, the PYTHIA simulation was reweighted to accurately reflect the data distribution. We find that the total charm jet tagging efficiency does not change very much at all, from 13.0% to 12.96%. The uncertainty was obtained using a toy MC simulation, where the contribution from each hadron type was varied within its uncertainty, assuming a gaussian distribution. The total uncertainty has a standard deviation of about 1.5% of the nominal tagging efficiency which is factored into the total systematic uncertainty.

Finally, the tagging efficiency dependence on the meson decay multiplicity is also tested. Table 8.2 shows the mean decay multiplicities from the three most common D meson types in data and in PYTHIA, where we observe that PYTHIA tends to overestimate the decay multiplicity of the D^+ meson. The uncertainty from this is tested in a similar way as was the hadron fraction, where the simulation is reweighted jet-by-jet to accurately reflect the world averaged decay multiplicity values. The tagging efficiency decreases by 4% after reweighting, largely due to the effects from the D^+ meson. Then, to estimate the uncertainty from this procedure, a similar toy MC study is performed, where the multiplicities are varied within their uncertainties. The distributions of tagging efficiency for each D meson type is shown in Fig. 8.8 where

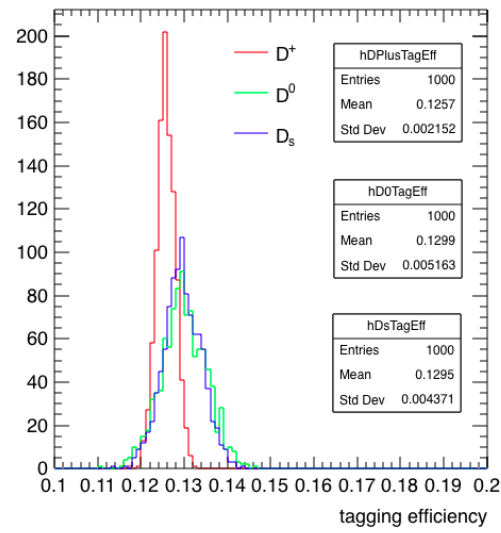


Figure 8.8. The distribution of tagging efficiencies obtained by variation of the charged decay multiplicities of the D^+ , D^0 and D_s^+ .

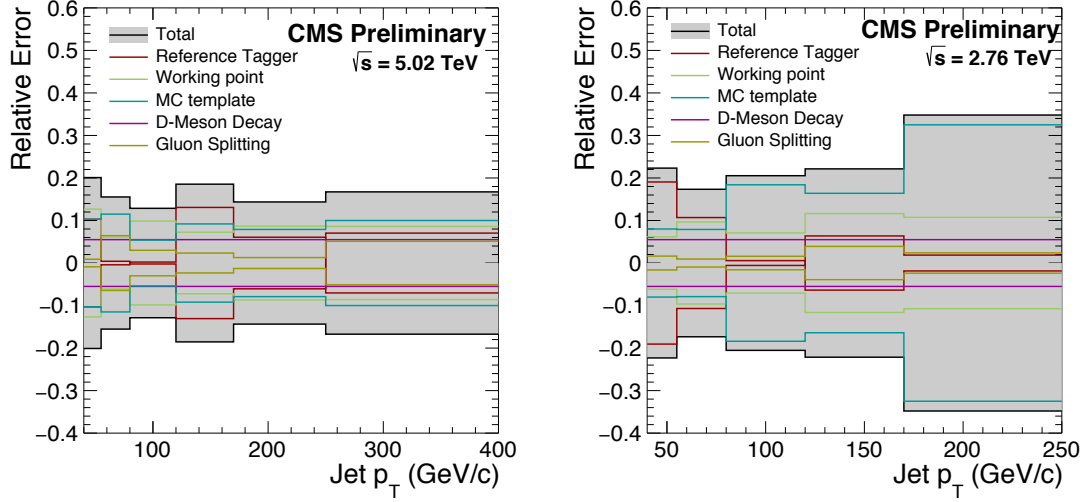


Figure 8.9. The total c-tagging systematic uncertainties are shown for pPb at 5.02 TeV (left) and pp at 2.76 TeV (right) using various colored histograms. The quadratic sum is shown as a solid black histogram.

we find that the standard deviations of tagging efficiency are 1.7%, 4.0%, and 3.4% for the D^+ , D^0 and D_s^+ mesons, respectively. Adding all these effects in quadrature leads to a total tagging uncertainty of 5.5%.

The last tagging uncertainty is from the uncertainty of the gluon splitting fraction in simulation. As in the b-jet analysis, the jets resulting from a fragmentation process including $g \rightarrow c\bar{c}$ contributions are varied in simulation up and down by 50%. This is generally a small effect and is less than roughly 5% for all p_T bins.

The total tagging uncertainty is shown in Fig. 8.9, with contributions from all previously mentioned sources. We find that the total uncertainty fluctuates around roughly 20% for both pPb and pp, except for large values of jet p_T in pp collisions, where the statistical uncertainties from the MC templates dominate the uncertainties. This of course can be corrected with the inclusion of additional statistics, so we expect the published measurement to be relatively more accurate, especially for the high- p_T pp result.

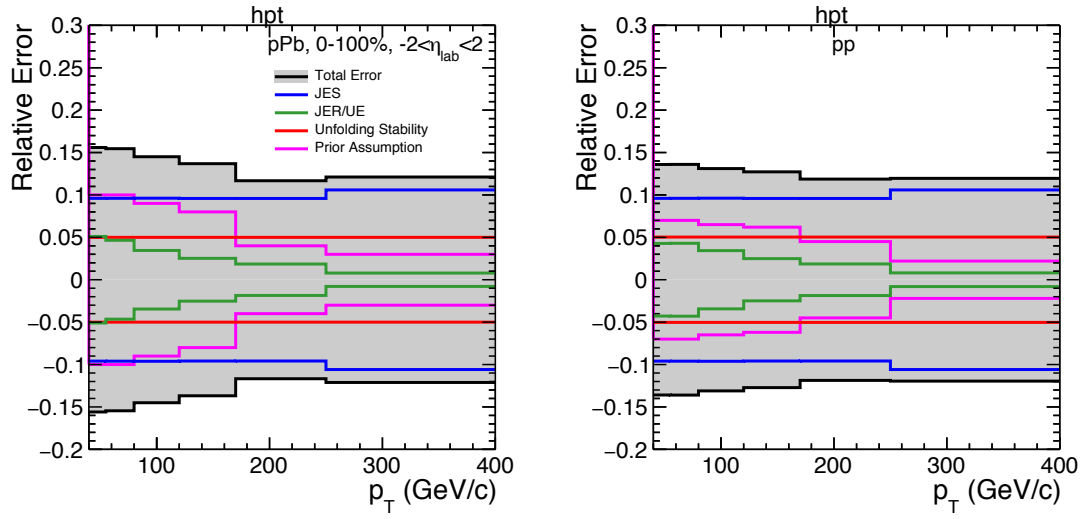


Figure 8.10. The total jet reconstruction systematic uncertainties are shown for pPb at 5.02 TeV (left) and pp at 2.76 TeV (right) using various colored histograms. The quadratic sum is shown as a solid black histogram.

In addition to tagging uncertainties, jet reconstruction uncertainties must be quantified. Since the datasets are the same, with an identical reconstruction procedure and event selection as the b-tagged jet analysis, the jet reconstruction uncertainties can simply be taken from the b-tagged analysis, and rederived for the correct p_T binning scheme. As in the b-jet analysis, the contributions to the b-tagging uncertainty are shown in Fig. 8.10 for pPb (left) and pp (right).

9. SUMMARY

We have measured the b-jet and c-jet nuclear modification factors in a variety of collision systems and nearly completed the global high- p_T picture of heavy flavored energy loss in heavy-ion collisions.

The analysis of b-tagged jets in PbPb and pp collisions at 2.76 TeV and pPb collisions at 5.02 TeV described in Chapter 7 shows strong b-jet energy modification in PbPb collisions as a function of centrality, where we observe large quenching effects in the most central events for jets with p_T between 80 and 100 GeV/ c . The b-jet energy loss is equivalent to that of inclusive jets in both the pPb and PbPb collision systems within the systematic uncertainties, indicating that the mass-dependence of energy loss is negligible above jet p_T values of roughly 80 GeV/ c . This conforms to theoretical predictions made using perturbative QCD and tend to disfavor models that are based in AdS/CFT correspondence (though the latter models are currently being revised). Through additional comparisons with pQCD models, our results favor a model with strong gluon-to-medium coupling strength as we observe significant jet suppression out to very large values in jet p_T . We additionally observe that jet energy modification in pPb collisions relative to simulated pp events using the PYTHIA generator are consistent with unity within uncertainties. Fitting a constant to this R_{pA}^{PYTHIA} distribution gives a value of 1.22 ± 0.15 (stat.+syst. pPb) ± 0.27 (syst. PYTHIA), indicating that initial-state suppression is negligible and that nuclear initial state effects may actually lead to a minor enhancement of jet production, rather than suppression. Finally, we probed for the presence of large effects due to nuclear parton distribution functions (nPDFs) based on the correspondence of jet pseudorapidity distribution and parton Björken- x values. While the correspondence is relatively weak, the b-jet production depends primarily on the gluon nPDFs, so large gluon

nPDF effects, if any, should be visible. Within our systematic uncertainties, however, no such large effects are observed.

Charm-tagged jets are measured for the first time in a heavy-ion collision system, as described in Chapter 8. We presented a robust tagging methodology that can be applied to a number of collision systems (hopefully including PbPb in the near future). We observe a charm jet production that is consistent with PYTHIA predictions for both pPb at 5.02 TeV and for pp at 2.76 TeV. Again, fitting a constant to these data/PYTHIA ratios gives 1.00 ± 0.19 for pPb and 1.15 ± 0.27 in pp.

Future work will aim to complete the single-jet energy loss picture with the inclusion of charm-tagged jet energy loss in PbPb. Initial energy loss measurements of D mesons in PbPb collisions show similar suppression as measurements of inclusive charged particles [83], so it will be interesting to see if these effects extend to high- p_T and persist in the jet sector. In addition, it is important to begin to identify the production mechanism of heavy flavored jets in data in order to deconvolute gluon splitting effects from measurements of b-quark energy suppression. A measurement of back-to-back b-tagged jets would reject large contributions from next-to-leading order effects and remove ambiguities based on the presence of multiple b-quarks in the jet cone. In addition, the ability to tag heavy flavored jets in a jet-by-jet fashion will open the door to many advanced measurements, including b-jet fragmentation functions and b-jet to hadron correlations. Such studies ought to be possible using statistics from future run periods of the LHC.

REFERENCES

REFERENCES

- [1] PHENIX Collaboration. Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX Collaboration. *Nucl. Phys. A*, 757:184, 2005.
- [2] PHOBOS Collaboration. The PHOBOS perspective on discoveries at RHIC. *Nucl. Phys. A*, 757:28, 2005.
- [3] BRAHMS Collaboration. Quark-Gluon plasma and color glass condensate at RHIC? The perspective from the BRAHMS experiment. *Nucl. Phys. A*, 757:1, 2005.
- [4] STAR Collaboration. Experimental and theoretical challenges in the search for quark-gluon plasma: The STAR Collaboration’s critical assessment of the evidence from RHIC collisions. *Nucl. Phys. A*, 757:102, 2005.
- [5] CMS Collaboration. Observation and studies of jet quenching in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. C*, 84:024906, 2011.
- [6] ALICE Collaboration. Elliptic flow of charged particles in Pb-Pb collisions at 2.76 TeV. *Phys. Rev. Lett.*, 105:252302, 2010.
- [7] ATLAS Collaboration. Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS Detector at the LHC. *Phys. Rev. Lett.*, 105:252303, 2010.
- [8] Partha Bagchi and Ajit M. Srivastava. Quarkonia Disintegration due to time dependence of the $q\bar{q}$ potential in Relativistic Heavy Ion Collisions. *Mod. Phys. Lett. A*, 30:1550162, 2015.
- [9] T. Celik, F. Karsch, H. Satz. A Percolation Approach to Strongly Interacting Matter. *Phys. Lett. B*, 97:128, 1980.
- [10] Samuel S. M. Wong. *Introductory Nuclear Physics*. Wiley-VCH Verlag Gmbl & Co., Weinheim, Germany, 2 edition, 2008.
- [11] J. W. Cronin, H. J. Frisch, M. J. Shochet, J. P. Boymond, R. Mermod, P. A. Piroue and R. L. Sumner. Production of hadrons at large transverse momentum at 200, 300, and 400 gev. *Phys. Rev. D*, 11:3105, 1975.
- [12] CMS Collaboration. Measurement of inclusive jet production and nuclear modifications in ppb collisions at $\sqrt{s_{NN}} = 5.02$ tev. *Eur. Phys. J. C*, page Submitted, 2016.
- [13] ATLAS Collaboration. Centrality and rapidity dependence of inclusive jet production in $\sqrt{s_{NN}} = 5.02$ TeV proton-lead collisions with the ATLAS detector. *Phys. Lett. B.*, 748:392, 2014.

- [14] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.
- [15] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22:579, 1961.
- [16] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264, 1967.
- [17] UA1 Collaboration. Experimental observation of isolated large transverse energy electrons with associated missing energy at $s=540$ GeV. *Phys. Lett. B*, 122:103, 1983.
- [18] UA2 Collaboration. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider. *Phys. Lett. B*, 122:476, 1983.
- [19] CDF Collaboration. Observation of Top Quark Production in $P\bar{p}$ -P Collisions. *Phys. Rev. Lett.*, 74:2626, 1995.
- [20] D0 Collaboration. Observation of the Top Quark. *Phys. Rev. Lett.*, 74:2632, 1995.
- [21] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012.
- [22] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012.
- [23] George R. Blumenthal, S.M. Faber, Joel R. Primack, and Martin J. Rees. Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311:517, 1984.
- [24] Vernon Barger, Danny Marfatia, and Kerry Lewis Whisnant. *The Physics of Neutrinos*. Princeton University Press, Princeton, New Jersey, 1 edition, 2012.
- [25] BCDMS Collaboration. Test of QCD and a measurement of Λ from scaling violations in the proton structure function $F_2(x, Q^2)$ at high Q^2 . *Phys. Lett. B*, 223:490, 1989.
- [26] NNPDF Collaboration. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244–289, 2013.
- [27] V.A. Khoze, et. al. Proceedings, Conference on High energy physics. In *Conference on High energy physics*, Tbilisi, USSR, 1976.
- [28] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nucl. Phys. B*, 126:298, 1977.
- [29] Jorge Casalderrey-Solana and Carlos Salgado. Introductory Lectures on Jet Quenching in Heavy Ion Collisions. *Acta Phys. Polon. B*, 38:3731, 2007.
- [30] Bjorken, J.D. Energy Loss of Energetic Partons in Quark-Gluon Plasma: Possible Extinction of High-pT Jets in Hadron-Hadron Collisions. Technical report, Fermilab, Batavia, IL., 1982.

- [31] Adams, J. and others (STAR Collaboration). Evidence from d + Au measurements for final state suppression of high p_T hadrons in Au+Au collisions at RHIC. *Phys. Rev. Lett.*, 91:072304, 2003.
- [32] Adler, S. S. and others (PHENIX Collaboration). Suppressed π^0 production at large transverse momentum in central Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 91:072301, 2003.
- [33] ATLAS Collaboration. Jet size dependence of single jet suppression in lead-lead collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector at the LHC. *Phys. Lett. B*, 719:220, 2013.
- [34] M. Djordjevic and M. Djordjevic. Heavy flavor puzzle at RHIC: Analysis of the Underlying Effects. *Phys. Rev. C*, 90:034910, 2014.
- [35] Sidney D. Drell and Tung-Mow Yan. Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies. *Phys. Rev. Lett.*, 25:316, 1970.
- [36] ATLAS Collaboration. Measurement of the flavour composition of dijet events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 73:2301, 2013.
- [37] H. van Hees and R. Rapp. Thermalization of heavy quarks in the quark-gluon plasma. *Phys. Rev. C*, 71:034907, 2005.
- [38] G.D. Moore and D. Teaney. How much do heavy quarks thermalize in a heavy ion collision? *Phys. Rev. C*, 71:064904, 2005.
- [39] W.A. Horowitz. Fluctuating Heavy Quark Energy Loss in Strongly-Coupled Quark-Gluon Plasma. *Phys. Rev. D*, 91:085019, 2015.
- [40] CMS Collaboration. Nuclear effects on the transverse momentum spectra of charged particles in pPb collisions at a nucleon-nucleon center-of-mass energy of 5.02 TeV. *Eur. Phys. J. C*, 75:237, 2015.
- [41] CMS Collaboration. Studies of dijet pseudorapidity distributions and transverse momentum balance in pPb collisions at $\sqrt{s_{NN}}=5.02$ TeV. *Eur. Phys. J. C*, 74:2951, 2014.
- [42] Hung-Liang Lai et al. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.
- [43] CMS Collaboration. Transverse momentum spectra of inclusive b jets in ppb collisions at $\sqrt{s_{NN}} = 5.02$ tev. *Phys. Lett. B*, 754:59, 2016.
- [44] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. Technical report, CERN, Geneva, 2006.
- [45] CMS Collaboration. Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays. *JINST*, 5:T03009, 2010.
- [46] CMS Collaboration. Tracking and Vertexing Results from First Collisions. *European Physics Journal C*, 70:1165, 2010.
- [47] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9:P10009, 2014.

- [48] CMS Collaboration. Missing transverse energy performance of the CMS detector. *JINST*, 6:P09001, 2011.
- [49] CMS Collaboration. Radiation background with the CMS RPCs at the LHC. *JINST*, 10:C05031, 2015.
- [50] L. Kumar (for the STAR Collaboration). STAR Results from the RHIC Beam Energy Scan-I. In *Proceedings of the XXIII International Conference on Ultra-relativistic Nucleus-Nucleus Collisions*, volume 904, page 256c, Washington, DC, 2013.
- [51] R.J. Glauber and G. Matthiae. High-energy scattering of protons by nuclei. *Nucl. Phys. B*, 21:135, 1970.
- [52] Michael L. Miller, Klaus Reygers, Stephen J. Sanders, and Peter Steinberg. Glauber Modeling in High Energy Nuclear Collisions. *Annual Review of Nuclear and Particle Science*, 57:205, 2007.
- [53] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.
- [54] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton fragmentation and string dynamics. *Phys. Rept.*, 97:31, 1983.
- [55] Xin-Nian Wang and Miklos Gyulassy. HIJING 1.0: A Monte Carlo program for parton and particle production in high energy hadronic and nuclear collisions. *Comp. Phys. Comm.*, 83:307, 1994.
- [56] Zi-Wei Lin, Che Ming Ko, Bao-An Li, Bin Zhang, and Subrata Pal. Multiphase transport model for relativistic heavy ion collisions. *Phys. Rev. C*, 72:064901, 2005.
- [57] B. Zhang. ZPC 1.0.1: A parton cascade for ultrarelativistic heavy ion collisions. *Comp. Phys. Comm.*, 111:276, 1998.
- [58] I.P. Lokhtin, et. al. Heavy ion event generator HYDJET++ (HYDrodynamics plus JETs). *Comp. Phys. Comm.*, 180:779, 2009.
- [59] R. Baier, Yu. L. Dokshitzer, A.H. Mueller, and D. Schiff. Angular dependence of the radiative gluon spectrum. *Phys. Rev. C*, 64:057902, 2001.
- [60] T. Pierog, Iu. Karpenko, et. al. EPOS LHC: test of collective hadronization with LHC data. *Phys. Rev. C*, 92:034906, 2015.
- [61] S. Agostinelli, et. al. GEANT4 - a simulation toolkit. *Nucl. Instr. Meth. A*, 506:250, 2003.
- [62] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nucl. Instr. Meth. A*, 262:444, 1987.
- [63] Gregory Soyez Gavin P. Salam. A practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 0705:086, 2007.
- [64] Ryan Atkin. Review of jet reconstruction algorithms. *Journal of Physics: Conference Series*, 645(1):012008, 2015.

- [65] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.
- [66] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [67] M. Bahr, M.A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer, P. Richardson, M.H. Seymour, A. Sherstnev, J. Tully, and B.R. Webber. Herwig++ Physics and Manual. *Eur. Phys. J. C*, 58:639, 2008.
- [68] CMS Collaboration. Studies of jet quenching using isolated photon+jet correlations in PbPb and pp collisions at $\sqrt{s[\text{NN}]} = 2.76$ TeV. *Phys. Lett. B*, 718:773, 2013.
- [69] G. D’Agostini. A multidimensional unfolding method based on Bayes’ theorem. *Nucl. Instr. Meth. A*, 362:487, 1995.
- [70] A. Höcker, V. Kartvelishvili. SVD approach to data unfolding. *Nucl. Instr. Meth. A*, 372:469, 1996.
- [71] Tim Adye. Unfolding algorithms and tests using RooUnfold. In Harrison Prosper and Louis Lyons, editors, *PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, page 313, Geneva, Switzerland, 2011.
- [72] CMS Collaboration. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.
- [73] W. Verkerke and D. Kirkby. The RooFit toolkit for data modeling. In *2003 Computing in High Energy and Nuclear Physics (CHEP03)*, La Jolla, Ca., 2003.
- [74] Gary John Barker. b-quark physics with the LEP collider. *Springer Tracts Mod. Phys.*, 236:170, 2010.
- [75] LHCb Collaboration. Identification of beauty and charm quark jets at LHCb. *JINST*, 10:P06013, 2015.
- [76] CMS Collaboration. Evidence of b-jet quenching in PbPb collisions at $\sqrt{s[\text{NN}]} = 2.76$ TeV. *Phys. Rev. Lett.*, 113:132301, 2014.
- [77] J. Huang, Z. B. Kang, and I. Vitev. Inclusive b-jet production in heavy ion collisions at the LHC. *Phys. Lett. B*, 726:251, 2013.
- [78] Rishi Sharma and Ivan Vitev. High transverse momentum quarkonium production and dissociation in heavy ion collisions. *Phys. Rev. C*, 87:044905, 2013.
- [79] Constantin Loizides. Experimental Overview on small collision systems at the LHC. In *Proceedings of the XXVth International Conference on Ultrarelativistic Nucleus-Nucleus Collisions*, Kobe, Japan, 2016.
- [80] E.R. Cazaroto, V.P. Goncalves, and F.S. Navarra. Heavy quark production and gluon saturation in double parton scattering at lhc. *Phys. Rev. D*, 88:034005, 2013.
- [81] Serguei Chatrchyan et al. Inclusive b-jet production in pp collisions at $\sqrt{s}=7$ TeV. *JHEP*, 1204:084, 2012.

- [82] CMS Collaboration. Charm-tagged jet production in pPb collisions at 5.02 TeV and pp collisions at 2.76 TeV. CMS Physics Analysis Note CMS-HIN-15-012, CMS, 2015.
- [83] ALICE Collaboration. Suppression of high transverse momentum D mesons in central Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 09:112, 2012.
- [84] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A Positive-Weight Next-to-Leading-Order Monte Carlo for Heavy Flavor Hadroproduction. *JHEP*, 0709:126, 2007.
- [85] ATLAS Collaboration. Measurement of the inclusive and dijet cross-sections of b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 71:1846, 2011.
- [86] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 0411:040, 2004.
- [87] Andy Buckley, Jonathan Butterworth, David Grellscheid, Hendrik Hoeth, Leif Lönnblad, James Monk, Holger Schultz, and Frank Siegert. Rivet user manual. Technical Report MCnet/10/03, 2013.

APPENDIX

A. COORDINATE SYSTEM

The origin of the CMS frame of reference is the collision point, where the z -axis is along the beamline, pointing west, the x -axis is horizontal, pointing south (toward the center of the LHC) and the y -axis is pointing vertically upwards. These axes are shown schematically in Fig. A.1.

In addition to an (x, y, z) coordinate system, a more natural coordinate system is used that is partially cylindrical. Along with the usual z -axis, two vectors are defined as rapidity (y) and azimuth (ϕ), such that rapidity is defined as:

$$y = \frac{1}{2} \ln \frac{E + p_L}{E - p_L} \quad (\text{A.1})$$

where E and p are the energy and three-momentum of the particle, respectively, and the azimuth is defined as an opening angle from the x -axis, where positive azimuth

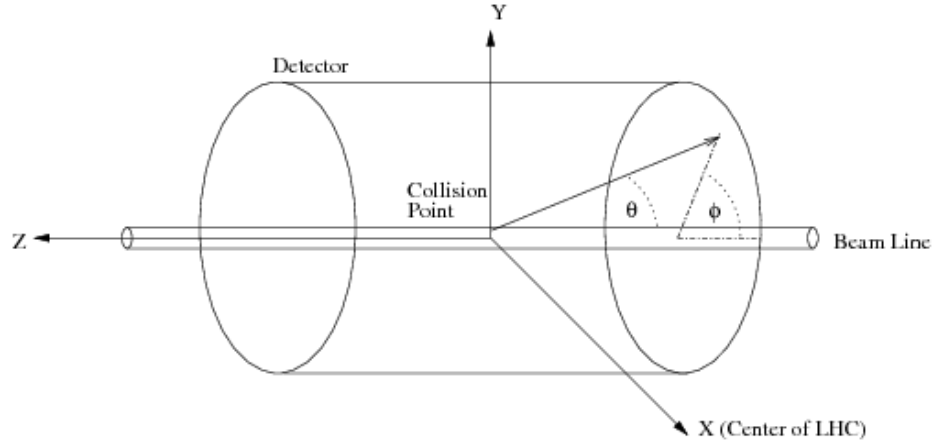


Figure A.1. Schematic of the x, y, z coordinate system, along with visual representations of the θ and ϕ angles.

is in the counterclockwise direction around the positive z -axis. An alternative to rapidity is the pseudorapidity, defined as:

$$\eta = \frac{1}{2} \ln \frac{p + p_L}{p - p_L} = -\ln \tan \frac{\theta}{2} \quad (\text{A.2})$$

where θ is the vertical angle relative to the beam pipe. Pseudorapidity is useful because particle momentum is measured more often than energy and is equivalent to rapidity so long as the particle mass is negligible ($E \gg m$ and $E \approx p$). The advantage to the η, ϕ coordinate system rather than an x - y system is that η is Lorentz-invariant for boosts along the beam axis so long as the mass of the particle is negligible and ϕ is Lorentz-invariant for all boosts along the beam axis.

B. CONSISTENCY OF HEAVY-FLAVOR TAGGER INPUT BETWEEN SIMULATION AND DATA

This appendix shows the consistency of inputs to the flavor taggers between MC and data. The majority of the systematic uncertainty in the b- and c-jet results derive implicitly from discrepancies seen here. While the PYTHIA simulation predicts most kinematic quantities accurately, PYTHIA is only known to predict jet fragmentation behavior to leading-order and, as such, inherently next-to-leading-order processes like heavy flavor production will not be modeled especially accurately in these simulations. Therefore, a great many data-to-simulation checks were made to ensure generator coherence with data. While they do not merit inclusion in the body of this dissertation, they are important to include for completeness. We also note that many of the generator discrepancies from data have been reweighted to match distributions of data as closely as possible.

B.1 Secondary Vertex

These plots reflect the coherence of secondary vertex distributions between data and MC in pp, pPb and PbPb. Both pp and pPb share identical reconstruction procedures, so plots are very similar between the two collision systems, modulo underlying event fluctuations. Above a jet p_T of 50 GeV, underlying event fluctuations do not contribute much to jet kinematics. We observe good consistency over a wide range of vertex variables, including distributions of vertex flight distance and number of tracks per secondary vertex. We observe minor differences in the number of secondary vertices found for each jet, though these differences are negligible and do not dramatically affect the template shapes. Most importantly, we observe distributions of the Simple Secondary Vertex (SSV) tagger discriminators, both high-efficiency

(SSVHE) and high-purity (SSVHP) show good closure, within about one standard deviation over the entire discriminator range.

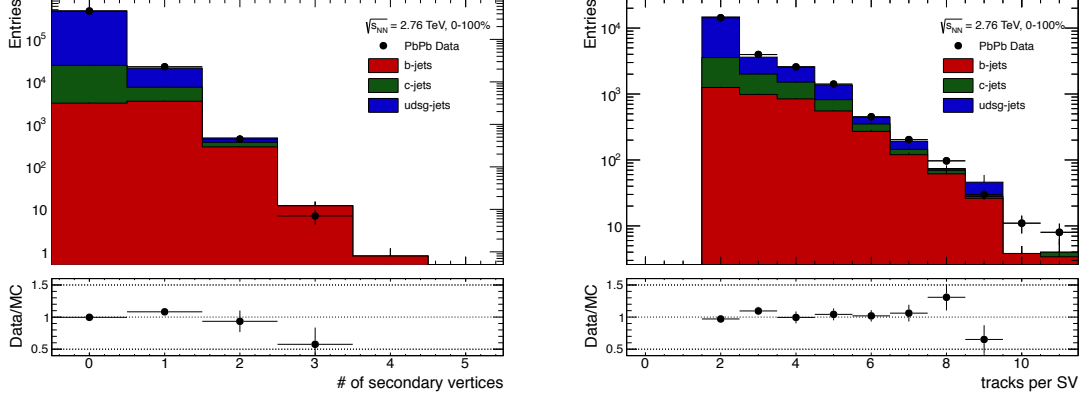


Figure B.1. The distributions of the number of reconstructed secondary vertices (left) and number of associated tracks per secondary vertex (right) in PbPb collisions.

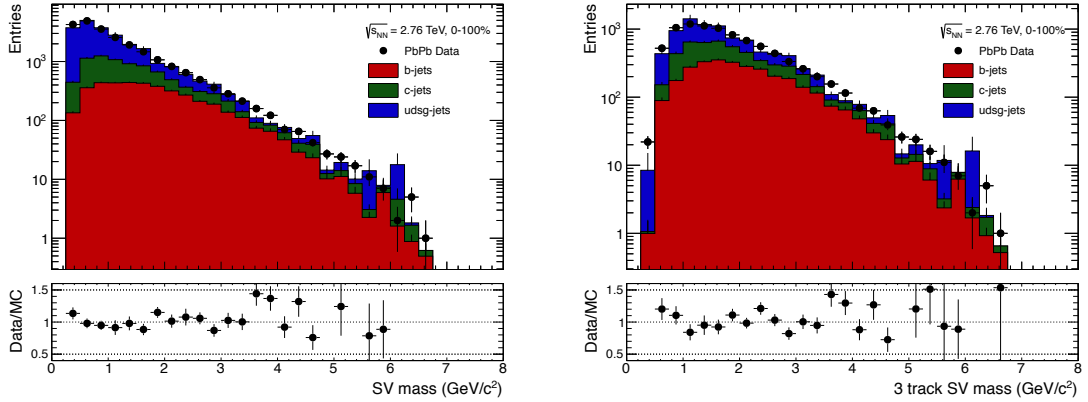


Figure B.2. Secondary vertex mass distributions for all vertices (left) and those vertices reconstructed from at least three tracks (right) in PbPb collisions.

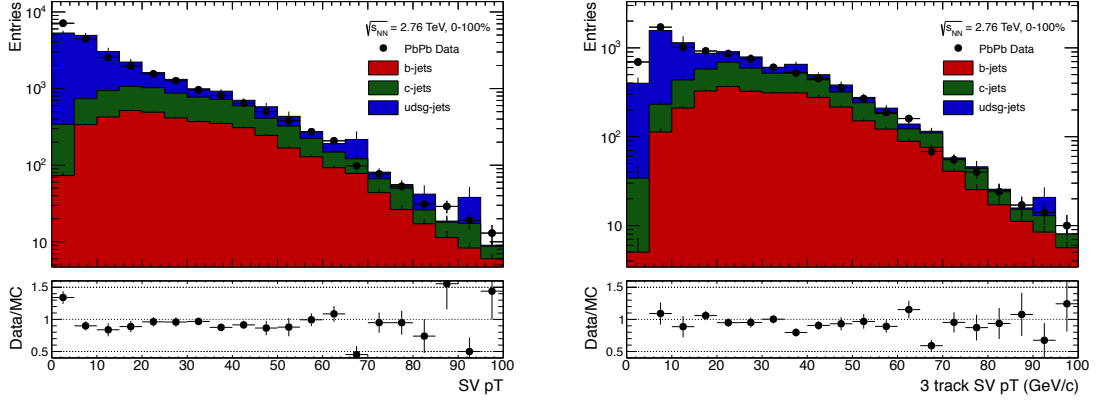


Figure B.3. The p_T distributions of secondary vertices for all reconstructed vertices (left) and for vertices reconstructed from at least three tracks (right) in PbPb collisions.

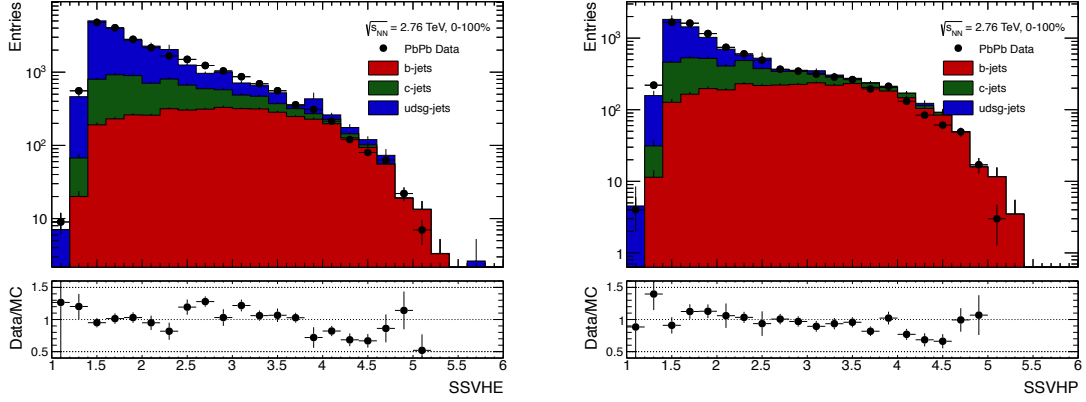


Figure B.4. The distributions of the Simple Secondary Vertex High Efficiency (left) and the Simple Secondary Vertex High Purity (right) discriminators in PbPb collisions.

B.2 Single-Track Displacement

These plots reflect the coherence of individual displaced track kinematics between data and MC in PbPb collisions and also in pp and pPb collisions. We observe that most track kinematic quantities are in good agreement with simulation over a wide

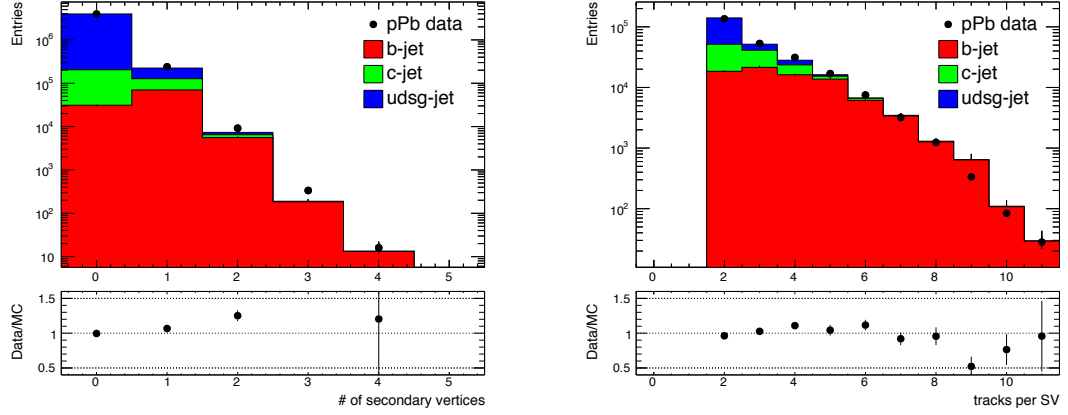


Figure B.5. The distributions of the number of reconstructed secondary vertices (left) and number of associated tracks per secondary vertex (right) in pPb collisions.

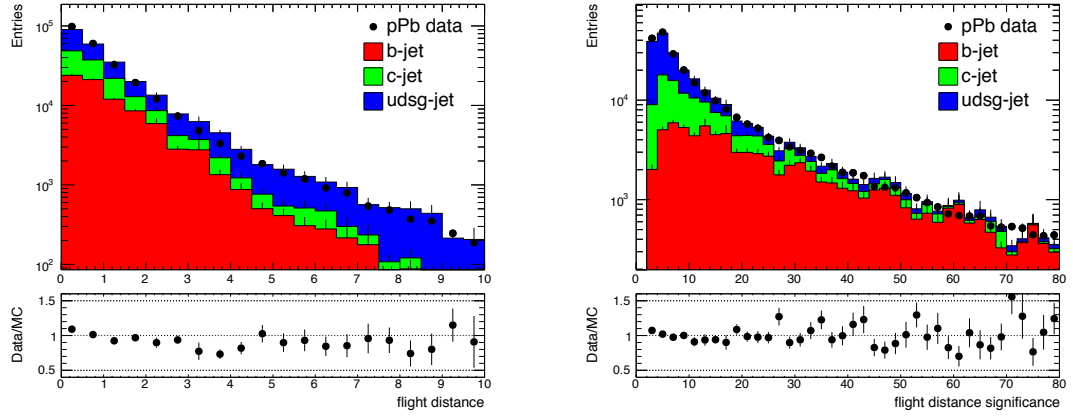


Figure B.6. Flight distance (left) and flight distance significance (right) distributions of reconstructed secondary vertices.

range of single track variables. We especially note reasonable coherence between data and simulation for the distributions of track decay length and the track distance to the jet axis which feed into the Jet Probability (JP) tagger. Distributions of these quantities are not as important as the secondary vertex distributions, however, as

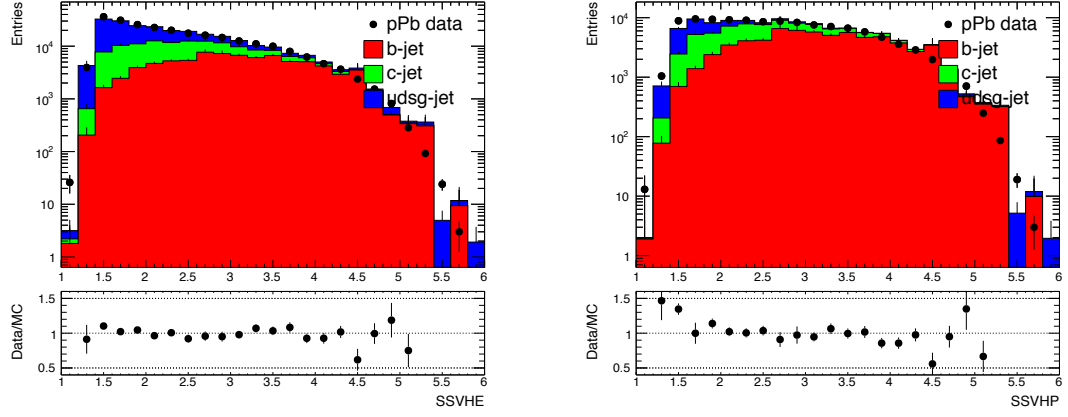


Figure B.7. The distributions of the Simple Secondary Vertex High Efficiency (left) and the Simple Secondary Vertex High Purity (right) discriminators.

the JP tagger can still be reweighted using data to account for any deviation of the simulated sample from distributions of single-track decay kinematics in data.

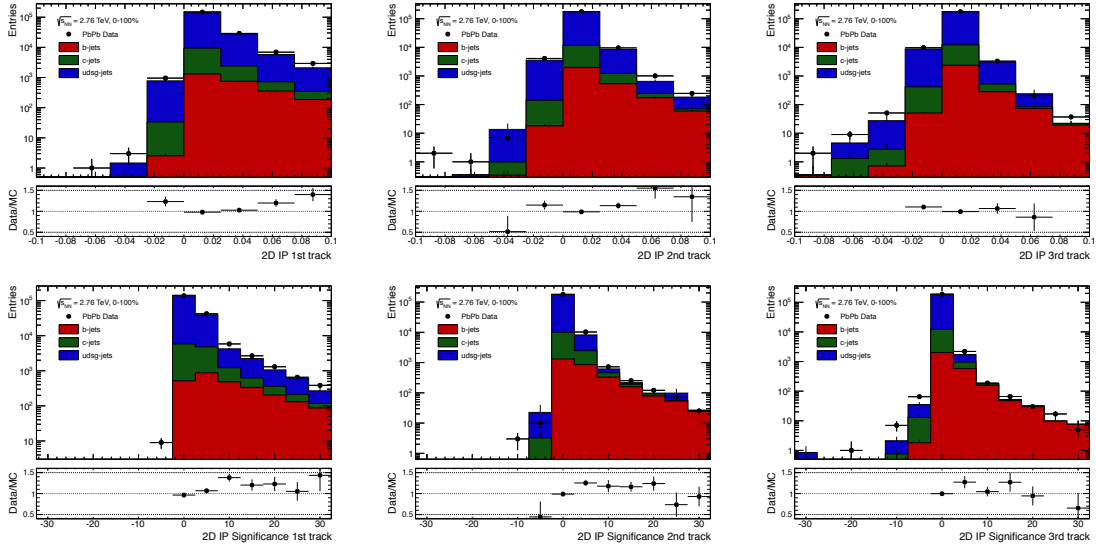


Figure B.8. The 2-D impact parameter (top) and 2-D impact parameter significance (bottom) for the first (left), second (middle) and third (right) most significant track in PbPb collisions.

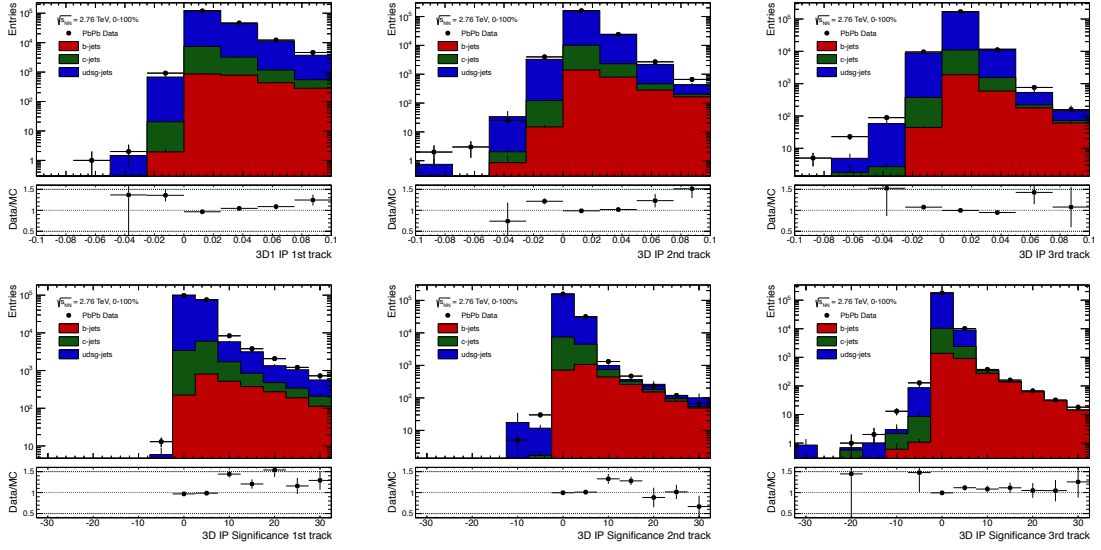


Figure B.9. The 3-D impact parameter (top) and 3-D impact parameter significance (bottom) for the first (left), second (middle) and third (right) most significant track in PbPb collisions.

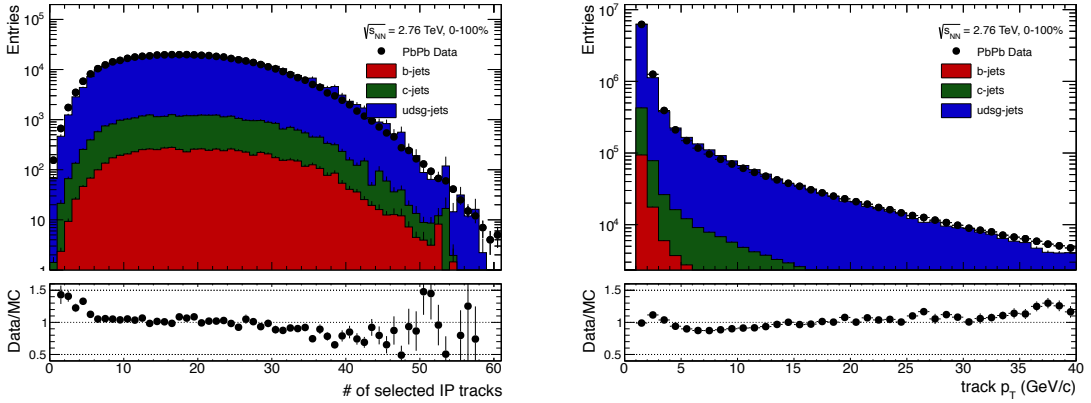


Figure B.10. The distribution of the number of selected jet-associated tracks (left) for the impact parameter-based taggers and their p_T distribution (right) in PbPb collisions.

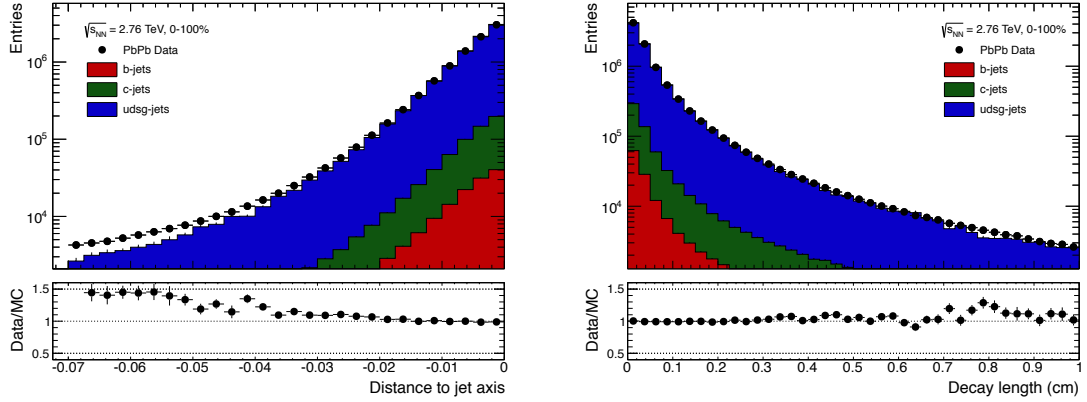


Figure B.11. The distributions of the distance from tracks to the axis of their associated jet and track decay length in PbPb collisions.

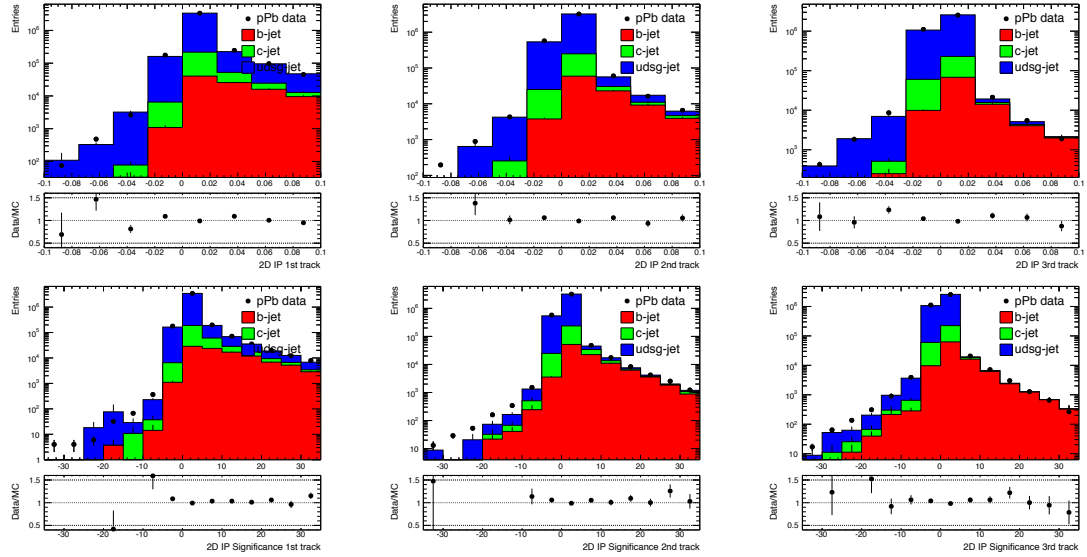


Figure B.12. The 2-D impact parameter (top) and 2-D impact parameter significance (bottom) for the first (left), second (middle) and third (right) most significant track in pPb collisions.

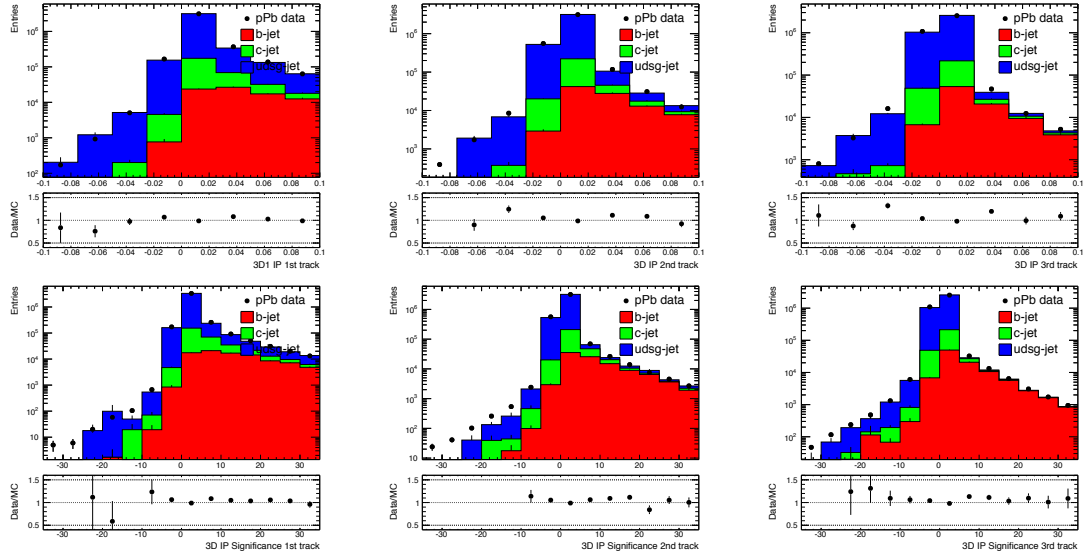


Figure B.13. The 3-D impact parameter (top) and 3-D impact parameter significance (bottom) for the first (left), second (middle) and third (right) most significant track in pPb collisions.

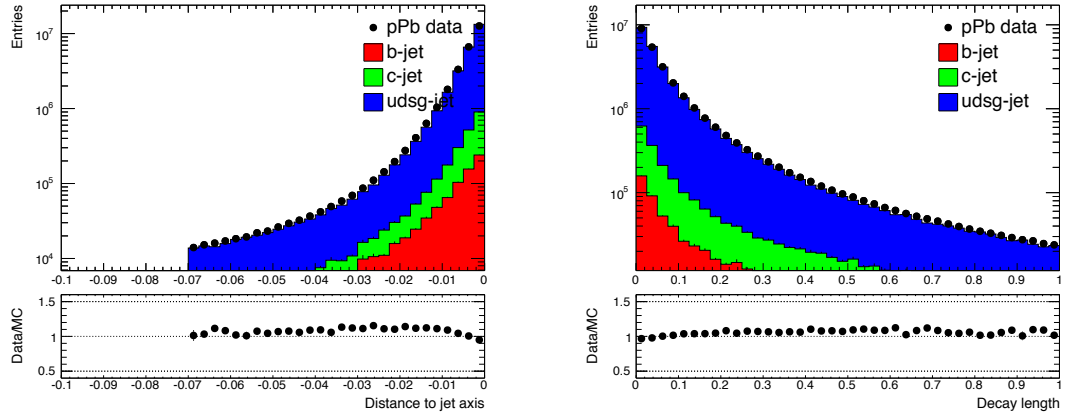


Figure B.14. The distributions of the distance from tracks to the axis of their associated jet and track decay length in pPb collisions.

C. ATTEMPTS TO INCLUDE NLO GENERATOR COMPARISONS

It is unfortunate that comparisons to generators other than PYTHIA were not shown in the analyses described in this dissertation, however an attempt was made to try and include comparisons to a next-to-leading order (NLO) generator known as POWHEG [84]. Analyses of b-jets in pp at 7 TeV confirm that the generators POWHEG and HERWIG match the b-jet spectrum more accurately than does PYTHIA, as these generators are inherently next-to-leading order and include a consistent treatment of NLO corrections during jet production. While some shortcuts are taken during fragmentation and hadronization, the jet yields are often times more accurate than LO or log-leading order generators.

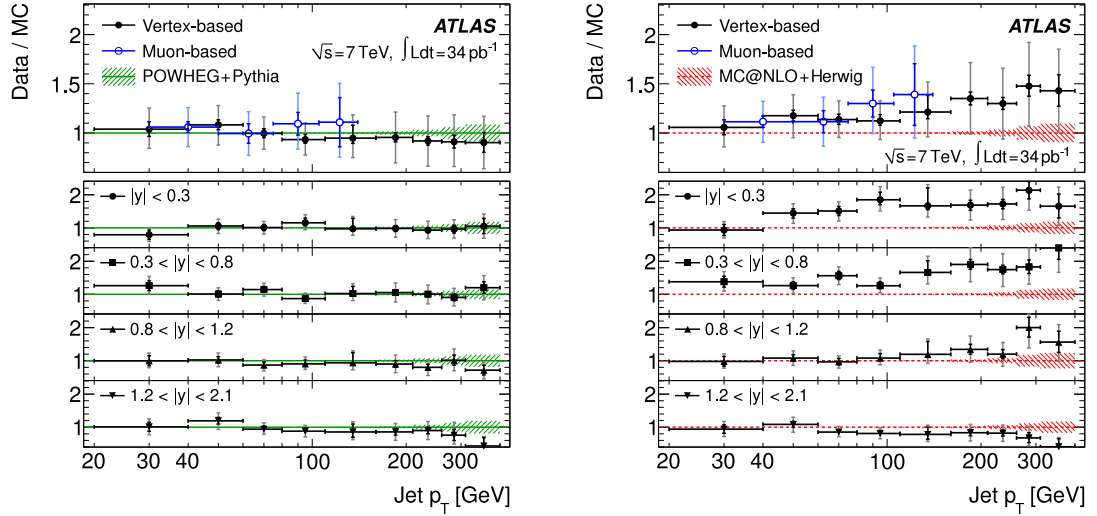


Figure C.1. The cross-section of b-jets at 7 TeV is compared to predictions from the POWHEG (left) and MC@NLO (right) generators [85]. The adherence of POWHEG to the data is much better than MC@NLO.

During the b-jet study, POWHEG was targeted as the preferred model to use, due to an ATLAS study of the b-jet cross-section [85]. Figure C.1 shows the coherence of POWHEG (left) and MC@NLO (right) to 7 TeV pp data. ATLAS observes a much better description of the data from the POWHEG generator than from MC@NLO, so it is expected that the generator ought to represent the 5 TeV pp b-jet cross section well.

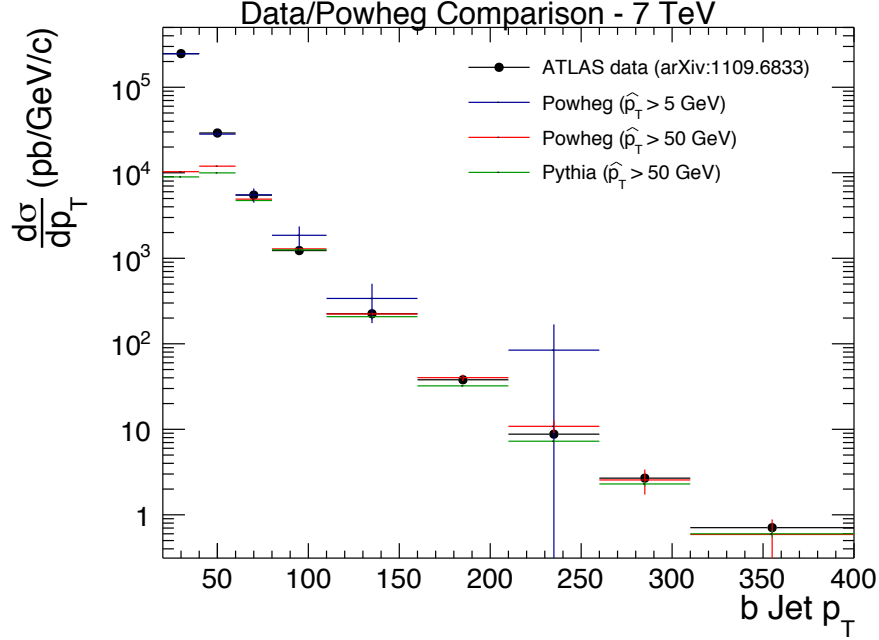


Figure C.2. The cross-section of b-jets at 7 TeV is compared to predictions from the POWHEG and PYTHIA generators, as made by the author. POWHEG matches the data to within 5%.

The technical procedure to generate a POWHEG event invokes jet fragmentation using the Lund string fragmentation from PYTHIA. Therefore, the standard procedure is to use the POWHEG box [86], HVQ module [84] to generate stand-alone b-jet events that are then externally integrated into the PYTHIA generator. Finally, a RIVET procedure [87] is used to run the FastJet jet reconstruction package and properly weight the jet production by the cross-section obtained from the POWHEG simulation. As POWHEG is inherently NLO, a parton distribution function is re-

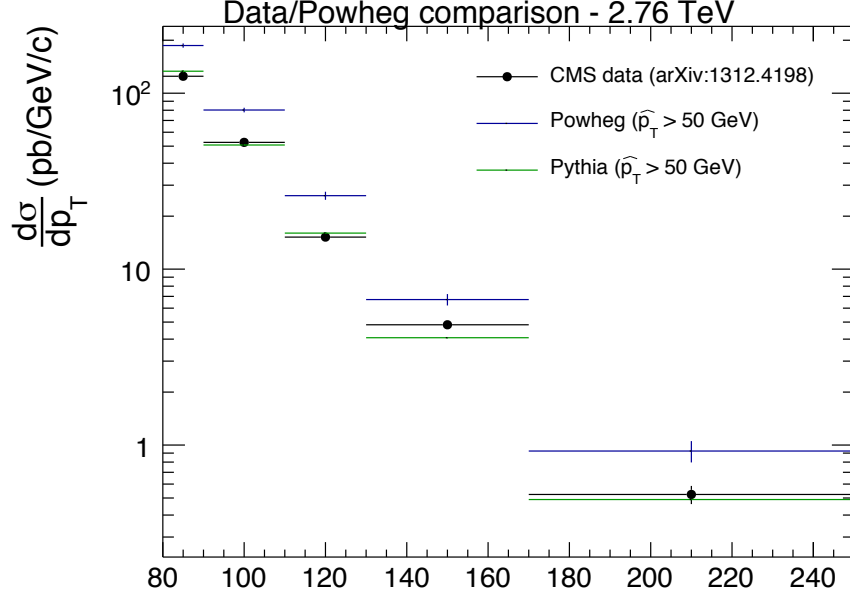


Figure C.3. The cross-section of b-jets at 2.76 TeV is compared to predictions from the POWHEG and PYTHIA generators, as made by the author. POWHEG predictions deviate significantly from the 2.76 TeV data.

quired as an input to the program. A wide variety of parton distribution functions were tried, but none of them were able to fit the pp b-jet cross section at 7 and 2.76 TeV simultaneously. The PDFs used included the CT10, MSTW2008nlo and NNPDF3.0 sets. Examples of the distributions of the b-jet cross section in data, PYTHIA and POWHEG are shown in Figs. C.2 and C.3 for 7 TeV and 2.76 TeV, respectively. Both plots have the same generation settings and procedures, so it is fairly surprising that while the 7 TeV data matches within 5%, the b-jet cross section at 2.76 TeV is overpredicted by almost a factor of 2.

Finally, there were some technical issues with weighted event generation, such that a minimum-bias type b-jet production needs to be run, since the usual method of generating weighted jets with POWHEG (the use of a Born suppression factor) is not implemented in the heavy quark module of POWHEG. Due to these complications and the non-coherence with b-jet data at 2.76 TeV and 7 TeV simultaneously, POWHEG's

ability to reproduce data at 5 TeV is not well-constrained and therefore cannot be trusted as an accurate representation of 5 TeV data.

VITA

VITA

Kurt Jung was born in Iowa City, IA in 1988. He has always had a love of science, primarily spurred on by an awe of the achievements of the US space program. This continued through high school, where he was particularly interested in physics. After high school graduation in 2006, he attended the University of Notre Dame, where he decided to major in physics. An undergraduate research project in particle physics under Dr. Michael Hildreth convinced him to continue to graduate school in this area. After graduation with a B.S. in Physics in 2010, he attended Purdue University, where he joined the experimental nuclear high-energy group under Dr. Wei Xie, as a member of the STAR Collaboration and later the CMS Collaboration. He will graduate from Purdue University with a Ph.D. in Nuclear Physics in 2016.