Measurement of Charged Hadron Distributions in Beauty and Charm Quark Tagged Jets at the Large Hadron Collider beauty Experiment

by

Kara R. Mattioli

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2022

Doctoral Committee:

Professor Christine Aidala, Chair
Professor Myron Campbell
Professor Aaron Pierce
Associate Professor Thomas Schwarz
Professor Alexander Thomas

Kara R. Mattioli

krmatt@umich.edu

ORCID iD: 0000-0003-2222-7727

To my parents, Steven and Dorothy Mattioli, who have encouraged and supported my love of science for my entire life. I could not have done this without you.

ACKNOWLEDGEMENTS

It is no exaggeration to say that there are truly countless people who have helped me along the path to my PhD. I will name several here, but this is by no means an exhaustive list. First and foremost, I am so grateful to have had the opportunity to study QCD with Christine. I feel incredibly lucky and grateful to have had a PhD advisor who has not only taught me so much about science, but who has also helped me become a better person. She is an excellent role model not only as a phenomenal scientist but also as one of the most kind-hearted people I have ever had the pleasure of knowing, who always is thinking of the needs of others. The example she set for me is something that extends far beyond my professional life. I'm also very thankful that she took a chance on me as a graduate student when I switched from a previous research career in condensed matter physics to particle physics - that is not something most advisors would have done!

I would also like to thank the Aidala research group at the University of Michigan, who feel more like family than colleagues at this point. Joe Osborn, Sookhyun Lee, Nicole Lewis, Catherine Ayuso, Desmond Shangase, Dillon Fitzgerald, Dylan Manna, and Cynthia Nunez have all been so supportive and have taught me so much over the years. Joe and Nicole helped me get started doing research in Christine's group and patiently answered my (many) questions. I'm thankful for Dylan's optimism and confidence in me which has encouraged me on many occasions. I'm also very grateful that I had the chance to know Jordan Roth. In addition to teaching me so much about QCD and helping me get started on LHCb, Jordan was also a friend who truly

brightened my life. While I am devastated that he is no longer with us, I will always treasure the memories I have with him, which include so many happy times such as traveling to CERN and Long Island, NY together, kayaking, and practicing French together. I miss him so much.

I'm also so grateful for the friends, professors, and staff at the University of Michigan Physics Department who have had a significant impact on my life during my time in Ann Arbor. Elizabeth Drueke, Rachel Owen, Johnathon Jordan, Owen Puls, Garrett Merz, and Eric Gonzalez were all wonderful friends to go through graduate school with. I'm especially thankful for Matthew Day, who shared many of the highs and lows of graduate school with me and was a constant source of encouragement to me. I would like to thank Professor Liuyan Zhao for mentoring me early in my graduate career, and Professor Aaron Pierce, from whom I learned so much during his graduate quantum and particle physics classes and who was supportive when I changed fields from condensed matter to particle physics. I would like to thank Ramon Torres-Isea for inspiring me to pursue a career in nuclear physics. And last but certainly not least, I am so grateful for the staff who have helped me navigate the PhD requirements and provided so much encouragement along the way: Chrissy Zigulis, Elise Bodei, Nitesh Singh, and Lauren Segall all helped me so much.

It has truly been a highlight of my life to work on the LHCb experiment and I have greatly enjoyed working with my wonderful LHCb collaborators. I would especially like to thank Stephen Farry, Lorenzo Sestini, and James Mead for their help, feedback, and support as I have progressed with my analysis. Eliane Epple was a kind and welcoming office mate to me during my time at CERN and I think I am indebted to her for many coffees. I would also like to thank Emilie Maurice for her kindness and support during the difficult time in my life corresponding to Jordan's passing and me finishing my PhD thesis.

I'm also thankful for the many friends and family who have supported me during my PhD. In particular I would like to thank Becca Huff, Jamie Larimer, Kyle Lee, and Nima Zardoshti. Nima especially provided me with endless encouragement during the last stretch of my PhD, which I am so thankful for and will always remember. And finally, I would like to thank my parents and my sister Rachel for their immense love and support for me. They have been my biggest cheerleaders, and I am so grateful to have had them with me through every step of this process.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Hadronization, the process by which a color-charged quark or gluon becomes bound within a color-neutral hadron, remains a fundamental yet poorly understood process in quantum chromodynamics. Current theoretical descriptions of hadronization rely on the use of fragmentation functions, which parameterize the probability for a quark or gluon to become bound within a specific hadron. While fragmentation functions parameterize the hadronization process beginning with the initial state parton, experiments can only detect the final state hadrons produced. One solution to address the current discrepancy between theoretical and experimental approaches to hadronization is to study systems in which information about both the initial state parton and final state hadrons is experimentally accessible. Heavy-flavor-tagged jets, collimated sprays of high-energy particles originating from the hadronization of a beauty or charm quark, are one example of such a system as they provide access to the flavor content of both the initial and final state particles. Distributions of the longitudinal momentum fraction z, transverse momentum j_T, and radial profile r of charged hadrons in heavy-flavor-tagged jets were measured using 13 TeV proton-proton collision data taken by the Large Hadron Collider beauty (LHCb) experiment. The distributions were compared to those measured in Z-tagged jets, which are primarily light-quark-initiated, to study differences between the hadronization mechanisms of heavy and light quarks. This measurement complements previous measurements of single-hadron heavy flavor fragmentation functions and constitutes another approach with which to gain insight into the mechanisms underlying heavy flavor hadronization.

CHAPTER I

Introduction

1.1 Quantum Chromodynamics

Quantum chromodynamics (QCD) is the quantum field theory of the strong nuclear force, one of the four fundamental forces in nature. QCD describes the properties and interactions of quarks and gluons, fundamental particles which combine to form a multitude of bound states in QCD called hadrons. The most common hadrons are protons and neutrons, which themselves comprise the nuclei of atoms. The strong nuclear force is responsible for keeping quarks and gluons tightly bound within hadrons and protons and neutrons tightly bound within nuclei, therefore its dynamics play a vital role in giving rise to the vast majority of the visible matter in our universe. Several features of QCD distinguish it from the other quantum field theories of the Standard Model, yielding interesting and complex phenomena.

1.1.1 Confinement

QCD is so named for the "color" charge that quarks and gluons, collectively called partons, have in addition to the familiar electric charge; electric charge is carried by quarks but not gluons. Quarks can have a color charge of red, green, or blue, while gluons carry one of these colors plus an anticolor. The use of color to describe QCD interactions was chosen as an analogy to describe the fact that bare

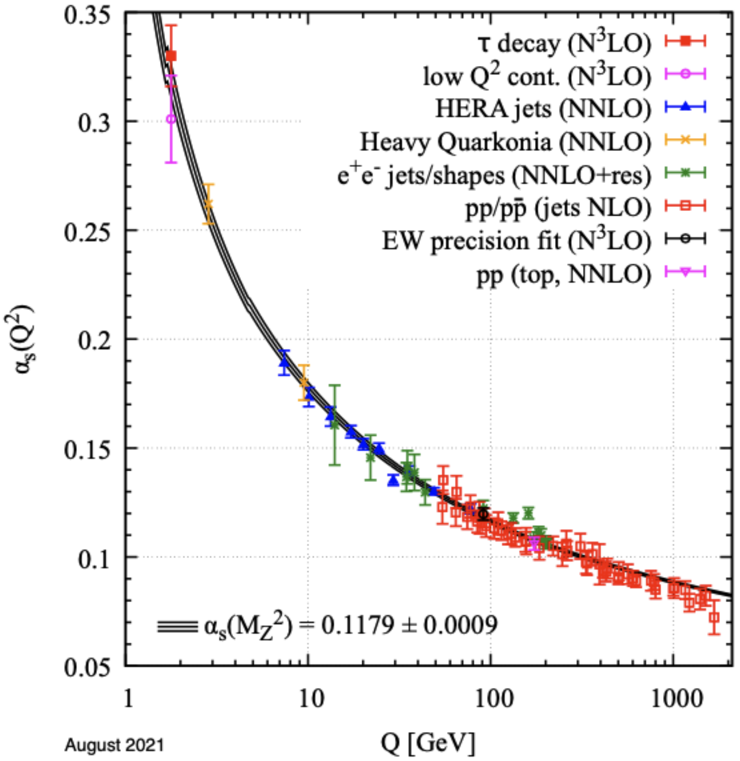

Figure 1.1: The running of the strong coupling constant α_S as a function of the momentum transfer Q [1].

color charges are never observed, and instead quarks and gluons are only observed in color-neutral bound states consisting of all three color charges, or a color and anti-color. This principle is known as confinement in QCD. Six different types or “flavors” of quarks exist in QCD, all of which can be confined within hadrons except for the heaviest quark, called the top quark, which decays faster than the characteristic hadron formation time. The different quarks, in order from heaviest to lightest, are the top ($m_t \approx 173$ GeV), beauty (also known as bottom, $m_b \approx 4.18$ GeV), charm ($m_c \approx 1.27$ GeV), strange ($m_s \approx 93$ MeV), down ($m_d \approx 4.67$ MeV), and up ($m_u \approx 2.16$ MeV). The beauty and charm quarks are called heavy flavor quarks because they have significantly larger masses than the strange, down, and up quarks, which are collectively called light flavor quarks.

1.1.2 Asymptotic Freedom

QCD exhibits asymptotic freedom, the property that the strong force becomes weak at short distances, corresponding to large energy scales, and extremely strong at large distances, corresponding to small energy scales. Asymptotic freedom is quantitatively described by the “running” of the strong coupling constant α_S as a function of energy [21, 22]. Figure 1.1 displays the measured constant α_S as a function of the momentum transferred in a QCD scattering event, Q [1]. With increasing Q, α_S becomes smaller, facilitating perturbative expansions used to compute cross sections and other observables in QCD. At small values of Q, α_S becomes too large for perturbative expansions to converge. QCD processes occurring below an energy scale of approximately 1 GeV are considered to be in this non-perturbative regime [1]. Hadronic structure, the description of hadrons in terms of the positions and momenta of their constituent partons, and hadronization, the process by which partons combine to form hadrons, are examples of QCD phenomena that cannot be described directly with perturbative expansions.

Since confinement implies that only hadrons rather than quarks or gluons can be detected experimentally, and as hadron structure and formation cannot be calculated perturbatively, every QCD cross section contains a non-perturbative component. This difficulty is circumvented by the use of factorization theorems in QCD, which enable cross section calculations by separating the non-perturbative components from those that can be calculated perturbatively. Different factorization theorems exist for different collision systems [23]. In pp collisions, the factorization theorem for inclusive hadron production takes the following form:

$$d\sigma^{pp\rightarrow hX} = \sum_{a,b,c} f_a(x_a,\mu) \otimes f_b(x_b,\mu) \otimes H^c_{ab}(x_a, x_b, z, \mu) \otimes D^h_c(z,\mu) \tag{1.1}$$

In Equation 1.1, $f_a(x_a,\mu)$ and $f_b(x_b,\mu)$ are functions describing the probability that a parton with momentum fraction x_a is selected from one of the protons and a parton with momentum fraction x_b is selected from the other proton. At leading order, these Parton Distribution Functions (PDFs) describe the probability that a parton within a hadron carries a certain fraction of the hadron's momentum. The object H_{ab}^{c} is the hard partonic cross section for partons a and b to produce parton c and is directly calculable in perturbation theory. Finally, $D_c^h(z)$ is a function that at leading order describes the number density of hadrons in parton c. The $D_c^h(z)$ is known as a Fragmentation Function (FF) and describes the formation of the final-state hadron h, produced as a consequence of the fact that parton c must be confined within a hadron, with fraction z of parton c's momentum. FFs will be described in more detail in Section 1.2.1. In all functions in Equation 1.1, μ is the factorization energy scale, typically taken approximately to be the transverse momentum (p_T) of the produced hadron h. The PDFs and FF in Equation 1.1 are the nonperturbative components of the cross section and must be extracted from fits to data. They are "universal" functions in the sense that they can be extracted in one collision process and used to calculate predicted cross sections for other processes. The extractions of the PDFs and FFs are performed in global fits to data collected by many experiments and encompassing a wide variety of kinematic regimes and collision systems [24–28].

1.1.3 Jets

A jet is a unique QCD phenomenon that occurs when a high-energy parton is briefly free from being confined within a hadron. A free parton can be produced from the radiation of another parton, or when its initial confining hadron is broken apart in a high-energy particle collision. The high-energy free parton will begin to radiate gluons, which can themselves split into additional gluons and quark-antiquark pairs. This process is called a "parton shower" or fragmentation. The resulting partons eventually combine into hadrons, forming a collimated spray of final-state particles called a "jet." Jets are detected experimentally with the use of jet algorithms, which identify hadrons that are close together in phase space and therefore consistent with originating from the hadronization of a single parton. Due to the conical nature of the radiation emitted by a high-energy parton, jets are defined with a "jet radius" R that describes the radius of the cone encapsulating the produced hadrons. The jet radius is defined in (η, ϕ) space, with the pseudorapidity η being proportional to the polar angle θ in the lab frame according to $\eta = -\ln[\tan(\frac{\theta}{2})]$ and ϕ being the azimuthal angle in the laboratory frame. The jet radius defines the distance scale between two particles in (η, ϕ) space for which they can be considered as part of the same jet. Jets are extremely useful tools for studying QCD, as they provide a proxy for the parton produced in the high-energy scattering event which is otherwise inaccessible experimentally.

1.2 Hadronization in QCD

Despite the abundance of hadrons formed in high-energy particle collisions, the process by which individual quarks and gluons, or partons, become confined in hadrons, called hadronization, remains a poorly understood area of QCD. A better understanding of hadronization would greatly increase our knowledge of QCD dynamics, specifically the interplay between partons and hadrons, and would have far-reaching implications for understanding the visible matter in our universe. Current knowledge of hadronization consists primarily of parameterizations of experimental measurements and many open questions. Qualitative descriptions of hadronization describe the color field between a splitting quark-antiquark pair stretching as the distance between the quarks increases until the energy density of the color field is sufficient to produce new quark-antiquark pairs that then combine with the splitting quarks to form color-neutral states [1]. Fundamental details about the dynamics of hadronization remain unclear from parameterizations and qualitative pictures, such as the distance scale over which it occurs, the dependence of the flavor composition of the produced hadrons on the initial parton or partons that hadronize, and the various physical mechanisms by which a color-neutral bound state can form.

1.2.1 Fragmentation Functions

Fragmentation functions (FFs) parameterize the probability of a parton to fragment into a specific hadron [29]. The FF describing the fragmentation of a parton q into a hadron h is commonly denoted as $D_q^h(z, \mu)$, with z defined as the fraction of the parton momentum carried by the hadron produced in the fragmentation process, and μ the factorization scale at which the FF is evaluated [29,30]. Collinear FFs depend only on z and μ, while FFs that depend additionally on the transverse momentum of the produced hadron relative to the fragmenting parton are known as transverse-momentum-dependent (TMD) FFs. TMD FFs are sensitive to the nonperturbative transverse momentum, with values of the order of the nonperturbative QCD scale Λ_{QCD}. Larger values of the transverse momentum can be generated perturbatively by hard gluon radiation. As they describe the nonperturbative hadronization process, FFs generally cannot be measured directly and instead must be extracted from fits to data. Jet FFs and heavy quark FFs, which will be discussed in Sections 1.2.2 and 1.3, respectively, are notable exceptions to this generalization.

Electron-positron collisions provide an ideal environment in which to measure FFs, as there are no PDFs in the initial state [29]. However, they also provide little sensitivity to the gluon FF, necessitating the measurement of FFs in the significantly more complex environment of pp collisions [29, 30]. Cross section measurements of identified hadrons in pp collisions are sensitive to FFs, but the additional convolution of the initial-state PDFs with the hard scattering cross section make extraction of the FFs more difficult.

1.2.2 Fragmenting Jet Functions

Pioneering theoretical and experimental work turned to jets as objects in which FFs could be more easily measured in pp collisions [30–34]. A new QCD factorization theorem was developed to describe the production of hadrons within jets, with the FF describing the inclusive production of a single hadron, $D_q^h(z,\mu)$, replaced by a Fragmenting Jet Function (FJF), $G_q^h(z, z_h, \omega_J R, \mathbf{j}_\perp, \mu)$ describing the production of a hadron within a jet [30, 31, 33, 35, 36]:

$$\frac{d\sigma^{pp\longrightarrow(jeth)X}}{dp_T d\eta dz_h d^2\mathbf{j}_\perp} = \sum_{a,b,c} f_a(x_a,\mu)\otimes f_b(x_b,\mu)\otimes H_{ab}^c(x_a,x_b,\eta,p_T/z,\mu)\otimes G_c^h(z,z_h,\omega_J R,\mathbf{j}_\perp,\mu) \tag{1.2}$$

In Equation 1.2, f_a and f_b are the PDFs of the initial colliding protons, and H_{ab}^c is the hard scattering cross section describing the production of parton c from partons a and b. In addition to depending on the z and μ found in single-hadron FFs, the FJF G_q^h also depends on the fraction of the jet momentum carried by hadron h in the jet, z_h and the large light-cone momentum of the jet multiplied by the jet radius, $\omega_J R$ [35,36]. In the FJFs, z describes the fraction of the fragmenting parton's momentum carried by the jet, while z_h refers to the fractional jet momentum carried by the hadron in the jet. As with the single hadron FFs, the FJFs can also be transverse-momentum-dependent by describing the production of a hadron in the jet with momentum z and transverse momentum $\mathbf{j}_\perp$ relative to the jet axis [36]. The FJFs are related to the standard collinear FFs $D_q^h(z,\mu)$:

$$G_i^h(z, z_h, \omega_J, \mu) = \sum_j \int_{z_h}^{1} \frac{dz_h^{'}}{z_h^{'}} J_{ij}(z, z_h^{'}, \omega_J, \mu) D_j^h(\frac{z_h}{z_h^{'}}, \mu) \tag{1.3}$$

with the use of matching coefficients $J_{ij}(z, z_h^{'}, \omega_J, \mu)$ that have been calculated in perturbative QCD up to next-to-leading order [35]. With this theoretical framework for describing hadron-in-jet production, measurements of hadron distributions in jets,

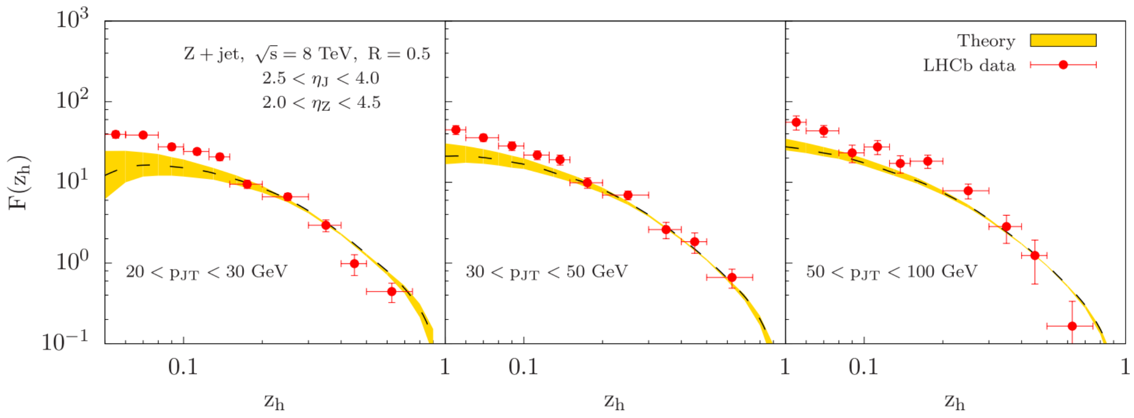

Figure 1.2: A comparison between perturbative QCD calculations and data for the jet fragmentation function describing the distribution of the longitudinal momentum fraction z of charged hadrons in Z-tagged jets [2].

known as Jet Fragmentation Functions (JFFs), can be used to extract collinear and TMD FFs with a higher precision than that obtained from single-hadron cross section measurements [35, 36]. The ATLAS and LHCb experiments at the LHC have measured charged hadron distributions in jets in pp collisions as a function of the longitudinal momentum fraction z, the transverse momentum relative to the jet axis j_T, and the radial distance r from the jet axis [32, 37]. Experimentally, these observables are defined by the following equations:

$$z = \frac{\mathbf{p}_{jet} \cdot \mathbf{p}_{hadron}}{|\mathbf{p}_{jet}|^2} \tag{1.4}$$

$$j_T = \frac{|\mathbf{p}_{jet} \times \mathbf{p}_{hadron}|}{|\mathbf{p_{jet}}|} \tag{1.5}$$

$$r = \sqrt{(\phi_{jet} - \phi_{hadron})^2 + (\eta_{jet} - \eta_{hadron})^2} \tag{1.6}$$

with $\mathbf{p}$ denoting the momentum of the jet or hadron, ϕ the azimuthal angle, and η the pseudorapidity [32, 37].

The theoretical framework using FJFs has been successfully used to describe experimental measurements, as seen in Figure 1.2 which compares perturbative QCD theoretical calculations to LHCb data [2]. Figure 1.2 displays $\mathrm{F}(z_h)$, which is defined theoretically as the differential cross section for the production of a Z boson and an associated jet with a hadron h measured within the jet divided by the differential cross section for the production of a Z boson and an associated jet, with no additional measurement of hadrons within the jet [2]. The numerator of $\mathrm{F}(z_h)$ depends on the FJF $G_i^h(z, z_h, \omega_J, \mu)$, which in turn depends on the standard collinear FFs $D_j^h(z_h, \mu)$ as shown in Equation 1.3 to provide the non-perturbative inputs to the cross section calculation.

In addition to providing better constraints on the collinear and TMD FFs, measurements of charged hadron distributions in jets also provide additional means with which to probe hadronization in QCD. Comparisons of the z, j_T, and r distributions measured by ATLAS and LHCb, for example, provided insight into the differences between gluon-dominated jets and quark-dominated jets [37]. LHCb measured charged hadron distributions in jets produced in association with a Z boson, which already tends to select more quark-initiated than gluon-initiated jets at the LHC because the $qg \rightarrow Zq$ partonic process dominates over the $q\bar{q} \rightarrow Zg$ process. In addition, due to the momentum imbalance needed in the hard scattering event to produce a jet within the forward acceptance of LHCb, the quark involved in the hard scattering is more likely to be a light quark as their PDFs are peaked at large fractional values of the initial proton momentum. In contrast to the predominantly light-quark-initiated Z-tagged jets at LHCb, inclusive jets measured by ATLAS at midrapidity are mostly gluon-initiated jets because the $gg \rightarrow gg$ process dominates. The comparisons of the charged hadron distributions between these samples revealed that the light-quark-initiated jets were more collimated than the gluon-initiated jets in both z and r [37].

1.3 Heavy Flavor Hadronization

Heavy flavor hadrons serve as an excellent probe for hadronization studies as several unique features of their decays facilitate their experimental detection, and their heavy mass makes them predominantly produced directly in the hard scattering. Consequently, heavy flavor hadrons provide direct access to the final state of the hadronization of a beauty or charm quark produced in the hard scattering. This is in contrast to light flavor hadrons, which can be produced from decays of heavier particles and from the hadronization of light quarks and gluons. In this section, previous experimental studies of heavy flavor hadronization in e^+e^-, $p\bar{p}$, and pp collisions are reviewed, with a specific focus on measurements involving jets. Studies probing the fragmentation of a heavy quark to a single beauty or charm hadron are reviewed, followed by a review of studies probing heavy quark fragmentation to multiple charged hadrons, and finally the heavy flavor dead cone and the current status of theoretical descriptions for hadronization in heavy flavor jets are discussed.

1.3.1 Fragmentation to Beauty Hadrons

The ability to experimentally identify heavy flavor hadrons, and the unique kinematic features allowed by the clean environment of e^+e^- collisions allowed for the inclusive B-hadron fragmentation function to be experimentally measured at the Large Electron-Positron (LEP) collider and the Stanford Linear Accelerator (SLAC) [3, 38–40]. With no parton distribution functions in the initial state, and in the case where the electron and positron beam energies are equal, the beam energy is a good approximation to the energy of the b-quark produced in an $e^+e^- \to b\bar{b}$ event. Therefore, the fraction of the initial b-quark energy carried by the final-state B-hadron, denoted x_B^{weak} by the LEP experiments, is simply determined by the energy of the B-hadron divided by the beam energy, according to the following equation [3]:

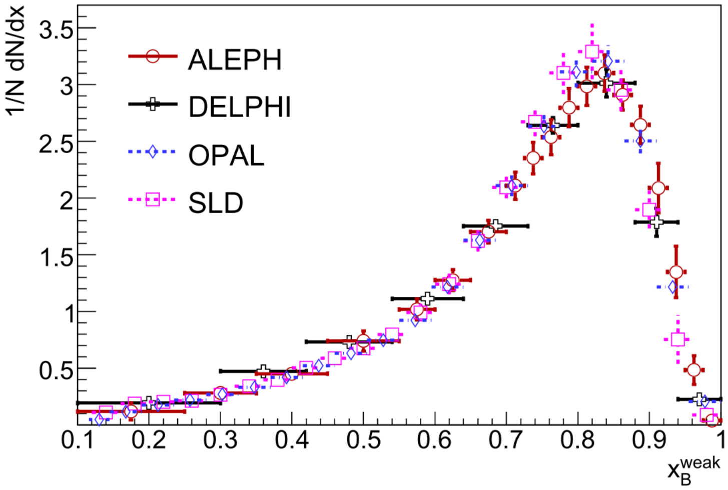

Figure 1.3: The inclusive B-hadron collinear fragmentation function measured by the ALEPH, DELPHI, and OPAL experiments at LEP, and the SLD experiment at SLAC [3].

$$x_B^{weak} = \frac{2E_B^{weak}}{\sqrt{s}} = \frac{E_B^{weak}}{E_{beam}} \tag{1.7}$$

The inclusive B-hadron fragmentation function measured in e^+e^- collisions is shown in Figure 1.3. The distributions measured at the separate experiments are all consistent with each other, and show a strong peak at a z value of approximately 0.84, indicating that the B-hadron produced during the high-energy hadronization process carries most of the momentum of the initial fragmenting b-quark. Global fits have been performed to the LEP and SLAC data to extract the b-hadron FF [4, 41]. Figure 1.4 shows the b-hadron FF extracted from a recent global fit on the left and the theoretical comparison to the data using the extracted FF on the right [4].

Measurements of B-hadron fragmentation functions in jets in pp collisions present a greater challenge than those in e^+e^- collisions due to the more complex initial state. However, they provide an important opportunity to test the universality of fragmentation functions. The ATLAS experiment at CERN recently measured the fragmentation function in jets for a b-quark fragmenting into a $B^{+/-}$ meson [5] by

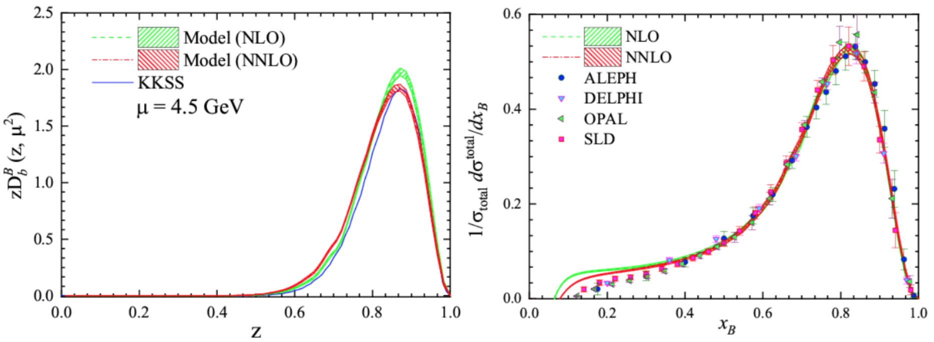

Figure 1.4: The extracted b-hadron FF from fits to e^+e^- data (left) and the theoretical comparison to the data using the extracted FF (right) [4]

fully reconstructing $B^{+/-} \rightarrow J/\psi\ K^{+/-}$ decays within jets [5]. The momentum of the fully reconstructed $B^{+/-}$ mesons was used to calculate the longitudinal momentum fraction z and transverse momentum relative to the jet axis, p_T^{rel}, as defined by the equations below:

$$z = \frac{\vec{p_B} \cdot \vec{p_{jet}}}{|\vec{p_{jet}}|^2} \tag{1.8}$$

$$p_T^{rel} = \frac{\vec{p_B} \times \vec{p_{jet}}}{|\vec{p_{jet}}|} \tag{1.9}$$

The observable z is used to measure the collinear $B^{+/-}$ fragmentation function, which is shown in the left plot of Figure 1.5. The collinear fragmentation function shows that the $B^{+/-}$ carries most of the jet momentum, consistent with the measurements of inclusive B-hadron production in e^+e^- collisions shown in Figure 1.3. To probe the scale dependence of the fragmentation function, the measurement was repeated in two additional bins of jet p_T. The peak position at a z of approximately 0.8 remained constant across all three jet p_T bins probed, but was observed to broaden to lower z values with increasing jet p_T, which was attributed to the increase of $g \rightarrow b\bar{b}$ splittings [5]. The right plot in Figure 1.5 shows the p_T^{rel} distribution for $B^{+/-}$ mesons in jets. While measurements at lower p_T^{rel} values are needed in order to extract the TMD $B^{+/-}$ fragmentation function, the ATLAS measurement shown in Figure 1.5

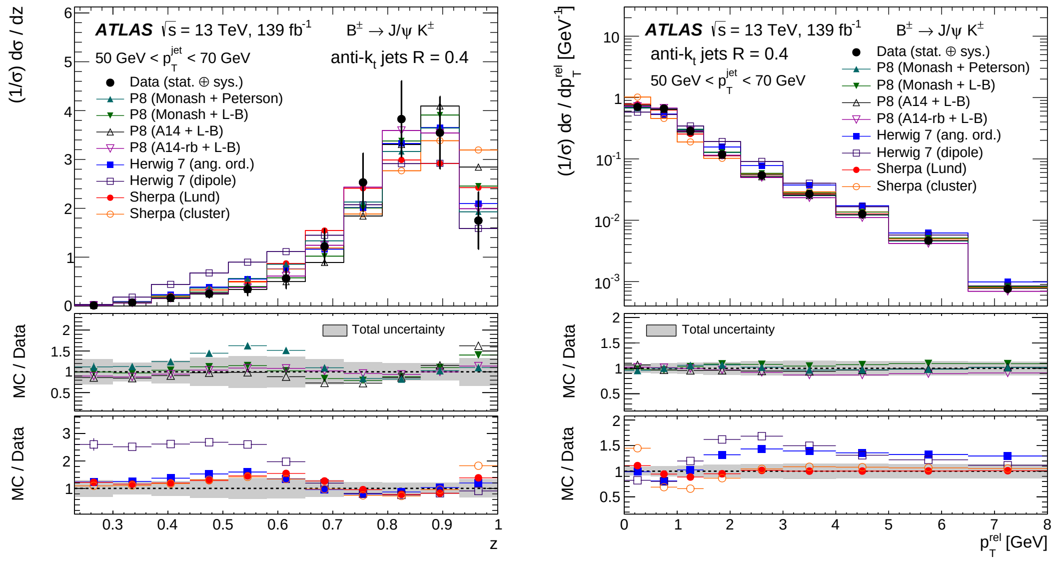

Figure 1.5: The longitudinal momentum fraction z (left) and transverse momentum relative to the jet axis p_T^{rel} (right) for fully reconstructed $B^{+/-}$ hadrons in jets. The data are compared to a variety of different MC generators and hadronization models [5].

lays the groundwork for future multi-dimensional studies of single beauty hadron fragmentation in jets. The ATLAS and CMS experiments have also studied inclusive B hadronization in jets by tagging b-jets from $t\bar{t}$ decays [42, 43]. Instead of fully reconstructing B hadrons in the jets, ATLAS used secondary vertices in the jets as proxies for the B-hadrons, while CMS used charm hadrons decayed from B-hadrons. ATLAS measured the z distribution and the fraction of the jet p_T carried by the SV in track-based jets, $p_{T,b}^{chg}/p_{T,jet}^{chg}$, while CMS measured the fraction of the charged particle jet p_T carried by the reconstructed charm hadron, $p_{T,c}/\sum p_T^{chg}$ [42, 43]. All distributions were strongly peaked towards high fractional values, consistent with the collinear B-hadron fragmentation functions described above.

1.3.2 Fragmentation to Charm Hadrons

More experimental measurements of fragmentation functions exist for charm hadrons than for beauty hadrons, primarily because charm is more abundantly produced than

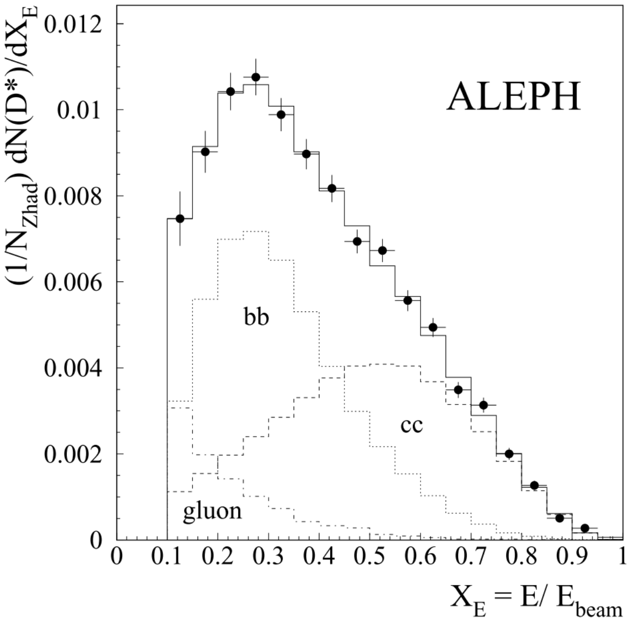

Figure 1.6: The scaled energy distribution of $D^{*+/-}$ mesons in e^+e^- collisions, compared to a fit of the three fragmentation contributions [6].

beauty and its simpler decays facilitate the reconstruction of charm hadrons in jets. Early measurements of charm FFs were performed in e^+e^- collisions at LEP and DESY and in jets in $p\bar{p}$ collisions at CERN and the Tevatron [6, 7, 44–46]. The production of $D^{*+/-}$ mesons was extensively studied in these early measurements as full reconstruction in the decay channel $D^{*+/-} \to D^0\pi^{+/-}$ provided a good signal to background ratio, and the reconstruction of an excited D-meson state provided a more accurate measurement of the energy of the charm hadron produced directly after hadronization [44]. As with the B-hadron FF measurements, the FF was measured in e^+e^- collisions as a function of the scaled energy of the $D^{*+/-}$ meson relative to the beam energy. Figure 1.6 shows the $D^{*+/-}$ FF measured by the ALEPH experiment at LEP, compared to the results of a template fit to determine the regions of the FF originating from $e^+e^- \to b\bar{b}$ events, $e^+e^- \to c\bar{c}$ events, and from gluon splitting to heavy quark pairs [6]. The $b\bar{b}$ template was compared to the X_E distribution from B-tagged events in data and found to describe the shape well, therefore strengthening

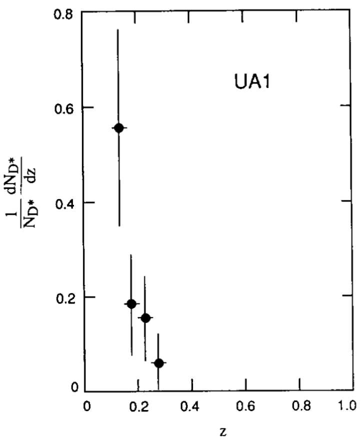

Figure 1.7: The collinear fragmentation function for $D^{*+/-}$ mesons measured in jets in $p\overline{p}$ collisions by the UA1 experiment at CERN [7]

the conclusion that the tail of the measured distribution in Figure 1.6 at high values of the scaled energy X_E is primarily from the hadronization of promptly produced charm quarks [6]. The shape of the $D^{*+/-}$ fragmentation function was found to differ between e^+e^- and $p\overline{p}$ collisions, as can be seen from comparing the $p\overline{p}$ result in Figure 1.7 and the e^+e^- result in the rightmost figure in Figure 1.8. The difference was attributed to the different processes dominating charm production between the collision systems [7]. In e^+e^- collisions, charm quarks are predominantly produced directly from the hard scattering, while in $p\overline{p}$ collisions the dominant charm production mechanism is from gluon splitting to $c\overline{c}$ or $b\overline{b}$ pairs [46]. Charm produced from gluon splitting was observed to have a softer fragmentation spectrum, with the peak of the fragmentation function shifted to a z value of approximately 0.1 [7].

Fragmentation functions of additional charm mesons were measured in e^+e^- collisions at DESY at a lower center of mass energy, 10.6 GeV, and were later repeated with higher statistics using data collected by the CLEO experiment at the Cornell Electron-positron Storage Ring (CESR) [8, 45]. Figure 1.8 shows the fragmentation functions measured by CLEO for promptly produced D^0, $D^{+/-}$, and $D^{*+/-}$ mesons as a function of the scaled momentum fraction $x_p = p_D/\sqrt{E_{beam}^2 - m_D^2}$ [8]. The

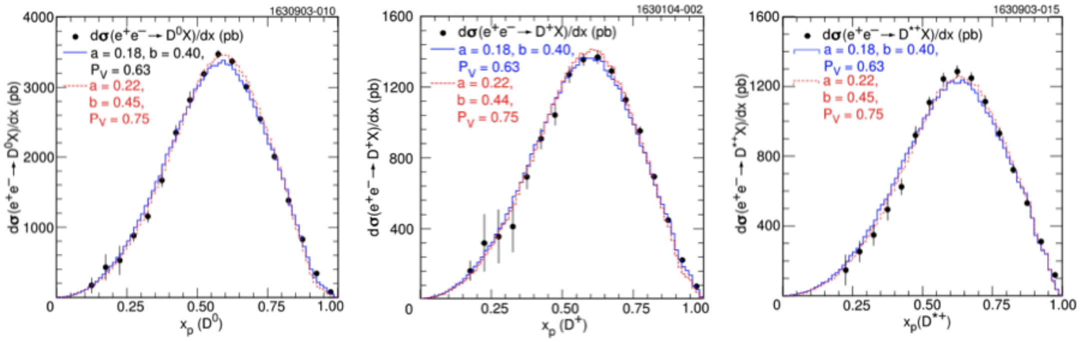

Figure 1.8: D^0 (left), $D^{+/-}$ (middle), and $D^{*+/-}$ (right) fragmentation functions measured in e^+e^- collisions [8].

charm FFs peak around an x_p of 0.6, indicating a hard fragmentation, but not as hard as the fragmentation observed for B hadrons which predominantly occurs at higher fractional momentum values.

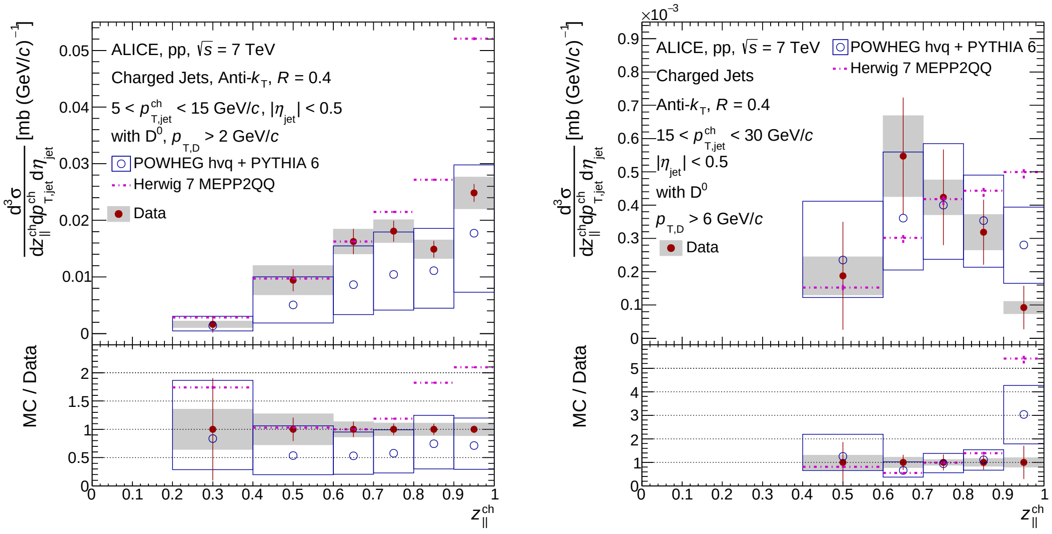

Figure 1.9: The longitudinal momentum fraction z of D^0 mesons in track-based jets with a transverse momentum between 5 and 15 GeV (left) and 15 and 30 GeV (right) [9]

Several charm FFs have also been measured in jets in pp collisions. The collinear $D^{*+/-}$ FF was measured in jets by the ATLAS experiment and found to display a soft fragmentation spectrum consistent with the $D^{*+/-}$ mesons primarily originating from gluon splittings [47]. The $D^{*+/-}$-in-jet data was recently included in global fits to extract $D^{*+/-}$ FFs [48]. The collinear D^0 and Λ_c FFs were measured in jets by the ALICE experiment, with the latter being the first known heavy flavor baryon FF measurement [9, 49]. Figure 1.9 shows the D^0 FF measured in track-based jets by ALICE in two different jet p_T bins. In the lower jet p_T bin, spanning 5 to 15 GeV, the D^0 exhibits a harder fragmentation, while in the higher jet p_T bin from 15 to 30 GeV the fragmentation spectrum peaks around a z of 0.65, closer to what was previously measured in e^+e^- decays. Heavy quarkonium fragmentation has also been studied in jets. The fraction of the jet p_T carried by J/ψ mesons, $p_T(J/\psi)/p_T(jet)$, was measured and found to be well-described for J/ψ mesons decayed from B hadrons but poorly described for prompt J/ψs [50].

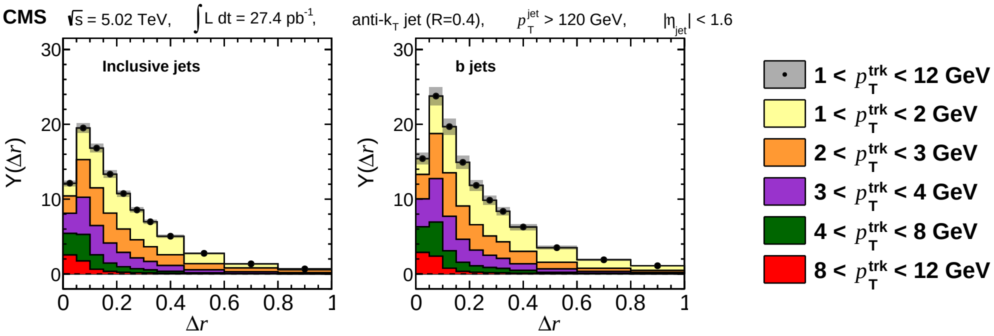

Figure 1.10: Charged particle yields $Y(\Delta r)$ as a function of the distance from the jet axis Δr and the track p_T, for inclusive jets (left) and b-jets (right) [10].

1.3.3 Fragmentation to Charged Hadrons

In addition to measuring single-hadron fragmentation functions, measuring distributions of the additional particles produced in the hadronization process is important to understand the mechanisms underlying hadronization. Few studies of heavy flavor fragmentation to charged hadrons exist, but those that have been performed revealed differences in the charged hadron multiplicity and distribution between light and heavy quark hadronization [10, 51]. A measurement of the average charged particle multiplicity in $e^+e^- \to b\bar{b}$ events compared to light-quark events, $e^+e^- \to q\bar{q}$,

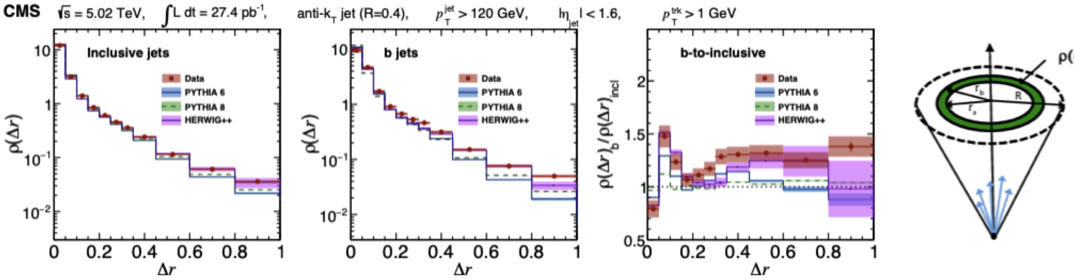

Figure 1.11: Measurements of the jet shape $\rho(\Delta r)$ for (from left) inclusive jets, b-jets, and the ratio in b-jets to inclusive jets [10], and a diagram illustrating the δr annular ring probed in the jet shape measurement [11].

determined that more charged particles are produced in $b\bar{b}$ events [10]. The charged particle multiplicity measurement was consistent with perturbative QCD predictions and inconsistent with predictions that assumed the charged particle multiplicity was independent of the quark mass [10].

The multiplicity of charged particles from heavy quark hadronization has also been studied in b-jets in pp collisions. The CMS experiment measured the b-jet charged particle multiplicity and jet shape, an observable that describes the energy distribution of particles within jets, and compared the measurements to those in inclusive jets, which are primarily gluon-jet dominated at midrapidity at the LHC [10]. Figure 1.10 shows the measured yields of charged hadrons in inclusive jets and b-jets as a function of distance from the jet axis Δr and charged hadron p_T. The b-jets were observed to have a higher charged hadron multiplicity than inclusive jets, with most of the additional charged hadrons being produced close to the jet axis and with low p_T [10]. Figure 1.11 shows the measured jet shape results for inclusive jets and b-jets, with the jet shape observable defined by the following equation:

$$\rho(\Delta r) = \frac{1}{\delta r} \frac{\sum_{jets} \sum_{particle \in (\Delta r_a, \Delta r_b)} p_T^{trk}}{\sum_{jets} \sum_{trk} p_T^{trk}} \tag{1.10}$$

Equation 1.10 describes how the jet energy is distributed in an annular ring with outer radius Δr_b and inner radius Δr_a and width $\delta r = \Delta r_b - \Delta r_a$. Comparisons of the jet shapes for inclusive jets and b-jets revealed that b-jets have a broader energy distribution, with more of the jet energy distributed at larger distances from the jet axis [10].

1.3.4 Dead Cone Effect

Observables related to the radial distributions of hadrons in heavy flavor jets are thought to be sensitive to the heavy flavor dead cone effect. The "dead cone" angle is so named because it describes a conical region in which a particle does not emit any radiation. The dead cone effect is a fundamental prediction of gauge theories that states that a charged particle does not emit radiation at angles smaller than that defined by its mass divided by its energy [12, 52, 53]. Since the QCD dead cone is proportional to the mass of the radiating quark, it is predicted to be significantly larger for heavy quarks than for light quarks [52, 53].

The beauty and charm dead cones were indirectly detected at LEP via the observation that fewer particles were produced at small angles relative to the jet axis in heavy flavor jets compared to inclusive jets [54]. By ordering the particles in the jet based on their angular separation, the history of partonic splittings in the jet can be approximately reconstructed. This technique, known as iterative declustering, enabled the first direct detection of the dead cone angle in charm jets by the ALICE collaboration, the results of which are shown in Figure 1.12 [12]. The observable $R(\theta)$ is defined as the ratio of the number of splittings observed in charm jets to the number of splittings in inclusive jets [12]. As seen in Figure 1.12, the number of splittings observed in charm jets is suppressed relative to inclusive jets below a certain angle θ. The red shaded regions in Figure 1.12 indicate where the charm dead cone is predicted to exist, and the suppression seen in data is clearly observed to coincide with this predicted dead cone region [12].

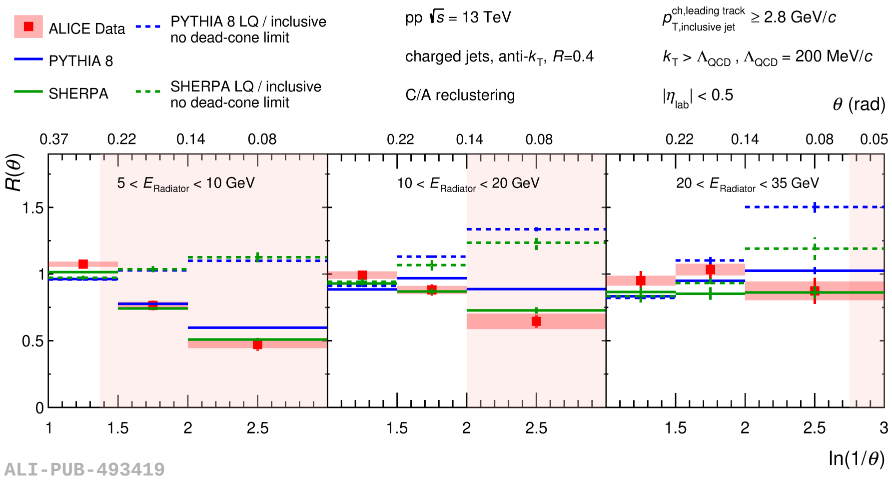

Figure 1.12: Experimental observation of the QCD dead cone in charm jets, observed as a suppression in the ratio $R(\theta)$ of the number of splittings in charm jets to the number in inclusive jets [12].

1.3.5 Theoretical Description

Heavy flavor jet production in pp collisions has been studied theoretically, and generally good agreement with data has been achieved [13, 55]. Figure 1.13 shows an example of theoretical calculations for the inclusive b-jet cross section compared to experimental measurements by the CMS experiment [13]. When describing the production of a jet without describing the production of specific hadrons within the jet, the FJF in the factorization theorem in Equation 1.2 is replaced by a function known as a jet function, which can be calculated in perturbative QCD [13]. In order to compare to experimental data, it is necessary to take into account contributions not only from a heavy quark fragmenting into a heavy-flavor jet, but also contributions from light-quark and gluon fragmentation into heavy flavor jets as well [13, 56]. The combination of all three contributions is described as the fragmentation function to a b-jet, denoted FFbJ in Figure 1.13 and Reference [13].

Less work has been done to theoretically describe the production of hadrons within heavy flavor jets and provide comparisons to measurements in data described in the previous sections. However, several aspects of the theoretical framework needed for

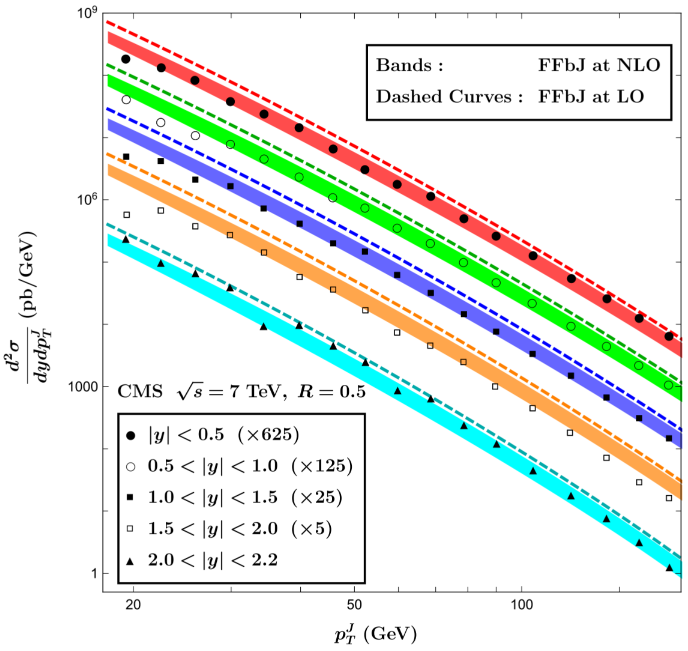

Figure 1.13: Theory predictions for the inclusive b-jet cross section compared to CMS data [13].

such comparisons have been studied. One feature of perturbative QCD calculations with heavy quarks is that due to the large heavy quark mass relative to the nonperturbative QCD scale, the heavy quark FFs can be computed perturbatively [56]. Since the FJFs depend on the collinear FFs as shown in Equation 1.3, the heavy quark FJFs can be computed perturbatively as well. This is in contrast to the light-parton FJFs which depend on FFs that cannot be calculated perturbatively and must be extracted from fits to data. FJFs initiated by a heavy quark have been calculated perturbatively up to order $O(\alpha_S)$, while those initiated by light quarks and gluons have been calculated to order $O(\alpha_S^2)$ [56]. The calculation of the heavy quark FJFs should enable future theoretical comparisons for the measurements of charged hadron distributions in heavy flavor jets presented in this thesis.

1.4 Aims and Outline of this Thesis

This thesis aims to expand on the previously discussed studies of heavy flavor hadronization in jets by measuring charged hadron distributions sensitive to the heavy flavor fragmentation functions in beauty and charm jets. This thesis specifically describes the measurement of the longitudinal momentum fraction z, transverse momentum relative to the jet axis j_T, and radial distance from the jet axis r of charged hadrons in beauty and charm jets. These measurements are complementary to the measurements of single-hadron heavy flavor fragmentation functions discussed previously, as they probe the additional particles produced from the hadronization of a single heavy quark. The outline for this thesis is as follows: Chapter 2 describes the experimental setup used to collect the data analyzed for this measurement, Chapter 3 describes the measurement tools used for the analysis, Chapter 4 discusses the steps of the data analysis in detail, and the measured results are presented in Chapter 5. Chapter 6 presents studies performed for a proposed detector at a future Electron-Ion Collider that are relevant for future measurements of heavy flavor hadronization, and finally Chapter 7 provides a summary of the thesis and highlights some prospects for future measurements.

CHAPTER II

Experimental Setup

The experimental setup used for this thesis measurement is described in this chapter. The Large Hadron Collider accelerator facility is introduced, followed by a detailed introduction to the Large Hadron Collider beauty (LHCb) experiment detector. The components of the LHCb detector relevant for the analysis presented in this thesis are discussed, including the vertex detector, tracking system, calorimeter system, muon system, and the trigger system.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a hadron-hadron collider located at the European Centre for Nuclear Research (CERN) in Geneva, Switzerland. The LHC consists of two rings that accelerate beams of protons or lead ions in opposite directions and collide them at four interaction points along the ring. With a ring circumference of 26.7 km, and maximum collision center of mass energy $\sqrt{s}$ of 13 TeV, the LHC is the largest and highest energy particle collider in the world.

Each interaction point on the LHC ring is home to one of four large particle physics experiments - ATLAS, CMS, ALICE, and LHCb. ATLAS and CMS are general purpose detectors with a physics program strongly focused on searching for physics beyond the Standard Model. ALICE is dedicated primarily to studying heavy-ion

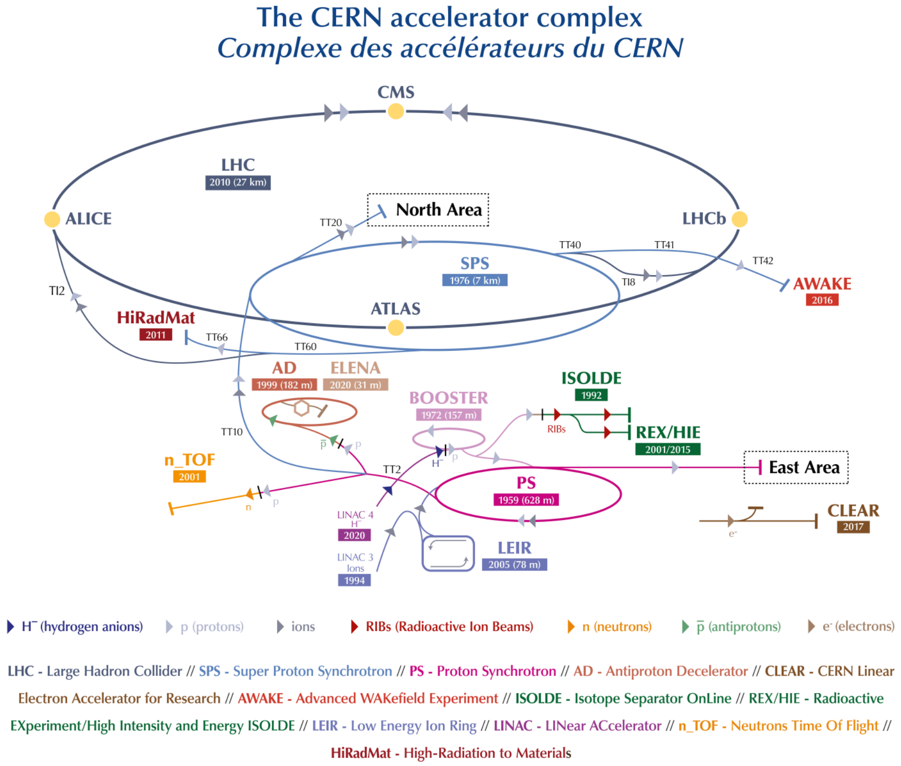

Figure 2.1: The CERN Accelerator Complex [14]

physics with the goal of understanding the formation and properties of the Quark-Gluon Plasma formed in heavy-ion collisions. LHCb is dedicated to studying the production and decay of particles containing heavy (beauty or charm) quarks and searching for new origins of CP violation. For the first data-taking run of the LHC, Run 1, the LHC delivered proton-proton collisions to the experiments at a center of mass energy of 7 TeV during 2010-2011, and 8 TeV during 2012. For the second data-taking run, Run 2, the center of mass energy was increased to 13 TeV for the entire run period, from 2015 to 2018. In between the run periods, the LHC enters Long Shutdown periods to allow for maintenance and upgrades of the accelerator and experiments.

2.1.1 The LHC Accelerator Complex

Several stages of beam acceleration are required to obtain the 6.5 TeV beams that produce a collision center of mass energy of 13 TeV. Therefore in addition to the LHC ring, CERN hosts several other facilities. These facilities are also used to produce high-energy particle beams for many other nuclear and particle physics experiments based at CERN. A diagram of the CERN accelerator complex and the numerous experiments it hosts is shown in Figure 2.1. The components of the accelerator complex necessary for accelerating beams to LHC energies are the linear accelerators (LINAC), the proton synchrotron booster (BOOSTER), the proton synchrotron (PS) and the super proton synchrotron (SPS) [57]. For Run 1 and Run 2, protons were first accelerated in CERN's LINAC2, which has since been replaced with LINAC4 for Run 3. Hydrogen atoms from a bottle of hydrogen gas were passed through an electric field to remove their electrons, leaving the protons to enter LINAC2 which accelerated them to 50 MeV. The resulting proton beam was then directed to the proton sychrotron booster, which accelerated the protons to 1.4 GeV before injecting them into the PS. The PS further accelerated the protons to 25 GeV, and then injected them into the SPS, which accelerated them to 450 GeV before finally injecting them into the LHC ring [57].

2.2 The Large Hadron Collider beauty (LHCb) Experiment

The Large Hadron Collider beauty (LHCb) experiment is specifically designed to study the production and decay of beauty and charm hadrons. The unique focus on studying heavy flavor hadrons is reflected in key features of the LHCb detector, shown in Figure 2.2. Unlike the other LHC experiments, which predominantly cover the midrapidity region, the LHCb detector is a single arm spectrometer spanning the forward pseudorapidity region of $2 < \eta < 5$. The forward acceptance and detector

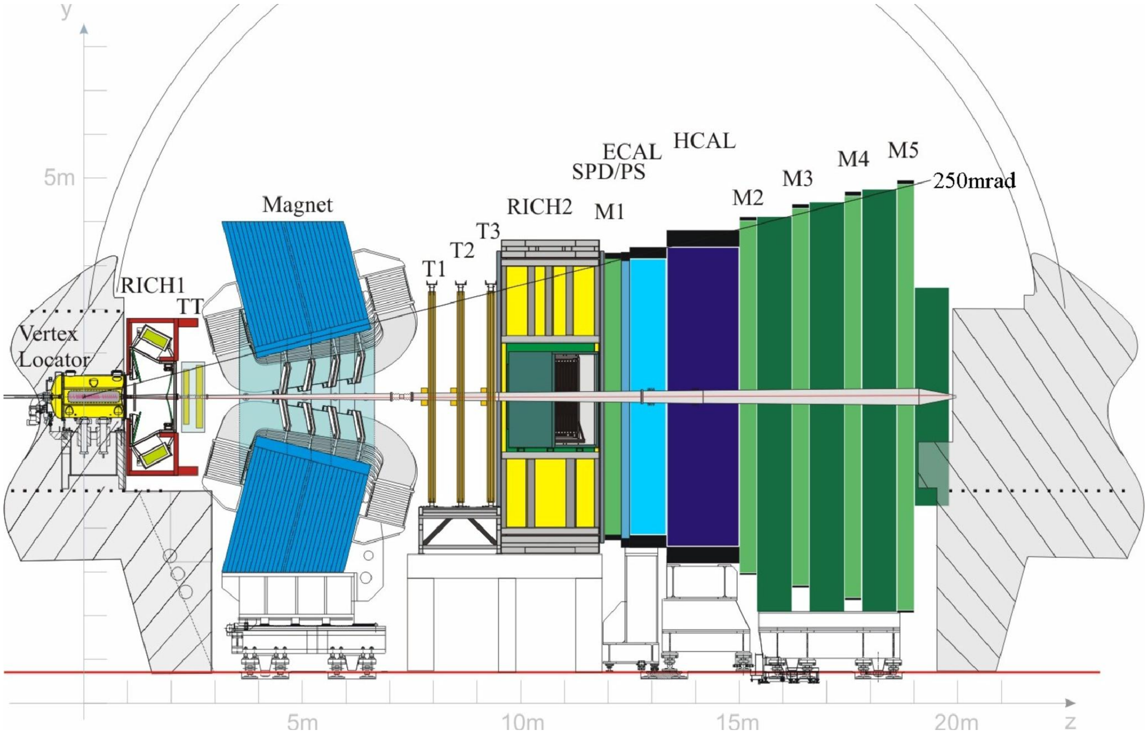

Figure 2.2: A side view of the LHCb detector, with the subdetector components labeled. RICH1 and RICH2 are the Ring-Imaging Cherenkov Detectors, TT is the Tracker Turicensis, T1-T3 are the Tracking Stations 1-3, SPD/PS is the Scintillating Pad Detector and Pre-Shower Detector, ECAL is the Electromagnetic Calorimeter, HCAL is the Hadronic Calorimeter, and M1-M5 are the Muon Stations 1-5 [15]

geometry of LHCb was chosen to maximize the number of b and $\bar{b}$ quarks produced within the detector volume. As shown in Figure 2.3, more b and $\bar{b}$ quarks are produced within the LHCb acceptance than in the acceptance of midrapidity general purpose detectors. At LHC energies, the boosted b and $\bar{b}$ quarks are often produced in the same forward or backward cone, which motivates the conical-like design of LHCb shown in Figure 2.2 [17].

Features of heavy flavor hadron decays influenced key aspects of the LHCb detector specifications. Heavy flavor hadrons typically decay after traveling several hundreds of microns in the lab frame. $B^{+/-}$ and $D^{+/-}$ mesons, for example, have decay lengths $c\tau$ in the lab frame of approximately 461 μm and 312 μm, respectively [1]. These characteristic secondary decay vertices can be reconstructed with precise charged particle tracking capabilities. LHCb's tracking system features a dedicated detector, the Vertex Locator, around the LHC beam interaction region dedicated to detecting charged particle tracks from secondary vertices of heavy flavor hadron decays. Due to their larger mass relative to light-flavored hadrons, heavy flavor hadron decays also typically produce a larger number of decay products than light hadron decays. The higher multiplicity of heavy flavor hadron decays often results in the need to combine four or more particles to reconstruct a mass peak. This fact, coupled with the large number of particles produced in a *pp* collision, represents a combinatoric challenge that, if not resolved, can easily overwhelm the signal peak. In order to reduce the combinatorial background and cleanly reconstruct the mass spectra of heavy flavor hadrons, LHCb features excellent particle identification capabilities to differentiate between pions, kaons, protons, muons, and electrons. LHCb's two Ring-Imaging Cherenkov (RICH) detectors provide charged hadron identification for hadrons with momenta between 2 and 100 GeV/c. In the following sections, key subdetectors of the LHCb detector will be discussed, with a focus on those especially relevant for measuring heavy flavor jets: the VELO, the tracking system, the calorimeter system, and the muon system.

2.2.1 The Vertex Locator (VELO)

The Vertex Locator (VELO) is a silicon microstrip detector surrounding the LHCb interaction region, where the two proton beams from the LHC collide to produce an event which can then be detected by LHCb. A diagram of the VELO is shown in Figure 2.4. The purpose of the VELO is to detect charged particle tracks produced from the secondary vertices of heavy flavor hadron decays. VELO tracks are critical for reconstructing secondary vertices, which when present in a jet are used to identify the jet as originating from a heavy flavor quark. The VELO consists of 42 circular silicon modules which measure the r and ϕ coordinates of tracks produced near the

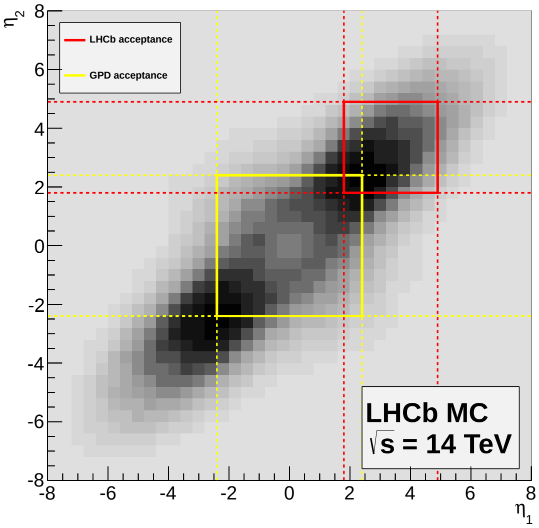

Figure 2.3: The pseudorapidity of b (η_1) and $\bar{b}$ (η_2) quarks produced in the MC processes $q\bar{q} \rightarrow b\bar{b}$, $gg \rightarrow b\bar{b}$, $b\bar{b} \rightarrow b\bar{b}g$, $gg \rightarrow b\bar{b}g$, and $q\bar{q} \rightarrow b\bar{b}g$ at $\sqrt{s} = 14$ TeV. The yellow region highlights the (η_1, η_2) range accessible with a general purpose detector (GPD), while the red region highlights the range accessible to LHCb. [16]

interaction region. One side of each module has sensors segmented to measure ϕ (shown as blue sensors in Figure 2.4), and the other side of the module has sensors segmented for measuring r (shown as red sensors in Figure 2.4). The hit resolution of the VELO sensors varies as a function of the strip pitch and the projected track angle, and was measured to range from 5 - 25 μm [15]. The geometry of the VELO was designed so that a track within the LHCb acceptance must cross at least three VELO modules [17]. Due to its proximity to the beam, the VELO is divided into two halves that can be retracted during the injection of the LHC beams to avoid damaging the sensors [17].

2.2.2 The Turicensis Tracker (TT), Inner (IT) and Outer Tracker (OT) and Dipole Magnet

The Turicensis Tracker (TT), Inner Tracker (IT), Outer Tracker (OT) and dipole magnet comprise the rest of LHCb's tracking system after the VELO. The main

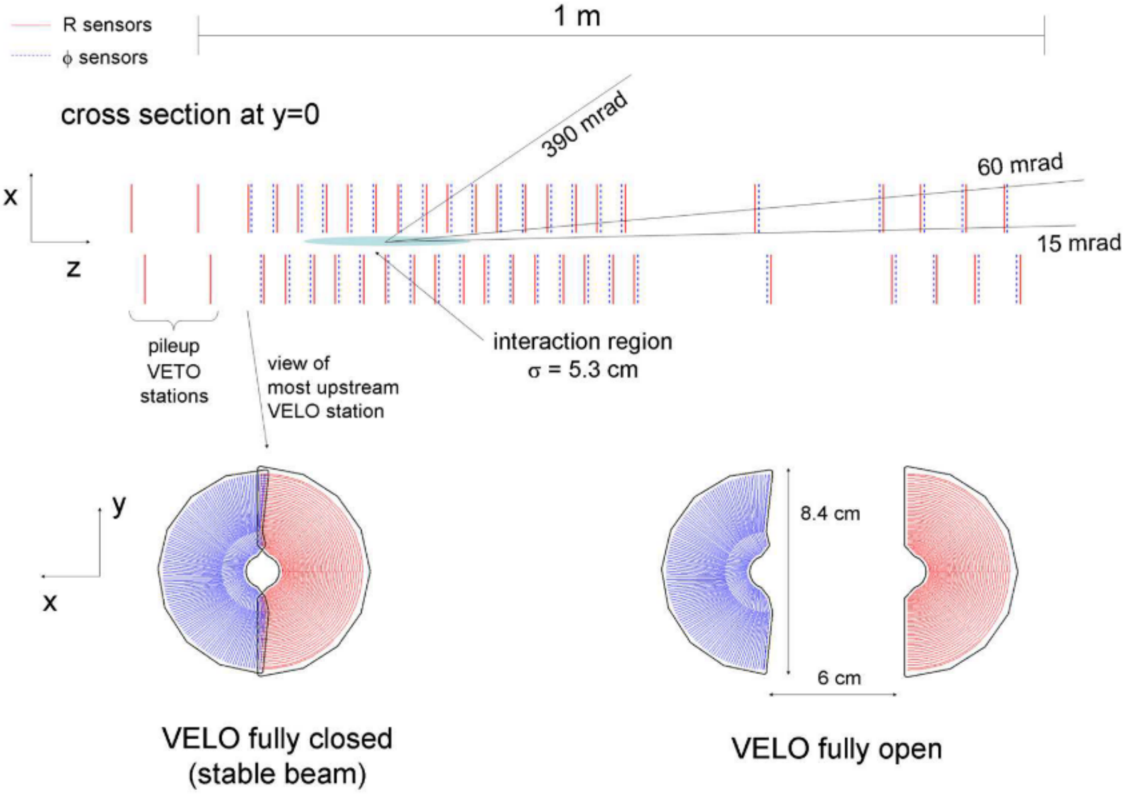

Figure 2.4: A diagram of the VELO detector, showing the 42 VELO modules located at various points along the beam axis. The proximity of the VELO modules to the interaction region (light blue) is also shown. [17]

purpose of the tracking system after the VELO is to provide a measurement of the momentum of charged particles produced in the event. The dipole magnet provides a magnetic field with a field strength of 4 Tm that bends the charged particle tracks, and the TT, IT and OT provide measurements of the particle's position before and after it traverses the magnetic field. By reconstructing the trajectory of a charged particle using the hits in the tracking detectors, the momentum of the particle can be determined. The TT is a planar silicon microstrip detector with a strip pitch of 183 μm located before the dipole magnet [17]. The IT and OT detectors form three tracking stations (T1-T3 in Figure 2.2) located after the dipole magnet. The IT is a silicon microstrip detector with a strip pitch of 198 μm surrounding the beam pipe, while the OT is a drift-tube gas detector around the IT. Figure 2.5 shows the TT,

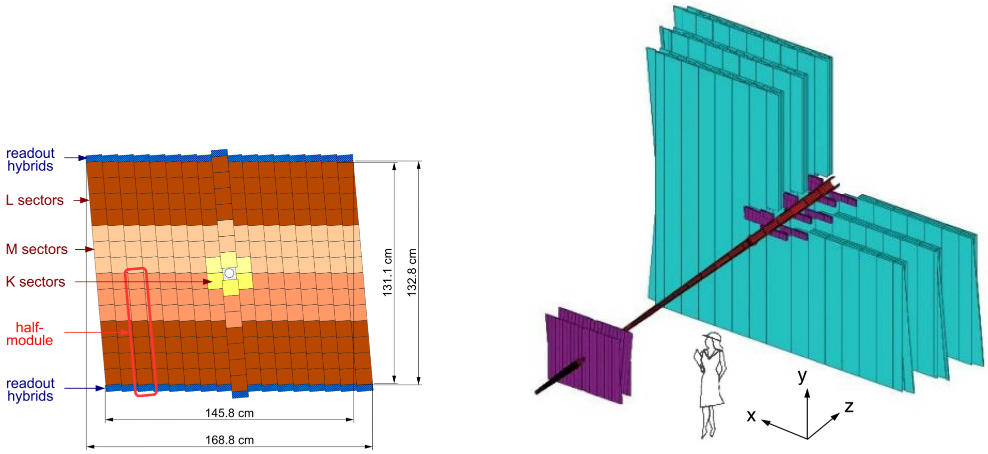

Figure 2.5: (Left) The third layer of the TT. (Right) A diagram of the TT, IT, and OT, with the silicon detectors (TT and IT) shown in purple and the drift tube detectors (OT) shown in teal. [17]

IT, and OT detectors [17]. The OT consists of 200 straw tube modules, with each module containing two layers of drift tubes with a gas mixture of 70% argon, 28.5% CO_2, and 1.5% O_2 [17]. The hit resolution in the TT and IT is approximately 50 μm, while in the OT it increases to approximately 200 μm [17].

All of the tracking detectors - the TT, IT, and OT - consist of four detection layers. The first and last layers of the TT and IT have vertically oriented silicon strips, while the second and third layers have strips rotated by a stereo angle of -5 degrees and +5 degrees, respectively, forming an x-u-v-x geometry. The OT has four layers of drift tubes arranged in the same x-u-v-x geometry [17].

2.2.3 Calorimeter System

LHCb's calorimeter system consists of preshower (PS) and scintillating-pad (SPD) detectors, an electromagnetic calorimeter (ECAL), and a hadronic calorimeter (HCAL). The purpose of the calorimeter system is to provide a measurement of the energy and position of charged and neutral particles, as well as to discriminate between photons, electrons, and hadrons. The ECAL and HCAL measure the energy deposits of charged and neutral particles by detecting the "showers" of particles produced when a particle interacts with the calorimeter material. The PS and SPD detectors assist with differentiating between the energy deposits of electrons, charged pions, and neutral particles in the ECAL. Both the PS and SPD consist of nearly-identical planes of scintillating pads with a 15 mm layer of lead absorber between them [17]. The SPD helps to discriminate between energy deposits of π^0s and electrons in the ECAL, as electrons will produce scintillation light within the SPD while photons from π^0 decays will not. The PS detector, located directly in front of the ECAL, helps to discriminate between energy deposits of charged pions and electrons in the ECAL by providing additional longitudinal segmentation for measuring the charged particle showers.

The ECAL is a shashlik-type calorimeter, consisting of discrete cells each made of alternating layers of scintillating tiles as an active material and lead as an absorber material. The scintillation light generated by particles interacting with the active material is collected via wavelength-shifting fibers and read out using photo-multiplier tubes [17]. The particle hit density falls off sharply as a function of lateral distance from the beam pipe, and varies by two orders of magnitude over the surface of the ECAL [17]. Therefore, the lateral segmentation of the ECAL is such that the finest lateral resolution is near the beam pipe and the coarsest lateral resolution is farther away from the beam pipe. The inner cells of the calorimeter are 40.4 square mm in size, the cells in the middle section are 60.6 square mm in size, and the cells in the outer section are 121.2 square mm in size, as shown in Figure 2.6. The total thickness of the ECAL, which characterizes the resolution on the longitudinal size of the electromagnetic shower, is 25 radiation lengths (X_0). The Moliere radius of the ECAL, which characterizes the resolution on the transverse size of the electromagnetic shower, is 3.5 cm. The energy resolution of the ECAL is given by $\sigma(E)/E = 10\%/\sqrt{E} \oplus 1\%$, with the first term describing the stochastic energy resolution and the second term describing the constant energy resolution.

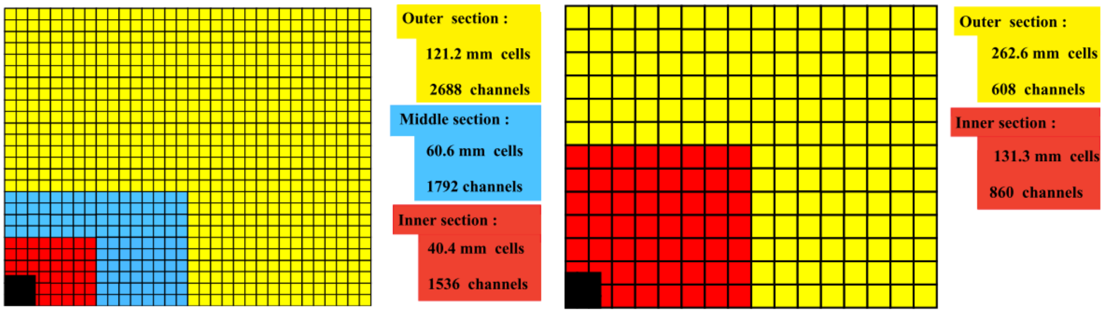

Figure 2.6: A diagram of the lateral segmentation of the SPD, PS, and ECAL (left) and HCAL (right). A quarter of the detector face is shown for each.

The HCAL is also a sampling calorimeter, consisting of alternating layers of scintillating tiles as an active material and iron as an absorber material. Unlike the ECAL, in which the scintillating tiles are perpendicular to the beam axis, the scintillating tiles in the HCAL cells are oriented parallel to the beam axis. As with the ECAL, scintillation light in the HCAL is collected via wavelength-shifting fibers and read out with photo-multiplier tubes [17]. The HCAL also has a varying lateral cell segmentation, but with larger cell sizes than those of the ECAL due to the larger dimensions of hadronic showers compared to electromagnetic showers. The inner cells of the HCAL are 131.3 square mm in size, while the outer cells are 262.6 square mm in size, as shown in Figure 2.6. The total thickness of the HCAL is limited by space constraints and corresponds to 5.6 nuclear interaction lengths (λ_i) [17]. The energy resolution of the HCAL is given by $\sigma(E)/E = (69 \pm 5\%)/\sqrt{E} \oplus (9 \pm 2)\%$ [17].

2.2.4 The Muon System

Alternating iron absorbers and five stations of multiwire proportional chambers (MWPCs), as well as gas electron multiplier (GEM) detectors in the regions of highest particle density, comprise LHCb's muon system [17]. The purpose of the muon system is to identify muons and measure their $\mathrm{p_T}$ for use in the Level 0 trigger. A diagram of the muon system is shown in Figure 2.7. The first muon station, M1, is located

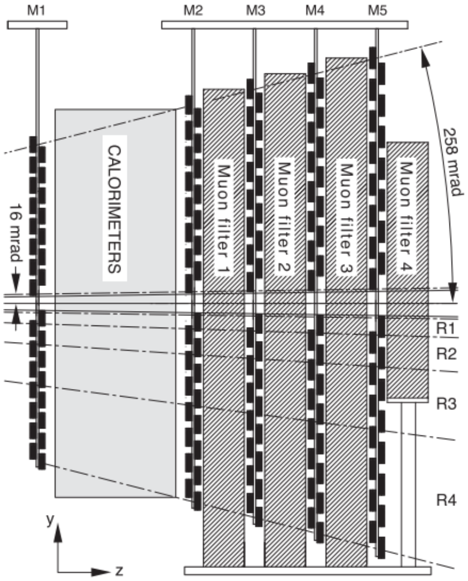

Figure 2.7: A side view of the LHCb Muon System. M1-M5 are the five muon stations composed of MWPCs and GEMs. The four muon filters are iron absorbers. [17]

before the ECAL and HCAL in order to improve the measurement of the muon p_T [17]. The remaining muon stations, M2-M5, are located after the calorimeters and are separated by iron absorbers. All stations are composed solely of MWPCs except for M1, which also has triple-GEM detectors in the inner region close to the beam pipe due to the high particle density before the calorimeters [17]. Stations M1-M3 have a higher spatial resolution than stations M4-M5 and are used to measure the muon p_T for the Level 0 trigger. Stations M4-M5 are primarily used to identify penetrating muons. In order to pass through all five muon stations, a muon must have a minimum momentum of 6 GeV/c [17].

2.2.5 The Trigger System

LHCb's trigger system serves a very important role in rapidly identifying events of interest during data-taking and saving them for offline analysis. It consists of three different levels, or stages: Level 0 (L0), the High Level Trigger 1 (HLT1) and the High Level Trigger 2 (HLT2) [17]. The complexity of the decisions made increases with each level of the trigger. The L0 trigger coarsely identifies events of interest, the HLT1 trigger adds additional selections to increase the likelihood of events of interest being saved for offline analysis, and the HLT2 trigger selects events with specific topologies, reconstructed decay chains, or even jets (see Section 3.1 for more details about jet reconstruction in the trigger). The L0 trigger is a hardware trigger, using information from the calorimeter and muon systems to identify high-energy deposits in the calorimeters or high-energy muons. The L0 trigger reduces the 40 MHz pp collision rate delivered by the LHC to 1 MHz, the frequency at which all detector subsystems of the LHCb detector can be read out [15, 17]. The HLT1 and HLT2 triggers then use information from the LHCb detector subsystems to quickly make selections or reconstruct physics objects of interest. The HLT1 and HLT2 triggers are both software triggers, which enables the more sophisticated decision-making capabilities needed in these triggers.

CHAPTER III

Measurement Tools

Several tools used for the measurement described in this thesis are introduced in this chapter. The jet reconstruction algorithm used at LHCb is described, followed by a detailed discussion of how heavy flavor jet tagging is performed at LHCb. Finally, a brief overview of the theory behind the Bayesian unfolding technique to remove detector smearing effects from measured distributions is provided.

3.1 Jet Reconstruction at LHCb

In order to reconstruct jets, a selection of input particles for the jet finding algorithm must be identified, and a type of jet finding algorithm specified. At LHCb, a particle flow algorithm is used to select both charged and neutral particle inputs for jet finding, and jets are reconstructed using the anti-k_T algorithm [58]. In Run 2, the jet reconstruction algorithm was modified so that the algorithm used during real-time ("online") data processing in the High Level Trigger (HLT) was identical to the jet reconstruction algorithm used during offline data processing. Jets reconstructed in the HLT are called HltJets, and HltJets corrected for jet energy corrections and quality cuts offline are called Standard HltJets, abbreviated StdHltJets. The next sections detail the steps involved with reconstructing StdHltJets and describe the HLT particle flow algorithm, jet building algorithm, and offline quality cuts.

3.1.1 The High Level Trigger Particle Flow Algorithm

The particle flow algorithm uses information from the LHCb detector subsystems, in particular the tracks and calorimeter clusters, to reconstruct charged and neutral particles for input to the jet reconstruction. Extrapolated tracks that overlap with the location of a calorimeter cluster are "matched" to the cluster and used to reconstruct a charged particle. Clusters with no associated tracks are used to reconstruct neutral particles. The track-cluster matching procedure ensures that a single particle is not double counted in the input to the jet reconstruction. Tracks and clusters that pass loose cuts for specific particles are used to construct particle candidates. Photon candidates are identified using information about energy deposits in the preshower detector and single ECAL clusters that are not matched to any tracks. Pairs of well-separated photons with an invariant mass within 30 MeV of the π^0 mass are identified as resolved π^0 candidates. Pairs of charged pions with an invariant mass near the K_s^0 mass are identified as K_s^0 candidates in the $K_s^0 \to \pi^+\pi^-$ decay channel, while pairs of a charged pion and a proton near the Λ mass are identified as Λ candidates in the $\Lambda^0 \to p\pi^-$ decay channel. The photon, π_0, K_s^0, and Λ candidates, along with the unidentified charged and neutral particle candidates from the track-cluster matching procedure, are passed as inputs to the jet building algorithm.

3.1.2 The High Level Trigger Jet Building Algorithm

Jets at LHCb are built using the anti-$\mathrm{k_T}$ jet finding algorithm [58], which groups particles together based on their transverse momentum p_T and the distance in $\eta - \phi$ space between them, $\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$. For each input particle i, the following distance measures are calculated [58]:

$$d_{ij} = min(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2})\frac{\Delta_{ij}^2}{R^2} \tag{3.1}$$

$$d_{iB} = \frac{1}{p_{T,i}^2} \tag{3.2}$$

In Equations 3.1 and 3.2, d_{ij} is the distance measure between two particles i and j and d_{iB} is the distance measure between particle i and the beam. R is the radius of the conical jet. If d_{ij} is smaller than d_{iB}, then the momenta of particles i and j are added to form a combined "particle." If d_{iB} is smaller than d_{ij} the particle is called a jet and is removed from the inputs to the jet building algorithm. The distance measures are recalculated with the combined particles and the process repeated until there are no input particles remaining. At LHCb, jets are reconstructed using a radius of $R = 0.5$.

3.2 Heavy Flavor Jet Tagging at LHCb

LHCb's algorithm for tagging heavy flavor jets consists of two components: the first is a Secondary Vertex (SV) Tagger algorithm that reconstructs a secondary vertex consistent with originating from a heavy flavor hadron decay in the jet, and the second is two Boosted Decision Trees that provide optional further jet flavor separation. The next sections will describe each of these components in further detail.

3.2.1 The Secondary Vertex (SV) Tagger Algorithm

The Secondary Vertex (SV) Tagger algorithm begins by selecting tracks in the event that are displaced from the primary vertex. Displaced tracks are selected with a large χ^2_{IP} value, which is the difference in χ^2 from reconstructing the primary vertex with and without the track. The displaced tracks are not required to be in a jet, but must have a momentum p of at least 5 GeV and a transverse momentum p_T of at least 0.5 GeV. No hadron identification is used to identify the tracks, and all tracks are assigned the pion mass [18]. The displaced tracks passing the χ^2_{IP}, p, and p_T requirements are used as inputs to the second stage of the algorithm. In the second stage, all possible combinations of two-body secondary vertices are constructed from the selected displaced tracks. The mass of each two-body secondary vertex is required to be within 0.4 - 5.2794 GeV. The lower mass limit of 0.4 GeV removes most secondary vertices due to strange hadron decays, while the upper mass limit restricts the two-body secondary vertex mass to be consistent with originating from a beauty or charm hadron decay [18]. The distance of closest approach between the tracks in the two-body SV is required to be less than 0.2 mm. In the third and final stage of the algorithm, two-body SVs that share tracks are merged into N-body SVs. Additional cuts are applied to the N-body SVs to reject SVs from strange hadron decays and to enhance the beauty and charm tagging performance. The N-body SV z position must be less than 200 mm, and the minimum radial flight distance of the two-body SVs used to construct the N-body SV must be less than 15 mm. Both position requirements restrict the N-body SV to be within a region consistent with that of beauty and charm hadron decays [18]. To suppress strange hadron decays, cuts are placed on two additional variables: the SV corrected mass, M_{cor}, and a proxy for the hadron lifetime, τ. The SV corrected mass is the minimum mass a long-lived hadron can have that is consistent with the flight direction of the SV, and is defined by the following equation [18]:

$$M_{cor} = \sqrt{M^2 + p^2 \sin\theta^2} + p \sin\theta \tag{3.3}$$

In Equation 3.3, M is the invariant mass of the N-body SV, p is the momentum of the N-body SV, and θ is the angle between the SV momentum and its flight direction. The corrected mass accounts for the fact that the heavy flavor decay vertex is often only partially reconstructed by the SV-tagger algorithm. The corrected mass is required to be greater than 0.6 GeV to suppress contributions from strange hadron decays [18]. The flight distance of the SV divided by its momentum is taken as a proxy for the

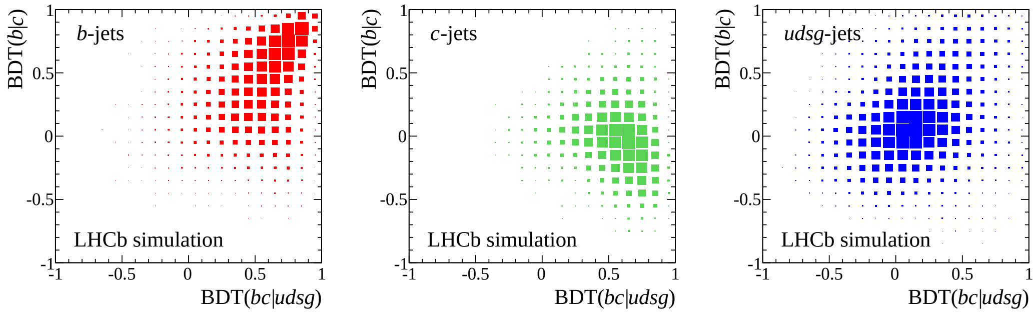

Figure 3.1: Two-dimensional BDT distributions for simulated beauty, charm and light parton jets [18]

hadron lifetime τ and required to be less than 1.5 mm/GeV. N-body SVs are required to have p_T greater than 2 GeV. If multiple N-body SVs passing the requirements are present in a jet, the N-body SV with the highest p_T is used to tag the jet.

3.2.2 Boosted Decision Trees

While the SV-tagger algorithm identifies jets containing secondary vertices consistent with originating from heavy flavor decays, it does not discriminate between beauty jets and charm jets. Therefore, for further jet flavor separation, information about the reconstructed SV in the jet is passed to two Boosted Decision Trees (BDTs). BDT(bc|udsg) discriminates between heavy and light flavor jets, while BDT(b|c) discriminates between beauty and charm jets. The BDTs were trained on Run 1 simulation samples of beauty, charm, and light parton jets [18]. The suitability of their use in Run 2 data is discussed in Section 3.2.3. Several characteristics of the N-body SV reconstructed by the SV-tagger algorithm are passed as inputs to the BDTs for jet flavor discrimination. Some of the inputs are independent of the jet properties, including the SV mass and corrected mass, the number of tracks in the SV and their net charge and χ^2_{IP}, the SV flight distance χ^2, and the transverse flight distance of the two-track SV closest to the primary vertex [18]. Other inputs depend on the SV characteristics relative to the jet, including the fraction of the jet p_T carried by the

SV, the ΔR between the SV flight direction and the jet axis, and the number of SV tracks within the jet radius. Each BDT outputs a response ranging from -1 to 1. For BDT(bc|udsg), a response closer to 1 indicates the jet has characteristics consistent with those of heavy flavor jets, while a response closer to -1 indicates that the jet more closely resembles a light-parton jet than a heavy flavor jet. For BDT(b|c), a response close to 1 indicates the jet is more b-jet like, while a response closer to -1 indicates the jet is more c-jet like. Figure 3.1 shows the 2D BDT responses for beauty, charm, and light-parton jets in simulation. Each jet flavor occupies a distinct region of the 2D BDT response phase space, indicating that the two BDTs provide good discrimination power between the three types of jets.

3.2.3 Performance in Data

The performance of the SV-tagging algorithm and BDTs for heavy flavor jet tagging and flavor separation were studied in detail in Run 1 data [18]. Data samples of events with fully reconstructed beauty and charm hadrons with a separation of $\Delta\phi > 2.5$ between the highest-p_T jet in the event were used to study the b- and c-jet tagging performance in data. Events with a W boson produced in association with a jet were used to study the rate at which light-parton jets were mistagged as heavy flavor jets, as the W-tagged jets are overwhelmingly light-parton jets [18]. Figure 3.2 shows an example of a fit to data using the 2D BDT templates for b-, c-, and light-parton jets shown in Figure 3.1. The data shown in Figure 3.2 is from the sample with a fully reconstructed charm hadron and a jet nearly back-to-back. The 2D BDT distribution in data shows that most of the data is concentrated in the region consistent with a charm jet response, however there is also a significant concentration in the region consistent with the beauty jet response, which is expected given that many beauty hadrons decay to charm hadrons. The fit to the data using the BDT templates in Figure 3.1 describes the data well, as seen from the fit projections for each BDT re-

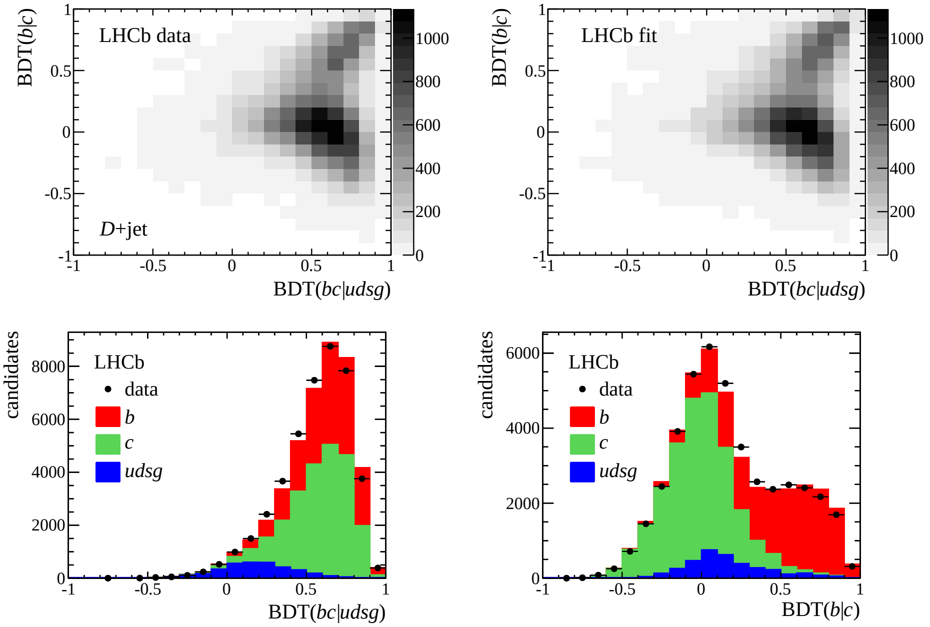

Figure 3.2: The 2D BDT distribution in data for events with a fully reconstructed charm hadron and a jet nearly back to back (top left), a 2D fit to the data using the BDT templates (top right), and the 1D fit projections for the BDT($bc|udsg$) response (bottom left) and BDT($b|c$) response (bottom right) [18].

sponse. The b- and c-jets both score high on the BDT discriminating between heavy and light jets, while the c-jets are shifted to lower scores than the b-jets on the BDT discriminating between beauty and charm jets.

Figure 3.3 shows the measured SV-tagging efficiencies for beauty and charm jets as a function of jet p_T. The efficiencies were measured in data for jets with $p_T > 20$ GeV and $2.2 < \eta_{jet} < 4.2$, and found to be consistent with the efficiencies in MC to within 10%. The probability of mistagging a light-parton jet as a heavy flavor jet was also measured in data and found to be 0.3% [18]. The SV-tagging efficiencies generally increase as a function of jet p_T. The beauty jets have a significantly higher SV-tagging efficiency than charm jets, primarily because the higher decay multiplicity of beauty hadrons facilitates the SV reconstruction within the jet. The SV-tagging efficiencies measured by LHCb are consistent with those measured by other LHC experiments [59–61].

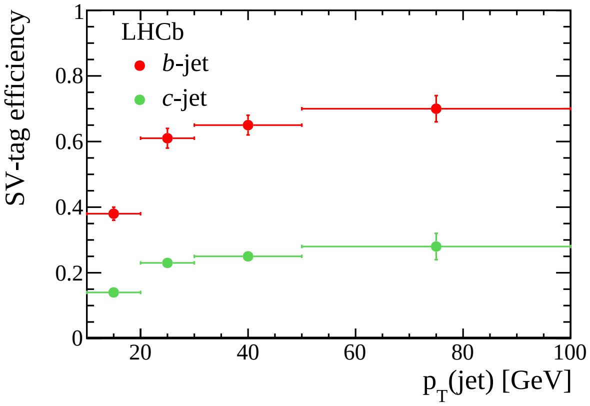

Figure 3.3: The SV-tagging efficiency measured in data as a function of jet p_T for beauty and charm jets [18].

The heavy flavor jet tagging performance has also been studied in Run 2 data and found to have degraded slightly compared to that measured in Run 1 data due to the higher track multiplicity in $\sqrt{s}$ = 13 TeV collisions compared to $\sqrt{s}$ = 8 TeV collisions. The b-jet tagging efficiency decreased from ∼65% in Run 1 data to 60% in Run 2 data, while the c-jet tagging efficiency decreased from ∼25% to 20% [19]. However, the SV-tagging algorithm and BDT templates still describe the Run 2 data well, as evidenced by their use in the LHCb measurement of $b\bar{b}$ and $c\bar{c}$ differential cross sections in Run 2 data [19]. For the dijet measurement, the sum of the BDT responses for each jet was used to perform a template fit to extract the flavor-tagged dijet yields. The fit results projected onto the t_0 and t_1 axes, with $t_0 = BDT_{bc|udsg}(\text{jet } 0) + BDT_{bc|udsg}(\text{jet } 1)$ and $t_1 = BDT_{b|c}(\text{jet } 0) + BDT_{b|c}(\text{jet } 1)$, are shown in Figure 3.4. Work is also ongoing within LHCb to improve the heavy flavor jet tagging tools developed for Run 1 data. A recent LHCb study calibrated the charm jet tagging efficiency in Run 2 data using exclusive charm meson decays, which resulted in the uncertainty on the charm SV-tagging efficiency being reduced from the value of 10% measured in Run 1 data to 6% in Run 2 data [62].

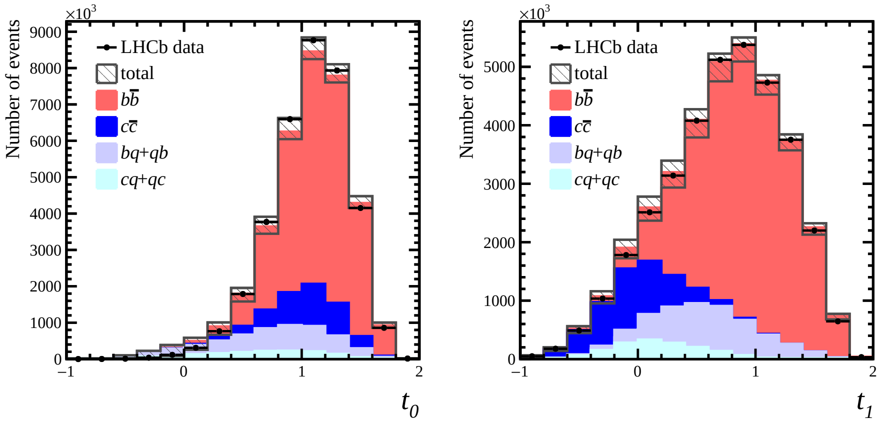

Figure 3.4: The result of a template fit to Run 2 data to extract flavor-separated heavy flavor dijet yields [19].

3.3 Bayesian Unfolding

Bayesian unfolding is a technique commonly used to remove the effects of detector smearing in measured distributions. By iteratively applying Bayes' Theorem to a measured distribution, one can actually extract the underlying "truth" distribution [63]. Bayes' Theorem relates the conditional probability that an event E will occur if a cause C_i happens, $P(E|C_i)$, to the conditional probability that given an event E, cause C_i was responsible for it occurring, $P(C_i|E)$, according to the following equation:

$$P(C_i|E)P(E) = P(E|C_i)P(C_i) \tag{3.4}$$

In Equation 3.4, $P(E)$ is the probability of event E occurring at all, and $P(C_i)$ is the probability of cause C_i occurring. If the possible causes of event E and their individual probabilities are known, then $P(E)$ can be calculated in the following way [63]:

$$P(E) = \sum_{l=1}^{n_l} P(E|C_l) \cdot P(C_l) \tag{3.5}$$

Replacing $P(E)$ in Equation 3.4 with its definition in Equation 3.5, and generalizing to the possibility that a single cause can produce multiple effects E_j, gives the following definition for the conditional probability $P(C_i|E_j)$ [63]:

$$P(C_i|E_j) = \frac{P(E_j|C_i)P(C_i)}{\sum_{l=1}^{n_l} P(E_j|C_l) \cdot P(C_l)} \tag{3.6}$$

The number of observed events attributed to cause C_i is given by the number of observed events, $n(E_j)$, multiplied by the probability of cause C_i to cause the event:

$$\hat{n}(C_i)|_{obs} = \sum_{j=1}^{n_E} n(E_j) \cdot P(C_i|E_j) \tag{3.7}$$

In Equation 3.7, n_E is the total number of effects observed. The number of observed events is related to the number of "true" events attributable to cause C_i via the efficiency for cause C_i to produce the observed effects:

$$\hat{n}(C_i)|_{obs} = \varepsilon_i \cdot \hat{n}(C_i)|_{true} \tag{3.8}$$

with the efficiency ε_i defined as the following [63]:

$$\varepsilon_i = \sum_{j=1}^{n_E} P(E_j|C_i) \tag{3.9}$$

The number of true events attributable to cause C_i can be determined using Equations 3.6-3.9 [63]:

$$\begin{aligned} \hat{n}(C_i)|_{true} &= \frac{\hat{n}(C_i)|_{obs}}{\sum_{j=1}^{n_E} P(E_j|C_i)} \\ &= \frac{\sum_{j=1}^{n_E} n(E_j)P(E_j|C_i)P(C_i)}{\sum_{j=1}^{n_E} P(E_j|C_i) \sum_{l=1}^{n_l} P(E_j|C_l) \cdot P(C_l)} \end{aligned} \tag{3.10}$$

In Equation 3.10, $n(E_j)$ is the measured experimental distribution, for example a jet p_T distribution. Each jet p_T bin would be one possible observed "effect" E_j. The $P(C_i)$ describe the probability for a certain true value or "cause" to exist, for example the probability that a jet is produced with a certain true jet p_T bin. The probability $P(E_j|C_i)$ describes the probability for the true cause C_i to produce a measured effect E_j. For the analogy to jet p_T, this probability describes the likelihood that a jet in a true p_T bin C_i is observed in bin E_j at the detector level. To unfold detector smearing effects, one needs to know how much the detector "smears" a truth-level MC distribution. By comparing truth-level information from an MC generator and detector-level information once the truth distributions have been passed through a detailed detector simulation, one can obtain the $P(E_j|C_i)$ as a 2D matrix known as a "response matrix" [63]. With the response matrix $P(E_j|C_i)$ and measured distribution E_j, the only unknown in Equation 3.10 is the truth-level distribution $P(C_i)$.

For an iterative Bayesian unfolding, one begins by choosing an initial guess for the truth distribution $P(C_i)$ and calculating the expected number of true events $\hat{n}(C_i)|_{true}$ according to Equation 3.10. A uniform distribution can even be used as the initial guess [63]. The expected total number of true events, N_{true} is also calculated:

$$N_{true} = \sum_{i=1}^{n_C} \hat{n}(C_i)|_{true} \tag{3.11}$$

The initial guess for the truth distribution can then be updated using the calculated $\hat{n}(C_i)|_{true}$ and N_{true} [63]:

$$P(C_i) = \frac{\hat{n}(C_i)|_{true}}{N_{true}} \tag{3.12}$$

Then another iteration is performed using the updated truth distribution, and the result of the second iteration is used to calculate another updated guess for the truth distribution. The iterations are performed until the difference between successive iterations is sufficiently small, indicating that the unfolding procedure is converging on the truth distribution [63]. Too many unfolding iterations introduce statistical uncertainties on the unfolded distribution, so generally the minimum number of iterations for which the unfolding converges is used.

CHAPTER IV

Analysis Details

The measurement presented in this thesis of charged hadron distributions in beauty and charm jets was performed using data collected by the LHCb experiment in 2016. In this chapter, the steps of the data analysis will be described in detail. The criteria used to select events in data for the analysis is discussed, followed by a description of how the hadronization observables are measured from the data events. The calculation of the relevant jet and track efficiencies is described, and the results of the Bayesian unfolding procedure to remove detector smearing effects from the distributions are shown. Finally, several studies to determine systematic errors on the measured distributions are presented.

4.1 Data Selection

This analysis of charged hadron distributions in heavy flavor jets was completed using the full 2016 LHCb pp collision dataset collected at a center of mass energy of $\sqrt{s} = 13$ TeV. The dataset corresponds to an integrated luminosity of 1.6 fb^{-1} [19]. Collision events containing heavy flavor dijet candidates were reconstructed during data-taking in the second stage of the High Level Trigger (HLT2) and saved for further processing offline. In the offline data processing, the jet clustering and SV-tagging algorithms described in Sections 3.1 and 3.2.1, respectively, were re-run over the events selected by the HLT2 trigger. A vertex fit was performed to identify SV-tagged dijets, pairs of SV-tagged jets originating from the same primary vertex. Events selected for physics analysis were required to have passed specific trigger lines in the Level 0 (L0) hardware trigger and the first stage of the High Level Trigger (HLT1) that were consistent with being fired by a heavy-flavor dijet event. The L0 lines for this analysis required the presence of a reconstructed high-p_T muon in the event or a hadron, photon, or electron in the event with a high transverse energy measured in the calorimeter system [19]. The presence of a high p_T muon in the event increased the likelihood of an event containing a semileptonic beauty or charm hadron decay, while the presence of a large energy deposit in the calorimeter increased the likelihood of an event containing a high-p_T jet. The HLT1 lines for this analysis required the presence of at least one high-p_T charged particle that was inconsistent with originating from a primary vertex in the event, indicating the presence of a secondary decay vertex in the event that could have been produced from a heavy-flavor hadron decay. Either jet in the dijet could satisfy the trigger requirements, but the same jet was required to fire both the L0 and HLT1 lines. An event was not selected if one jet fired a required L0 line, and the other jet in the dijet fired the required HLT1 line, but neither jet individually passed both the L0 and HLT1 requirements.

4.1.1 Jet Selection

Both jets in the dijet were required to have a transverse momentum p_T larger than 17 GeV and a pseudorapidity η within the range $2.5 < \eta < 4.0$. Jets with $17 < p_T < 20$ GeV were only used for the unfolding procedure, which is discussed in detail in Section 4.3. The final measurement was only performed with jets with a $p_T > 20$ GeV. While the nominal acceptance of the LHCb detector spans from $2 < \eta < 5$, the nominal track acceptance within LHCb falls within a slightly reduced pseudorapidity range, from $2 < \eta < 4.5$. The jet pseudorapidity range was chosen so that in a jet with a radius of 0.5 in (η, ϕ) space, all of the tracks in the jet fell within the nominal LHCb track acceptance that is known to have good track reconstruction and quality. Additionally, the absolute difference in azimuthal angle between the two jets in the dijet, $|\Delta\phi| = |\phi_{jet1} - \phi_{jet2}|$, was required to be larger than 1.5 radians. The jet p_T, η, and $\Delta\phi$ cuts summarize the fiducial region of this measurement. The fiducial region is almost identical to that used in the $b\bar{b}$- and $c\bar{c}$-dijet cross section measurement that LHCb published using the same 2016 dataset [19], with the only difference being the slightly narrower jet pseudorapidity range used in the present analysis. Jet flavor separation into samples of beauty-dominated and charm-dominated jet samples was achieved by imposing a cut on the output of the BDT($b|c$) classifier described in Section 3.2.2. A cut of BDT($b|c$) greater than 0.1 was used to select b-jets, and a cut less than 0.1 was used to select c-jets.

4.1.2 Track-in-Jet Selection

Several cuts were applied to the tracks within the jets to select high-quality tracks corresponding to charged hadrons. The tracks were required to have a momentum between 4 and 1000 GeV, transverse momentum greater than 250 MeV, and a good track fit quality, defined as a track fit χ^2 per degree of freedom of less than 3. The LHCb particle identification system was not used to identify individual species of charged hadrons for this analysis, and instead the tracks were required to be identified as charged hadrons in the HLT particle flow algorithm. Finally, the ΔR between each track and the jet axis was required to be less than 0.5, to confirm that the track is within the jet radius.

4.2 Calculating the Hadronization Observables

The hadronization observables measured in this analysis include the longitudinal momentum fraction z, the transverse momentum j_T, and the radial distance r of charged hadrons in heavy flavor jets. The hadronization observables are defined again below:

$$z \equiv \frac{\mathbf{p}_{jet} \cdot \mathbf{p}_{chg}}{|\mathbf{p}_{jet}|^2} \tag{4.1}$$

$$j_T \equiv \frac{|\mathbf{p}_{jet} \times \mathbf{p}_{chg}|}{|\mathbf{p}_{jet}|} \tag{4.2}$$

$$r = \Delta R(chg, jet) \equiv \sqrt{(\phi_{chg} - \phi_{jet})^2 + (\eta_{chg} - \eta_{jet})^2} \tag{4.3}$$

In Equations 4.1-4.3, $\mathbf{p}$ is the momentum vector, ϕ is the azimuthal angle and η is the pseudorapidity. The subscript "jet" refers to the jet quantities and the subscript "chg" refers to the quantities for a charged hadron in the jet. The z, j_T, and r observables were calculated using the jets and charged hadrons selected with the criteria detailed in the previous section, and then used to compute the normalized charged hadron distributions according to the following definition:

$$\frac{1}{N^{meas}_{(b,c)-jet}} \frac{dN^{meas}_{track}}{dz} = \frac{\epsilon_{jet}}{U_{jet} \times N^{reco}_{(b,c)-jet}} \times \frac{1 - \epsilon^{gained}_{track}}{\epsilon^{in\,jet}_{track}} \times \frac{dN^{reco}_{track}}{dz}. \tag{4.4}$$

The j_T and r distributions were measured in the same way as Equation 4.4, with z replaced by j_T or r, accordingly. In Equation 4.4, dN^{reco}_{track}/dz is the raw number of charged hadrons in a given z bin divided by the bin width, and $N^{reco}_{(b,c)-jet}$ is the raw number of reconstructed b-jets or c-jets in a given jet p_T bin. Two corrections were applied to the raw number of jets: ϵ_{jet} is the total efficiency for reconstructing and selecting a b- or c-jet, and U_{jet} is a correction factor to account for bin migrations caused by the jet p_T resolution. Similarly, the raw number of charged hadrons in a bin was corrected by two efficiencies: $\epsilon^{in\,jet}_{track}$ is the efficiency with which tracks are correctly reconstructed within the jet, and $\epsilon^{gained}_{track}$ is the efficiency with which tracks are misreconstructed inside of the jet. The efficiency-corrected and unfolded number of charged hadrons in a given z bin is denoted by dN^{meas}_{track}/dz, while $N^{meas}_{(b,c)-jet}$ denotes the efficiency-corrected number of beauty or charm jets in a given jet p_T bin. The following sections describe the calculation of the jet and track efficiencies in more detail.

In addition to calcuating the distributions in Equation 4.4 with all of the tracks in the jet that passed the selections, the distributions are also calculated without using the tracks from the secondary vertex in the jet. The tracks from the secondary vertex are produced from the decay of the heavy flavor hadron in the jet, and not by the hadronization of the heavy quark itself. Therefore, the "PV-only" track distributions directly probe the hadronization of the heavy flavor quark by only including tracks produced directly from the heavy quark that initiated the jet. The distributions with the SV tracks, the "PV+SV" track distributions, are of interest for improving jet flavor tagging algorithms, as the decay products from the heavy flavor hadron decay in the jet can provide valuable information for discriminating between beauty and charm jets.

4.2.1 Jet Efficiency Corrections

Since the z, j_T, and r distributions were measured as a function of jet p_T and normalized by the total number of jets in each jet p_T bin, several jet efficiencies were determined as a function of the jet p_T in order to correct the raw number of jets. The jet efficiencies needed for this analysis included the jet reconstruction efficiency, the jet acceptance correction or "unfold factor", the jet SV-tagging efficiency, the jet BDT cut efficiency, and the efficiencies associated with the trigger selections. The total jet efficiency ϵ_{jet} in Equation 4.4 included all of these efficiencies except the unfold factor and was defined in the following way:

$$\epsilon_{jet} = \epsilon_{jet}^{reco} \times \epsilon_{jet}^{SV} \times \epsilon_{jet}^{trig} \times \epsilon_{jet}^{BDT} \tag{4.5}$$

Since the same trigger selections were used for this analysis as for the previous LHCb measurement of the $b\bar{b}$- and $c\bar{c}$-dijet cross section in the same 2016 dataset, the trigger efficiencies for this analysis were taken from the cross section measurement [19]. The remaining efficiencies described in this section were determined using simulated MC events, as the number of heavy flavor jets at truth level was needed to compute most of the efficiencies. The MC events used in this analysis were simulated $\sqrt{s} =$ 13 TeV $pp \to b$-dijet and $pp \to c$-dijet events produced with the PYTHIA 8 event generator [64]. For each dijet flavor, three different event samples generated with different hard scattering scales $\hat{p_T}$ were used: a $15 < \hat{p_T} < 20$ GeV sample, a 20 $< \hat{p_T} <$ 50 GeV sample, and a $\hat{p_T} >$ 50 GeV sample. When used together, the MC events were weighted according to the ratios of the partonic cross sections in the different $\hat{p_T}$ bins. The two heavy-flavor dijets were required to be produced within the forward acceptance accessible to LHCb. Further details about the PYTHIA 8 configuration used in this analysis are listed in Appendix A.

The jet reconstruction efficiency, ϵ_{jet}^{reco}, is defined as the efficiency with which a truth-level MC jet is reconstructed within the LHCb detector. For this analysis, ϵ_{jet}^{reco} was defined as the fraction of truth-level MC heavy flavor dijets that were reconstructed at the detector-level. A truth-level dijet was considered reconstructed if there was a reconstructed-level jet within $\Delta R < 0.5$ of each truth-level jet in the dijet. Both the truth-level dijets and their corresponding reconstructed dijets in ϵ_{jet}^{reco} were required to pass the fiducial region cuts in jet p_T, η and $\Delta\phi$ listed in Section 4.1.1. Figure 4.1 shows ϵ_{jet}^{reco} for beauty and charm jets as a function of the jet p_T for both configurations of the LHCb dipole magnet polarity. The jet reconstruction efficiency increases as a function of the jet p_T, as higher energy jets tend to have both a higher particle multiplicity and higher energy particles, both of which facilitate jet-finding

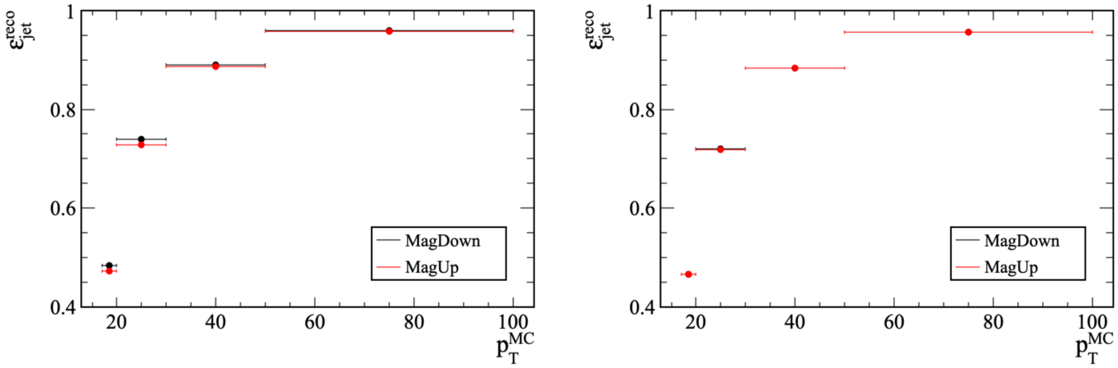

Figure 4.1: The jet reconstruction efficiency for b-jets (left) and c-jets (right) as a function of jet p_T for both LHCb magnet polarities.

at the detector-level.

The jet unfold factor, U_{jet}, is an acceptance correction that accounts for bin migrations between jet p_T bins. The unfold factor was defined as the ratio of the number of truth-matched reconstructed dijets that passed the analysis selection cuts and the number of reconstructed truth dijets that passed the fiducial region cuts. The matching condition between truth dijets and reconstructed dijets in both the numerator and denominator of U_{jet} is necessary as the jet reconstruction efficiency was applied as a separate correction. The value of U_{jet} indicates the fraction of the truth-level jets in a given jet p_T bin whose reconstructed jets were also correctly reconstructed in the same jet p_T bin. Previous LHCb analyses showed that correcting the jet p_T spectrum with the jet reconstruction efficiency and unfold factor is consistent with a full unfolding of the jet p_T spectrum to within 1% [37]. Figure 4.1 shows the jet unfold factor for beauty and charm jets as a function of the jet p_T. As with ϵ_{jet}^{reco}, U_{jet} increases as a function of jet p_T, as the majority of bin migrations occur at low jet p_T where the jet p_T spectrum is steeply falling.

The jet SV-tagging efficiency, ϵ_{jet}^{SV} was defined as the fraction of reconstructed dijets with two SV-tagged jets. Figure 4.3 shows ϵ_{jet}^{SV} as a function of the single jet p_T for beauty and charm jets, and as a function of the p_T of each jet in b-dijets.

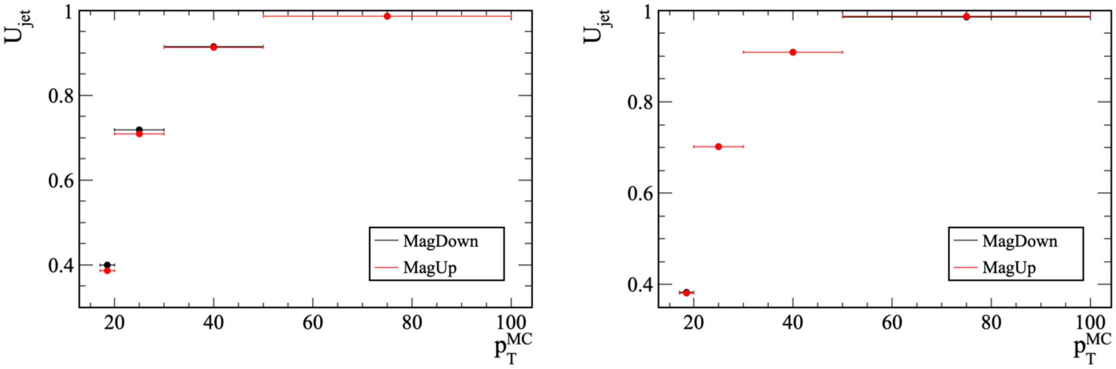

Figure 4.2: The jet unfold factor for b-jets (left) and c-jets (right) as a function of jet p_T for both LHCb magnet polarities.

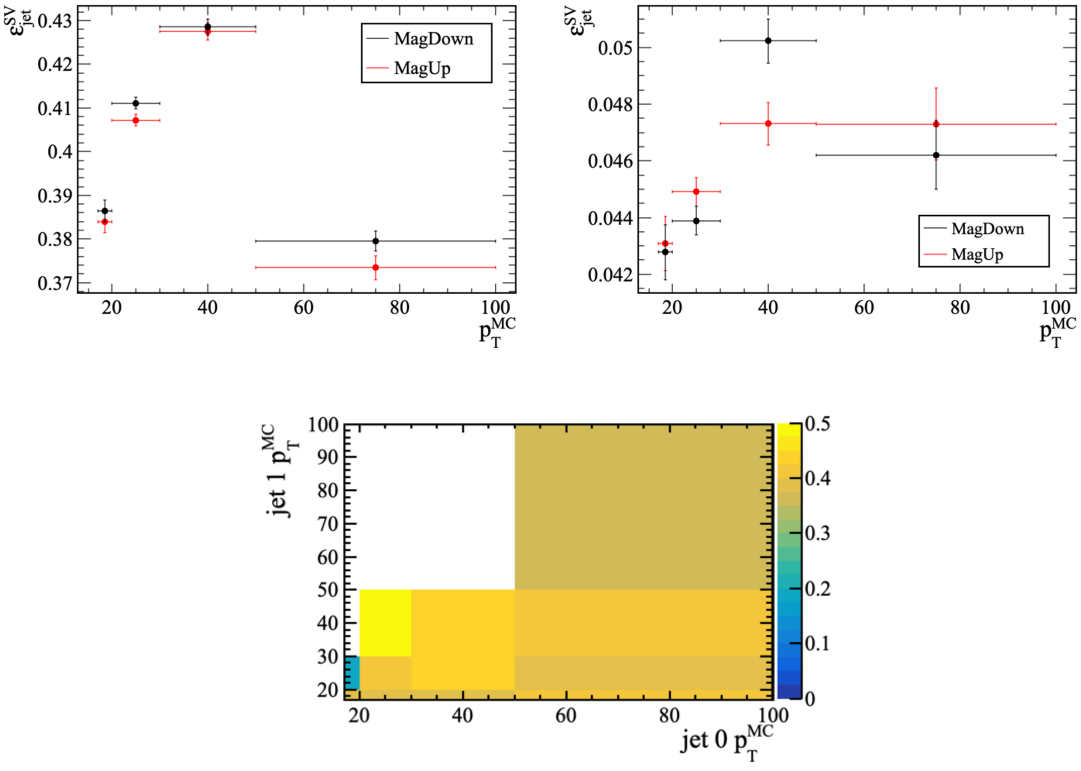

Figure 4.3: The jet SV-tagging efficiency as a function of the single jet p_T for b-jets (top left) and c-jets (top right) and as a function of the p_T of each jet in b-dijets (bottom). Note that in the bottom figure, the z-axis indicates ϵ_{jet}^{SV}.

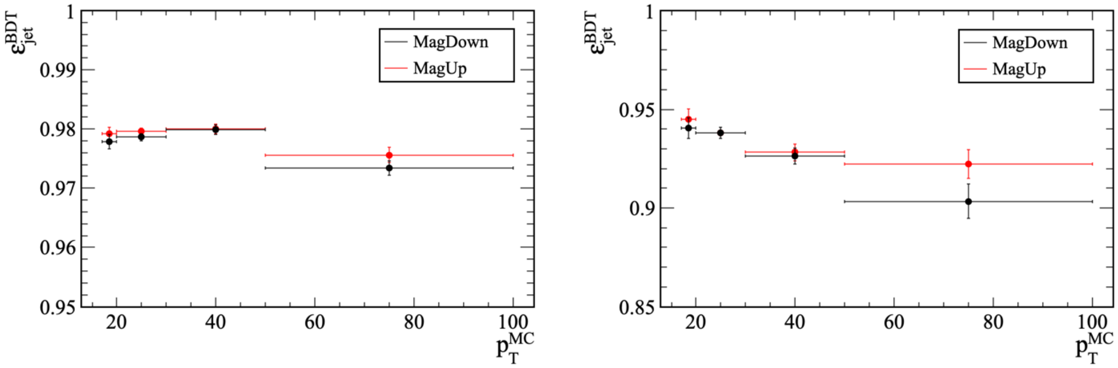

Figure 4.4: The jet BDT cut efficiency as a function of the single jet p_T for b-jets (left) and c-jets (right).

The two-dimensional SV-tagging efficiency is generally uniform across the jet p_T bins, ranging from approximately 37% to 44%. When binned as a function of the single jet p_T, the SV-tagging efficiency drops slightly in the highest jet p_T bin. The jet BDT cut efficiency, ϵ_{jet}^{BDT}, was defined as the fraction of SV-tagged dijets with at least one jet passing the applied BDT($b|c$) cut. Figure 4.4 shows ϵ_{jet}^{BDT} for beauty and charm jets as a function of the jet p_T. The fraction of b-jets that passed the BDT cut was nearly uniform, which is consistent with the data sample previously being measured to contain mostly b-jets, which score high on the BDT($b|c$) classifier [19]. The fraction of c-jets that passed the BDT cut was lower than that for b-jets, primarily because the BDT($b|c$) response for charm jets is not as localized as it is for b-jets.

4.2.2 Track Efficiency Corrections

Two track efficiencies were used to correct the raw distributions of hadrons in the jets. The track-in-jet efficiency, ϵ_{track}^{injet}, accounts for tracks present in a truth-level jet that were not reconstructed within the corresponding reconstructed jet at the detector-level. The track-in-jet efficiency was calculated as the fraction of truth-level tracks in a given (p, η) bin that were successfully reconstructed within the reconstructed jet at detector-level. The track-gained efficiency accounts for truth-

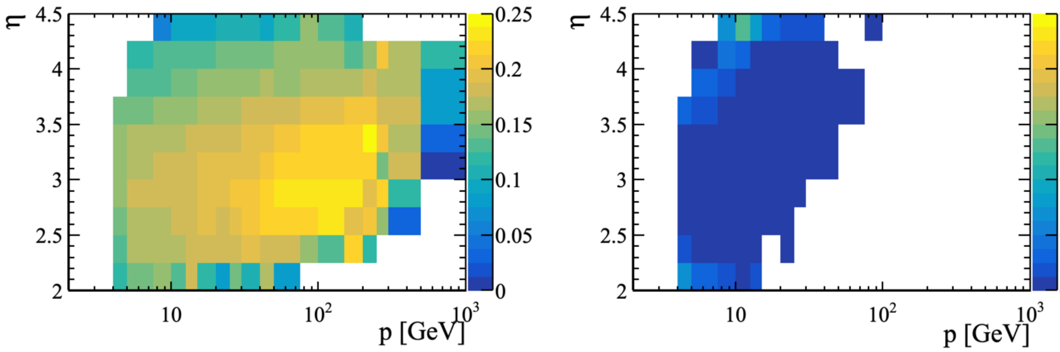

Figure 4.5: The track-in-jet efficiency (left) and the track-gained efficiency (right) for b-jets as a function of the track momentum and pseudorapidity. The z axis labels the value of the efficiency, which includes the jet efficiencies described in the previous section.

level tracks that were not present in the truth-level jet but were reconstructed within the corresponding reconstructed jet. The track-gained efficiency was calculated as the fraction of truth-level tracks in a given (p, η) bin that were misreconstructed within the jet radius. For both track efficiencies, the truth-level tracks and their reconstructed-level counterparts were required to pass the track selections detailed in Section 4.1.2.

Figure 4.5 shows ϵ_{track}^{injet} and $\epsilon_{track}^{gained}$ for b-jets as a function of the track momentum and pseudorapidity, while Figure 4.6 shows the corresponding efficiencies for c-jets. Since the total number of jets used to normalize the charged hadron distributions was corrected with the jet efficiencies described in the previous section, the total number of charged hadrons also needed to be corrected to take into account the efficiency-corrected number of jets. Therefore, the denominator of both track efficiencies includes all truth-level dijets passing the fiducial cuts, regardless of whether or not the jets were reconstructed at the detector level. This definition of the track efficiencies accounts for all of the jet efficiencies described in the previous section when the truth-matched reconstructed jets are required to pass the analysis selection cuts. The efficiency values shown on the z axes in Figures 4.5 and 4.6 therefore describe

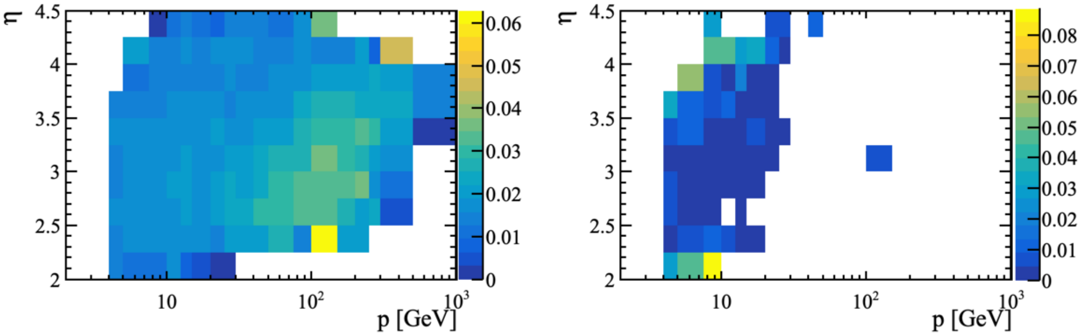

Figure 4.6: The track-in-jet efficiency (left) and the track-gained efficiency (right) for c-jets as a function of the track momentum and pseudorapidity. The z axis labels the value of the efficiency, which includes the jet efficiencies described in the previous section.

the efficiency with which a truth-level track in a truth-level b- or c-jet is reconstructed in a reconstructed, SV-tagged jet passing the BDT cuts, kinematic cuts, and trigger selections listed in Section 4.1.

4.3 Unfolding

Due to the large jet p_T resolution, there are a non-negligible amount of bin migrations in the z, j_T, and r distributions which must be corrected for with an unfolding procedure. A two-dimensional iterative Bayesian unfolding, the details of which are described in detail in Section 3.3, was used for this analysis. The reconstructed z, j_T, and r distributions measured in MC were efficiency-corrected using the track efficiencies described in Section 4.2.2 and then used to construct response matrices for the unfolding procedure.

4.3.1 Response Matrices

The response matrices were constructed using simulation, as both the truth-level MC generated in PYTHIA and the reconstructed-level MC passed through the detector simulation are needed in order to quantify how much the truth distributions

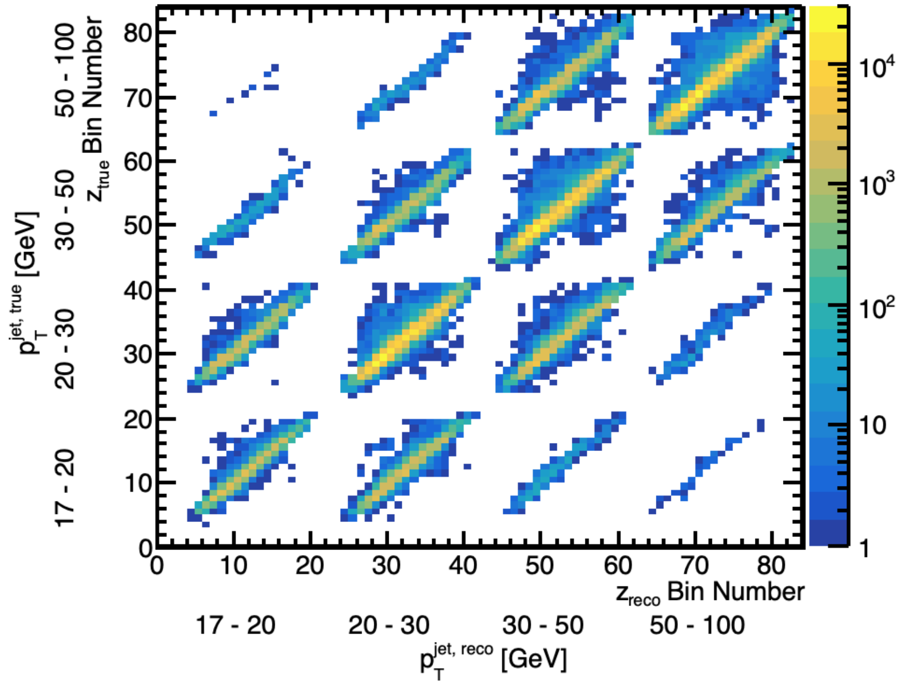

Figure 4.7: The 2D response matrix for the longitudinal momentum fraction, z.

are smeared after their reconstruction at the detector-level. Simulated b-dijet events were used to construct the response matrices, since the c-dijet events have a low reconstruction efficiency and the detector response is expected to be similar for b- and c-jets [19]. The truth-level MC events were required to pass the fiducial selection cuts, while the reconstructed MC events were subjected to the same selection cuts as applied to real data. Tracks in the reconstructed jets were matched to tracks in the truth jets via a standard LHCb algorithm that matches reconstructed MC particles to their corresponding truth MC particles, if they exist. The pairs of matched truth and reconstructed jets and truth and reconstructed tracks are used to construct a response matrix for each observable.

Figure 4.7 shows the response matrix constructed for the longitudinal momentum fraction z. The sixteen large bins in the matrix account for migrations between the truth-level and reconstructed-level jet p_T bins, while the smaller bins within a single large bin indicate migrations between the truth-level and reconstructed-level z bins. The four large bins along the diagonal of the response matrix indicate reconstructed jets that were correctly reconstructed within their true jet p_T bin. The concentration of most entries along the diagonal in Figure 4.7 indicates that most jets are reconstructed in the correct truth-level p_T bin. However, there are also a significant number of entries in the off-diagonal p_T bins, indicating that a non-negligible number of jets are reconstructed in a $p_T^{jet,reco}$ bin other than their $p_T^{jet,true}$ bin. Most of the jet p_T migrations are between neighboring p_T bins; the few entries in the large bins at the top left and bottom right of the response matrix indicate that there are very few jets whose true p_T is in the 17-20 GeV bin but whose reconstructed p_T is in the 30-50 GeV or 50-100 GeV bins, and vice versa. Within one of the large bins in the response matrix, a diagonal line indicates the number of charged hadrons whose reconstructed z was correctly reconstructed in the true z bin. Off-diagonal entries within a large bin indicate charged hadrons whose z^{reco} was reconstructed in a bin other than the z^{true}. Figure 4.7 shows that within a ($p_T^{jet,reco}$, $p_T^{jet,true}$) bin, most of the reconstructed z values are reconstructed in the correct truth-level z bin, even in off-diagonal ($p_T^{jet,reco}$, $p_T^{jet,true}$) bins. However, there are still a significant number of entries in off-diagonal (z^{reco}, z^{true}) bins, indicating detector smearing effects on z that the unfolding will correct. The response matrix shows expected kinematic effects on z due to the jet p_T smearing; for jets that have a reconstructed p_T larger than their truth p_T, the z smearing is concentrated above the diagonal, indicating that the reconstructed z values are smaller than the true z values. For jets that have a reconstructed p_T smaller than their truth p_T, the opposite is true: the smearing in z is concentrated below the diagonal, indicating that the reconstructed z values are larger than the true z values.

The response matrices for the transverse momentum relative to the jet axis, j_T, and the radial distance of hadrons from the jet axis, r, are shown in Figures 4.8 and

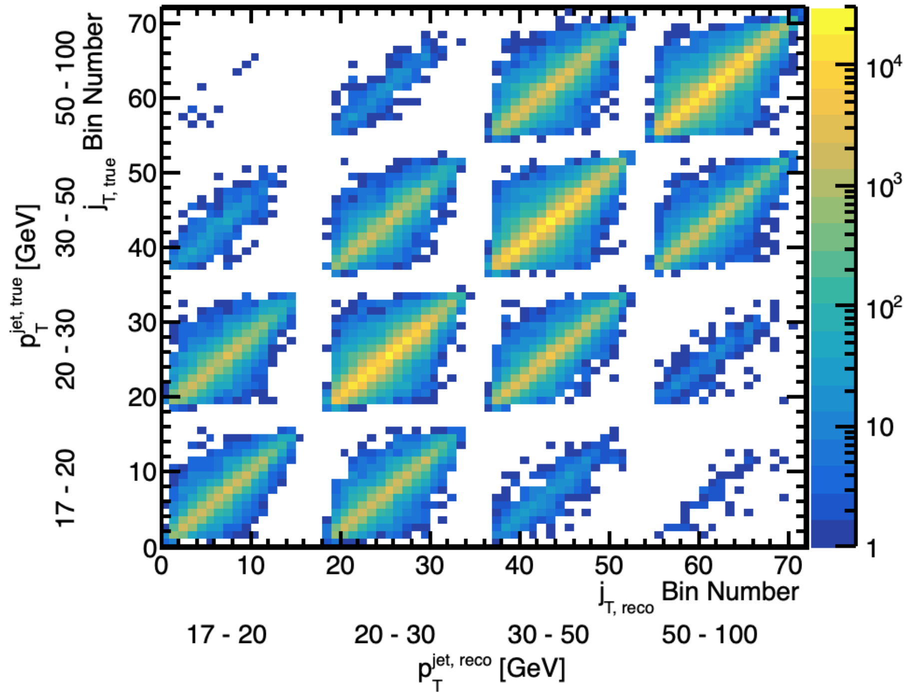

Figure 4.8: The 2D response matrix for the transverse momentum with respect to the jet axis, j_T.

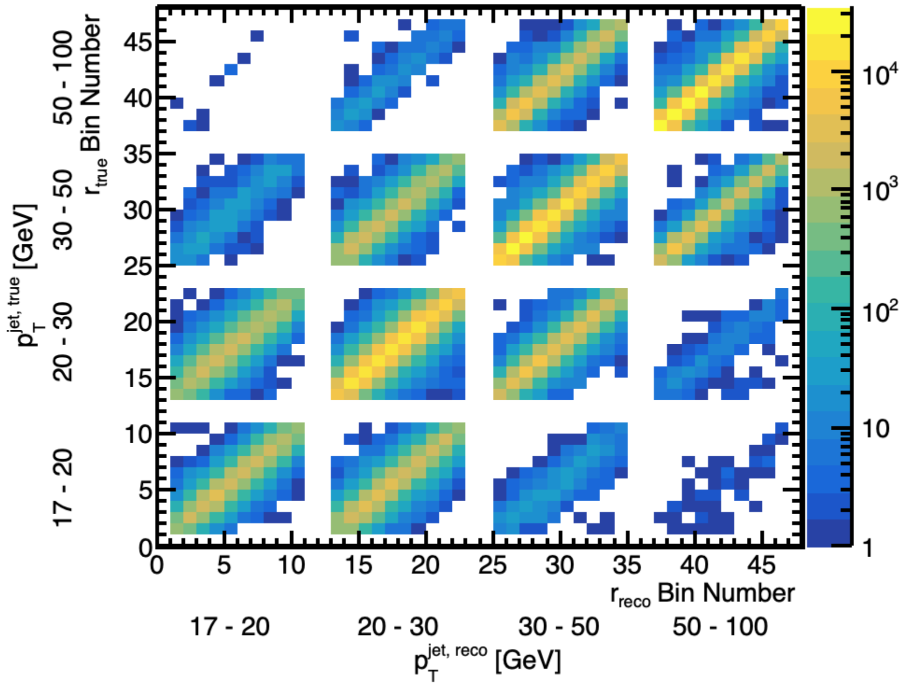

Figure 4.9: The 2D response matrix for the radial distance from the jet axis, r.

4.9, respectively. The j_T and r response matrices show a larger amount of smearing than observed in the z matrix, as indicated by the broader spread of entries above and below the diagonal in each ($p_T^{jet,reco}$, $p_T^{jet,true}$) bin. As neither the j_T nor the r observables explicitly depend on the jet p_T, the smearing observed in their response matrices is due primarily to the jet ϕ and η resolution, which is larger than the j_T and r bin sizes.

4.3.2 Closure Test

The raw distributions measured in data were efficiency-corrected and then unfolded seven times. The unfolding iteration number was determined by performing an additional iteration on the efficiency-corrected MC distributions and observing the change in the z, j_T, and r distributions relative to the previous iteration. The

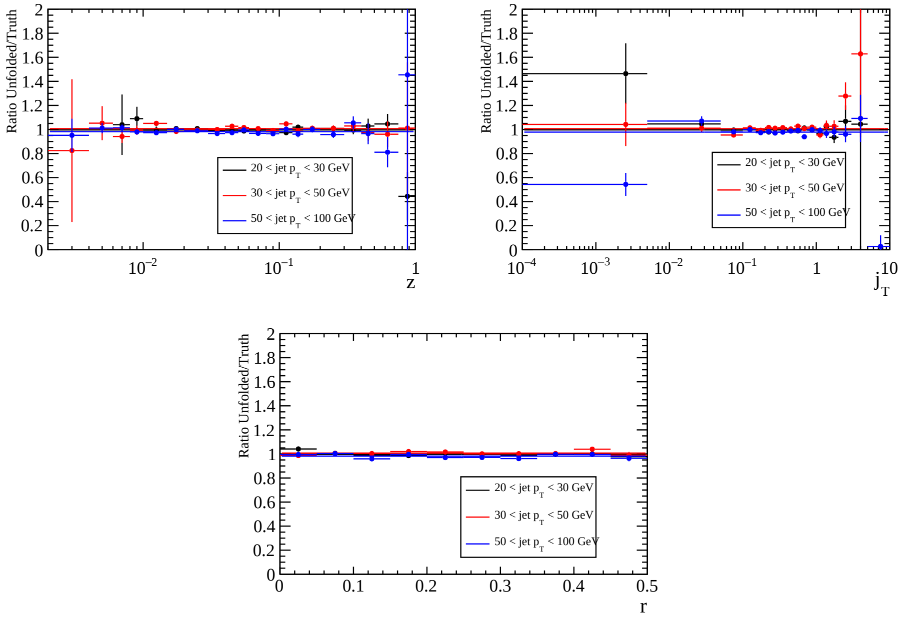

Figure 4.10: The results of the unfolding closure test in each jet p_T bin for the z, j_T, and r distributions. The straight lines are a fit to the distribution in each jet p_T bin to determine the deviation from unity.

iteration number was chosen when an additional iteration changed the distributions by less than 0.5%, excluding the edge bins which can vary by up to 3%.

To test that the unfolding recovered the truth distributions in the selected number of iterations, a closure test was performed using the MC. Since the MC was generated for each configuration of the LHCb dipole magnet, with the magnet polarity in the $+\hat{y}$ direction (Magnet Up) or in the $-\hat{y}$ direction (Magnet Down), these two independent MC samples were used to perform the closure test. The Magnet Down MC sample was arbitrarily used to generate response matrices, which were then used to unfold the reconstructed z, j_T, and r distributions measured in the Magnet Up MC. The ratio of the unfolded distributions to the truth distributions was calculated and a fit was performed to determine the deviation from unity in each jet p_T bin. Figure 4.10 shows the results of the closure test. All bins are consistent with unity, indicating they pass the closure test, except for the highest j_T bin which is dropped from the final measurement.

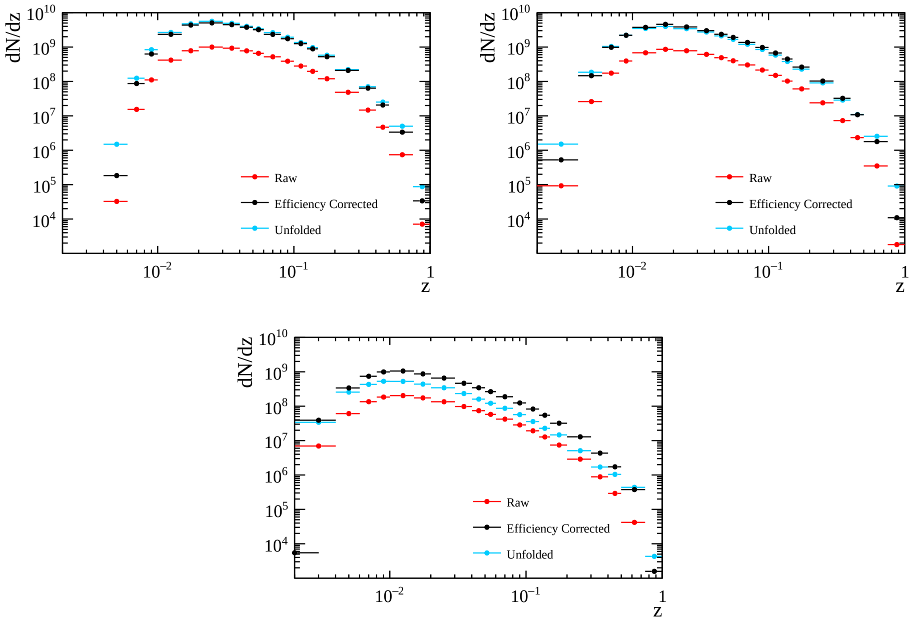

Figure 4.11: The raw (red), efficiency-corrected (black), and unfolded (blue) z distributions for b-jets in real data in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

Figures 4.11, 4.12, and 4.13 show the raw, efficiency-corrected, and unfolded distributions for b-jets in data for z, j_T, and r, respectively, while Figures 4.14-4.16 show the corresponding distributions for c-jets in data. As will be discussed in more detail in Section 4.4.1, several bins at high and low z and j_T are combined for the charm distributions due to the low MC statistics as a result of the small charm dijet SV-tagging efficiency. The low statistics available for the charm MC present a difficulty when assigning systematic uncertainties for the charm distributions. The charm binning shown in Figures 4.14-4.16 is currently used for assigning systematic uncertainties on the final measured distributions, however it should be noted that the charm statistics available in data are significantly higher and the limited charm dijet MC statistics can be improved upon. The distributions show that the unfolding is a small correction relative to the efficiency correction. As expected, the unfolding also tends to migrate counts from the higher jet p_T bins to the lowest jet p_T bin. Due to the steeply falling jet p_T spectrum, the smearing at the detector level tends to result in more truth-level jets being shifted into higher reconstructed jet p_T bins than into lower jet p_T bins. The unfolding corrects this migration in the jet p_T bins, as well as between the individual z, j_T, and r bins.

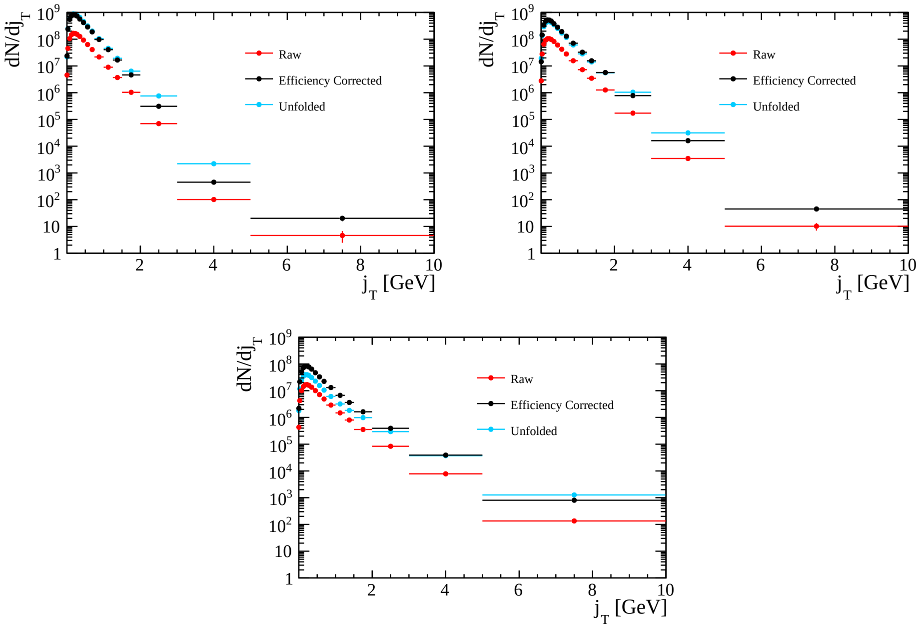

Figure 4.12: The raw (red), efficiency-corrected (black), and unfolded (blue) j_T distributions for b-jets in real data in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

4.4 Systematic Uncertainties

4.4.1 Bias from heavy flavor jet tagging

The SV-tagging algorithm used to tag the jets uses information about the tracks in the jets and applies cuts to them in order to reconstruct a secondary vertex consistent with originating from a heavy flavor hadron decay, as described in detail in Section 3.2.1. The displaced track selection and non-negligible inefficiency of the SV-tagging algorithm could potentially result in the z, j_T, and r distributions being biased towards heavy flavor jets that pass the SV-tagging requirement, rather than describing the general features of hadronization in inclusive heavy flavor jets. A systematic un-

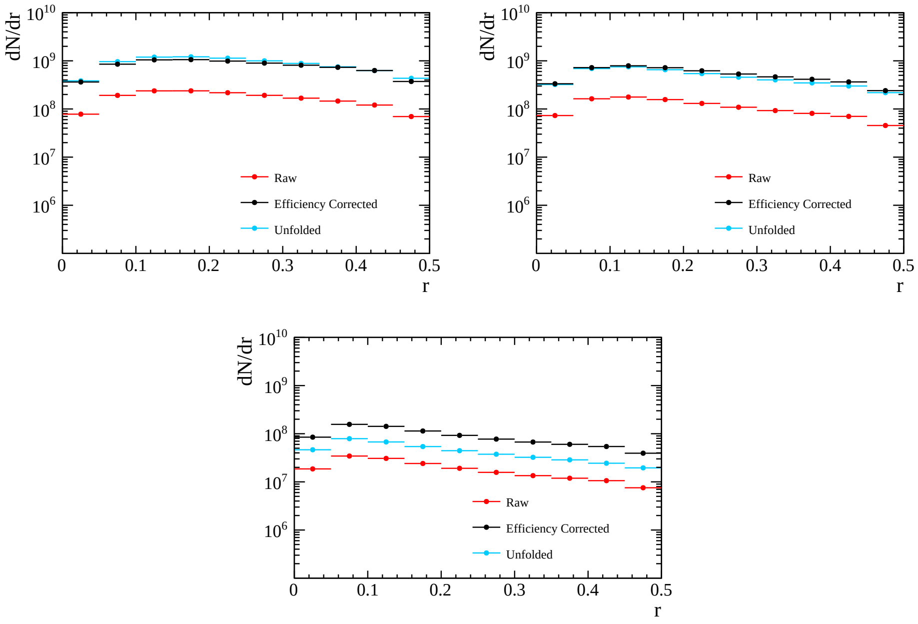

Figure 4.13: The raw (red), efficiency-corrected (black), and unfolded (blue) r distributions for b-jets in real data in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

certainty was assigned for this bias by studying the effect of the SV-tagging on the z, j_T, and r distributions in simulation. Reconstructed MC dijets were selected with the same requirements as applied to real data, with the exception of the SV-tag and BDT cut requirements. The reconstructed MC dijets were matched to their corresponding truth-level MC dijets, and the truth-level charged hadron distributions were measured. The distributions in the truth-level MC jets, rather than the distributions in the reconstructed MC jets, were used for this study to avoid including bin migration effects in the determination of the heavy flavor tagging bias, which were corrected for separately with the unfolding. The truth charged hadrons were selected with the requirements listed in Section 4.1.2, with the exception of the track χ^2 requirement. The charged hadron distributions were measured in three different categories of jets: all jets in the selected dijets, SV-tagged jets in the selected dijets, and BDT-tagged jets in the selected dijets. When selecting SV-tagged jets for this study, both jets in the dijet were required to pass the SV-tag requirement, as is the case in real data.

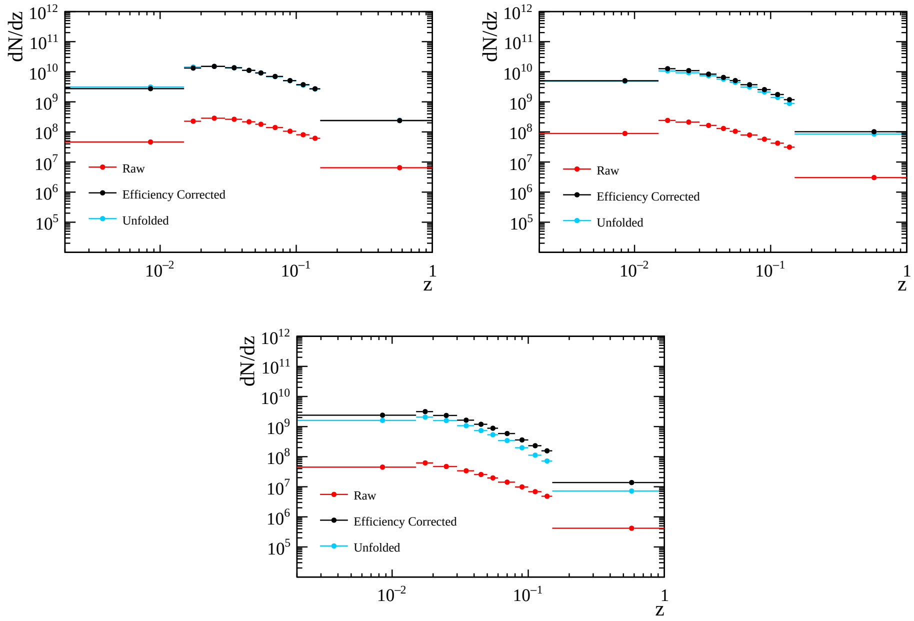

Figure 4.14: The raw (red), efficiency-corrected (black), and unfolded (blue) z distributions for c-jets in real data in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

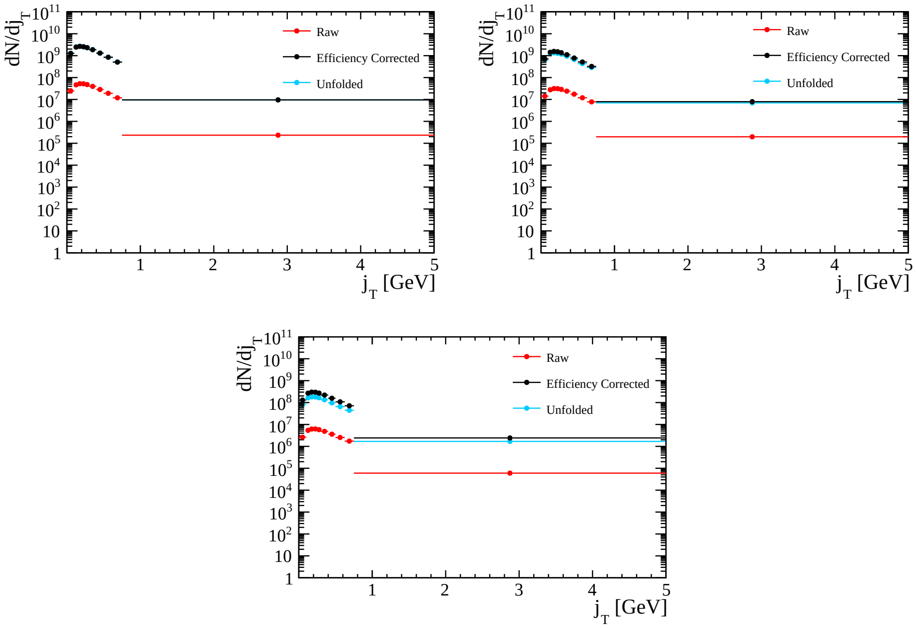

Figure 4.15: The raw (red), efficiency-corrected (black), and unfolded (blue) j_T distributions for c-jets in real data in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

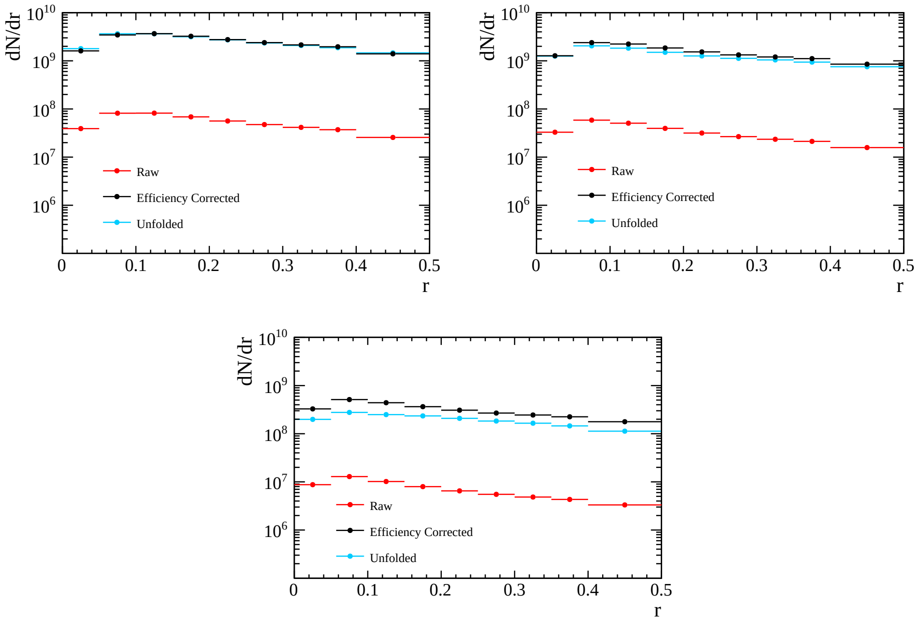

Figure 4.16: The raw (red), efficiency-corrected (black), and unfolded (blue) r distributions for c-jets in real data in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

When selecting BDT-tagged jets, only one of the jets in the dijet was required to pass the BDT requirement.

The ratio of each charged hadron distribution in SV-tagged jets to the jets with no heavy flavor tag requirement applied was measured to study how the SV-tagging algorithm affected the shape of the charged hadron distributions. To study the effect of the applied BDT cut on the distributions, the ratio of each distribution in BDT-tagged jets to SV-tagged jets was measured. Comparing the distributions in BDT-tagged jets to those in SV-tagged jets was necessary to study the effect of the BDT cut alone since the BDT tag requirement can only be applied to jets that have already passed the SV-tagging algorithm. The bias of the combined SV-tag and BDT cut requirements used to tag heavy flavor jets in this analysis was also studied. The ratio of distributions with both the SV-tagging and BDT cut applied to distributions with no heavy flavor tagging applied was measured to study the bias of the heavy flavor jet selection on the charged hadron distributions. For each bias study, the SV-tag bias, BDT cut bias, and the combined SV+BDT tag bias, the relevant ratios were calculated for distributions with and without the SV tracks. Each bias study was performed in the jet $\mathrm{p_T}$ bins used in the final measurement.

The results of the combined SV-tagging and BDT cut bias on the z, j_T, and r distributions in b-jets are shown in Figures 4.17, 4.19, and 4.21, respectively, as a function of jet p_T and with and without the SV tracks. The corresponding distributions in c-jets are shown in Figures 4.18, 4.20 and 4.22, respectively. Due to the low charm dijet reconstruction efficiency, the c-dijet MC begins to run out of statistics in the highest jet p_T bin and at high and low z and j_T. The bins with low statistics are combined to assign a final heavy flavor tagging systematic uncertainty on the charm distributions. For all results, a ratio of one would indicate no bias. Therefore deviations from one indicate where the heavy flavor tagging biases the shape of the charged hadron distributions, and where a larger systematic uncertainty must be applied. Most of the observed bias is due to the SV-tagging and not the applied BDT cut, as can be seen from the individual SV-tag and BDT cut bias studies shown in Appendix B. The b-jet z distributions, shown in Figure 4.17, all exhibit an enhancement in the number of charged hadrons selected with a z value of approximately 0.15 relative to inclusive b-jets. From previous measurements of heavy flavor fragmentation functions, the SV in the jet is expected to carry most of the jet momentum and therefore would be expected to have a high z value. However, the decay products from the SV have smaller momenta and as a result would be expected to have smaller z values than the SV itself. For SV decays involving more than two final decay products, which are common for beauty hadron decays, the z values of the SV decay products would be expected to be even smaller. The c-jet z distributions, shown in Figure 4.18, confirm this interpretation as they are strongly peaked towards higher z values than the b-jet distributions. Charm hadron decays tend to include fewer decay products than beauty decays, therefore charged hadrons from these decays should

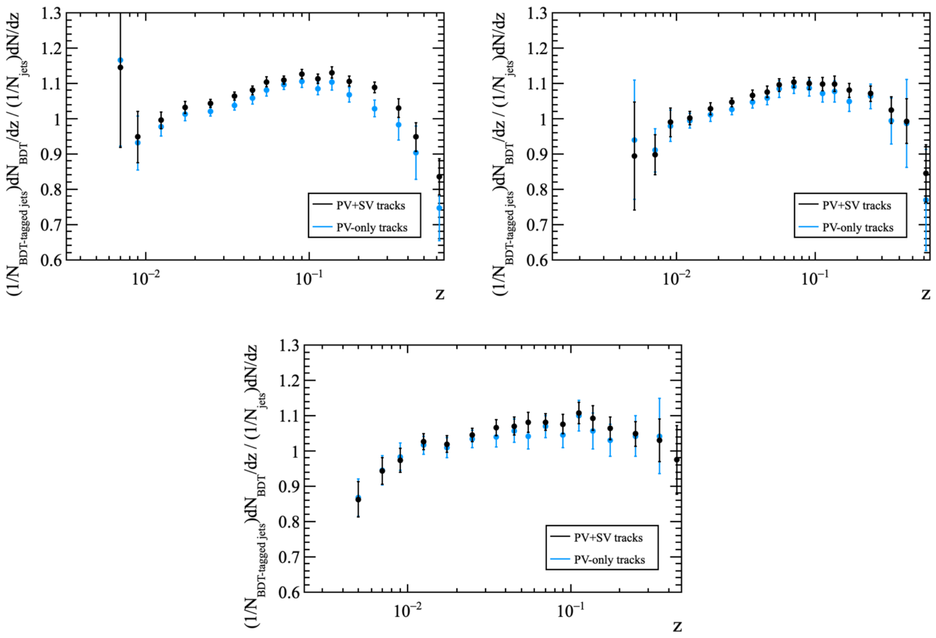

Figure 4.17: The combined bias due to the SV-tagging and BDT requirements on the z distributions in beauty jets. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

carry a higher fraction of the heavy flavor hadron momentum, consistent with the peak observed at large z values.

The b-jet j_T distributions in Figure 4.19 show an enhancement at high j_T values relative to the distributions for inclusive b-jets. The enhancement is more pronounced for the distributions containing the SV tracks, indicating that these tracks tend to have higher j_T values. A similar enhancement, slightly less pronounced due to the lower charm statistics, is observed in the c-jet j_T distributions in Figure 4.20. The enhancement is likely due to the explicit requirement in the SV-tagging algorithm that requires the SV in the jet to have a p_T of at least 2 GeV. The b-jet r distributions, shown in Figure 4.21, show a slight peak relative to inclusive b-jets at low r values. The broad peak narrows slightly and shifts to lower r values with increasing jet p_T, both of which are features that are consistent with the dead cone effect which is simulated in Pythia. A similar trend is observed in the lowest jet p_T bins of the c-jet r distributions, shown in Figure 4.22, although the peak at low r is much more

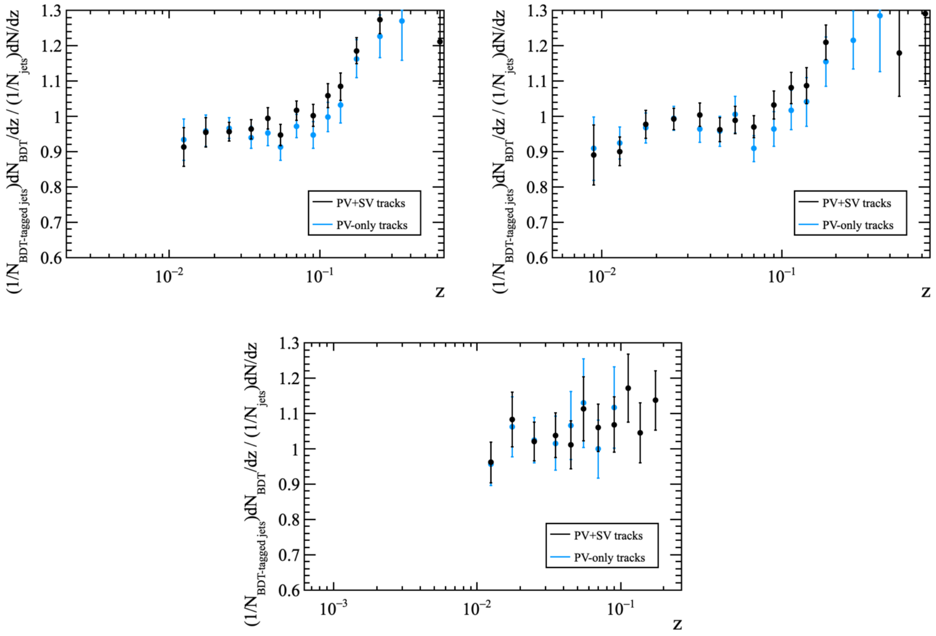

Figure 4.18: The combined bias due to the SV-tagging and BDT requirements on the z distributions in charm jets. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

pronounced for the charm distributions. The sharper peak observed for c-jets could be attributed to the fewer decay products from charm decays, which would tend to localize the approximate position of the charm hadron more than the decay products of beauty hadrons, which effectively smear out the r distribution. While the dead cone is also present in the inclusive heavy flavor jets, it is not surprising that partially or fully reconstructing the SV in the jet, which can be interpreted as a proxy for the radiating heavy flavor quark, would tend to enhance the selection of tracks sensitive to the dead cone.

For all ratios, the PV-only distributions have a slightly smaller bias than the PV+SV distributions. The smaller bias is consistent with tracks in the SV being more sensitive to specific kinematic cuts that are applied to the SV in the jet. However, the bias in the PV-only distributions is not zero because the jets are still SV-tagged and required to pass the BDT cut. While the SV tracks are not included in calculating the PV-only z, j_T, and r distributions, they are still present in the jet and therefore still

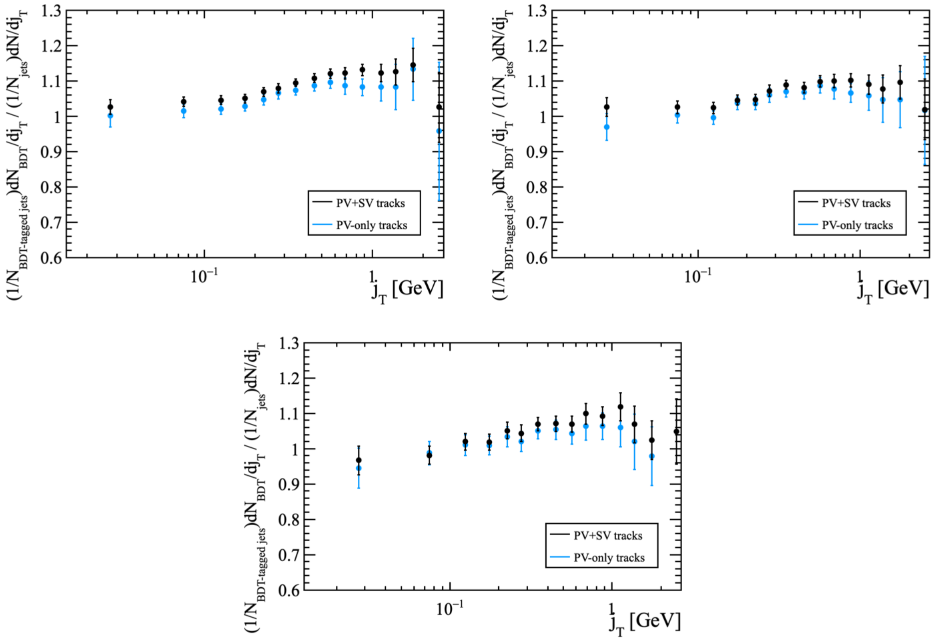

Figure 4.19: The combined bias due to the SV-tagging and BDT requirements on the j_T distributions in beauty jets. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

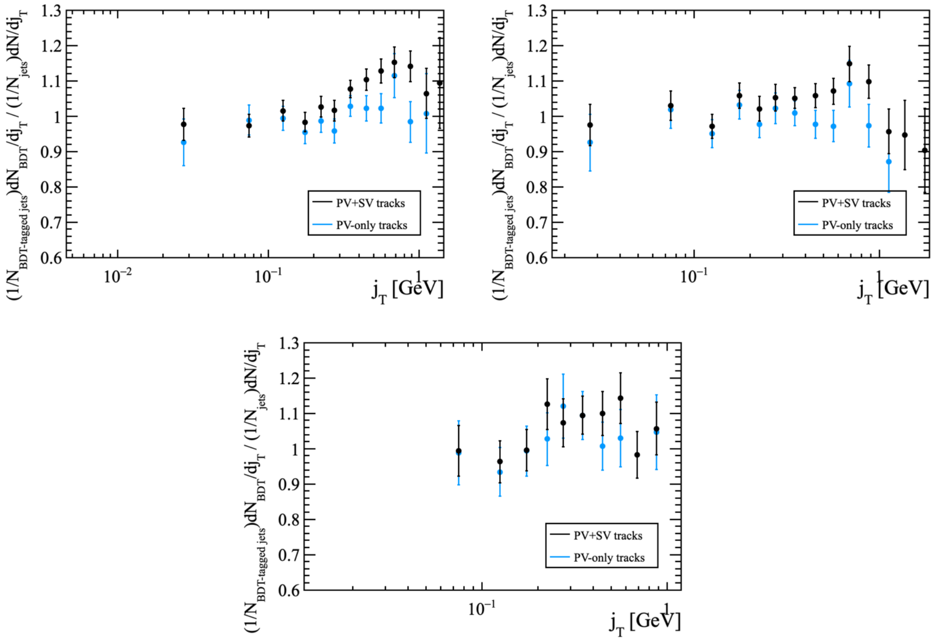

Figure 4.20: The combined bias due to the SV-tagging and BDT requirements on the j_T distributions in charm jets. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

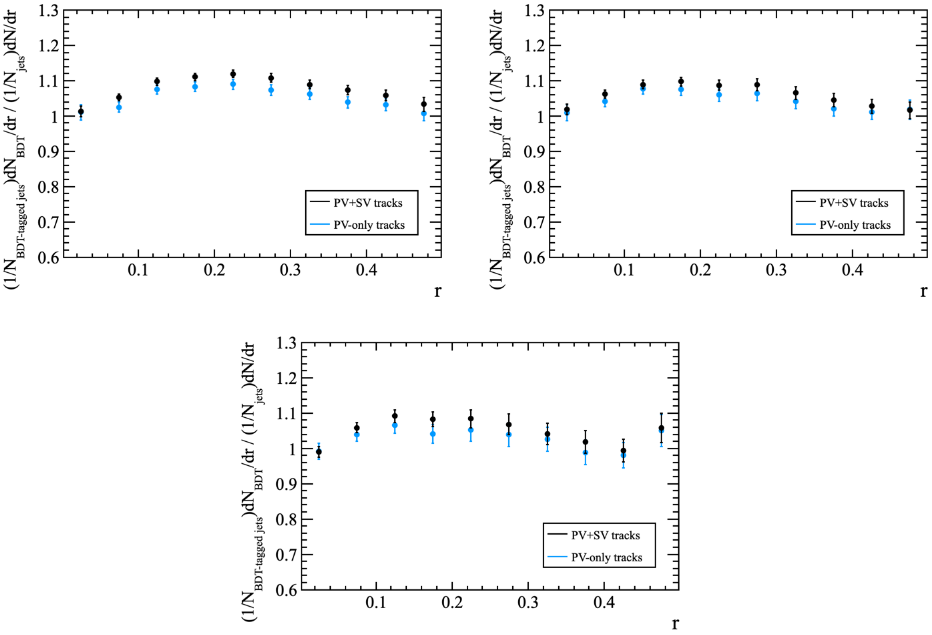

Figure 4.21: The combined bias due to the SV-tagging and BDT requirements on the r distributions in beauty jets. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

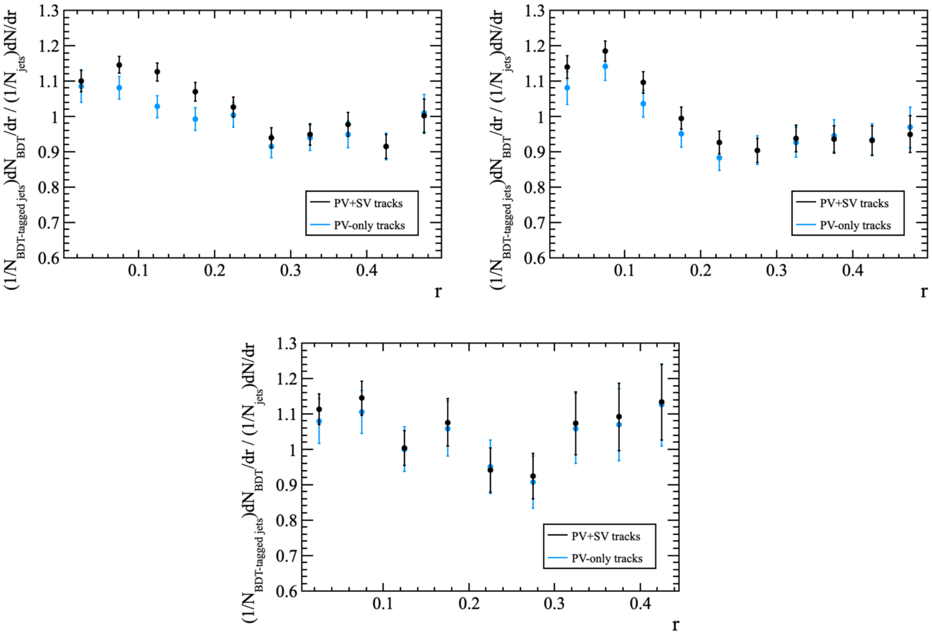

Figure 4.22: The combined bias due to the SV-tagging and BDT requirements on the r distributions in charm jets. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

induce a bias. For all distributions, the deviation from unity is taken as a systematic uncertainty, which is generally less than 10% for most bins.

4.4.2 Jet energy scale and resolution

The uncertainty with which the detector simulation accurately describes the detector response to jets is also considered as a systematic uncertainty. Studies of Z-tagged jets in Run 2 data were used to compare the detector response in data and simulation [19]. When the Z boson and the jet are produced nearly back-to-back, the p_T of the Z is to a good approximation the p_T of the produced jet. Therefore measuring the ratio of the $p_T(\text{jet})/p_T(Z)$ in data and simulation allows for quantitative comparisons between the detector response to jets in data and simulation. The jet energy resolution (JER) is determined as the maximum smearing that must be applied to the jet p_T in simulation in order for the width of the $p_T(\text{jet})/p_T(Z)$ distribution to agree with data within 1σ. The jet energy scale (JES) is determined as the scale factor by which the jet p_T in simulation must be multiplied in order to agree with the value at which $p_T(\text{jet})/p_T(Z)$ is peaked in data to within 1σ. In Run 2 data, the JER was found to be 20%, and the JES factor to be 1.05 [19]. To evaluate the effects of the JER uncertainty on the z, j_T, and r distributions, each component of the jet four-momentum vector in simulation is smeared by a Gaussian with a width of the JER uncertainty. The efficiencies and response matrices are recalculated and the ratio of the unfolded smeared distributions to the nominal distributions is measured. The JER and JES systematic uncertainties are calculated with the b-dijet simulation, as it has higher statistics than the c-dijet simulation and the detector response is expected to be very similar for beauty and charm jets. Figure 4.23 shows the ratio of the smeared to nominal distributions used to determine the JER uncertainty. The JER uncertainty generally decreases with increasing jet p_T. With the exception of the z distributions, the uncertainties due to the JER resolution in the two highest jet

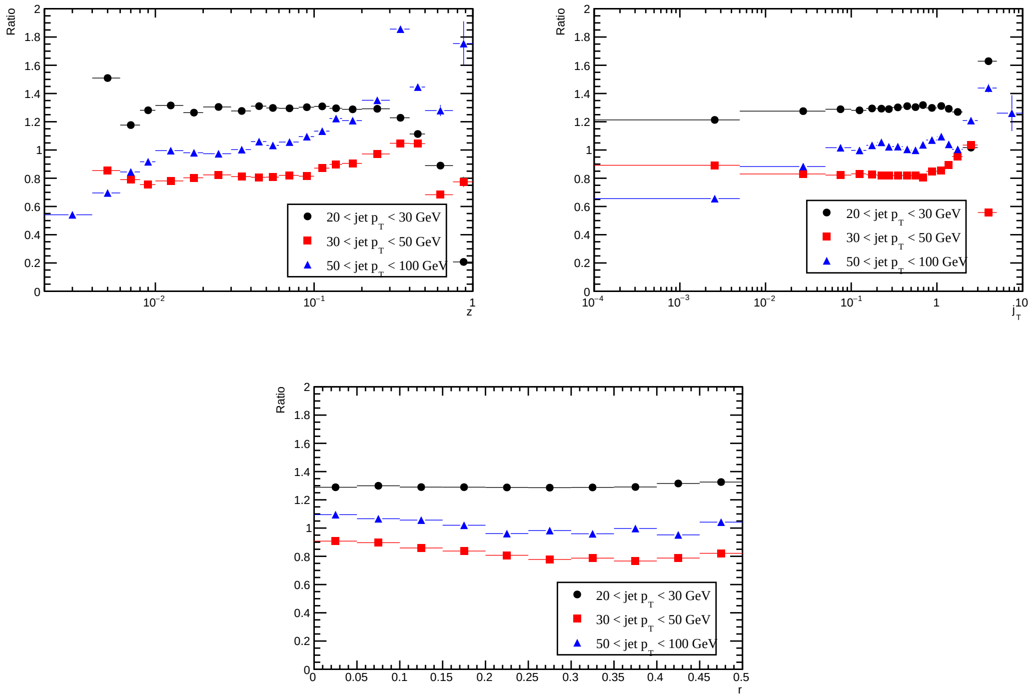

Figure 4.23: The ratios of the z, j_T, and r distributions smeared with the JER uncertainty to the nominal distributions.

p_T bins are generally on the order of 10%. The JES systematic uncertainty is calculated in a similar way to the JER; all components of the jet four-momentum vector in simulation are scaled by the JES factor, the efficiencies and response matrices are regenerated, and the ratio of the distributions with the JES scaling applied to the nominal distributions is determined. Figure 4.24 displays the ratios of the smeared to nominal distributions used to determine the JES uncertainty.

4.4.3 Track purity and efficiency

Contamination from "ghost" tracks, tracks that do not actually correspond to a charged particle trajectory, but were combinatorially reconstructed from uncorrelated hits in the tracking detectors, could affect the measured charged hadron distributions. To assign a systematic uncertainty for the ghost track contamination, a tight track purity cut was applied to the charged hadrons selected for the analysis and the entire

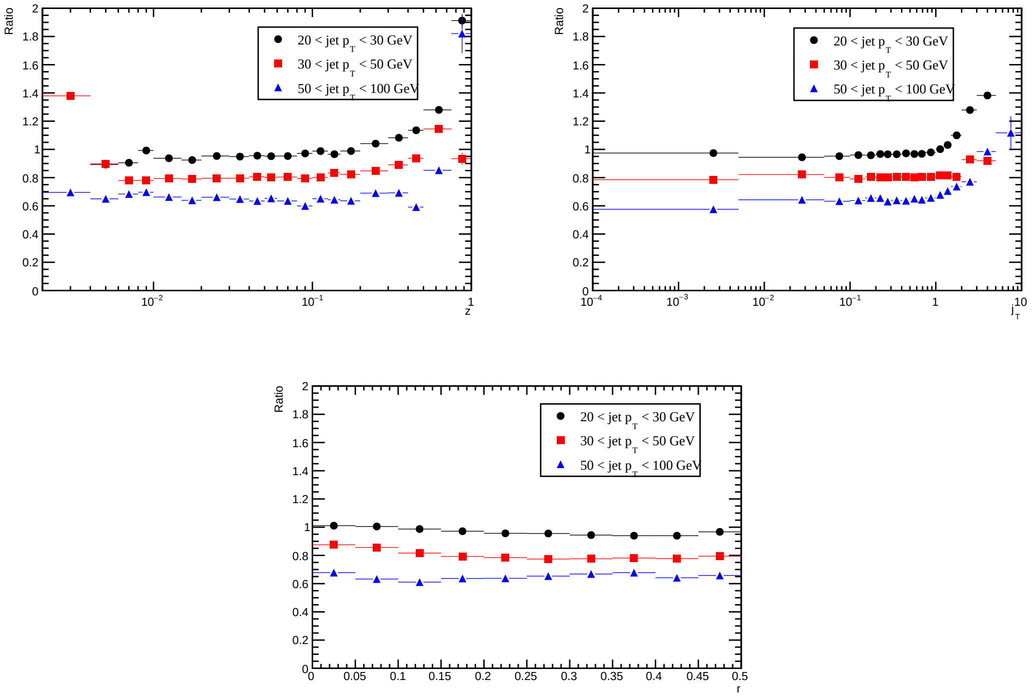

Figure 4.24: The ratios of the z, j_T, and r distributions smeared with the JES uncertainty to the nominal distributions.

analysis was repeated. The b-dijet simulation was again used for the track systematic uncertainty studies, as the track efficiencies for b-jets and c-jets display similar trends and the b-dijet simulation has higher statistics. The ratio of the distributions in real data with and without the tight purity cut applied was calculated, and the deviation from unity was assigned as a systematic error. Figure 4.25 shows the track purity ratios for the z, j_T, and r distributions as a function of jet p_T. The uncertainty is generally less than 10% for all distributions except at high and low z and at low j_T.

A systematic uncertainty was also assigned for the track efficiency determination. Instead of using a bin-by-bin method to compute the track efficiencies, as was described in Section 4.2.2, the nominal track efficiencies shown in Figure 4.5 were smoothed to yield a continuous distribution. The analysis was repeated using the smoothed track efficiencies, and the ratio of the nominal to smoothed distributions was calculated. Figure 4.26 shows the resulting ratios for the z, j_T, and r distribu-

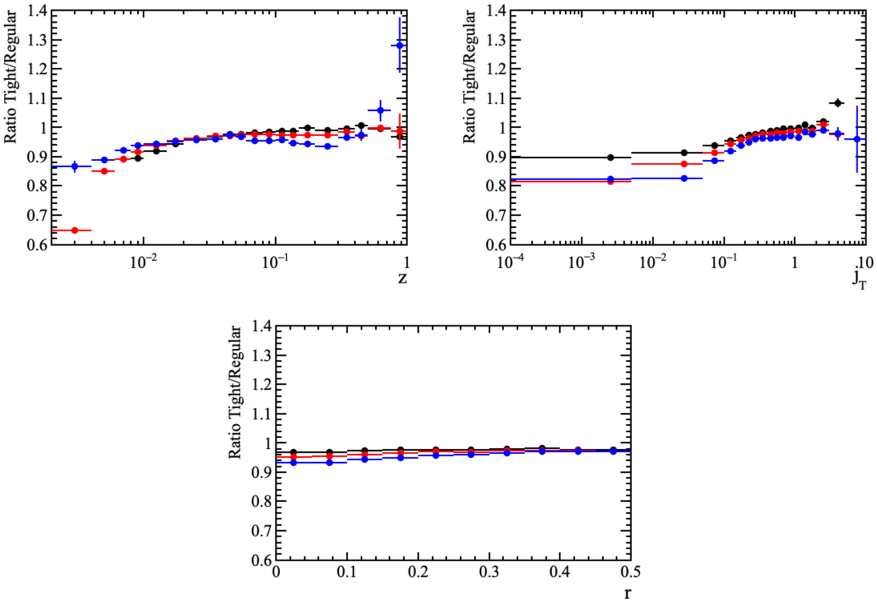

Figure 4.25: The ratios of the z, j_T, and r distributions in real data with the tight track purity cut applied to those with the regular track purity cut applied. The black distributions are the $20 <$ jet $p_T < 30$ GeV bin, while the red and blue are the $30 <$ jet $p_T < 50$ GeV and $50 <$ jet $p_T < 100$ GeV bins, respectively

tions. The uncertainties are generally less than 10%, with the exception of those at low z where the tracking efficiency is less well known, and are a few percent for the j_T and r distributions.

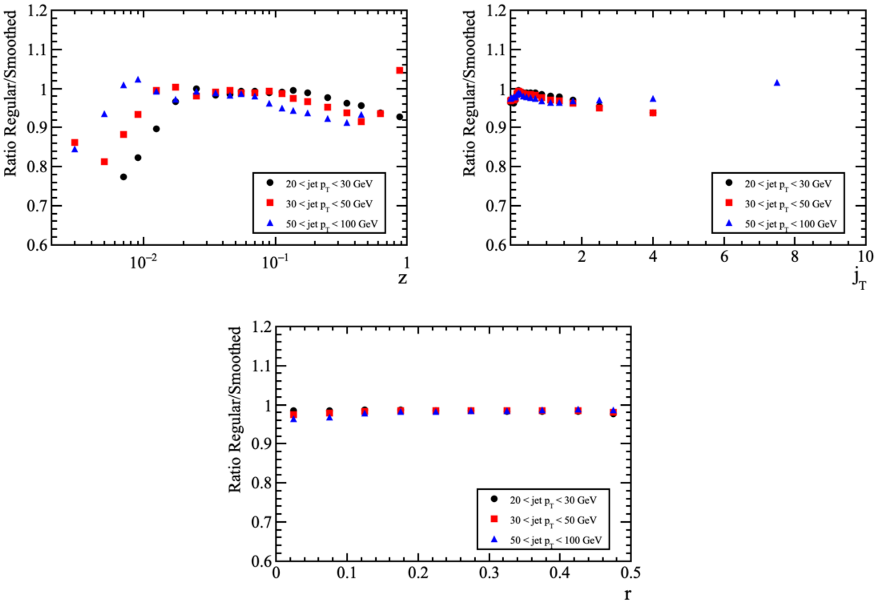

Figure 4.26: The ratios of the z, j_T, and r distributions corrected with the nominal track efficiencies to those corrected with the smoothed track efficiencies.

CHAPTER V

Results

This chapter presents the results of the measurement performed for this thesis. The measured charged hadron distributions of the longitudinal momentum fraction z, transverse momentum relative to the jet axis j_T, and radial distance from the jet axis r are presented for beauty and charm jets. The measured distributions are compared to predicted distributions in simulated pp collisions generated with the PYTHIA event generator [64]. The PYTHIA generator simulates hadronization with the Lund string fragmentation model, which parameterizes the color field between a quark and anti-quark pair produced in a high-energy collision as a tube or "string" [65]. The energy stored in the color field, or "string", increases linearly as a function of distance between the quark and anti-quark, until it reaches an energy threshold at which the string "breaks" and a new $q\bar{q}$ pair is formed, producing two strings [65]. The process continues until the individual strings no longer have enough energy to produce $q\bar{q}$ pairs. Comparing the measured charged hadron distributions to those from PYTHIA therefore constitutes a comparison to a model for a proposed hadronization mechanism that is motivated by the confining nature of QCD.

The measured beauty and charm distributions are presented with and without tracks from the secondary vertex in the jet included, and the potential implications of the differences between the distributions are discussed. After discussion of the beauty and charm distributions separately, they are then directly compared to probe differences in hadronization between beauty and charm quarks. Finally, the beauty and charm distributions are compared to previously measured distributions in Z-tagged jets at LHCb, which are predominantly light-quark-initiated, and the differences between the heavy and light flavor distributions are discussed.

5.1 Beauty

The results for the measured z, j_T, and r distributions are shown for b-jets in Figure 5.1. The data are compared to generator-level predictions from PYTHIA 8. A general trend that can be observed from the PYTHIA comparisons is that PYTHIA underpredicts the data across nearly all distributions in the two lowest jet p_T bins. The discrepancy between data and PYTHIA decreases with increasing jet p_T; the agreement in the highest jet p_T bin is significantly better across all three charged hadron distributions. For the z distributions, the agreement between data and PYTHIA is best near the middle of the measured z range, between approximately 0.01 and 0.03. PYTHIA underpredicts the distributions at higher and lower values of z. PYTHIA also underpredicts the j_T values over most of the measured range, with the discrepancy worsening with increasing j_T. PYTHIA describes the shape of the r distributions well, capturing the narrowing of the distributions toward small r with increasing jet p_T. The shape of the r distribution displays a broad peak in the $20 <$ jet $p_T < 30$ GeV bin, then begins to display a slight peak around $r \sim 0.13$ in the $30 <$ jet $p_T < 50$ GeV bin, and further sharpens into a peak centered around $r \sim 0.8$ in the $50 <$ jet $p_T < 100$ GeV bin. This trend in r is consistent with that expected from the beauty dead cone effect. As the dead cone angle is proportional to the mass of the quark divided by its energy, a higher energy b-quark would have a smaller dead cone than a lower energy b-quark. Since the jet p_T is proportional to the energy of the initial fragmenting b-quark, higher p_T b-jets would be expected to have a smaller

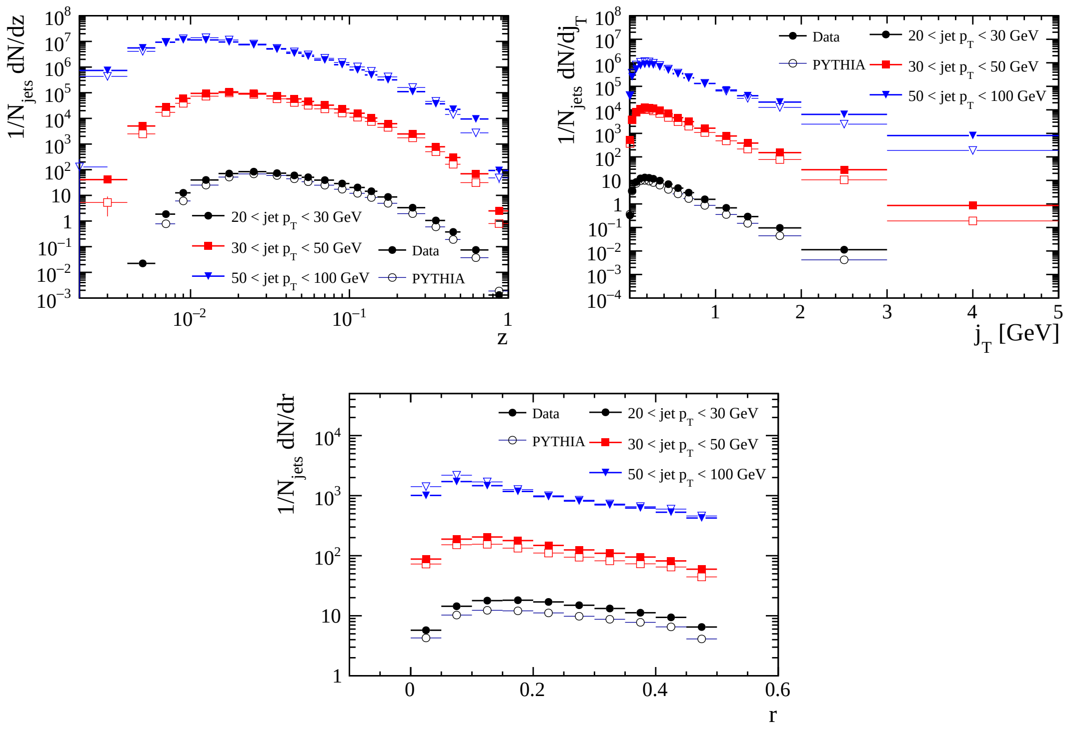

Figure 5.1: The b-jet z, j_T, and r distributions measured in pp collisions in $\sqrt{s} = 13$ TeV data collected by LHCb. The distributions in different jet p_T bins are arbitrarily scaled for readability.

dead cone angle. This is qualitatively consistent with the observed shift of the "peak" in the r distributions to smaller r with increasing jet p_T.

Figures 5.2-5.4 show the comparison of the b-jet z, j_T, and r distributions with and without the tracks from the secondary vertex in the jet included. The z distributions in Figure 5.2 show the largest discrepancy between the PV+SV track distributions and PV-only distributions in the lowest jet p_T bin. At low jet p_T, the SV tracks comprise a significant amount of the tracks at high z. With increasing jet p_T, an increasing number of tracks from the primary vertex contribute to the observed distributions at high z. The increasing fraction of PV-only tracks at high z with increasing jet p_T would indicate a harder fragmentation within the jet, with the additional high z particles being produced directly from the hadronization of the b-quark. The j_T distributions, shown in Figure 5.3 show a similar trend. In the lowest jet p_T bin,

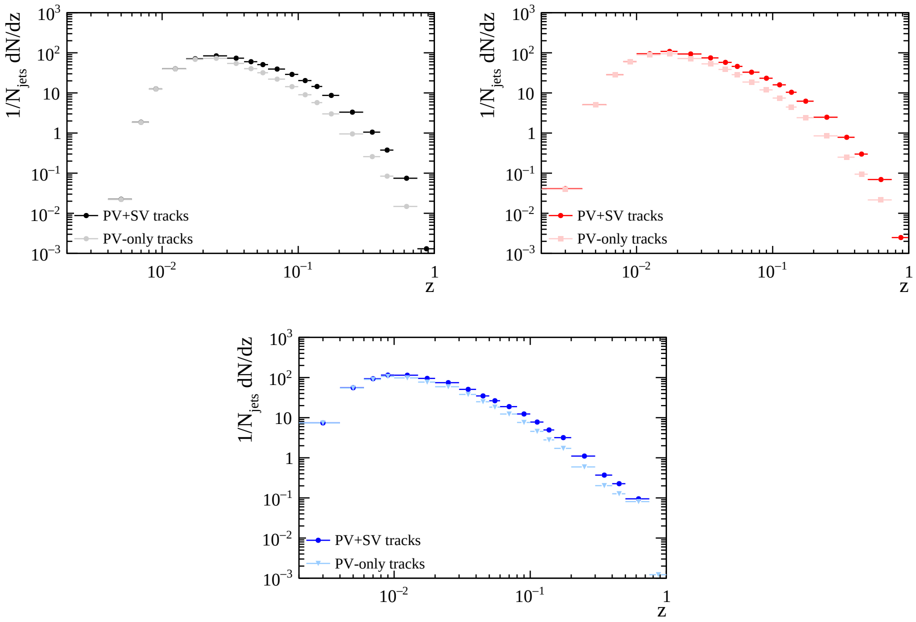

Figure 5.2: Comparison of the b-jet z distributions with and without SV tracks included for $20 <$ jet $p_T < 30$ GeV (top left), $30 <$ jet $p_T < 50$ GeV (top right), and $50 <$ jet $p_T < 100$ GeV (bottom).

the SV tracks comprise a large fraction of the distribution at large j_T. The fraction decreases with increasing jet p_T. Interestingly, the two highest jet p_T bins appear to show that very few SV tracks contribute to the largest j_T bin. This is consistent with the radiation at large j_T being due to hard gluon radiation, which is not expected to arise from the SV. The r distributions shown in Figure 5.4 exhibit significant differences between the PV+SV and PV-only distributions. The slight peak at small r described above is most clearly observed in the PV+SV track distributions, but the same trend is also observed in the PV-only track distributions, albeit significantly more muted. With increasing jet p_T, the SV tracks tend to be concentrated at smaller r values. The inclusion of the SV tracks likely serves to better localize the position of the beauty hadron produced within the jet, thus providing a better proxy for sensitivity to the beauty dead cone.

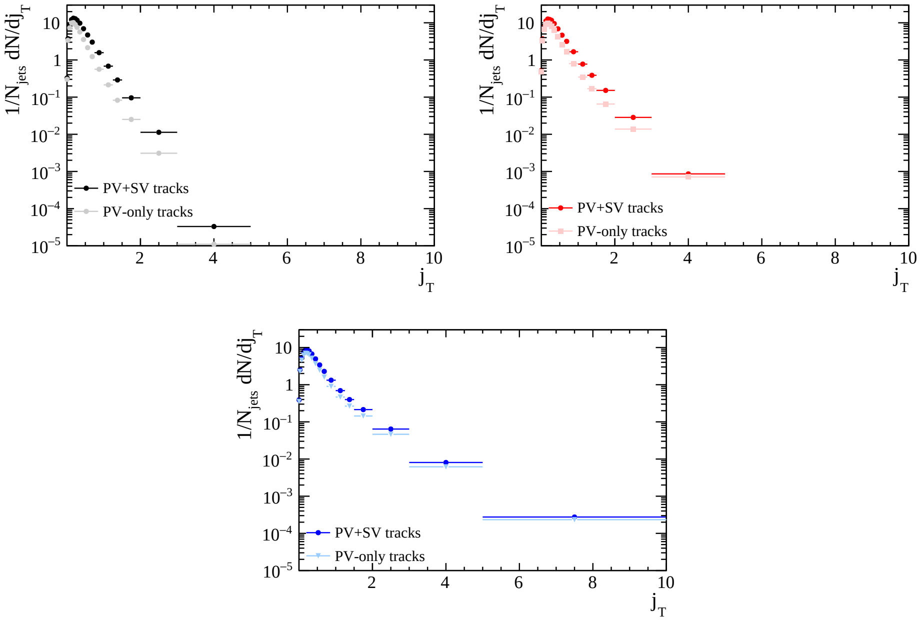

Figure 5.3: Comparison of the b-jet j_T distributions with and without SV tracks included for $20 <$ jet $p_T < 30$ GeV (top left), $30 <$ jet $p_T < 50$ GeV (top right), and $50 <$ jet $p_T < 100$ GeV (bottom).

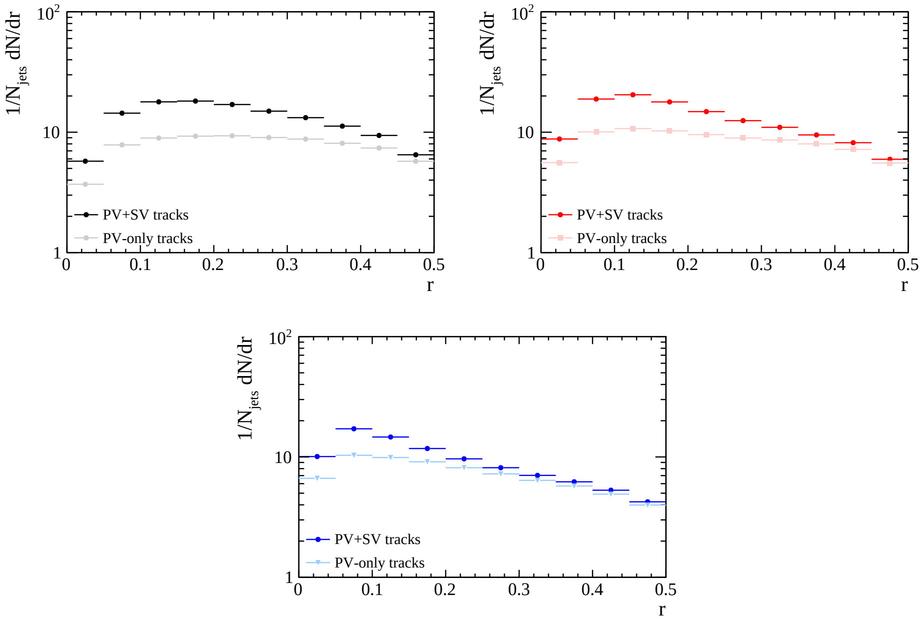

Figure 5.4: Comparison of the b-jet r distributions with and without SV tracks included for $20 <$ jet $p_T < 30$ GeV (top left), $30 <$ jet $p_T < 50$ GeV (top right), and $50 <$ jet $p_T < 100$ GeV (bottom).

5.2 Charm

The results for the measured z, j_T, and r distributions are shown for c-jets in Figure 5.5. As with the b-jets, the data are compared to generator-level predictions from PYTHIA 8. PYTHIA does not describe the charm distributions as well as the beauty distributions; it underpredicts the data over nearly the entire measured range for all three distributions. The PYTHIA comparisons also generally do not improve with increasing jet p_T. The agreement with data improves slightly with increasing jet p_T in the $z \sim 0.1$-0.2 range of the z distributions and in the lowest r bin. This is in stark contrast with the PYTHIA comparisons to the b-jet distributions shown in Figure 5.1, which with the exception of the j_T distributions generally improved with increasing jet p_T across the entire measured range. Also notable is that PYTHIA does not describe the shape of the charm r distributions as well as it did for the beauty. The difference is most clearly seen in the highest jet p_T bin, which PYTHIA described well in terms of the shape and overall normalization for b-jets, but which fails to capture the shape and multiplicity for r larger than 0.1 in c-jets.

Figures 5.6-5.8 show the comparisons of the c-jet distributions with and without the SV tracks included. The same general features noted in the b-jet distributions are also observed in the c-jet distributions, which is not surprising given that many of the features discussed above are common to heavy flavor production in general, and are therefore not expected to differ significantly between beauty and charm jets. However, while the general features are the same as for b-jets, some flavor dependence is still observed. The SV tracks in b-jets contribute to the observed PV+SV distributions over a wider range of z values, beginning from $z \sim 0.03$. In c-jets, the SV tracks begin contributing around $z \sim 0.05$ in the lowest jet p_T bin, with the minimum SV track z shifting to higher values with increasing jet p_T. This would indicate that in higher p_T c-jets, the SV tracks tend to have higher z values. The observed difference between b- and c- jets is consistent with that observed in the comparison between

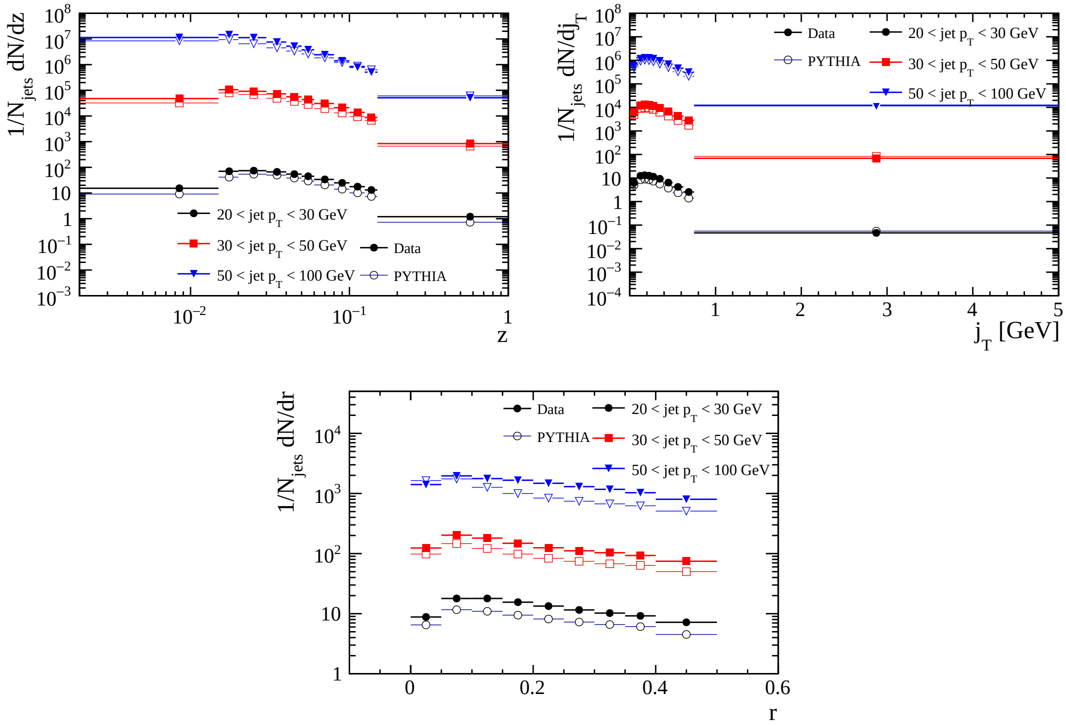

Figure 5.5: The c-jet z, j_T, and r distributions measured in pp collisions in $\sqrt{s} = 13$ TeV data collected by LHCb. The distributions in different jet p_T bins are arbitrarily scaled for readability.

the b-jet and c-jet heavy flavor tagging z bias studies described in Section 4.4.1. As described in that section, the difference in average multiplicity between beauty and charm decays could explain this difference. Differences between b- and c-jets can also be seen when comparing the r distributions with and without the SV tracks, shown for c-jets in Figure 5.8. The SV tracks in c-jets tend to be more concentrated at low r even at low jet p_T compared to those in b-jets. The charm r distributions also display a slight narrowing at low r with increasing jet p_T, although the expected shift due to the dead cone and seen in the beauty r distributions is not seen here, likely due to the significantly smaller size of the charm dead cone compared to the beauty dead cone. While hints of the narrowing of the distributions at small r could be seen in the PV-only b-jet r distributions, the same trend in the PV-only c-jet distributions is barely perceptible in the lowest two jet p_T bins and appears absent in the highest jet p_T bin. This is also likely due to the decreased sensitivity to the dead cone effect in c-jets. The beauty dead cone appears large enough to be observed, even if less clearly, in the PV-only distributions.

To further examine the differences in the hadronization distributions between beauty and charm jets, the PV+SV and PV-only z, j_T, and r distributions are compared directly in Figures 5.9 and 5.10, respectively. The same binning is used for both the beauty and charm distributions, since only statistical uncertainties, not the charm statistics-limited systematic uncertainties, are shown here. The use of the same binning facilitates comparison between the distributions and also exhibits the statistics available in the data for the charm distributions. The PV+SV distributions for the lowest two jet p_T bins display similar features for both beauty and charm jets. The charm jets tend to have slightly more particles produced at high z and fewer particles produced at high j_T relative to beauty jets. More particles are also produced at small r in charm jets compared to beauty jets, a difference which is likely attributable to the dead cone effect. The largest differences between the beauty and charm PV+SV distributions are seen in the $50 <$ jet $p_T < 100$ GeV bin. The charm jets appear to have a higher particle multiplicity than beauty jets in this bin, as the charm distributions are higher than the beauty distributions across nearly all z, j_T, and r. In the PV-only distributions, shown in Figure 5.10, the beauty and charm j_T distributions in the lowest two jet p_T bins look more similar than in the PV+SV case. This indicates that the j_T distribution for particles produced directly from the hadronization of the heavy quark does not differ strongly between beauty and charm hadronization. The $50 <$ jet $p_T < 100$ GeV bin is an exception to this and displays a similar behavior to that seen in the PV+SV distributions, indicating that the differences observed between the beauty and charm distributions in this jet p_T bin are likely due to flavor-dependent differences in the hadronization of the heavy quark.

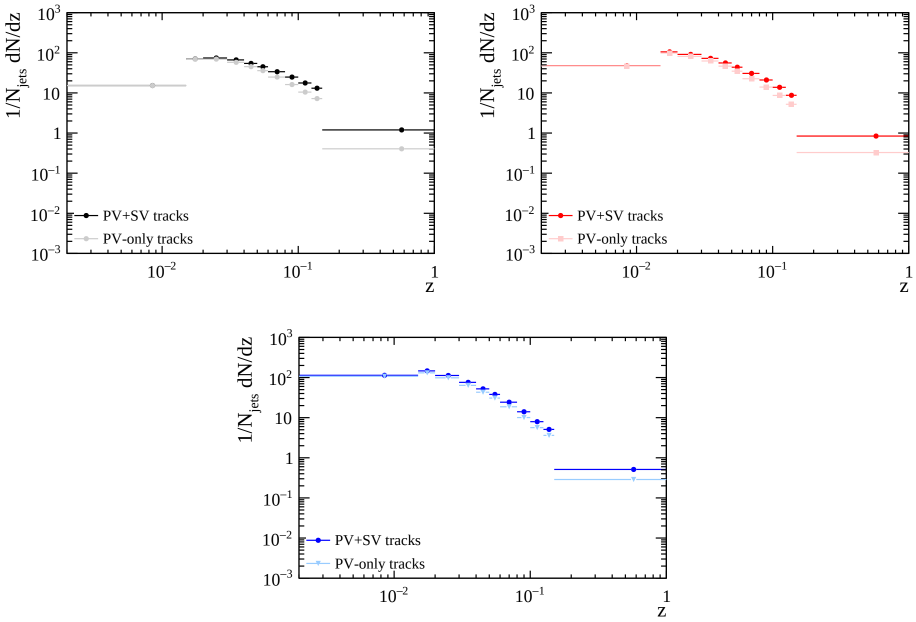

Figure 5.6: Comparison of the c-jet z distributions with and without SV tracks included for $20 <$ jet $p_T < 30$ GeV (top left), $30 <$ jet $p_T < 50$ GeV (top right), and $50 <$ jet $p_T < 100$ GeV (bottom).

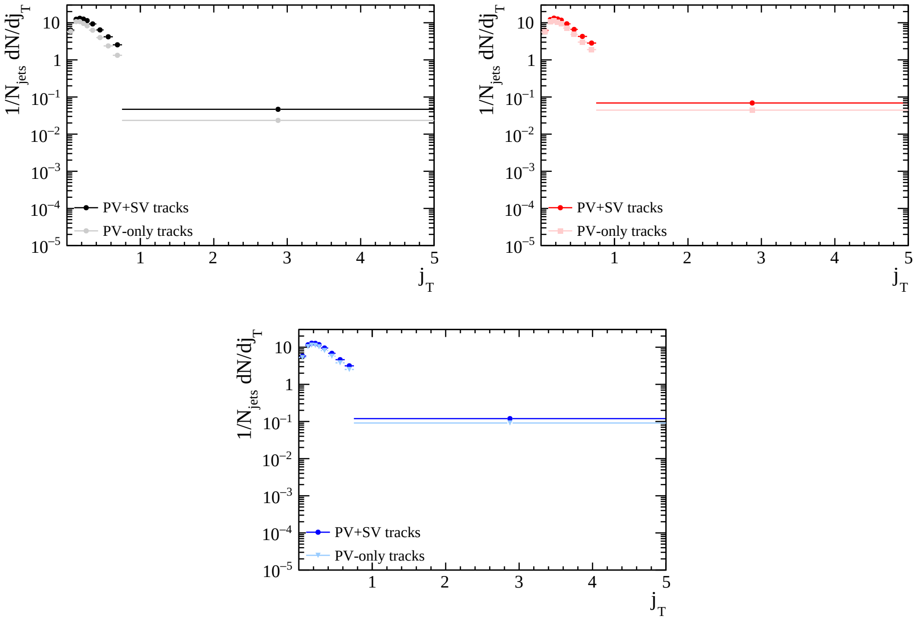

Figure 5.7: Comparison of the c-jet j_T distributions with and without SV tracks included for $20 <$ jet $p_T < 30$ GeV (top left), $30 <$ jet $p_T < 50$ GeV (top right), and $50 <$ jet $p_T < 100$ GeV (bottom).

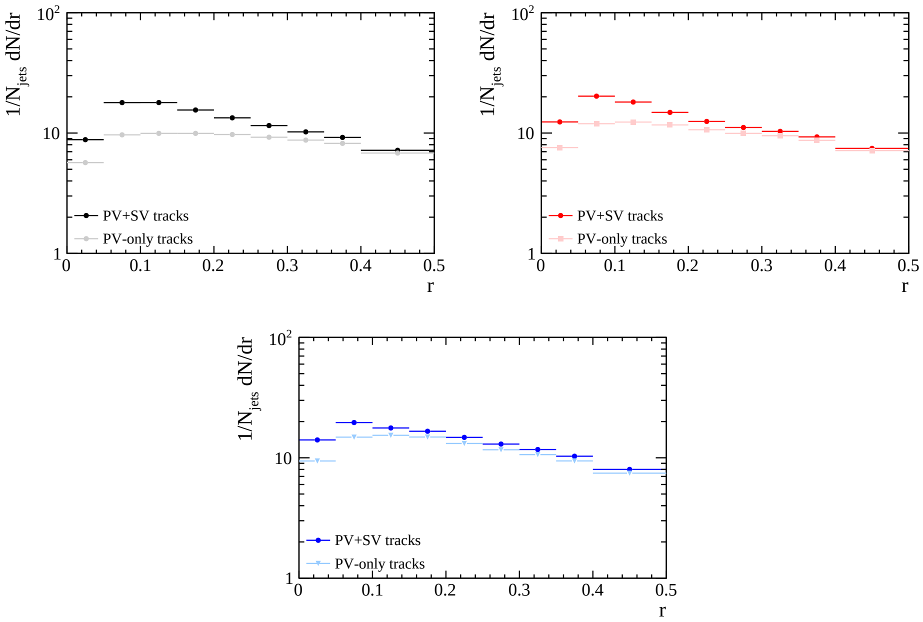

Figure 5.8: Comparison of the c-jet r distributions with and without SV tracks included for $20 <$ jet $p_T < 30$ GeV (top left), $30 <$ jet $p_T < 50$ GeV (top right), and $50 <$ jet $p_T < 100$ GeV (bottom).

5.3 Comparisons to Z-tagged jets

The measured distributions in beauty and charm jets are compared to previously measured distributions in Z-tagged jets at LHCb [37]. As previously discussed, Z-tagged jets in the forward region accessible to LHCb are primarily light-quark-initiated jets. Therefore comparing the Z-tagged distributions to those in heavy flavor jets can reveal differences between heavy and light quark hadronization. One caveat for the comparisons in this section is that the Z-tagged jet distributions were measured in pp collisions at $\sqrt{s} = 8$ TeV, while the heavy flavor jet distributions presented in this thesis were measured in pp collisions at $\sqrt{s} = 13$ TeV. However, the distributions are measured in the same jet p_T bins, which should select a similar hard scale for probing the hadronization. Charged hadron distributions in Z-tagged jets are currently being measured in pp collisions at $\sqrt{s} = 13$ TeV at LHCb, which will provide comparisons to the distributions in heavy flavor jets at the same center

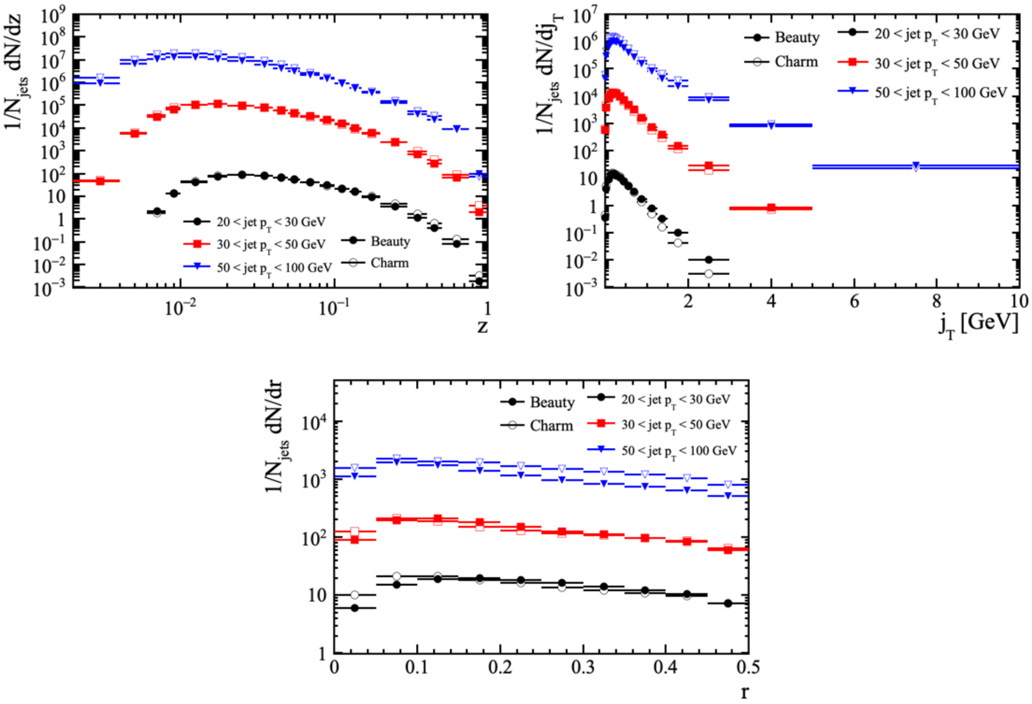

Figure 5.9: Comparison of the beauty and charm PV+SV distributions for z (top left), j_T (top right) and r (bottom).

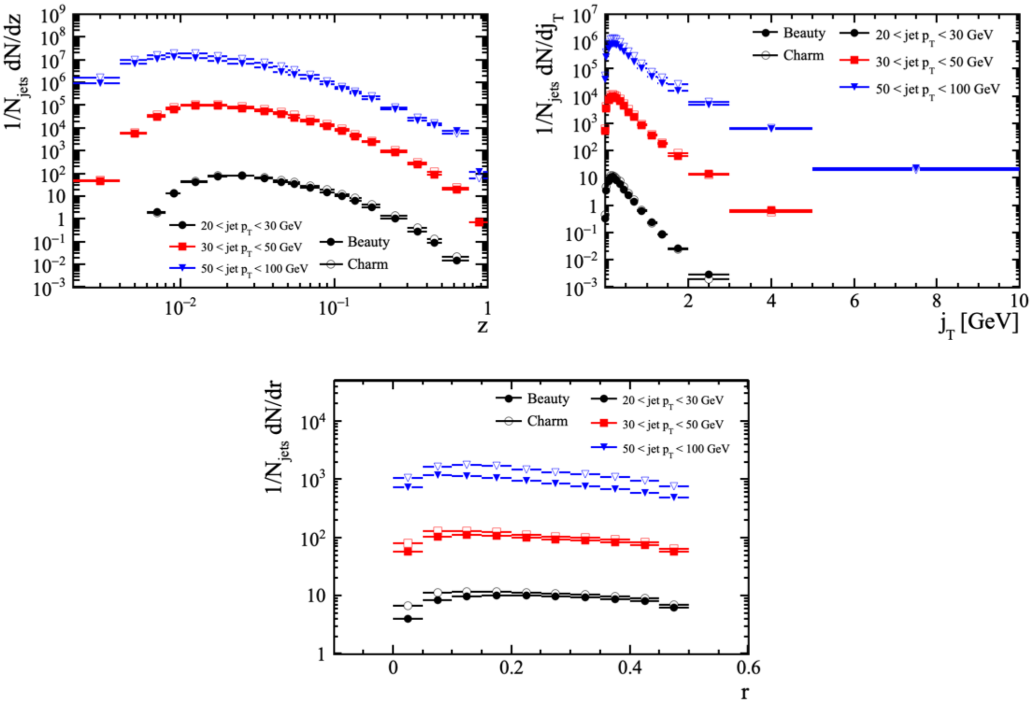

Figure 5.10: Comparison of the beauty and charm PV-only distributions for z (top left), j_T (top right) and r (bottom).

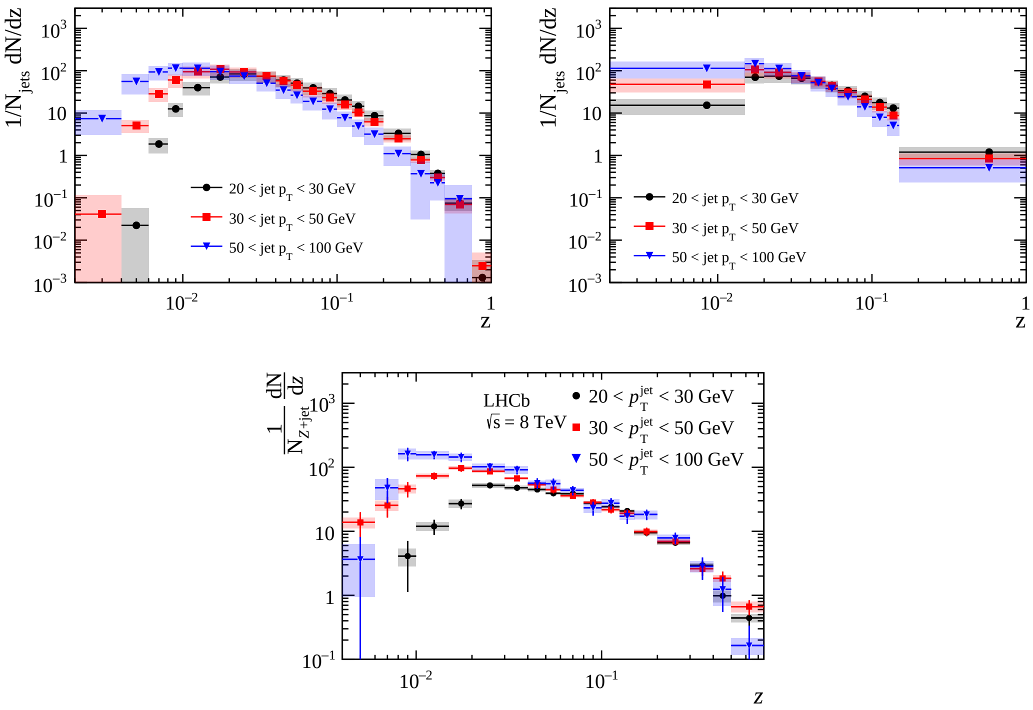

Figure 5.11: The z distributions in b-jets (top left), c-jets (top right), and Z-tagged jets (bottom).

of mass energy. Figures 5.11 - 5.13 show the z, j_T, and r distributions, respectively, among the three different jet flavors. The z distributions exhibit the same scaling as a function of jet p_T at low z values, which is largely due to the effect of the same minimum track p cut applied in all jet p_T bins [37]. At high z, the Z-tagged jet distributions are approximately constant as a function of jet p_T, while the heavy flavor jet distributions are not within the uncertainties shown. The heavy flavor jet data samples have much higher statistics than the Z-tagged jet data sample used for the comparisons here, so it is possible that the b- and c-jet distributions can probe the evolution of the fragmentation functions as a function of the hard scale. Theoretical comparisons will be needed to confirm whether the variation seen here is due to the FF evolution.

For the j_T distributions, shown in Figure 5.12, the most significant difference between the three jet flavors over the range where the measurements overlap is ob-

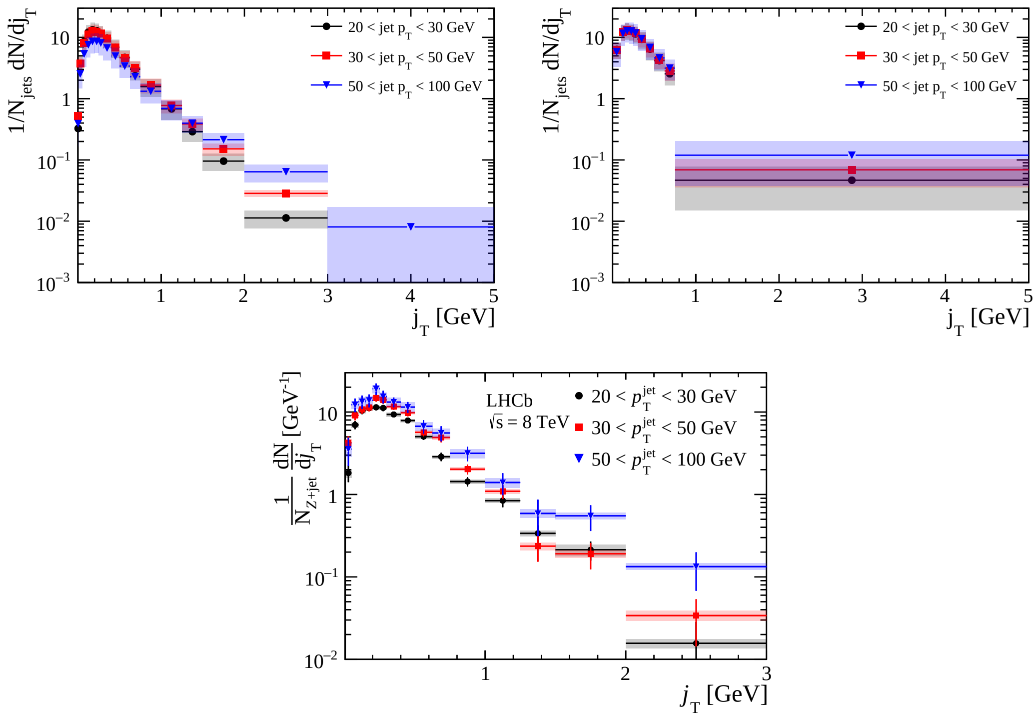

Figure 5.12: The j_T distributions in b-jets (top left), c-jets (top right), and Z-tagged jets (bottom).

served at small j_T. The ordering of the distributions with jet p_T seen in the charm and Z-tagged jet distributions appears to differ relative to that seen in the beauty jet distributions. The jet p_T ordering in the beauty distributions appears inverted relative to the charm and Z-tagged jet distributions. The beauty jets seem to have fewer particles in the highest jet p_T bin at low j_T, while the opposite is observed in the charm and Z-tagged jets.

The r distributions shown in Figure 5.13 display several differences between the heavy flavor and Z-tagged jets. The r distributions in Z-tagged jets have a more sharply falling distribution than that observed in the heavy flavor jets. The Z-tagged r distributions also exhibit a strikingly different behavior as a function of jet p_T at small r compared to the heavy flavor jets. In the lowest jet p_T bin, the Z-tagged jets have a depletion of particles at small r, similar to the trend observed in the beauty and charm jets. However with increasing jet p_T, the depletion at small r disappears. This

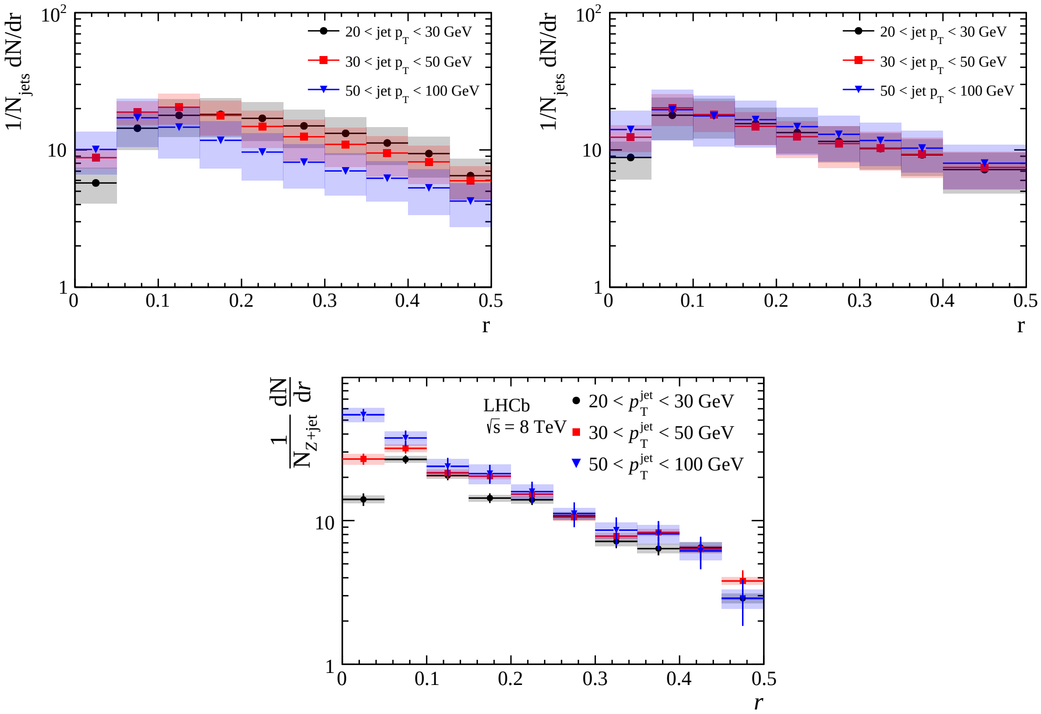

Figure 5.13: The r distributions in b-jets (top left), c-jets (top right), and Z-tagged jets (bottom).

r dependence is also qualitatively consistent with the dead cone effect. As previously discussed, for a given quark mass the dead cone angle decreases with increasing jet p_T. This could explain the increase in particles in the lowest r bin as a function of jet p_T, as the dead cone angle becomes smaller than the resolution of the lowest r bin.

CHAPTER VI

Heavy Flavor Hadronization Studies at a Future Electron-Ion Collider

An electron-ion collider (EIC) is planned to be built at Brookhaven National Laboratory (BNL) in the United States in the next decade. The incident electron, a point-like particle that interacts electroweakly but not via the strong force, will resolve scattering events with significantly higher precision than that available in $p+p$ collisions. With beams of spin-polarized electrons, protons, and light nuclei, and additional unpolarized nuclear beams up to uranium available in mulitiple energy configurations, and with hadron identification capabilities planned for an EIC detector, the EIC will offer opportunities for a rich hadronization program currently not possible at any existing facility [66]. Collisions with electron and nuclear beams provide access to studying how color charges propagate in nuclear matter, which is intimately connected to the distance scale of hadronization. By measuring inclusive cross sections of light and heavy hadron production in $e+A$ collisions, where A is a nucleus up to uranium, and comparing them to the corresponding cross sections measured in $e+p$ collisions, one can determine how color charge neutralization varies with nucleus size [66]. The EIC will also provide access to different regions of fragmentation known as the "current" and "target" fragmentation regions. The current fragmentation region refers to the hadronization of the struck parton, while the target region refers to the hadronization of the colored remnant from which the struck parton was removed [67,68]. Previous experiments with an initial lepton beam on nuclei or on polarized protons were fixed target experiments, which had difficulty separating the current and target regions. In a collider geometry, however, the struck parton and its remnant typically move in different directions, facilitating the detection and study of each region independently.

Studies were performed to study the feasibility of re-using the solenoid of the sPHENIX detector [69], located at BNL, in an EIC detector [20]. Figure 6.1 shows the detector schematic for an EIC detector based on sPHENIX, denoted “EIC-sPHENIX”. The detector design reflects the asymmetry of *ep* collisions; in Figure 6.1 the hadron beam travels from left to right, while the electron beam travels from right to left. Most of the particle identification and calorimetry is located on the hadron-going side of the detector, as the proton remnants will be boosted in that direction. Two performance studies for EIC-sPHENIX will be discussed in detail in this chapter: the first study focuses on the performance of a proposed combination of particle identification detectors for meeting the needs of the EIC physics program, and the second study investigates the feasibility of identifying charm hadrons without a dedicated vertex detector in EIC-sPHENIX.

6.1 Particle Identification

Particle identification (PID) is important for multiple facets of the EIC physics program. Hadron PID is relevant for studying the flavor dependence of the spin- and transverse-momentum-dependent PDFs in the nucleon, and for studying the flavor dependence of hadronization. Electron-pion separation at backward and central rapidities is furthermore important in correctly identifying the scattered electron for event kinematics reconstruction [20]. Hadron PID is also crucial in order to reconstruct heavy flavor hadrons, as it significantly reduces the combinatorial background

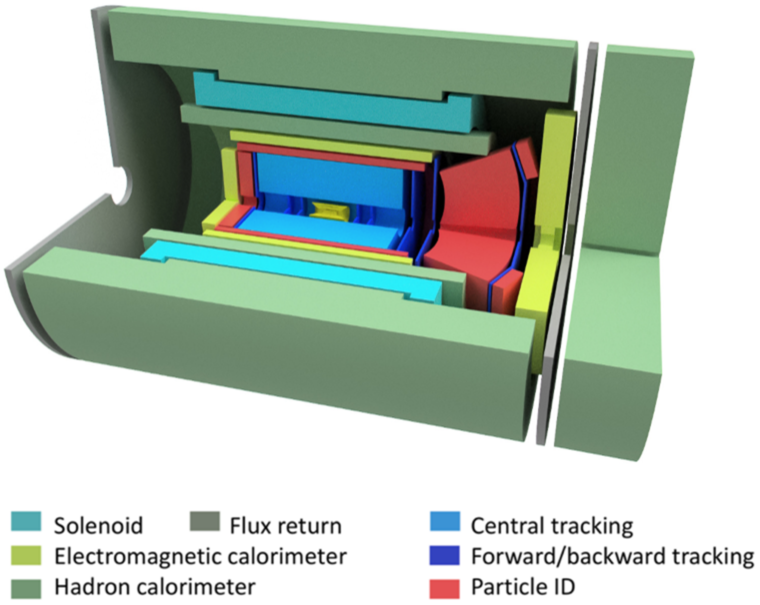

Figure 6.1: The schematic for an EIC detector re-using the sPHENIX solenoid [20]

in the invariant mass reconstruction.

A combination of four PID detectors was studied for EIC-sPHENIX: A Detection of Internally Reflected Cherenkov (DIRC) light detector, a gas Ring Imaging Cherenkov (RICH) detector, and two modular RICH (mRICH) detectors [20]. The purpose of the DIRC is to identify low-momentum charged hadrons produced at midrapidity. The gas RICH and hadron-side mRICH serve to identify charged hadrons produced from the hadronization of the proton remnant at forward rapidities. The electron-side mRICH helps discriminate between the scattered electron, which is needed to reconstruct Deep Inelastic Scattering (DIS) kinematics, from charged pions at backward rapidities.

Table 6.1 summarizes the pseudorapidity and momentum ranges in which a 3-sigma separation of pions and kaons and of pions and electrons is achievable for the EIC-sPHENIX reference design described above. A fast detector smearing package that parameterized the detector response was used to simulate the performance of the EIC-sPHENIX PID detectors. The simulation parameterized the performance of the

Table 6.1: Momentum and pseudorapidity coverage for the EIC-sPHENIX reference design [20]

Detector	pseudorapidity	K/π 3σ (GeV/c)	e/π 3σ (GeV/c)
DIRC	(-1.4, 1.24)	$\lesssim$6	
gas RICH	(1.24, 3.95)	(15,50)	(5, 15)
h-side mRICH	(1.10, 1.85)	(3,9)	$\lesssim$2
e-side mRICH	(-3.9, -1.4)	(3,9)	$\lesssim$2

DIRC, gas RICH, h-side mRICH, and e-side mRICH detectors for both K/π and e/π discrimination. Detailed parameterizations of PID efficiencies as a function of particle momentum were implemented for both mRICH detectors, while the parameterizations for the DIRC and the gas RICH were determined according to the momentum ranges listed in Table 6.1 for which the K/π and e/π separation is at least 3σ.

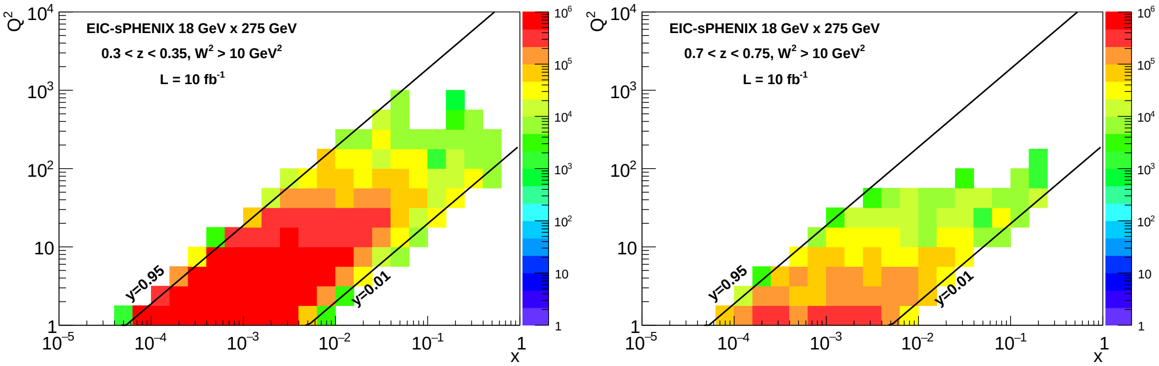

Figure 6.2: x and Q^2 distributions of events with K^+ at low z (left) and high z (right) identified with EIC-sPHENIX PID detectors in the eic-smear fast simulation. Both plots are scaled to an integrated luminosity of 10 fb^{-1}. [20]

Ten million minimum bias events were generated with PYTHIA 6 for the PID studies [70]. The events were generated at the highest ep beam energy configuration studied, with an 18 GeV electron beam and a 275 GeV proton beam. The generated event kinematics were restricted to be within $1 < Q^2 < 20000$ GeV^2, $0.01 < y < 0.95$, and $10^{-5} < x < 0.99$. In DIS events, x is the usual momentum fraction of the proton that the scattered parton carries and Q^2 is the square of the momentum transferred in the collision, defined as the negative of the squared difference between the initial and final momentum of the scattered electron. The variable y is known as the inelasticity, and is defined as the fractional energy of the photon exchanged between the electron and proton with respect to the incoming electron [71]. The fast detector smearing simulation identified pions and kaons in the PYTHIA 6 events only if their momentum and pseudorapidity fell within the ranges of one of the parameterized PID detectors. Figure 6.2 shows the expected statistics for positively charged kaons identified with EIC-sPHENIX particle ID detectors normalized to an integrated luminosity of 10 fb^{-1}, the expected amount of data to be collected by EIC-sPHENIX in one year of operation. The left and right plots show the kaon statistics in expected binning at low and high z, respectively. In order to better estimate the kaon ID efficiency with the chosen configuration of particle ID detectors, the ratio of the kaons identified with the smearing parameterization to the total number of kaons in the PYTHIA 6 events was also studied. Figure 6.3 shows the ratio of the identified kaons to truth kaons binned in (x, Q^2) and at low and high z.

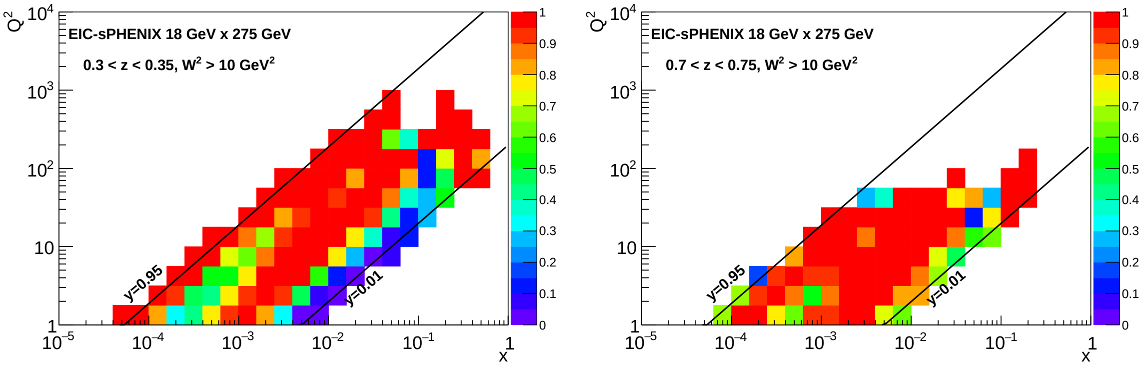

Figure 6.3: x and Q^2 distributions of the ratio of events with identified K^+ (as determined by EIC-SPHENIX PID detectors in the eic-smear fast simulation) to the total number of PYTHIA 6 events with K^+ at low z (left) and high z (right). [20]

The non-uniform features in Figure 6.3 are due to the varying momentum cuts in Table 6.1 for identification of different particles. However, it is important to note here that the parameterizations used for the DIRC and gas RICH are only those over which the K/π and e/π separation is better than 3σ. In contrast to the kaon ID ratio plots, Figure 6.4 shows the ratio of positively charged pions identified in the fast simulation to the total number of pions in the PYTHIA 6 events. The pion ratio plots are more uniform, as is expected since the PID detector parameterizations used in this study have fewer gaps in pion acceptance than in kaon acceptance in momentum space.

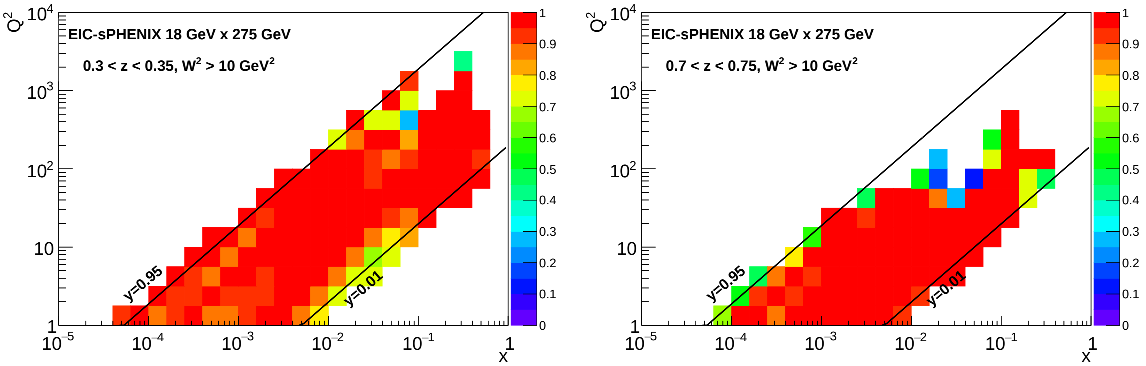

Figure 6.4: x and Q^2 distributions of the ratio of events with identified π^+ (as determined by EIC-SPHENIX PID detectors in the eic-smear fast simulation) to the total number of PYTHIA 6 events with π^+ at low z (left) and high z (right)

The effect of removing particle ID detectors at forward, central, and negative pseudorapidity on the kaon identification efficiency was also studied. Figure 6.5 shows the kaon ID efficiency binned in x and Q^2 without forward PID (top plots), without central PID (middle plots) and without PID at negative pseudorapidity (bottom plots), at low z (left plots) and high z (right plots). Removing forward PID consisted of removing the h-side mRICH and the gas RICH from the smearing simulation, while removing central PID consisted of removing the DIRC, and removing negative pseudorapidity PID consisted of removing the e-side mRICH. As shown in the top row of plots in Figure 6.5, removing the gas RICH and h-side mRICH results in a loss of events in the high x region. Removing the DIRC at central pseudorapidity, the effect of which is shown in the middle plots, results in the loss of the majority of events at moderate x and Q^2. Finally, removing the e-side mRICH, shown in the bottom plots, results in a loss of events at low x. Figure 6.5 therefore shows that in order to have a broad x and Q^2 coverage, particle ID is needed in the electron-going region (corresponding to negative pseudorapidity), the central region, and the forward pseudorapidity region of EIC-sPHENIX.

6.2 Charm Tagging

Charm tagging is important for several goals of the EIC physics program. Access to the gluon PDF in both the proton and in nuclei, as well as the gluon single-spin asymmetry arising from the Sivers effect [72], the spin-momentum correlation between the transverse momentum of a parton and the proton spin, relies on tagging the photon-gluon fusion process in $e + p$ and $e + A$ collisions [71, 73]. In photon-gluon fusion, a radiated photon from the incident electron interacts with a gluon from the proton or nucleus to create a quark-antiquark pair. Identifying the pair production of a charm and anti-charm quark serves as a more sensitive method of tagging the photon-gluon fusion process, as pair production of up, down and strange quarks are abundant in DIS events. The fragmentation of the pair-produced charm and anti-charm quarks to open charm observables provides one method of tagging the photon-gluon fusion process.

Reconstructing exclusive decays of charmed hadrons is a frequently used method of tagging open charm observables. The D^0 meson, which decays to a pion and kaon pair with a branching ratio of 3.89 ± 0.04 % [1], has already been studied as an open charm observable to tag the photon-gluon fusion process [71, 73]. Simulations were performed to determine the expected D^0 signal that EIC-sPHENIX could detect from reconstructing exclusive D^0 decays to pion and kaon pairs.

Ten million PYTHIA 6 events were generated at the highest beam energy configuration, 18 x 275 GeV, for a variety of DIS processes, including photon-gluon fusion.

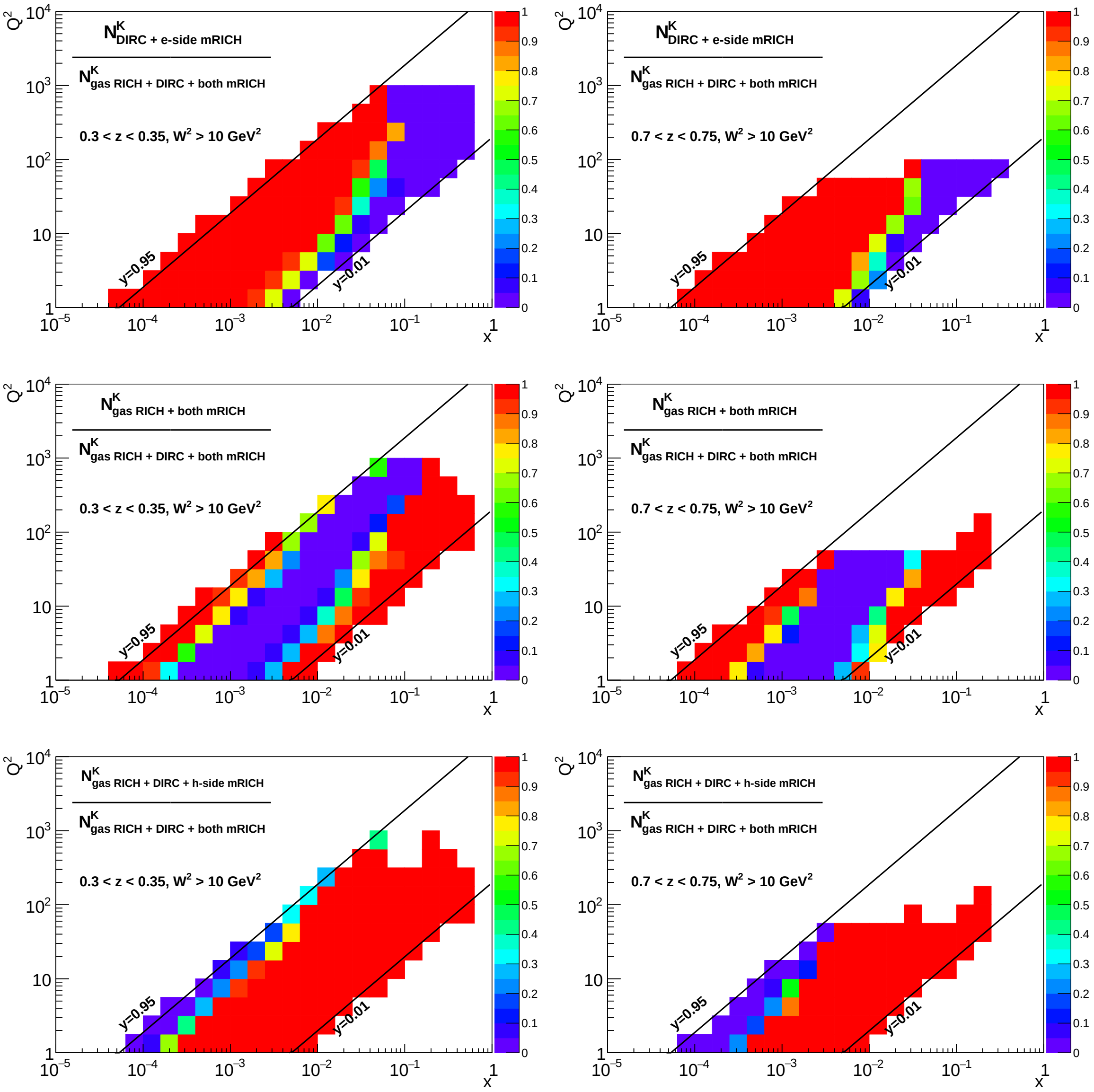

Figure 6.5: K^+ identification efficiency as a function of x and Q^2 when comparing to one possible EIC-sPHENIX PID detector configuration (e-side mRICH, DIRC, gas RICH, and h-side mRICH) when the detector(s) at forward (top plots), central (middle plots), or negative (bottom plots) pseudorapidity are removed. The left plots show the fraction of events binned in (x, Q^2) at low z, and the right plots show the fraction of events at high z. [20]

Events were generated with the same event kinematics as for the PID studies described in the previous section: $1 < Q^2 < 1000$ GeV^2, $0.01 < y < 0.95$, and $10^{-5} < x < 0.99$. To simulate realistic detector effects on the determination of particle momenta and PID, the PYTHIA 6 events were run through a fast detector smearing simulation of the EIC-sPHENIX detector. The detector parameterization included current estimates of uncertainties in the energy resolution and PID efficiencies. Uncertainties in the tracking resolution were estimated from previously determined EIC-sPHENIX tracking parameterizations [20]. PID parameterization was implemented as described in Section 6.1.

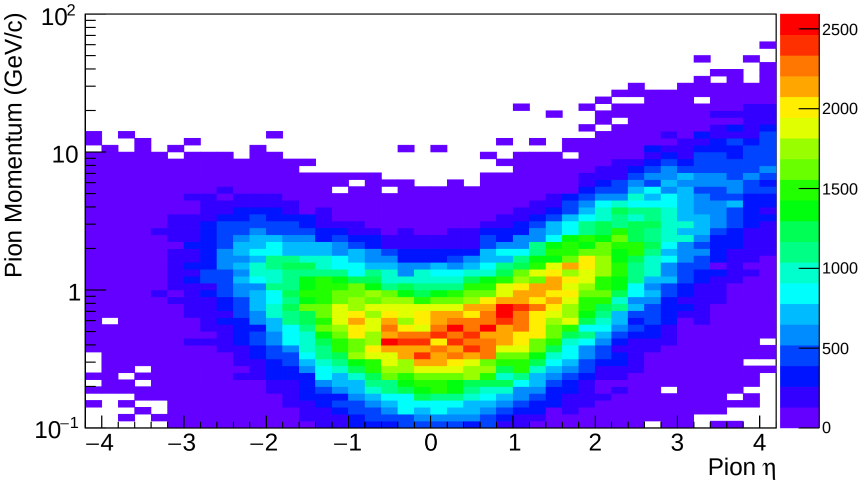

Figure 6.6: Momentum vs pseudorapidity distribution for pions decayed from D^0 mesons produced in the 18 x 275 GeV beam energy configuration from 10 million PYTHIA 6 events. [20]

Figure 6.6 shows the 2D momentum and pseudorapidity distribution of pions decayed from D^0 mesons obtained from 10 million PYTHIA 6 events at a beam energy configuration of 18 x 275 GeV. Kaons decayed from D^0 mesons have a similar distribution. Due to most of the pions and kaons from D^0 decays being at central and forward rapidities, the DIRC and h-side mRICH PID detectors provide critical hadron identification necessary to detect D^0 decays. Only pions and kaons identified by EIC-sPHENIX PID detectors, with pseudorapidity -2.5 $< \eta <$ 2.5, and with transverse momentum p_T greater than 0.1 GeV/c were used to calculate the invariant mass spectrum of $\pi - K$ pairs from the 10 million smeared PYTHIA 6 events. The pseudorapidity cut was implemented to correspond to the range over which EIC-sPHENIX tracking parameterizations were available. As seen in Figure 6.6, the range -2.5 $< \eta <$

2.5 covers the majority of pions produced from D^0 decays. In this simulation, the energies of identified pions and kaons were recalculated using the smeared momenta and known particle masses. Figure 6.7 shows the D^0 mass peak reconstructed from smeared pions and kaons. A peak near the D^0 mass at approximately 1.865 GeV is clearly distinguished above the fit to the combinatorial background. The fit estimates an expected yield of approximately 560,000 D^0 mesons detected by EIC-sPHENIX in one year of EIC operation.

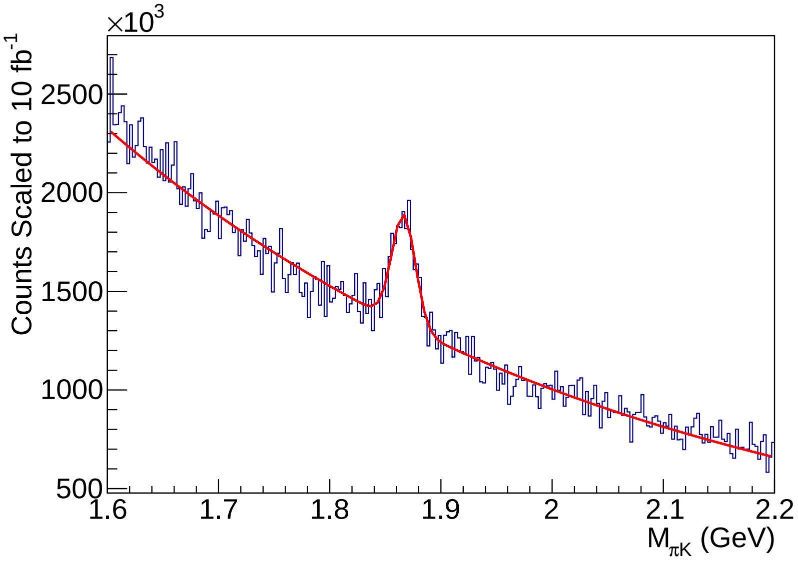

Figure 6.7: A fit to the D^0 mass peak from reconstruction of the exclusive decay $D^0 \rightarrow K^-\pi^+$ using smeared PYTHIA 6 events [20]

CHAPTER VII

Summary and Future Prospects

The charged hadron distributions for the longitudinal momentum fraction z, transverse momentum relative to the jet axis j_T, and radial distance from the jet axis, r, have been measured for the first time in beauty and charm jets using the 2016 pp collision dataset collected by LHCb at $\sqrt{s} = 13$ TeV. The distributions have been measured as a function of jet p_T, in the bins $20 <$ jet $p_T < 30$ GeV, $30 <$ jet $p_T < 50$ GeV, and $50 <$ jet $p_T < 100$ GeV, and also with and without the tracks from the tagged secondary vertex in the jet. The distributions reveal differences between the hadronization of beauty and charm quarks in jets, and also reveal differences between heavy- and light-quark hadronization when compared to the previously measured charged hadron distributions in Z-tagged jets at LHCb [37]. The largest difference between the beauty and charm jets, and between the heavy and light jets, is seen in the r distributions. The observed variations in the r distributions as a function of jet p_T and quark flavor may be due to the dead cone effect, however theoretical comparisons are needed in order to determine whether this is the case. The differences observed in the distributions that include the SV tracks are potentially promising for future use in improving the jet flavor tagging discrimination between beauty and charm jets. The distributions without SV tracks probe particles produced from the heavy quark hadronization in the jet more directly, providing complementary information about heavy quark hadronization relative to the previously measured heavy hadron fragmentation functions. The z and j_T distributions presented here should also provide additional constraints for extractions of the collinear and transverse-momentum-dependent heavy flavor fragmentation functions.

The measurement described in this thesis lays the groundwork for more detailed studies of heavy flavor hadronization in jets at LHCb. Measurements of the joint z and j_T and z and r distributions, as well as measurements of the z, j_T, and r as a function of the particle multiplicity in the jets, would further probe the hadronization dynamics within heavy flavor jets. Further studies of the dead cone in heavy flavor jets are also in progress at LHCb. A better understanding of the dead cone effect in jets is needed to help discriminate between fragmentation processes and hadronization mechanisms. The excellent particle identification capabilities at LHCb also allow for measurements of identified hadron distributions in heavy flavor jets. This is a particularly exciting prospect as in heavy flavor jets one knows the flavor of the initial fragmenting quark, and with identified hadrons in the jet one could in principle identify nearly all of the particles produced from the initial b or c quark hadronization, providing unprecedented insights into hadronization in QCD.

Hadronization will also be a main focus of the physics program at the Electron-Ion Collider. The clean environment of ep collisions relative to pp collisions and the variable center of mass energies available make the EIC ideal for hadronization studies. Heavy flavor jet production and hadronization measurements within jets at the EIC have already been studied and remain areas of active exploration in preparation for the first EIC data projected in 2031 [74–76]. In addition, the z, j_T, and r distributions measured in this thesis would also be interesting to measure in heavy flavor jets at the EIC, as the cleaner collision environment will have fewer fragmentation particles from the proton remnant in the jet and will therefore more directly probe the hadronization of the heavy quark.

APPENDICES

APPENDIX A

PYTHIA Event Generator Settings

The PYTHIA event generator was used to generate Monte Carlo events for comparison with the distributions measured in real data [64]. PYTHIA 8.183 with CT09MCS parton distribution functions was used for event generation of $pp \to b\bar{b}$ and $pp \to c\bar{c}$ events at a center of mass energy $\sqrt{s} = 13$ TeV. The settings used to configure the PYTHIA 8 generator are listed in Tables A.1-A.2. Table A.1 displays the beam and general event settings used, Table A.2 displays the settings used for multiparton interactions, and Table A.3 displays the parameters used for flavor selection and excited state meson generation.

The process settings HardQCD:hardbbbar and HardQCD:hardccbar were used to generate $pp \to b\bar{b}$ and $pp \to c\bar{c}$ events, respectively. For both the beauty and charm events, three different samples with [PhaseSpace:pTHatMin, PhaseSpace:pTHatMax] values of [15, 20] GeV, [20, 50] GeV, and PhaseSpace:pTHatMin $>$ 50 GeV (with no PhaseSpace:pTHatMax cut applied) were generated in order to simulate a range of interaction hard scales. When the three samples were used together, the PYTHIA events were weighted using previously determined weights obtained by computing the ratio of the cross section in each $\hat{p_T}$ bin relative to that in the [10,15] GeV bin [19]. For the $b\bar{b}$ events, the weights used were 0.25 for the $15 < \hat{p_T} < 20$ GeV bin, 0.15 for

Beam and PDF Settings	General Event Settings
Beams:frameType = 3	ParticleDecays:mixB = off
Beams:idA = 2212	HardQCD:hardbbbar(*) = on
Beams:idB = 2212	PhaseSpace:pTHatMin(*) = 15.0
Beams:allowMomentumSpread = on	PhaseSpace:pTHatMax(*) = 20.0
PDF:useLHAPDF = on	SigmaProcess:alphaSorder = 2
PDF:LHAPDFset = CT09MCS	SpaceShower:rapidityOrder = off
PDF:LHAPDFmember = 0	SpaceShower:alphaSvalue = 0.130

Table A.1: Beam settings and general event settings used to generate PYTHIA 8 events. The (*) indicates settings that were varied to generate different event samples, see text for details.

Multiparton Interaction Settings
MultipartonInteractions:bProfile = 1
MultipartonInteractions:alphaSvalue = 0.130
MultipartonInteractions:ecmRef = 7000
MultipartonInteractions:pT0Ref = 2.742289
MultipartonInteractions:ecmPow = 0.238

Table A.2: Settings used for multiparton interactions in PYTHIA 8 events

the $20 < \hat{p_T} < 50$ GeV bin, and 0.009 for the $\hat{p_T} > 50$ GeV bin. The corresponding weights for the $\hat{p_T}$ bins for the $c\bar{c}$ events were 0.21, 0.12, and 0.007.

Flavor Selection and Excited State Settings
StringFlav:mesonUDvector = 0.6
StringFlav:mesonSvector = 7.474387e-01
StringFlav:mesonCvector = 3.0
StringFlav:mesonBvector = 3.0
StringFlav:probQQtoQ = 1.615701e-01
StringFlav:probStoUD = 3.501613e-01
StringFlav:probSQtoQQ = 0.4
StringFlav:probQQ1toQQ0 = 0.05
StringFlav:mesonUDL1S0J1 = 0.0989
StringFlav:mesonUDL1S1J0 = 0.0132
StringFlav:mesonUDL1S1J1 = 0.0597
StringFlav:mesonUDL1S1J2 = 0.0597
StringFlav:mesonSL1S0J1 = 0.0989
StringFlav:mesonSL1S1J0 = 0.0132
StringFlav:mesonSL1S1J1 = 0.0597
StringFlav:mesonSL1S1J2 = 0.0597
StringFlav:mesonCL1S0J1 = 0.0990
StringFlav:mesonCL1S1J0 = 0.0657
StringFlav:mesonCL1S1J1 = 0.2986
StringFlav:mesonCL1S1J2 = 0.2986
StringFlav:mesonBL1S0J1 = 0.0990
StringFlav:mesonBL1S1J0 = 0.0657
StringFlav:mesonBL1S1J1 = 0.2986
StringFlav:mesonBL1S1J2 = 0.2986
StringFlav:etaSup = 1.
StringFlav:etaPrimeSup = 0.4

Table A.3: Settings for flavor selection and generation of excited states in PYTHIA 8 events

APPENDIX B

Individual SV-tagging and BDT Bias Studies

As mentioned in Section 4.4.1, the bias on the charged hadron distributions due to the SV-tagging and applied BDT cut was studied for the SV-tagging and BDT cut separately and when both requirements were applied together. Section 4.4.1 shows the bias study results when both the SV-tagging and BDT cut requirements were applied to the jets. In this section, results are shown separately for the SV-tagging bias and the BDT cut bias, to directly study the effect each of these requirements has on the measured distributions.

Individual Studies for *b*-jets

Figures B.1-B.3 show the results from the SV-tagging bias study for *b*-jets, while Figures B.4-B.6 show the results from the BDT cut bias study for *b*-jets.

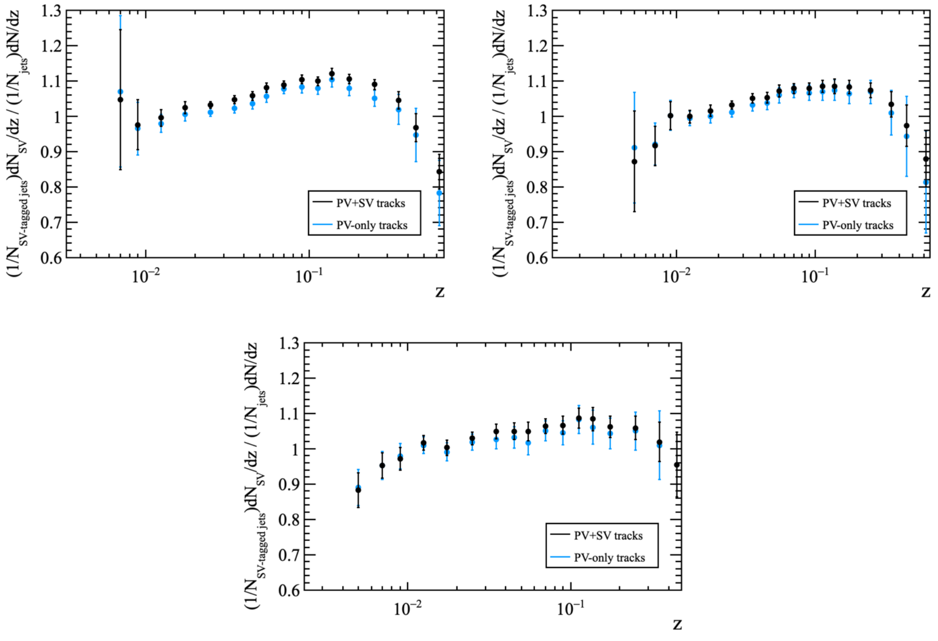

Figure B.1: The bias due to the SV-tagging requirement on the b-jet z distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

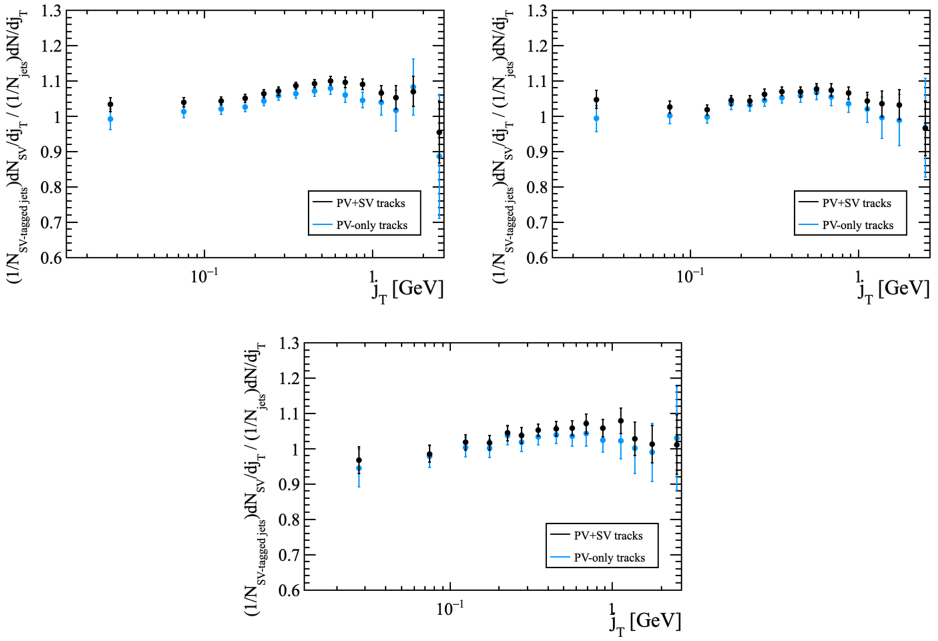

Figure B.2: The bias due to the SV-tagging requirement on the b-jet j_T distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

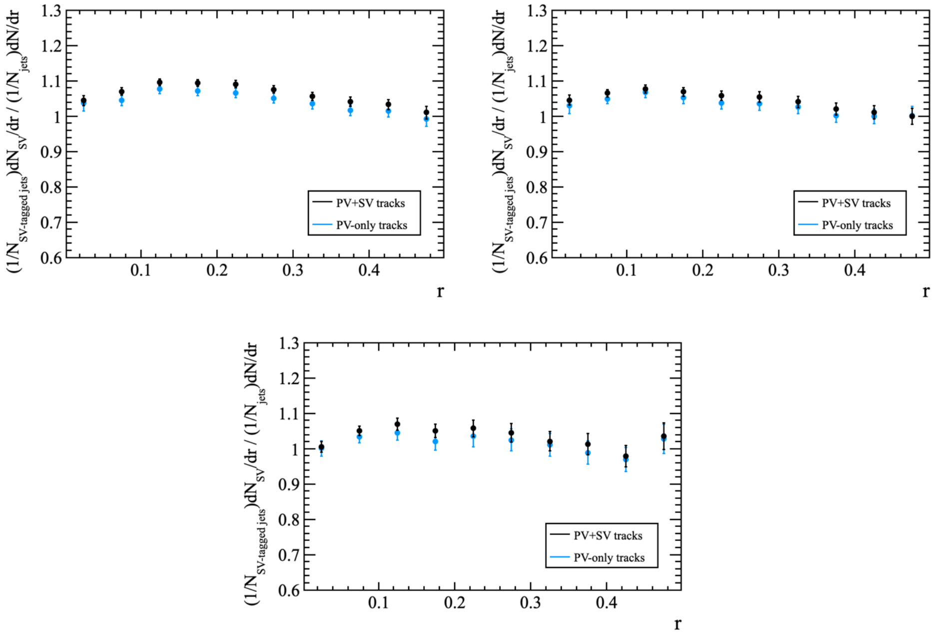

Figure B.3: The bias due to the SV-tagging requirement on the b-jet r distributions. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

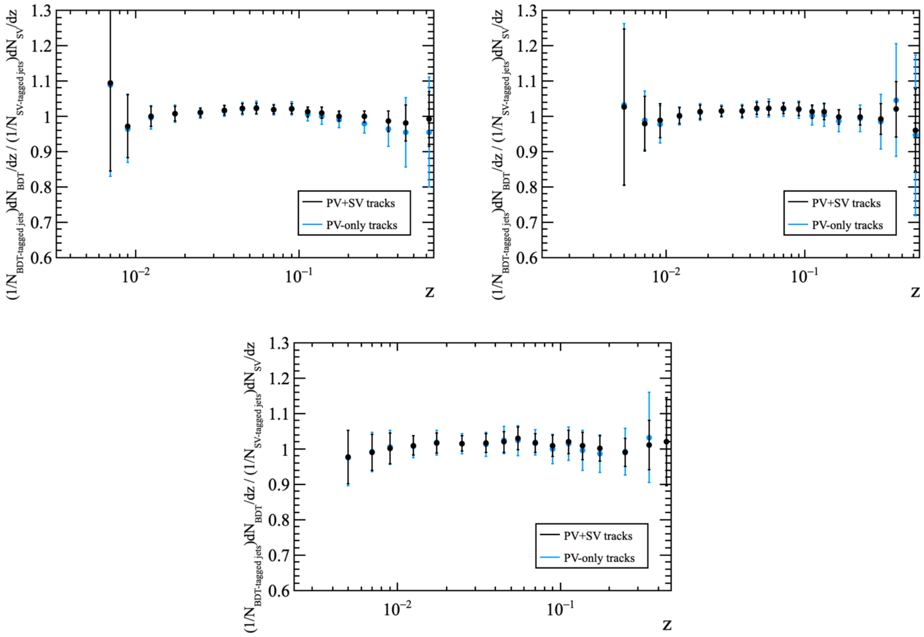

Figure B.4: The bias due to the BDT cut requirement on the b-jet z distributions. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

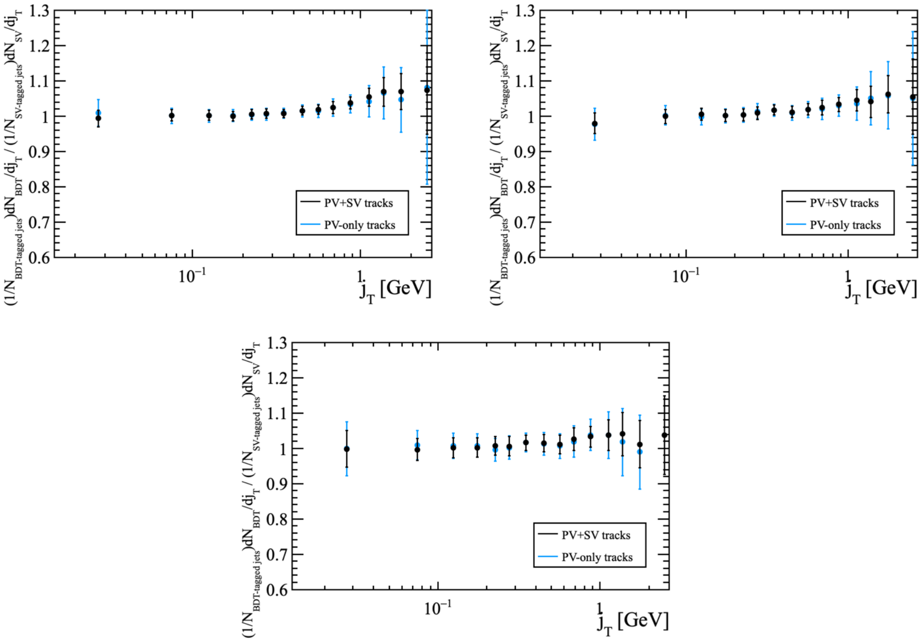

Figure B.5: The bias due to the BDT cut requirement on the b-jet j_T distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

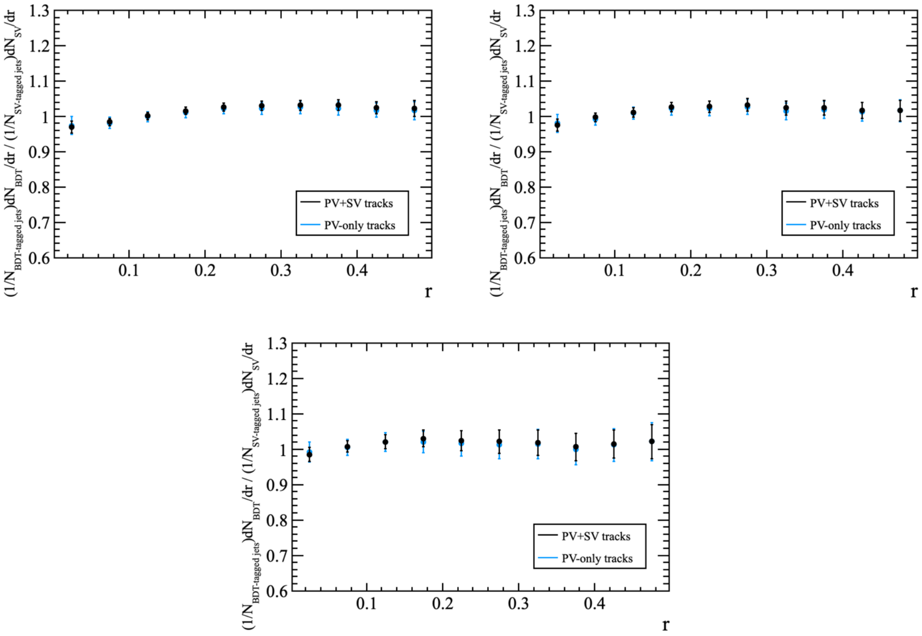

Figure B.6: The bias due to the BDT cut requirement on the b-jet r distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

Individual Studies for c-jets

Figures B.7-B.9 show the results from the SV-tagging bias study for c-jets, while Figures B.10-B.12 show the results from the BDT cut bias study for c-jets.

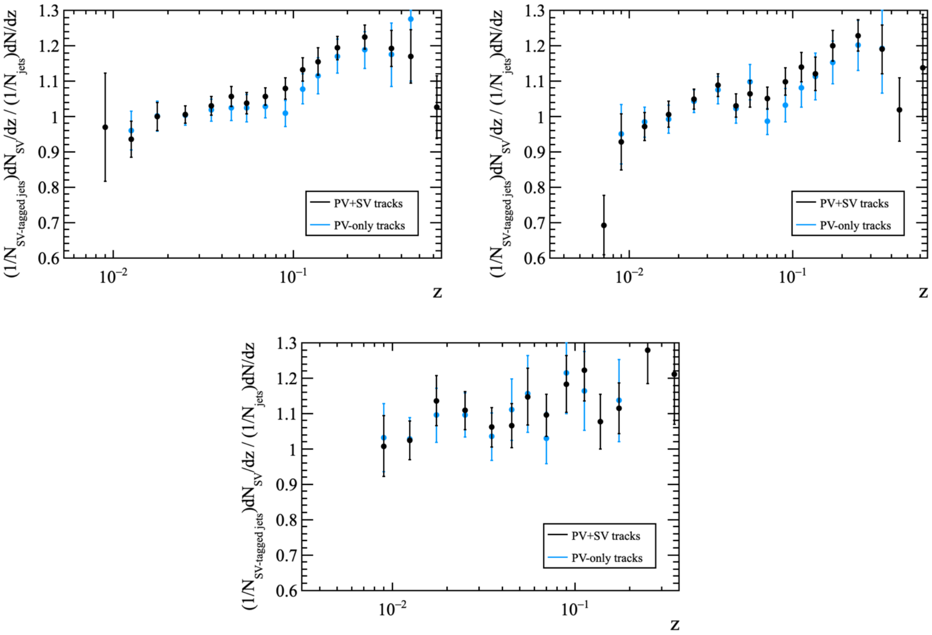

Figure B.7: The bias due to the SV-tagging requirement on the c-jet z distributions. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

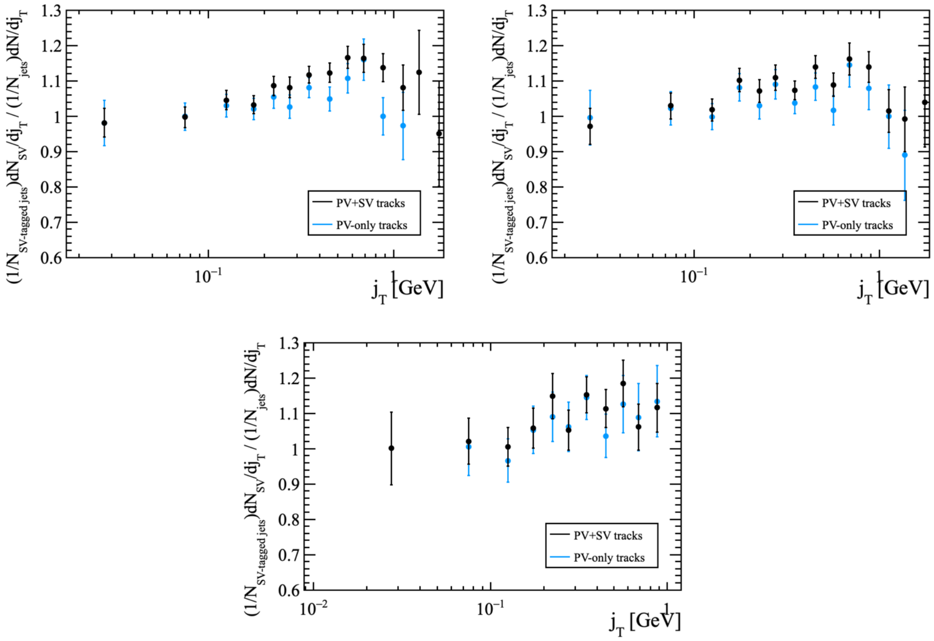

Figure B.8: The bias due to the SV-tagging requirement on the c-jet j_T distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

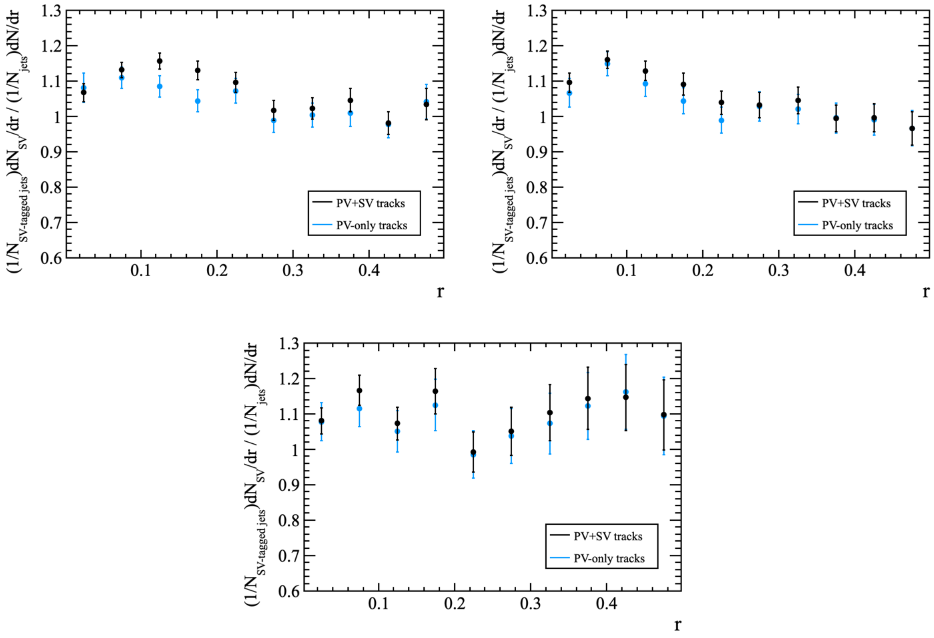

Figure B.9: The bias due to the SV-tagging requirement on the c-jet r distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

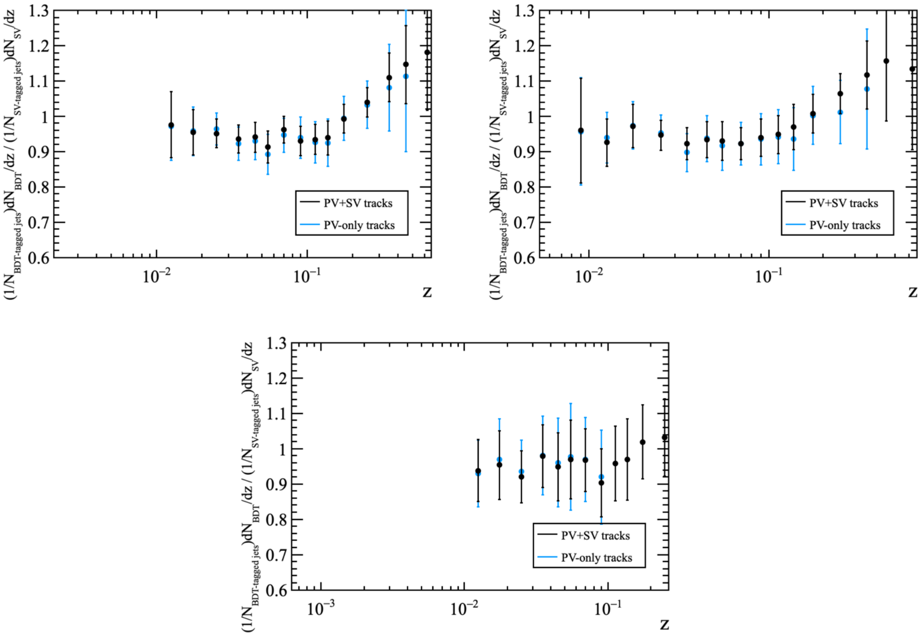

Figure B.10: The bias due to the BDT cut requirement on the *c*-jet z distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

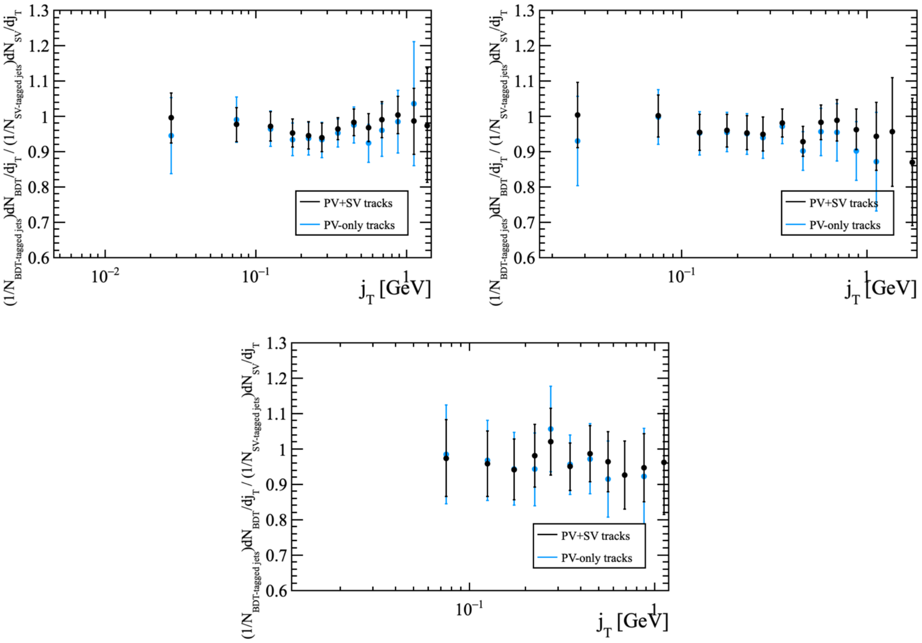

Figure B.11: The bias due to the BDT cut requirement on the *c*-jet j_T distributions. The bias is calculated in bins of jet $\mathrm{p_T}$: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

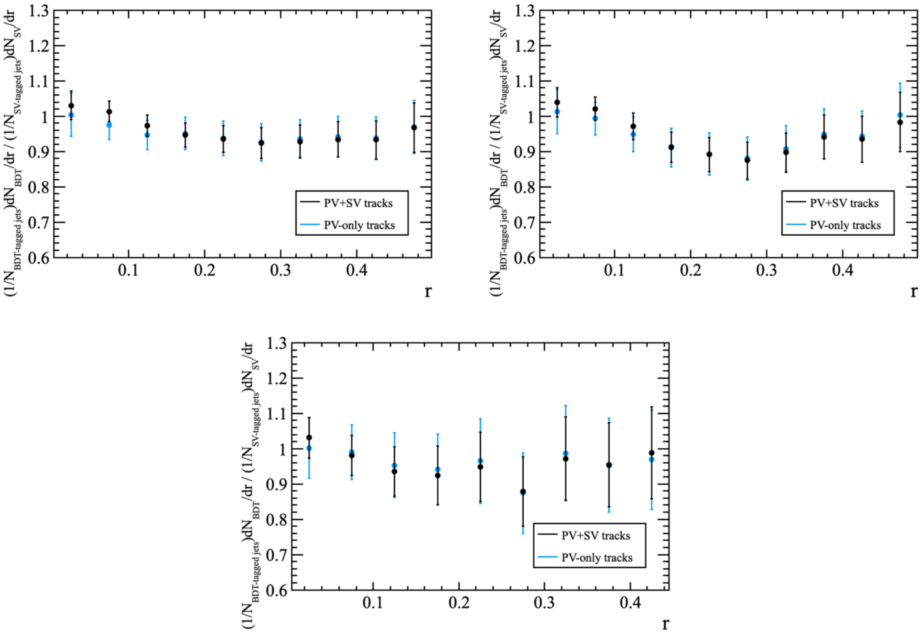

Figure B.12: The bias due to the BDT cut requirement on the *c*-jet r distributions. The bias is calculated in bins of jet p_T: 20-30 GeV (top left), 30-50 GeV (top right), and 50-100 GeV (bottom).

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] P. A. Zyla *et al.*, “Review of Particle Physics,” *PTEP*, vol. 2020, no. 8, p. 083C01, 2020.

[2] Z.-B. Kang, K. Lee, J. Terry, and H. Xing, “Jet fragmentation functions for Z-tagged jets,” *Phys. Lett. B*, vol. 798, p. 134978, 2019.

[3] J. Abdallah *et al.*, “A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole,” *Eur. Phys. J. C*, vol. 71, p. 1557, 2011.

[4] M. Salajegheh, S. M. Moosavi Nejad, H. Khanpour, B. A. Kniehl, and M. Soleymaninia, “B-hadron fragmentation functions at next-to-next-to-leading order from a global analysis of e^+e^- annihilation data,” *Phys. Rev. D*, vol. 99, no. 11, p. 114001, 2019.

[5] G. Aad *et al.*, “Measurement of b-quark fragmentation properties in jets using the decay $B^\pm \to J/\psi K^\pm$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” *JHEP*, vol. 12, p. 131, 2021.

[6] R. Barate *et al.*, “Study of charm production in Z decays,” *Eur. Phys. J. C*, vol. 16, pp. 597–611, 2000.

[7] C. Albajar *et al.*, “A Study of the D^* content of jets at the CERN $p\bar{p}$ collider,” *Phys. Lett. B*, vol. 244, pp. 566–572, 1990.

[8] M. Artuso *et al.*, “Charm meson spectra in e^+e^- annihilation at 10.5 GeV c.m.e.,” *Phys. Rev. D*, vol. 70, p. 112001, 2004.

[9] S. Acharya *et al.*, “Measurement of the production of charm jets tagged with D^0 mesons in pp collisions at $\sqrt{\mathrm{s}} = 7$ TeV,” *JHEP*, vol. 08, p. 133, 2019.

[10] P. Abreu *et al.*, “Hadronization properties of b quarks compared to light quarks in $\mathrm{e}^+ \mathrm{e}^- \to q\bar{q}$ from 183 GeV to 200 GeV,” *Phys. Lett. B*, vol. 479, pp. 118–128, 2000. [Erratum: Phys.Lett.B 492, 398–398 (2000)].

[11] S. Chatrchyan *et al.*, “Shape, Transverse Size, and Charged Hadron Multiplicity of Jets in pp Collisions at 7 TeV,” *JHEP*, vol. 06, p. 160, 2012.

[12] S. Acharya *et al.*, "Direct observation of the dead-cone effect in QCD," arXiv:2106.05713[nucl-ex], 2021.

[13] L. Dai, C. Kim, and A. K. Leibovich, "Heavy Quark Jet Fragmentation," *JHEP*, vol. 09, p. 109, 2018.

[14] E. Mobs, "The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019," Jul 2019. https://cds.cern.ch/record/2684277, Accessed 11-01-2021.

[15] R. Aaij *et al.*, "LHCb Detector Performance," *Int. J. Mod. Phys. A*, vol. 30, no. 07, p. 1530022, 2015.

[16] C. Elsässer, "$\bar{b}b$ production angle plots." https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html, Accessed 11-01-2021.

[17] A. A. Alves, Jr. *et al.*, "The LHCb Detector at the LHC," *JINST*, vol. 3, p. S08005, 2008.

[18] R. Aaij *et al.*, "Identification of beauty and charm quark jets at LHCb," *JINST*, vol. 10, no. 06, p. P06013, 2015.

[19] R. Aaij *et al.*, "Measurement of differential $b\bar{b}$- and $c\bar{c}$-dijet cross-sections in the forward region of pp collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 02, p. 023, 2021.

[20] C. Aidala *et al.*, "An EIC Detector Built Around the sPHENIX Solenoid: A Detector Design Study," https://indico.bnl.gov/event/5283/attachments/20546/27556/eic-sphenix-dds-final-2018-10-30.pdf, 2018.

[21] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories," *Phys. Rev. Lett.*, vol. 30, pp. 1343–1346, 1973.

[22] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?," *Phys. Rev. Lett.*, vol. 30, pp. 1346–1349, 1973.

[23] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD," *Adv. Ser. Direct. High Energy Phys.*, vol. 5, pp. 1–91, 1989.

[24] J. Gao, L. Harland-Lang, and J. Rojo, "The Structure of the Proton in the LHC Precision Era," *Phys. Rept.*, vol. 742, pp. 1–121, 2018.

[25] V. Bertone, N. P. Hartland, E. R. Nocera, J. Rojo, and L. Rottoli, "Charged hadron fragmentation functions from collider data," *Eur. Phys. J. C*, vol. 78, no. 8, p. 651, 2018.

[26] J. J. Ethier and E. R. Nocera, "Parton Distributions in Nucleons and Nuclei," *Ann. Rev. Nucl. Part. Sci.*, vol. 70, pp. 43–76, 2020.

[27] E. Moffat, W. Melnitchouk, T. C. Rogers, and N. Sato, “Simultaneous Monte Carlo analysis of parton densities and fragmentation functions,” *Phys. Rev. D*, vol. 104, no. 1, p. 016015, 2021.

[28] R. A. Khalek, V. Bertone, and E. R. Nocera, “Determination of unpolarized pion fragmentation functions using semi-inclusive deep-inelastic-scattering data,” *Phys. Rev. D*, vol. 104, no. 3, p. 034007, 2021.

[29] A. Metz and A. Vossen, “Parton Fragmentation Functions,” *Prog. Part. Nucl. Phys.*, vol. 91, pp. 136–202, 2016.

[30] T. Kaufmann, A. Mukherjee, and W. Vogelsang, “Hadron Fragmentation Inside Jets in Hadronic Collisions,” *Phys. Rev. D*, vol. 92, no. 5, p. 054015, 2015. [Erratum: Phys.Rev.D 101, 079901 (2020)].

[31] M. Procura and I. W. Stewart, “Quark Fragmentation within an Identified Jet,” *Phys. Rev. D*, vol. 81, p. 074009, 2010. [Erratum: Phys.Rev.D 83, 039902 (2011)].

[32] G. Aad *et al.*, “Measurement of the jet fragmentation function and transverse profile in proton-proton collisions at a center-of-mass energy of 7 TeV with the ATLAS detector,” *Eur. Phys. J. C*, vol. 71, p. 1795, 2011.

[33] A. Jain, M. Procura, and W. J. Waalewijn, “Parton Fragmentation within an Identified Jet at NNLL,” *JHEP*, vol. 05, p. 035, 2011.

[34] M. Procura and W. J. Waalewijn, “Fragmentation in Jets: Cone and Threshold Effects,” *Phys. Rev. D*, vol. 85, p. 114041, 2012.

[35] Z.-B. Kang, F. Ringer, and I. Vitev, “Jet substructure using semi-inclusive jet functions in SCET,” *JHEP*, vol. 11, p. 155, 2016.

[36] Z.-B. Kang, X. Liu, F. Ringer, and H. Xing, “The transverse momentum distribution of hadrons within jets,” *JHEP*, vol. 11, p. 068, 2017.

[37] R. Aaij *et al.*, “Measurement of charged hadron production in Z-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *Phys. Rev. Lett.*, vol. 123, no. 23, p. 232001, 2019.

[38] G. Abbiendi *et al.*, “Inclusive analysis of the b quark fragmentation function in Z decays at LEP,” *Eur. Phys. J. C*, vol. 29, pp. 463–478, 2003.

[39] A. Heister *et al.*, “Study of the fragmentation of b quarks into B mesons at the Z peak,” *Phys. Lett. B*, vol. 512, pp. 30–48, 2001.

[40] K. Abe *et al.*, “Measurement of the b quark fragmentation function in Z^0 decays,” *Phys. Rev. D*, vol. 65, p. 092006, 2002. [Erratum: Phys.Rev.D 66, 079905 (2002)].

[41] M. Fickinger, S. Fleming, C. Kim, and E. Mereghetti, “Effective field theory approach to heavy quark fragmentation,” *JHEP*, vol. 11, p. 095, 2016.

[42] G. Aad *et al.*, "Measurements of jet observables sensitive to b-quark fragmentation in $t\bar{t}$ events at the LHC with the ATLAS detector," arXiv:2202.13901[hep-ex], 2022.

[43] B. R. Yates, "YSF Measurement of the shape of the b quark fragmentation function using charmed mesons produced inside b jets from $\mathrm{t\bar{t}}$ decays," in *14th International Workshop on Top Quark Physics*. 2022.

[44] R. Akers *et al.*, "A Measurement of the production of D^{*+-} mesons on the Z^0 resonance," *Z. Phys. C*, vol. 67, pp. 27–44, 1995.

[45] H. Albrecht *et al.*, "Inclusive production of D^0, D^+ and $D^*(2010)^+$ mesons in B decays and nonresonant e^+e^- annihilation at 10.6 GeV," *Z. Phys. C*, vol. 52, pp. 353–360, 1991.

[46] F. Abe *et al.*, "Measurement of D^* production in jets from $\bar{p}p$ collisions at $\sqrt{s} =$ 1.8 TeV," *Phys. Rev. Lett.*, vol. 64, p. 348, 1990.

[47] G. Aad *et al.*, "Measurement of $D^{*+/-}$ meson production in jets from pp collisions at $\sqrt{(s)} = 7$ TeV with the ATLAS detector," *Phys. Rev. D*, vol. 85, p. 052005, 2012.

[48] D. P. Anderle, T. Kaufmann, M. Stratmann, F. Ringer, and I. Vitev, "Using hadron-in-jet data in a global analysis of D^* fragmentation functions," *Phys. Rev. D*, vol. 96, no. 3, p. 034028, 2017.

[49] J. Kvapil, "Heavy-flavour jet production and correlations with ALICE," *Nucl. Phys. A*, vol. 1005, p. 121921, 2021.

[50] R. Aaij *et al.*, "Study of J/ψ Production in Jets," *Phys. Rev. Lett.*, vol. 118, no. 19, p. 192001, 2017.

[51] A. M. Sirunyan *et al.*, "Measurement of b jet shapes in proton-proton collisions at $\sqrt{s} = 5.02$ TeV," *JHEP*, vol. 05, p. 054, 2021.

[52] Y. L. Dokshitzer, V. A. Khoze, and S. I. Troian, "On specific QCD properties of heavy quark fragmentation ('dead cone')," *J. Phys. G*, vol. 17, pp. 1602–1604, 1991.

[53] F. Maltoni, M. Selvaggi, and J. Thaler, "Exposing the dead cone effect with jet substructure techniques," *Phys. Rev. D*, vol. 94, no. 5, p. 054015, 2016.

[54] M. Battaglia, R. Orava, and L. Salmi, "A Study of depletion of fragmentation particles at small angles in b-jets with the DELPHI detector at LEP," DELPHI-2004-037-CONF-712, 2004.

[55] R. Maciula, R. Pasechnik, and A. Szczurek, "Production of forward heavy-flavour dijets at the LHCb within k_T-factorization approach," arXiv:2202.07585[hep-ph], 2022.

[56] C. W. Bauer and E. Mereghetti, “Heavy Quark Fragmenting Jet Functions,” *JHEP*, vol. 04, p. 051, 2014.

[57] L. Evans and P. Bryant, “LHC Machine,” *JINST*, vol. 3, p. S08001, 2008.

[58] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-k_t jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008.

[59] S. Chatrchyan *et al.*, “Identification of b-Quark Jets with the CMS Experiment,” *JINST*, vol. 8, p. P04013, 2013.

[60] G. Aad *et al.*, “Performance of *b*-Jet Identification in the ATLAS Experiment,” *JINST*, vol. 11, no. 04, p. P04008, 2016.

[61] G. Aad *et al.*, “ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV,” *Eur. Phys. J. C*, vol. 79, no. 11, p. 970, 2019.

[62] R. Aaij *et al.*, “Identification of charm jets at LHCb,” *JINST*, vol. 17, no. 02, p. P02028, 2022.

[63] G. D’Agostini, “A Multidimensional unfolding method based on Bayes’ theorem,” *Nucl. Instrum. Meth. A*, vol. 362, pp. 487–498, 1995.

[64] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.

[65] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics,” *Phys. Rept.*, vol. 97, pp. 31–145, 1983.

[66] A. Accardi *et al.*, “Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all,” *Eur. Phys. J. A*, vol. 52, no. 9, p. 268, 2016.

[67] M. Boglione, J. Collins, L. Gamberg, J. O. Gonzalez-Hernandez, T. C. Rogers, and N. Sato, “Kinematics of Current Region Fragmentation in Semi-Inclusive Deeply Inelastic Scattering,” *Phys. Lett. B*, vol. 766, pp. 245–253, 2017.

[68] F. A. Ceccopieri, “Lambda production in the DIS target fragmentation region,” *Eur. Phys. J. C*, vol. 76, no. 2, p. 69, 2016.

[69] C. Aidala *et al.*, “sPHENIX: An Upgrade Concept from the PHENIX Collaboration,” arXiv:1207.6378[nucl-ex], 2012.

[70] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP*, vol. 05, p. 026, 2006.

[71] E. C. Aschenauer, S. Fazio, M. A. C. Lamont, H. Paukkunen, and P. Zurita, “Nuclear Structure Functions at a Future Electron-Ion Collider,” *Phys. Rev. D*, vol. 96, no. 11, p. 114005, 2017.

[72] D. W. Sivers, "Single Spin Production Asymmetries from the Hard Scattering of Point-Like Constituents," *Phys. Rev. D*, vol. 41, p. 83, 1990.

[73] L. Zheng, E. C. Aschenauer, J. H. Lee, B.-W. Xiao, and Z.-B. Yin, "Accessing the gluon Sivers function at a future electron-ion collider," *Phys. Rev. D*, vol. 98, no. 3, p. 034011, 2018.

[74] E. Chudakov, D. Higinbotham, C. Hyde, S. Furletov, Y. Furletova, D. Nguyen, M. Stratmann, M. Strikman, and C. Weiss, "Heavy quark production at an Electron-Ion Collider," *J. Phys. Conf. Ser.*, vol. 770, no. 1, p. 012042, 2016.

[75] Z.-B. Kang, J. Reiten, D. Y. Shao, and J. Terry, "QCD evolution of the gluon Sivers function in heavy flavor dijet production at the Electron-Ion Collider," *JHEP*, vol. 05, p. 286, 2021.

[76] Y.-T. Chien, A. Deshpande, M. M. Mondal, and G. Sterman, "Probing hadronization with flavor correlations of leading particles in jets," *Phys. Rev. D*, vol. 105, no. 5, p. L051502, 2022.