UNIVERSITY OF CALIFORNIA
RIVERSIDE

Measurement of the Shape of the b Quark Fragmentation Function Using Charmed Mesons in Proton-Proton Collisions at a Center of Mass-Energy of 13 TeV

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Brent R. Yates

September 2019

Dissertation Committee:

Dr. Stephen Wimpenny, Co-Chairperson
Dr. Robert Clare, Co-Chairperson
Dr. John Ellison

The Dissertation of Brent R. Yates is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my advisors, Steve Wimpenny and Bob Clare, for all the help and mentoring they provided.

I am grateful to the CMS collaboration for keeping the CMS experiment running, and collecting the data for my analysis.

I am grateful to my high school physics teacher, who sparked my interest in this field.

To my parents for always encouraging me, to my husband for always supporting me.

ABSTRACT OF THE DISSERTATION

Measurement of the Shape of the b Quark Fragmentation Function Using Charmed Mesons in Proton-Proton Collisions at a Center of Mass-Energy of 13 TeV

by

Brent R. Yates

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, September 2019
Dr. Stephen Wimpenny, Co-Chairperson
Dr. Robert Clare, Co-Chairperson

In this analysis we present the first results of the measurement of the shape parameter r_{b} in the Lund–Bowler fragmentation function for the b quark in a $\mathrm{t\bar{t}}$ environment. The analysis uses charmed mesons produced in the leptonic decays of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 13\,\mathrm{TeV}$ in the CMS detector using the full 2016 dataset with an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. The charged particle decays of $\mathrm{J/\psi} \to \mu^+\mu^-$ and $\mathrm{D}^0 \to \mathrm{K}^\pm\mu^\mp$ are used as proxies for the parent b quark via the ratio x_{B} (the p_{T} of the charmed meson divided by the $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ of all charged particles in the jet containing the meson) to measure the shape parameter r_{b}. The shape parameter r_{b} is measured to be $r_{\mathrm{b}} = 0.841 \pm 0.014\,(\mathrm{stat}) \pm 0.010\,(\mathrm{syst})^{-0.026}_{+0.033}(\mathrm{FSR})$.

Contents

List of Figures

List of Tables

Chapter 1

The Standard Model of particle physics

The Standard Model of particle physics (SM) is one of the most extensively tested theories in physics. For example the predictions for the anomalous magnetic moment of the electron agrees within ten parts per billion with experimental measurements.

The SM is an incomplete theory which provides a framework for three of the four fundamental forces: electromagnetism, the weak force, and the strong nuclear force. It currently does not include gravity, and newer phenomena such as dark matter and dark energy. Built into the SM are a number of parameters that must be measured, such as the particle masses and coupling constants. Once these constants are known, a rich tapestry begins to unfold.

1.1 Particles and forces

The SM contains 17 fundamental particles (Fig. 1.1). These are categorized in two main groups: bosons and fermions, which are classified by their quantum spin. The bosons have integer spin. All the force carrying particles are spin 1 vector gauge bosons, and the Higgs boson is a spin 0 scalar. The fermions are the fundamental particles of matter and are spin 1/2.

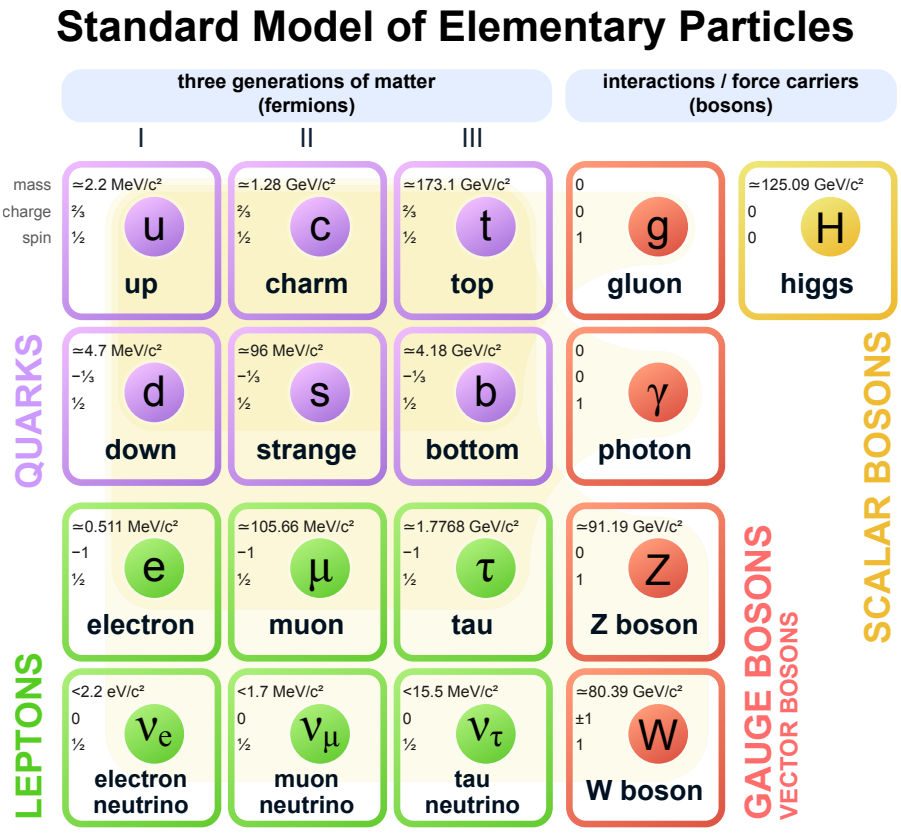

Figure 1.1: The fundamental particles of the SM

1.2 Fermions

The fermions are split into two groups: leptons, which may experience the electromagnetic and the weak forces, and quarks, which experience the electromagnetic force, the weak force, and the strong nuclear force. They come in isospin 1/2 pairs known as doublets which preserve $SU(2)$ symmetry, or isospin 0 singlets. The conserved charge corresponds to weak hypercharge $Y = 2(Q - I_3)$ where Q is the electric charge and I_3 is the third component of isospin. The first isospin doublet contains the electron (e) and the electron neutrino (ν_e). There are two additional higher mass copies, which contain the muon (μ) and tau (τ), and their respective neutrinos.

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L , \begin{pmatrix} \nu \\ \mu \end{pmatrix}_L , \begin{pmatrix} \nu \\ \tau \end{pmatrix}_L \tag{1.1}$$

The e, μ, and τ all have a charge of -1, while the three ν's have no charge. Antimatter leptons also exist with opposite charge (e.g. the anti-electron, or positron, has a charge of $+1$).

The quarks have a similar doublet structure, starting with the up (u) and down (d) quarks making up the first generation. The charmed (c), strange (s), top (t), and bottom (b) quarks make up the second and third generation respectively.

$$\begin{pmatrix} u \\ d \end{pmatrix}_L , \begin{pmatrix} c \\ s \end{pmatrix}_L , \begin{pmatrix} t \\ b \end{pmatrix}_L \tag{1.2}$$

The up type quarks (u, c, t) have $+2/3$ charge, and down type quarks (d, s, b) have $-1/3$ charge. The antimatter quarks have the opposite sign. Quarks also experience the strong nuclear force, which has three different color charges, denoted red (R), green (G), and blue

(B). Confinement Theory states quarks may only exist in color neutral singlets. These may be baryons (RGB) or mesons ($R\overline{R}$, $B\overline{B}$, $G\overline{G}$). Baryons and mesons are collectively known as hadrons. The most common baryon is the proton. It consists of three quarks: two up quarks and a down quark (uud), bound together by gluons. Due to the complicated interactions of the strong nuclear force, gluons may split into quark-antiquark pairs inside the proton. For this reason, the three defining quarks in the proton are called valence quarks. Other common hadrons include the neutron (udd) and the pion, π^+ ($u\overline{d}$), π^0 ($\frac{1}{\sqrt{2}}(u\overline{u}+d\overline{d})$), π^- ($d\overline{u}$).

Quark flavors mix together via the weak force, where the exact mixing is codified by the Cabibbo–Kobayashi–Maskawa (CKM) matrix,

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}. \tag{1.3}$$

The probability of a particular transition is proportional to its CKM matrix element, e.g. $P(t \to b) \sim |V_{tb}|^2$.

The L subscript in Equations (1.1) and (1.2) denotes the chiral symmetry of these doublets. Only left-handed particles (and right-handed antiparticles) experience the weak force. There are also right-handed singlets (left-handed for antiparticles) for the charged

particles*

$$\left(\mathrm{e}\right)_R, \left(\mu\right)_R, \left(\tau\right)_R \tag{1.4}$$

$$\left(\mathrm{u}\right)_R, \left(\mathrm{d}\right)_R, \left(\mathrm{c}\right)_R, \left(\mathrm{s}\right)_R, \left(\mathrm{t}\right)_R, \left(\mathrm{b}\right)_R. \tag{1.5}$$

The three ν's are not charged, and the weak force only acts on left-handed chiral particles, so the ν's are left-handed. Despite many searches, there is no experimental evidence for right-handed ν's.

1.3 Bosons

The gauge bosons are classified by the forces they carry. The photon (γ) mediates the electromagnetic interaction. It is generated by $U(1)$ symmetry, or complex number. It is a neutral, spin 1, massless particle.

Eight gluons (g) mediate the strong nuclear interaction, generated by $SU(3)$ symmetry, or complex 3×3 matrices. They are spin 1 and massless. Gluons carry color charge, making the strong interaction much more complicated than the electromagnetic as gluons self couple and radiate more gluons. The gluons are in superpositions of bi-colored charges, e.g.

$$\frac{\mathrm{R\overline{B}} + \mathrm{B\overline{R}}}{\sqrt{2}}. \tag{1.6}$$

The $\mathrm{W}^{\pm}$ and Z^0 bosons mediate the weak interaction, and are generated by the $SU(2)$ symmetry, or complex 2×2 matrices. The W boson has a measured mass of around

*This subtlety implies a total of four particles for each type. For example the electron is actually a mixture of $\mathrm{e}^-_L + \overline{\mathrm{e}^+_R}$ and the positron is a mixture of $\mathrm{e}^+_L + \overline{\mathrm{e}^-_R}$. The positive and negative, left and right chiral states are each a fundamental particle.

80 GeV, and the Z boson has a measured mass of around 91 GeV. However, these bosons are massless in the SM, and only gain mass through a broken symmetry. The massless weak force and electromagnetic force can be combined into a single framework known as the ELECTROWEAK FORCE, generated by $SU(2) \times U(1)$.

The Higgs (H) is a special boson. It is a spin 0 scalar particle with a rest mass of around 125 GeV. The Higgs boson is not a gauge boson, it is an excitation of the Higgs field which has a nonzero vacuum energy value (VEV) of 246 GeV. Its nonzero vacuum energy is responsible for the breaking of the electroweak symmetry, and gives the W and Z bosons their mass. Interactions between the Higgs field and the fermions can cause a change in fermion chirality ($R \leftrightarrow L$), and results in nonzero rest masses.

1.4 Gauge symmetry

The basis of the SM is in the preservation of local gauge symmetries in QUANTUM FIELD THEORY (QFT). A gauge transformation takes the form of

$$\Psi \rightarrow \Psi' = e^{ig\Lambda\theta}\Psi \tag{1.7}$$

where $i = \sqrt{-1}$ is the complex unit, g is a coupling constant, Λ is an $n \times n$ matrix, and θ is a vector of continuous parameters. The set of Λ matrices are known as the generators of the symmetry. The quantity Ψ is said to be gauge invariant if

$$\Psi'^{\dagger}\Psi' = \Psi^{\dagger}\Psi \tag{1.8}$$

where the dagger (†) in Equation (1.8) denotes the Hermetian conjugate. To preserve probabilities in QFT the generators Λ must belong to $SU(n)$, the group of special unitary $n \times n$ traceless matrices. To be a group the generators must satisfy

$$[\Lambda_i, \Lambda_j] = f_{ij}\Lambda_k, \tag{1.9}$$

where f_{ij} are structure constants, and the commutator $[A, B] = AB - BA$ is in general nonzero because matrices are nonabelian. The values i and j run from 1 to n. This is known as the Lie algebra, and the group is known as a Lie group. An $SU(n)$ group will in general have $n^2 - 1$ generators, as there are $n \times n$ matrices with the additional condition of being traceless.

The basic equation in the SM is the Lagrangian density ($\mathcal{L}$). Classically the Lagrangian (L) is the difference between the kinetic energy (T) and potential energy (V). The equations of motion are found by minimizing the action

$$S = \int L dt. \tag{1.10}$$

In quantum mechanics this only gives the most probable trajectory. QFT (and relativity in general) also replaces the Lagrangian with the Lagrangian density, so

$$S = \int \mathcal{L} d^4x \tag{1.11}$$

where x spans all four dimensions of space-time. The Lagrangian density will now be referred to as simply the Lagrangian.

1.4.1 The Dirac equation

The most important Lagrangian for fermions comes form the DIRAC EQUATION

$$\mathcal{L} = i\hbar c\overline{\psi}\not{D}\psi - mc^2\overline{\psi}\psi. \tag{1.12}$$

In Equation (1.13) $\hbar$ is the reduced Plank constant ($\frac{h}{2\pi}$), ψ is a Dirac spinor, $\overline{\psi} = \gamma^0\psi^\dagger$, $\not{D} = \gamma^\mu D_\mu$ where γ^μ are the Dirac matrices, m is the mass of the particle, and c is the speed of light. D is a gauge covariant derivative $\partial \to D = \partial - A$ where A is a gauge field. This term is required because the standard derivative breaks gauge invariance[†]. In high energy physics it is common to set $\hbar = c = 1$. This reduces the DIRAC LAGRANGIAN to

$$\mathcal{L} = i\overline{\psi}\not{D}\psi - m\overline{\psi}\psi. \tag{1.13}$$

Dirac spinors are four component vectors in spin space. A convenient form is the Weyl (chiral) spinors $\phi_{R,L}$ where

$$\psi = \begin{bmatrix} \phi_R \\ \\ \phi_L \end{bmatrix}, \tag{1.14}$$

$$\phi_R = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix}, \tag{1.15}$$

$$\phi_L = \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix}. \tag{1.16}$$

Inserting the Weyl spinors (Eqs. (1.14) to (1.16)) into the Dirac equation gives

$$\mathcal{L} = i\phi_R^\dagger\not{D}\phi_R + i\phi_L^\dagger\not{D}\phi_L - m(\phi_R^\dagger\phi_L + \phi_L^\dagger\phi_R). \tag{1.17}$$

[†]It is interesting to note that requiring gauge invariance be preserved necessitates the existence of the gauge bosons such as the photon.

Since only left-handed chiral fermions come in isospin doublets the third term in the Dirac equation (Eq. (1.17)) violates gauge invariance. This is because a left-handed fermion could transform (e.g. $e \rightarrow \nu_e$), changing the mass. Therefore, all fermions must be massless in the SM. A similar argument holds for bosons as well.

1.4.2 The Klein–Gordon equation

The analogue of the DIRAC EQUATION for bosons is the KLEIN–GORDON EQUATION. It takes the form of

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2. \tag{1.18}$$

In general ϕ is has multiple components. For example, the W boson is a four-vector W^μ denoting the polarizations. These indices will be suppressed in this dissertation. As mentioned in Section 1.4.1, the mass term violates gauge invariance, and the gauge bosons are massless in the SM.

1.5 Quantum Electrodynamics

QUANTUM ELECTRODYNAMICS (QED) is the simplest physical gauge theory in the SM and describes electromagnetic interactions. This theory preserves $U(1)_{em}$ symmetry, where $U(1)_{em}$ is the group of complex unitary numbers. The em term denotes electromagnetism. Gauge invariance in QED takes the form of

$$\Psi \rightarrow \Psi' = e^{ie\Lambda(x^\mu)}\Psi \tag{1.19}$$

$$A \rightarrow A' = A + \frac{1}{e}\partial_\mu\Lambda(x^\mu) \tag{1.20}$$

where e is the unit of electric charge, Λ is a real valued function, x^μ is the position four-vector in space-time, and A_μ is the electromagnetic four-potential. The covariant derivative term then becomes

$$D_\mu = \partial_\mu - ieA_\mu - i\partial_\mu\Lambda. \tag{1.21}$$

The second term in Equation (1.21) implies the existence of the photon (γ) which mediates the electromagnetic force. Noether's Theorem states all symmetries create conserved quantities. In QED the conserved quantity becomes electric charge.

An extra term must be added to the DIRAC EQUATION to account for the energy of the electromagnetic field. This term is the electromagnetic field tensor, and the Lagrangian becomes

$$\mathcal{L}_{\text{QED}} = \overline{\psi}(i\not{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.22}$$

The electromagnetic tensor is defined as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{1.23}$$

and is gauge invariant by construction.

1.6 Quantum Chromodynamics

QUANTUM CHROMODYNAMICS (QCD) is a generalization of QED to the symmetry $SU(3)_C$. Each of its 3×3 unitary matrices are associated with a different charge (C), which is known as the color charge. The symmetry takes on three distinct values, labeled red (R),

green (G), and blue (B)[‡]. The Lagrangian becomes

$$\mathcal{L}_{\text{QCD}} = \overline{\psi_i} i \not{D}_{ij} - m\delta_{ij}\overline{\psi_i}\psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{1.24}$$

where i and j index the group components, δ_{ij} is the Kroneker delta

$$\delta_{ij} = \left\{ \begin{array}{ll} 1, & \text{if} \quad i = j \\ \\ 0, & \text{else} \end{array} \right\}, \tag{1.25}$$

and $G^a_{\mu\nu}$ is the strong force equivalent to the electromagnetic tensor

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu. \tag{1.26}$$

A^a_μ is the gluon field, analogous to the electromagnetic field. The Lie algebra for QCD takes the form

$$\left[\tfrac{\lambda_i}{2}, \tfrac{\lambda_j}{2} \right] = i \sum_k f^{ijk} \frac{\lambda_k}{2}, \tag{1.27}$$

where λ_i are the eight Gell-Mann matrices, and f^{ijk} are the structure constants.

1.7 Electroweak theory

The concepts of QED can be generalized to include invariance under weak isopsin. The symmetry then becomes $SU(2)_L \times U(1)_Y$, where $SU(2)_L$ is the group of 2×2 special unitary matrices, and Y represents weak hypercharge. Only left chiral spinors experience the weak interaction (denoted by the L). This symmetry only holds in a massless theory, as stated in Section 1.4.1. The ELECTROWEAK THEORY has three W fields ($W^{1,2,3}$) for weak

[‡]Hadrons are not physically colored. The name is an analogy to the fact that mixing red, green, and blue gives white, and mixing the R, G, and B charges gives no charge.

isospin, and a B field for weak hypercharge.

In the physical world this symmetry is indeed broken, and $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$ indicating only electromagnetism remains invariant. The broken symmetry causes the W^3 and B fields to mix, resulting in the physical fields for the weak Z and electromagnetic A by

$$\begin{pmatrix} \mathrm{A} \\ \mathrm{Z}^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} \mathrm{B} \\ \mathrm{W}^3 \end{pmatrix}. \tag{1.28}$$

The weak mixing angle (also known as the Weinberg angle) is θ_W, the mass of the Z^0 boson becomes

$$m_{\mathrm{Z}} = \frac{m_{\mathrm{W}}}{\cos\theta_W}, \tag{1.29}$$

and the $\mathrm{W}^{\pm}$ bosons are identified as

$$\mathrm{W}^{\pm} = \frac{1}{\sqrt{2}}(\mathrm{W}^1 \mp i\mathrm{W}^2). \tag{1.30}$$

The Z boson is identified as

$$\mathrm{Z} = \frac{1}{\sqrt{g^2 + g'^2}}(g\mathrm{W}^3 - g'\mathrm{B}), \tag{1.31}$$

and the electromagnetic field becomes

$$\mathrm{A} = \frac{1}{\sqrt{g^2 + g'^2}}(g'\mathrm{W}^3 + g\mathrm{B}). \tag{1.32}$$

The terms g and g' are the weak force coupling constants. Precision measurements of the boson masses gives a value of $\sin^2\theta_W = 0.22343 \pm 0.00007$ [1].

1.8 The Higgs mechanism

In stark contrast to the SM, the observable universe contains massive particles. While the photon is indeed massless, the W and Z bosons are massive, implying the electroweak symmetry is somehow broken. As stated in Section 1.7, the symmetry becomes

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}. \tag{1.33}$$

In the SM this is due to the spontaneous symmetry breaking caused by the Higgs field. The Lagrangian for Higgs field takes the form of

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi), \tag{1.34}$$

where ϕ is a complex scalar field and $V(\phi)$ is a potential energy term. The Higgs potential is

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \tag{1.35}$$

where μ^2 is a negative constant, and λ is a positive constant. The sign of these two parameters forces a minimum at a nonzero location. A simplified $U(1)$ example of this potential can be seen in Figure 1.2. The Higgs field is an isospoin doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}_L . \tag{1.36}$$

The Higgs doublet has a weak hypercharge $Y = 1$. Since the Higgs potential is only a function of $\phi^\dagger \phi$ the expectation value of the field may be chosen to be

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.37}$$

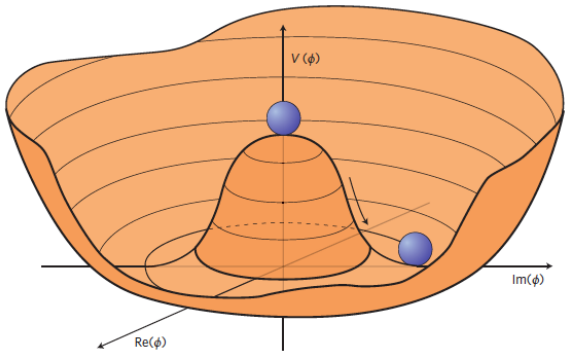

Figure 1.2: The Higgs potential with a nonzero vacuum energy

where $v = \frac{1}{\sqrt{2}}\frac{\mu^2}{\lambda}$. Keeping the electrically neutral component implies the $U(1)_{em}$ of electromagnetism remains unbroken, and the photon remains massless. The Higgs field may be rewritten as

$$\phi = \begin{pmatrix} 0 \\ \\ v + h \end{pmatrix}. \tag{1.38}$$

Using the same covariant derivative for the ELECTROWEAK THEORY, it can be shown that the Lagrangian gains an extra term[§] of

$$\mathcal{L} = \frac{1}{2}\left(\frac{gv}{2}\right)^2 \mathrm{W}^{\dagger}_{\mu}\mathrm{W}^{\mu}. \tag{1.39}$$

Goldstone's Theorem states local excitations of a scalar field will generate new bosons. There are four excitations in the Higgs field, one becoming the Higgs boson, and the other three are said to be eaten up by the vector bosons of the weak force. These three bosons become the longitudinal polarizations of the $\mathrm{W}^{\pm}$ and Z, resulting in the masses

$$m_{\mathrm{W}} = \frac{gv}{2} \tag{1.40}$$

[§]Terms of $\mathcal{O}(h)$ and higher are ignored.

$$m_{\mathrm{Z}} = \frac{v}{2}\sqrt{g^2 + g'^2}. \tag{1.41}$$

Yukawa couplings between the Higgs boson and the fermions results in a mixing of ϕ_L and ϕ_R chiral states

$$\mathcal{L}_{\mathrm{Yuk}} = \frac{f_i v}{\sqrt{2}}(\overline{\phi_L}\phi_R + \overline{\phi_R}\phi_L) \tag{1.42}$$

where f_i is a coupling constant for each fermion. The DIRAC EQUATION (Eq. (1.17)) shows this mixing can be interpreted as a mass term, with $m = \frac{f_i v}{\sqrt{2}}$.

Chapter 2

Top quark production and bottom quark fragmentation

2.1 The top quark

The top quark (t) was discovered in 1995 by the D0 and CDF collaborations [2] at the Tevitron collider at Fermilab. It is the most massive of the fundamental particles with a mass of $172.44 \pm 0.13\,(\text{stat}) \pm 0.47\,(\text{syst})$ GeV (Fig. 2.1:ref. [3]). This makes the top quark unique, and studying it is valuable for both the SM and in the search for physics beyond the Standard Model. The large mass means the top quark decays leptonically to an on-shell[†] W boson ($t \rightarrow qW$) on a time scale shorter than the QCD interaction. This means the top quark may be studied as a bare quark; something previously thought forbidden by Confinement Theory. The large mass also gives the top quark a Yukawa coupling constant of close

[†] $p^\mu p_\mu = E^2 - |\vec{p}|^2 c^2 = m^2 c^4$

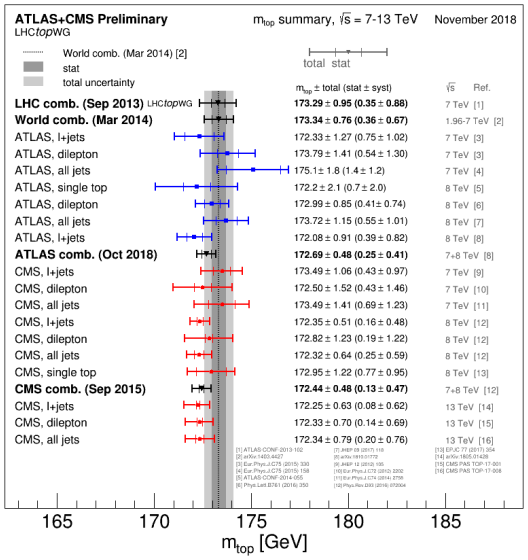

Figure 2.1: Measurements of the top quark mass by the CMS collaboration, the ATLAS collaboration, and the world average value.

to unity. This makes it vital to the study of the Higgs potential. Current experimental uncertainties on the top quark mass are the leading uncertainty in calculating the stability of the Higgs field and the electroweak vaccum (Fig. 2.2:ref [4]).

The most common form of top quark production is top-antitop pair ($t\bar{t}$) production. At the LHC the dominant mechanism is through gluon-gluon interactions (Fig. 2.3), with a small contribution coming from quark-antiquark annihilation (Fig. 2.4). The parton distribution functions (PDF) shown in Figure 2.5 indicate that the gluons carry a significant fraction of the proton momentum in the x and Q^2 ranges relevant to the LHC. Any anti-quarks in the proton only form from gluon to quark-antiquark pair splitting ($g \rightarrow q\bar{q}$). The quark-antiquark only receives a small fraction of the gluon's momentum (typically much

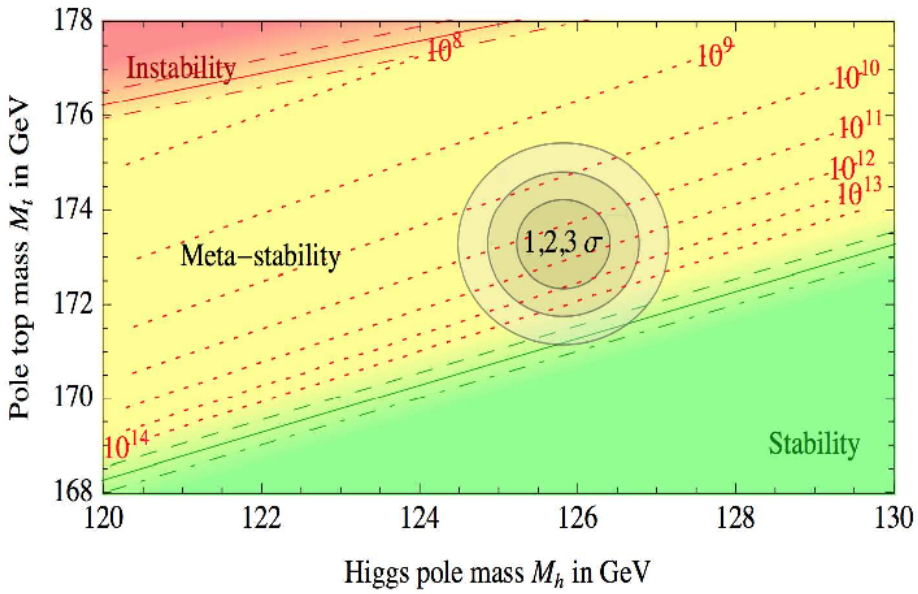

Figure 2.2: Stability of the vacuum based on the mass of the t quark and the mass of the Higgs boson

smaller than $m_{t\bar{t}}$), so top quark pair production by quark-antiquark annihilation is heavily suppressed.

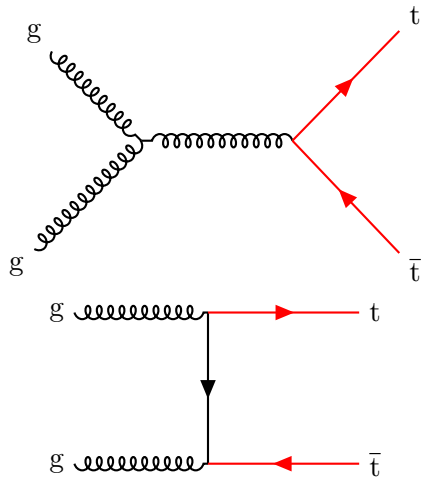

Figure 2.3: Gluon-gluon to $t\bar{t}$ pair production

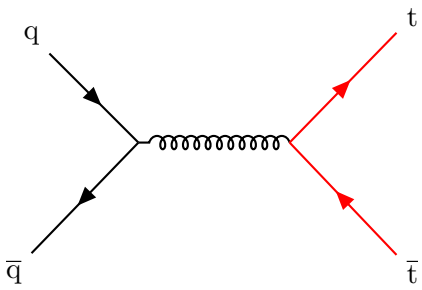

Figure 2.4: Quark-antiquark to $t\bar{t}$ pair production

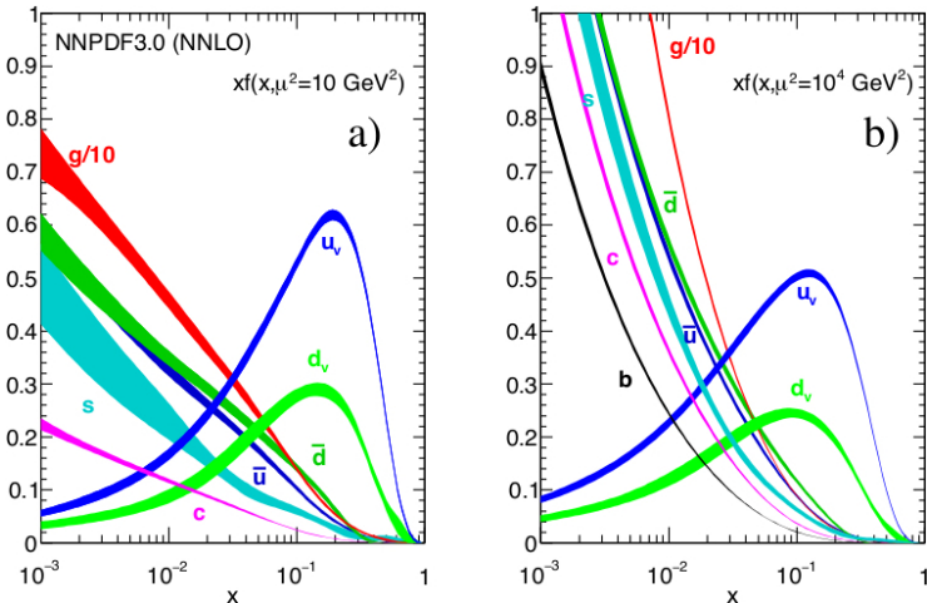

Figure 2.5: Parton distribution function (PDF) normalized to the expected number in the proton at the scale $Q^2 = 10\,\mathrm{GeV}^2$ (left) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (right)

.

Single top quark production is also possible through s-channel and t-channel (Fig. 2.6) decays, as well as W boson decay (Fig. 2.7), though it is much more rare than $t\bar{t}$ production.

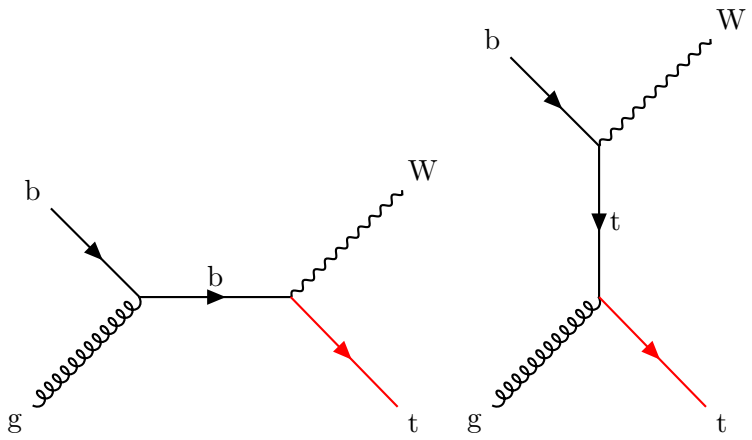

Figure 2.6: Single-t production through s-channel and t-channel decay

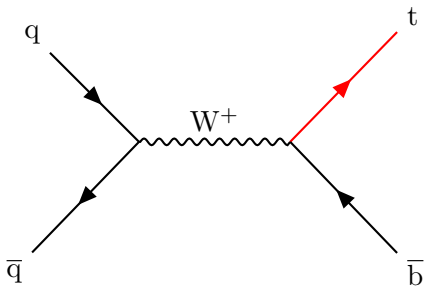

Figure 2.7: Single-t production though W boson decay

The top quark decays almost exclusively to the bottom (b) quark and a W boson. The CKM matrix elemnt from Equation (1.3) is $|V_{tb}| > 0.975$ [5] at the 95% confidence level. Decays giving the strange (s) and down (d) quarks are possible, but extremely unlikely. This allows top quark decays to be characterized by the decays of the W boson: leptonically ($W \to l\nu$) and hadronically ($W \to q\overline{q}'$) as shown in Figure 2.8. The large mass difference between the t quark and the b quark results in a large b quark momentum. This large

momentum, combined with a small probability of decay, gives the b quark a long lifetime , and allows events with a b quark to be flagged (b-tagging).

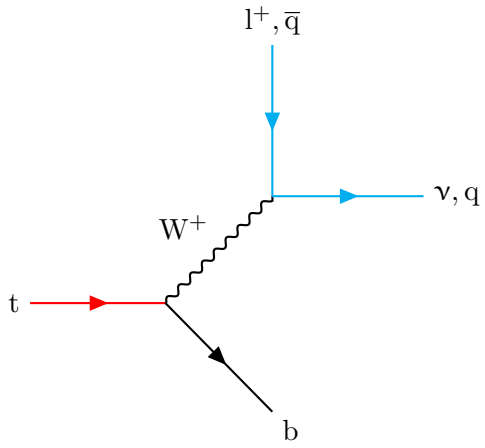

Figure 2.8: Characterizing the top quark through W boson decay

2.2 Bottom quark fragmentation

The high probability of decaying to the bottom quark also makes the top quark a very useful tool in studying the b quark decays. Fragmentation (or hadronization) is the processes in which partons decays into detecable particles. The exact fraction of momentum each particle carries is not exactly calculable, and must be fit with an empirical function. A commonly used model is the Lund–Bowler [6] fragmentation function

$$f(z) = \frac{1}{z^{1+r_b \cdot b \cdot m_b^2}} \cdot (1-z)^a \cdot exp(-\frac{b \cdot m_T^2}{z}), \tag{2.1}$$

where z is the fraction of momentum the particle receives from the fragmenting quark, a and b are general fit parameters, r_b and m_b are specific to the fragmenting b quark, and

$m_{\mathrm{T}}^2 = m_{\mathrm{had}}^2 + p_{\mathrm{T,had}}^2$. Extensive studies of light quarks $(\mathrm{u}, \mathrm{d}, \mathrm{s})$ have fixed the parameters a and b, while studies of the b quark have fixed m_{b}. The most recent value of r_{b} comes from Z pole decay in electron-positron colliders ($\mathrm{e^+e^-} \to \mathrm{Z^0} \to \mathrm{b\bar{b}}$) at the Large Electron Positron (LEP) collider [7, 8, 9] formerly at CERN, and the SLAC Linear Collider (SLC) [10] at Stanford. These colliders provided a very different environment from the color-rich $\mathrm{t\bar{t}}$ decays at the LHC. In order that the results can be safely used for LHC calculations, the fragmentation function needs to be re-measured and compared to the $\mathrm{e^+e^-}$ results. If the results agree then this would be direct prof of the environmental invariance of the fragmentation process. This is one of the goals of the analysis presented in this dissertation. A plot of the Lund–Bowler using the default CMS tune parameters [11] can be found in Figure 2.9.

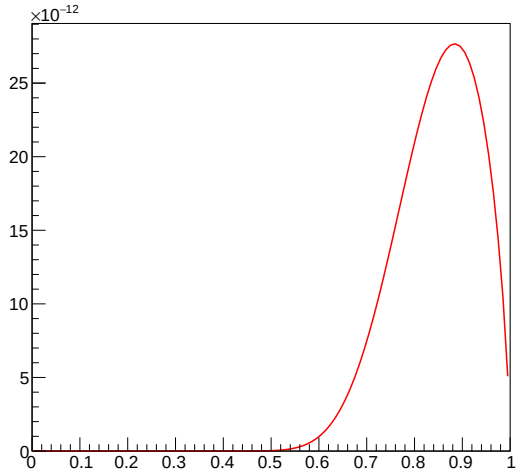

Figure 2.9: Lund–Bowler function using the default CMS tune parameters

Chapter 3

The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest scientific experiment ever created. It is an approximately circular collider with a circumference of 26.7 km, and it cuts across the Franco/Swiss border. It ranges from 50 m to 175 m underground, occupying tunnels originally built for the LEP collider. The LHC fist powered on in 2010 with a center of mass-energy ($\sqrt{s}$) of 7 TeV. In 2012 the LHC completed one of its major goals when the Higgs boson was experimentally observed for the first time. In 2015 the LHC moved to the higher energy of $\sqrt{s} = 13$ TeV.

The LHC consists of 1232 superconducting dipole magnets, as well as thousands of superconducting and normal multi-pole magnets for beam focusing, cleaning, and alignment. The LHC also has a superconducting Radio Frequency (RF) cavity to accelerate the protons to the required center of mass-energy.

The LHC ring has several INTERACTION POINTS (IP) which contain various detectors and other systems. IP1 (in Meyrin, Switzerland) and IP5 (in Cessy, France) house two complementary general purpose detectors: the ATLAS detector, and the COMPACT MUON SOLENOID (CMS) detector respectively. IP2 (in Saint–Genis–Pouilly, France) houses the ALICE detector for heavy-ion physics. IP8 (in Ferney–Voltaire, France) is home to the LHCb detector, which focuses on b quark physics. The RF cavities are located at IP4. A aerial image of the LHC may be found in Figure 3.1.

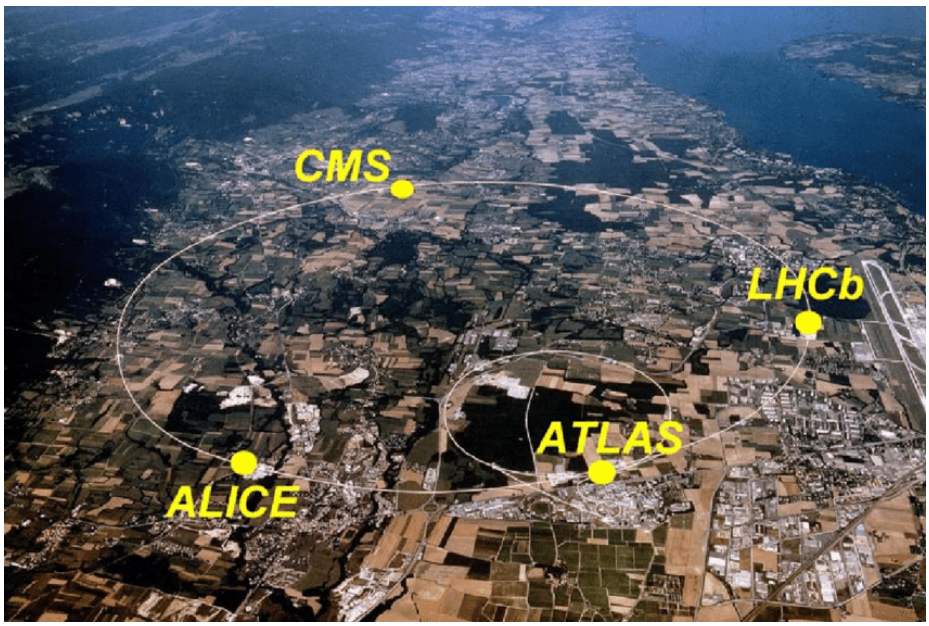

Figure 3.1: A aerial view of the LHC with the four main experiments labeled (Lake Geneva is located in the upper right corner)

The LHC collides protons at each of the IP around the ring. The number of collisions per second in particle physics is measured as the luminosity ($\mathcal{L}$) of an experiment, which can be calculated using

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{\pi \epsilon_n \beta^*} F. \tag{3.1}$$

Here N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is a relativistic factor, ϵ_n is the normalized transverse beam emittance, β^* is the betatron function at each a collision point, and F is the geometric luminosity reduction factor caused by crossing angle, bunch length, and the transverse beam size.

In circular colliders energy is lost in the form of synchrotron radiation. Charged particles deflected by a magnetic field lose an energy of

$$\Delta E = \frac{4\pi\alpha}{3r} \beta^3 \gamma^4, \tag{3.2}$$

where $\alpha = \frac{e^2}{4\pi}$[†] is the fine structure constant, r is the radius of curvature, $\beta = \frac{v}{c}$ is the relativistic velocity, and $\gamma = \frac{E}{m}$ is the relativistic Lorentz factor. The LHC employs protons as their large mass help to mitigate this last term.

The LHC has a peak luminosity of $\mathcal{L} = 1.4 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$. The luminosity can be combined with the cross section (σ), or probability of an event occurring, to give the total number of events produced

$$N = \sigma \mathcal{L}. \tag{3.3}$$

[†] $\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} \approx 1/137$ in SI unites.

A more useful unit at the LHC is the integrated luminosity ($\int \mathcal{L}$). This is the total luminosity over a given run period, and is expressed in inverse units of cross section. For example the LHC delivered a total integrated luminosity of $41\,\mathrm{fb}^{-1}$ during the 2016 data taking period, where the unit barn[†], $\mathrm{b} = 10^{-24}\,\mathrm{cm}^2$, is classically the cross sectional area.

In order to reach the desired energy, the LHC uses additional CERN accelerators for the initial acceleration of the protons. Starting at the linear accelerator (LINAC 2), hydrogen atoms are stripped of their electron. This isolates the protons, and they are accelerated to 50 MeV. Then LINAC 2 feeds into the Proton Synchrotron Booster (PSB), which is a circular accelerator that boosts the protons to 1.4 GeV. The PSB injects into the Proton Synchrotron (PS) accelerator which brings the protons to 25 GeV. The protons are then directed into the Super Proton Synchrotron (SPS) which accelerates them to 450 GeV. The protons finally enter the LHC and are split into two beam pipes, one flowing clockwise and the other counter-clockwise. The LHC RF cavity accelerates the protons to the final 6.5 TeV per beam, resulting in a center of mass-energy of 13 TeV when the beams collide. A single proton in the LHC is moving at about $0.999999991c$ at the time of collision. For heavy-ion runs, LINAC 2 is replaced by LINAC 3. A diagram of this injection chain can be seen in Figure 3.2.

[†]The barn is a rather large unit in nuclear physics terms. Originally coined from the expression"[they] can't hit the broad side of a barn."

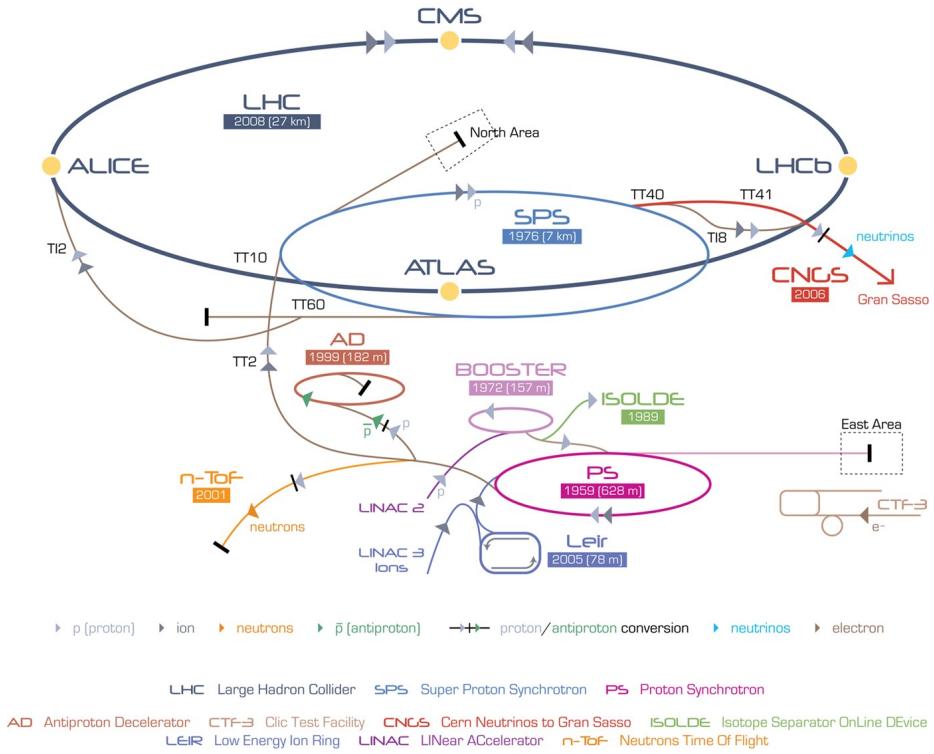

Figure 3.2: A diagram of the LHC including injection stages

While proton-proton interactions may have a large cross section in high energy terms, protons are still incredibly small. Crossing large bunches of protons at each IP helps ensure collisions actually occur, increasing the probability of a collision in each bunch crossing (bx). The bunch spacing is 25 ns, and a total of 2808 bunches are injected into the LHC per fill. Each bunch contains a total of roughly 10^{11} protons[†]. This also means the luminosity of a fill decreases over time as more protons in a bunch are successfully collided.

[†]It would take roughly 1 million years for the LHC to use up enough hydrogen to fill a single balloon.

Chapter 4

The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of the general purpose detectors at the LHC. It is located at IP5 in the town of Cessy, in the French country side. As the name implies, the CMS detector employs a large solenoid, which is used for particle identification. CMS has a large magnetic field of 3.8 T, one of the strongest in the world[†]. The field is used to bend the trajectory of charged particles as they pass through the detector, and the curvature gives a measure of the momentum each particle carries, as well as the sign of the particle's charge. The CMS detector is comprised of several smaller subsystems. The inner tracker is used to precisely measure the track momenta of charged particles using the relation

$$R = \frac{p_{\mathrm{T}}}{0.3eB}, \tag{4.1}$$

[†]There are magnets with a larger field strength, but the physical size of the CMS magnet means the field contains an enormous amount of energy.

where R is the radius of curvature in meters, p_{T} is the transverse momentum of the particle in GeV[†], and B is the magnetic field strength in Tesla. The electromagnetic calorimeter (ECAL) is used to measure the energy of electromagnetically interacting particles. The hadronic calorimeter (HCAL) is subsequently used to measure the energy of hadronic particles. The muon system lends its name to CMS, and is used to measure high energy, and therefore long lived, muons leaving the detector. A detailed cross sectional view of the CMS detector, along with its subsystems, may be found in Figure 4.1.

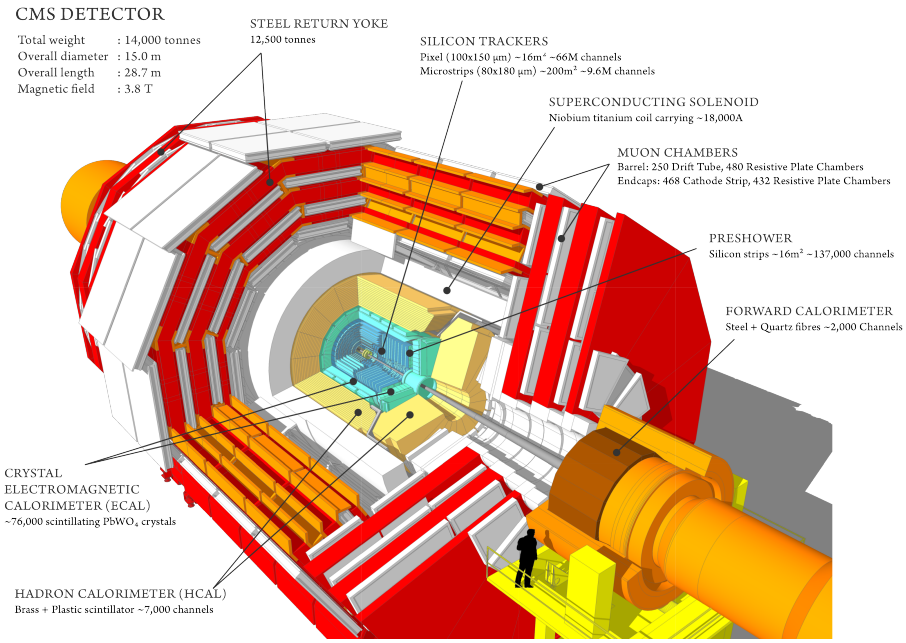

Figure 4.1: A diagram of the CMS detector, complete with subsystems

[†] GeV/c where $\hbar = c = 1$

4.1 CMS coordinate convention

The CMS experiment uses a standard right-handed coordinate system, with the $z-$axis pointing along the beam pipe, in the counterclockwise direction. The positive $x-$axis is defined to point inward towards the center of the LHC ring, resulting in the positive $y-$axis pointing upward. The r coordinate is measured in the transverse $(x-y)$ plane, and gives the radial distance from the interaction point. The azimuthal angle, ϕ, is the angle r makes with respect to the positive axis in the $x-y$ plane. The polar angle, θ, specifies the direction in the $y-z$ plane, but a more convenient quantity is the pseudorapidity η, which is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{4.2}$$

The quantity η is both a measure of the polar angle in the detector, and of how boosted the system is with respect to the detector. The definition of η as the limit of rapidity y ensures that $\Delta\eta$ is a Lorentz invariant. The r and ϕ quantities are important in measuring the transverse momentum (p_{T}) of the particles. On average the colliding partons should have zero net transverse momentum, making p_{T} a very useful quantity in analyses. The most common form of four-momentum used in CMS is codified by these quantities

$$p^{\mu} = (E, p_{\mathrm{T}}, \eta, \phi), \tag{4.3}$$

where the transverse momentum is defined as

$$|p_{\mathrm{T}}| = \sqrt{p_x^2 + p_y^2}. \tag{4.4}$$

The three spacial components of momentum may be recovered using the identities

$$p_x = p_{\mathrm{T}} \cos\phi \tag{4.5}$$

$$p_y = p_{\mathrm{T}} \sin \phi \tag{4.6}$$

$$p_z = p_{\mathrm{T}} \sinh \eta \tag{4.7}$$

$$|p| = p_{\mathrm{T}} \cosh \eta, \tag{4.8}$$

where cosh and sinh are the hyperbolic cosine and sine functions respectively. Using Equations (4.3) to (4.7) the standard definition for the momentum four-vector is recovered

$$p^{\mu} = (E, p^x, p^y, p^z). \tag{4.9}$$

4.2 CMS subsystems

The CMS detector is comprised of several subsystems working together. Each subsystem has a particular purpose, such as measuring particle momentum or identifying muons.

4.2.1 The inner tracker

Silicon Pixels

The pixel tracker is the inner most component of the CMS detector, and provides information on charged particle tracks. The silicon pixels are used for identifying interaction vertices of short lived particles. There are 65 million pixels in the detector ranging from 4 cm to 11 cm arranged in a barrel around the beam pipe, as well as endcap disks to cover the forward region. Due to the proximity to the interaction point the silicon pixels must be able to withstand a large amount of radiation, and have small dimensions to ensure a small track occupancy close to the beam pipe.

Silicon Strips

The silicon strips sit just outside of the silicon pixels. There are a total of 10 layers, with an overall radius of 130 cm. The silicon strips are divided into smaller regions: four tracker inner barrel (TIB) layers with two tracker inner disks (TID) as endcaps, six tracker outer barrel (TOB) layers, and finally two more tracker endcap (TEC) disks. The silicon strips contain a total of 15,200 modules with a total of 10 million detector strips. A single module includes a set of sensors, the support structure, and readout electronics. The silicon strips provide good spacial and time resolution suited for the high density region of particles. The strips provide precise measurements of the tracks of charged particles, and help measure the curvature as they pass through the magnetic field. They are also useful in identifying displaced, or secondary vertices. As the charged particles pass through the strip layers they ionize atoms in the silicon. This produces a small positive charge on the strips which is then amplified by an Analogue Pipeline Voltage (APV25) chip. The interactions happen on the scale of nanoseconds, but the APV25 stores the charge information for a few microseconds. The information is then sent down a 100 m fiber optic line for further processing in a radiation free zone. There are a total of 40,000 fiber optic connections in the silicon strips. An image of the silicon strips is in Figure 4.2.

4.2.2 ECAL

The ECAL is located outside the inner tracker and is used to measure the energy of all electromagnetically interacting particles, such as electrons and photons. Similar to the tracker, the ECAL is split into smaller regions: the ECAL barrel (EB) and endcaps (EE).

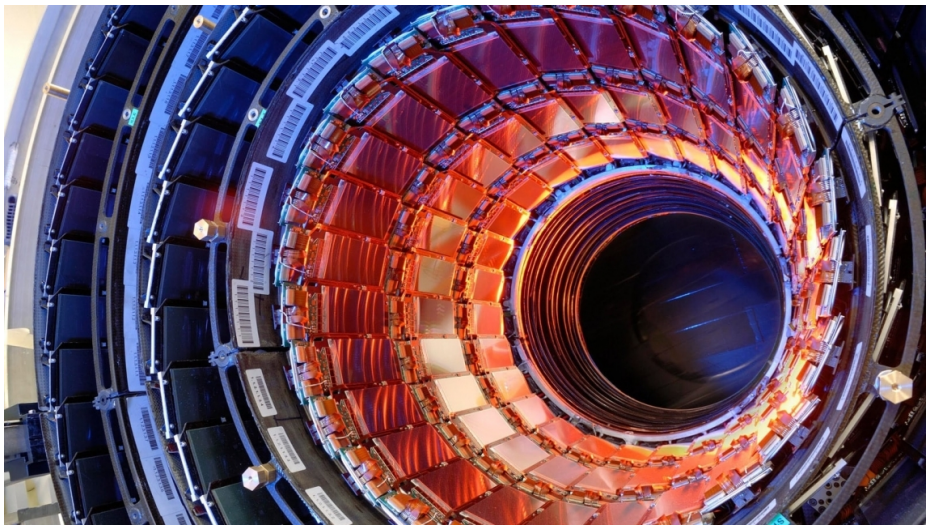

Figure 4.2: An image of the CMS silicon strips in the barrel region

The ECAL contains a total of 75,848 lead-tungstate ($PbWO_4$) scintillating crystals. These crystals are able to withstand the large amounts of radiation present at the ECAL. Photoamplifiers are attached to the end of each crystal which covert the scintillation photons to an electric signal for readout.

To help distinguish highly collinear photons (such as $\pi^0 \to \gamma\gamma$) from single high energy photons, the ECAL employs an ECAL Preshower (ES). The ES sits in front of the ECAL endcaps in the high η region.

4.2.3 HCAL

The HCAL is the final component of the inner CMS detector, and sits just outside the ECAL. The HCAL is used to measure the energy of strongly interacting particles, such as neutral and charged pions. Due to the rich nature of hadronic interactions the HCAL

is hermetic. This allows it to detect as many particles originating from the collisions as possible. The HCAL is divided into three regions: the HCAL barrel (HB and HO), the HCAL endacp (HE), and the HCAL forward (HF).

4.2.4 The Muon system

A special system designed for detecting muons sits outside the CMS magnet. Due to the unique conditions, muons are the only particles, other than neutrinos, which are expected to leave the inner detectors. The iron return yoke outside the CMS magnet helps provide a near constant magnetic field of 2 T. The muon system consists of three separate gas filled subsystems. A diagram of the muon system is in Figure 4.3.

Drift Tubes

The drift tubes (DT) cover the central region of the CMS detector. There are a total of 250 individual DT chambers. A single DT is 4 cm wide and contains several parallel wires inside a gas chamber. Muons passing through the chamber by will ionize the gas, and the ionized electrons will be attracted to the anode wires. This gives a single spacial coordinate for the muons. The original distance of a muon from a wire can be calculated using the drift speed of the electrons in the DT and the time for it to travel. This gives additional spacial coordinate. The DTs have a spacial resolution of about $250\mu m$.

Cathode Strip Chambers

The cathode strip chambers (CSC) cover the endcap disks of the CMS detector: $0.9 < |\eta| < 2.4$. Each CSC consists of positively charged wires (anodes) orthogonal to

negatively charged strips (cathodes) inside a gas chamber. When muons pass through a chamber the gas becomes ionized. As the ionized electrons travel towards the anodes they ionize more gas molecules, resulting in a avalanche of electrons. The positive gas ions are also attracted to the negatively charged cathode strips. Combining anode and cathode measurements allows the CSCs to resolve two spacial resolutions of the incoming muons to about $75\mu m$ for the inner chambers and $150\mu m$ for the outer chambers. The CSCs also have a fast timing of about 6 ns, making them useful for triggering (see Section 4.3).

Resistive Plate Chambers

The resistive plate chambers (RPC) overlap with the DTs and also cover a portion of the CSCs : $|\eta| < 2.1$; significant background (including from the beam pipe) is present at $|\eta| > 2.1$ which would diminish the triggering capabilities of the RPCs. A single RPC consists of a gas chamber sandwiched between two oppositely charged (anode and cathode), high resistivity insulated plates. Muons passing through the RPCs ionize the gas, resulting in an avalanche similar to how the CSCs operate, using electrodes read the avalanche signal. RPCs have good spacial resolution and a fantastic timing of about 1 ns, making them ideal for triggering.

4.3 Trigger and Data Acquisition System

The nominal design of the LHC results in a collision rate of 40 MHz, colliding roughly 20 protons per bunch crossing. This is simply too much data to save with current computing technology. To mitigate this restriction, CMS employs the Trigger and Data

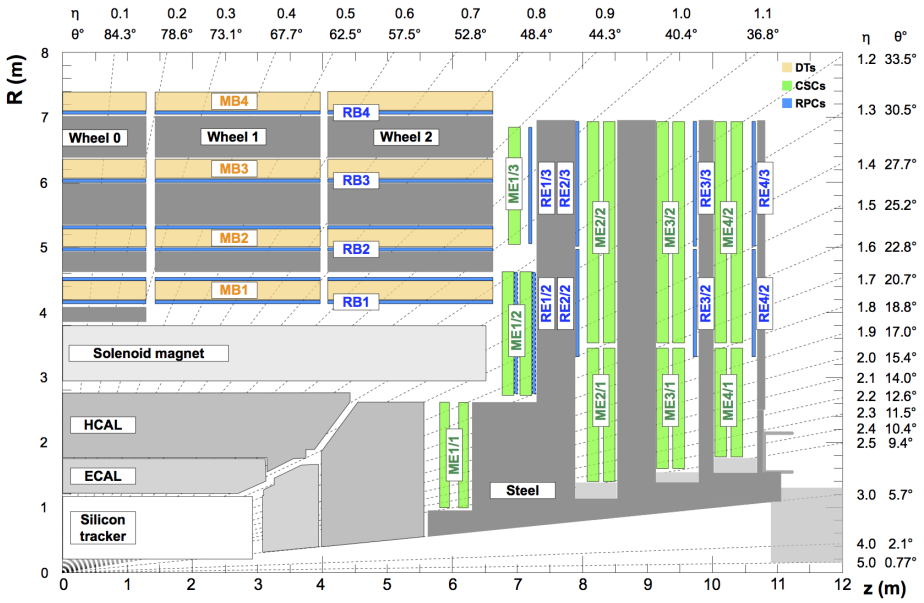

Figure 4.3: A side view of the CMS muon system in the $r - z$ plane

Acquisition System (TriDAS). A trigger is designed to dramatically lower the data rate; discarding less useful events, and keeping only the most interesting events. There are two main parts to the trigger: a quick hardware trigger known as the Level 1 (L1) Trigger, and a more sophisticated software trigger known as the High Level Trigger (HLT). The L1 trigger reduces the collision rate to about 100 kHz, and the HLT brings the final rate down to about 1 kHz.

Level 1 Trigger

The L1 Trigger has a decision time of about 3.2 μs, which is too short to reconstruct tracker tracks. Therefore, information from the silicon tracker is not used in L1 Trigger [†].

[†]The planned HL-LHC upgrade in 2026 will include an L1 Track Trigger. It will trigger on tracker particles with a minimum p_T of 2–3 GeV using FPGA hardware.

The ECAL and HCAL send particle information known as hits to the Calo Trigger Layer 1. Layer 1 constructs primitive object candidates which are passed to the Calo Trigger Layer 2. Layer 2 uses this information to construct high level physics objects, and decides whether an event is worth saving. Events which pass this layer are sent to both the Global Muon Trigger (GMT), and the Global Trigger.

The GMT uses hit information from the three muon subsystems to determine if an event is worth keeping. This trigger is split into three sub-regions, combining information known as trigger primitives from each region. The first is the Barrel Muon Track Finder (BMTF), using information from the DTs and RPCs from a region of $|\eta| < 0.85$. Next, the Overlap Muon Track Finder (OMTF) combines primitives from the DTs, RPCs, and CSCs covering a range of $0.8 < |\eta| < 1.25$. Finally, the Endcap Muon Track Finder (EMTF) combines primitives from the CSCs and RPCs in the region $1.25 < |\eta| < 2.4$.

The Global Trigger (μGT) combines the Calo Trigger Layer 2 and GMT information. Events which pass these triggers are labeled as L1 accept (L1A), and are passed to the HLT. All other events are rejected. A complete schematic of the L1 Trigger may be found in figure Fig. 4.4.

High Level Trigger

The HLT consists of a grid of 1,000 computers tasked with reconstructing physics objects. The HLT has a series of software triggers known as paths. There are several main

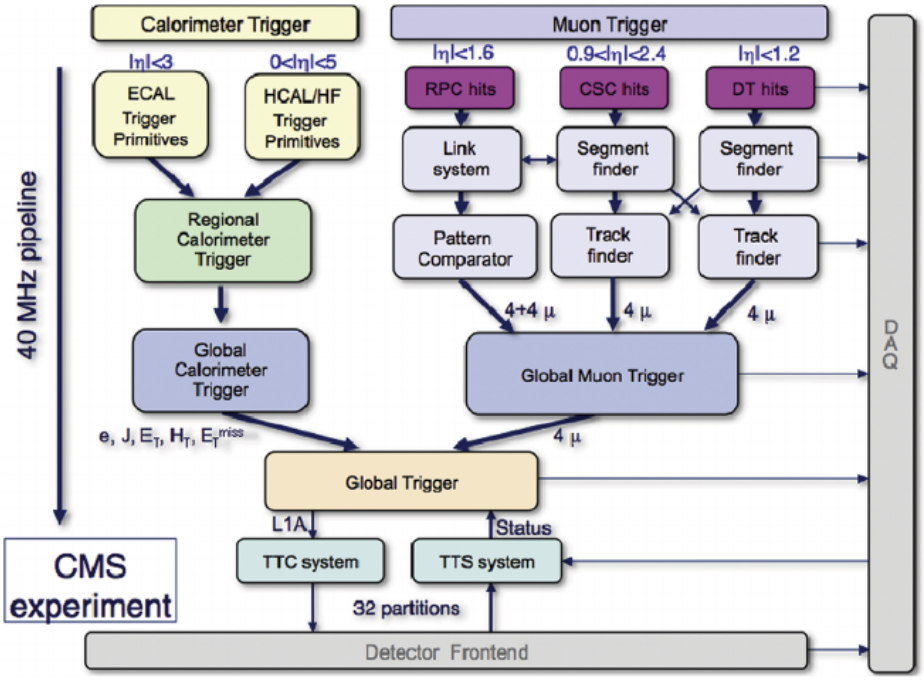

Figure 4.4: A schematic of the CMS L1 Trigger

groups for the paths: electrons/photons and muons, taus, jets, and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$ or MET). To increase the computational efficiency, a path is terminated as soon as a requirement is not met. Each path is split into multiple stages. In the first (actually called Stage 2) calorimeter information is used to determine if the event passes an energy threshold. The next is Stage 2.5, where the calorimeter information is combined with pixel hits. This ensures the tracker tracks line up with the correct energy tower in the ECAL and HCAL. Finally, Stage 3 uses the full detector to reconstruct charged-particle tracks. Since the HLT is a software trigger, the paths may be changed as needed in between runs[†].

[†]A single run is an uninterrupted period of data taking. A fill can consist of many runs.

Chapter 5

Object reconstruction

5.1 The Particle Flow algorithm

Particles in the CMS detector are reconstructed using the PARTICLE FLOW (PF) algorithm [12]. Reconstructed tracks are known as PF candidates or PF tracks. The PF algorithm uses the entire CMS detector; combining information from the tracker, the ECAL, the HCAL, and the muon systems. There are five main types of PF candidates: electrons, photons, charged hadrons, neutral hadrons, and muons. The muons are the easiest to reconstruct as they are the only charged particles which reach the muon systems. Charged hadrons are identified by matching tracks identified in the tracker with energy in the ECAL and HCAL. The PF algorithm cannot distinguish different types of charged hadrons (e.g. $\pi^{\pm}$ or $K^{\pm}$), so further distinction is required per analysis. Electrons are the most difficult to reconstruct as they experience a large amount of bremsstrahlung due to interactions with local electric fields; and a special algorithm was developed to properly identify them. Finally, any remaining energy in the ECAL is associated with photons, or minimum ionizing deposits

from pions, kaons, and muons. Any remaining energy in the HCAL is associated with charged and neutral hadrons, as well as minimum ionizing deposits from muons. The PF algorithm is an iterative process. PF tracks with the largest number of hits are reconstructed first, and then removed before the algorithm moves to the next iteration. A diagram of different tracks passing through the various sub-detector hits can be found in Figure 5.1.

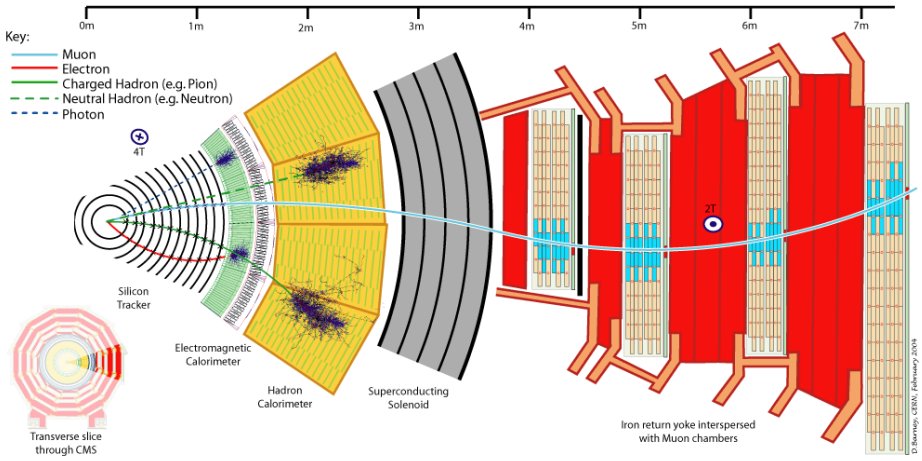

Figure 5.1: An example of tracks in the CMS detector including: electrons, photons, charged hadrons, neutral hadrons, and muons.

5.2 Track and vertex reconstruction

The silicon tracker combined with the 3.8 T magnetic field makes the tracker tracks the most accurately reconstructed objects in the CMS detector. These precise tracks are also used to reconstruct the primary interaction point and some secondary vertices. The reconstruction has in four main steps. The fist step requires tracks to be seeded. Hit information from the inner tracker is projected outward. A seed must consist of either three tracker hits, or two hits combined with the location of the beamspot.

Track candidates are are created by the Combinatorial Track Finder (CTF). Using momentum information from the first few hits, the CTF projects outwards in the helical path a charged particle would travel in the magnetic field. Interactions such as energy loss and multiple scattering are taken into account at this stage. The CTF may either find more tracker hits along the projected path, or flag a missing hit if a one was not found. Matching always begins closest to the interaction point, and hits are projected outward to avoid biasing the tracks.

A Kalman filter [13] is used by the CTF to determine a track probability. The probability decreases with the number of missing hits, and the distance from the track to the observed hits. The filter gives a χ^2 value for each track. Ambiguities may arise from hits corresponding to more than one track. Tracks containing these hits with are rejected based on the χ^2 value. The final tracks are fit again using a Kalman filter, and are smoothed with a least-squared smoother.

5.2.1 Vertex reconstruction

The primary vertex (PV) of an event can be reconstructed using tracker information. Tracks are clustered together to find a common origin. Because the LHC delivers protons in bunches, multiple interactions may happen in a single event. These, along with tracks from other hard collisons within the bunch crossing are known as pile-up (PU), and must be accounted for. A deterministic annealing (DA) algorithm [14] is used to correctly select the PV. To be classified as the PV, a given vertex must have a transverse impact parameter (d_{xy}) less than 2 cm, and a longitudinal impact parameter (d_z) less than 24 cm. The vertex must also contain a minimum of four tracks. If multiple vertices pass these requirements, the one with the largest scalar-p_T sum is selected as the PV.

5.3 Muon reconstruction

5.3.1 Standalone Muons

Muons are independently reconstructed in the moun system. Hits from the DTs and CSCs are used to construct track segments. The algorithm starts at the innermost chamber, and moves outwards. Gradients in the magnetic field, and multiple scattering in the return yoke, must be accounted for. Hits in the RPCs are also matched to track projections. Once a muon candidate is identified, it is labeled as a STANDALONE MUON.

5.3.2 Global Muons

Muons may also be reconstructed using the silicon tracker as stated in Section 5.1. These muons are labeled as Tracker Muons. If a Standalone Muon is matched with a Tracker Muon, it becomes a Global Muon.

5.3.3 Muon corrections

Muons identified by the PF algorithm may be corrected using the Rochester Corrections [15]. The muon momenta (muon momentum scale) are corrected in order to bring the invariant mass of the Z closer to the well established value of 91.2 GeV. The width of the peak (muon momentum resolution) is also corrected.

The Rochester corrections are derived using dimuon resonance decays at the Z pole ($Z^0 \to \mu^+\mu^-$) within a mass window of $70 - 110\,\text{GeV}$. The corrections consists of two stages by comparing a perfectly aligned MC sample with a data sample. In the first stage, the muon momenta are corrected by first using the mean $\left\langle 1/p_T^{\mu} \right\rangle$. The second stge involves correcting the mean invariant mass $\left\langle m_{\mu\mu}^{Z} \right\rangle$ as a function of η and ϕ. The width of the peak (muon momentum resolution) is also corrected to match the resolution in data. This helps remove any bias in the muon momentum due to detector misalignment, software reconstruction, and uncertinites in the magnetic field. Additional additive corrections are derived for μ^+ and μ^- due to misalignment and mis-modeling of the magnetic field.

5.4 Electron and photon reconstruction

Electrons are first reconstructed in the tracker using the standard PF algorithm (Section 5.1). These tracks are matched with ECAL groupings which are known as superclusters. As the electrons interact with the material in the CMS detector, they will radiate bremsstrahlung photons. A superclusters in the ECAL is defined in a specific η and ϕ range coinciding with bremsstrahlung, and other radiation emitted by the electrons. The energy lost by bremsstrahlung radiation is modeled, and the tracks are fit using a Gaussian Sum Filter (GSF). At each bremsstrahlung point a tangent line is projected to the ECAL in order to identify the radiated photons. If photons are found in the projected supercluster, their energy is added to the reconstructed electron. A multivariate estimator is also used to help reduce the level of fake electrons found by the GSF.

Any electromagnetic energy found in the ECAL that is not matched with an electron is usually identified as a photon — photons have no charge, and do not bend in the magnetic field. However, there is a probability of photons interacting with the silicon tracker, and producing electron/positron pairs ($\gamma \rightarrow e^+e^-$). These pairs will leave tracks, and deposit energy in the ECAL. These tracks are typically displaced, and will be properly accounted for by the photon reconstruction algorithm.

5.5 Jet reconstruction

Hadronic particles are clustered together in streams known as jets. Jets contain roughly 65% charged hadrons, 15% neutral hadrons, and 20% photons mainly from neutral meson decay, e.g. $\pi^0 \to \gamma\gamma$. Charged hadrons leave tracks in the silicon tracker, and are reconstructed similar to electrons and muons. Neutral hadrons deposit energy in the HCAL, or the ECAL when they decay electromagnetically. Neutral hadrons are reconstructed similar to photons. It is important to note that the size of the ECAL is roughly the first interaction length (λ_{int}) of the HCAL. This means charge exchange with the ECAL material (e.g. $\pi^+\text{p} \to \pi^0\text{n}$ followed by $\pi^0 \to \gamma\gamma$) can initiate electromagnetic showers. This results in all of the hadronic energy of said particles being measured by the ECAL.

Jets are reconstructed by using the anti-k_{T} algorithm. Charged hadrons identified as PU are removed form the PF candidates; this process is referred to as charged hadron subtraction or CHS. In general, k_{T} algorithms use the distance parameters

$$d_{ij} = \min(k_{\text{T}i}^{2p}, k_{\text{T}j}^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{5.1}$$

$$d_{iB} = k_{\text{T}i}^{2p}, \tag{5.2}$$

where $k_{\text{T}i}$ is the transverse momentum of the i^{th} particle, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, y_i is the rapidity[†] of the i^{th} particle, and R is the maximum width of a cone drawn around the jet. As long as d_{ij} is smaller than d_{iB}, the j^{th} track is added to the candidate jet. Once d_{iB} becomes smaller, the jet is promoted to a PF jet. All tracks in the current jet are removed from the algorithm, so they are not matched to any other jets. A choice of $R = 0.4$ is used

[†]Rapidity and pseudorapidity η are equivalent in the limit $m \to 0$.

to reject particles which are outside of the jet cone. The anti-k_{T} algorithm sets $p = -1$. These jets are identified as AK4 CHS (AK for anti-k_{T}, 4 for $R = 4$ and CHS for charged hadron subtracted). The anti-k_{T} algorithm provides excellent momentum resolution [16]. However, algorithms such as the Cambridge/Aachen algorithm ($p = 0$) have been shown to provide better performance in resolving jet substructure.

Chapter 6

Measuring the shape of the b quark fragmentation function

A good description of the fragmentation functions for each parton type is an essential component in modeling physical processes. The hadronization of parton level objects into final state, or detectable, particles requires a knowledge of the shape of the asociated fragmentation function. Previous studies have measured the parmeters in the Lund–Bowler function:

$$f(z) = \frac{1}{z^{1+r_{\mathrm{b}} \cdot b \cdot m_{\mathrm{b}}^2}} \cdot (1-z)^a \cdot exp(-\frac{b \cdot m_{\mathrm{T}}^2}{z}), \tag{6.1}$$

for light $(\mathrm{u}, \mathrm{d}, \mathrm{s})$ and heavy (c, b) quarks [11] to data from $\mathrm{e}^+\mathrm{e}^-$ colliders. This analysis focuses on the b quark case and the determination of r_{b} from LHC data. This parameter has been measured using $\mathrm{e}^+\mathrm{e}^-$ data from LEP and SLC [17] in a very different environment to $\mathrm{t\bar{t}}$ production at the LHC [18]. As previously stated in Section 2.2, the parameters a and b are treated as universal, and have been measured well with light quark data. At the level of preci-

sion offered by the LHC and the CMS detector, this appears to be a valid assumption. Tuning r_{b} is enough to reproduce the shape of current experimental data for b quark fragmentation.

To measure r_{b} this analysis studies the $\mathrm{t\bar{t}}$ environment produced at the LHC using the CMS detector [19]. Because the CKM matrix element $|V_{\mathrm{tb}}|$ is approximately unity the t quark can be used to provide a sample of b-jets. No measurements exist in this environment. The strategy is to find a reliable way of tagging the b-jets while ensuring a minimum amount of systematic uncertainty from sources such as the detector jet energy calibrations.

The analysis uses data in the lepton + jets (e, μ) and dilepton + jets ($\mathrm{ee}, \mu\mu, \mathrm{e}\mu$) decay chains and focuses on the use of the charmed mesons produced in the b quark fragmentation chain (Fig. 6.1). For this purpose the $\mathrm{D}^0 \to \mathrm{K}\pi$ and $\mathrm{J}/\psi \to \mu\mu$ decays are chosen as these can be fully reconstructed using only the precise momentum measurements from the central tracker. The branching ratio for these decays are quite small, hence the full $35.9\,\mathrm{fb}^{-1}$ of data from 2016 is used. This data is from proton-proton collisions at a center of mass-energy $\sqrt{s} = 13\,\mathrm{TeV}$. The data sample is divided into seven sub-samples labeled as epochs B–H.

The branching ratios for these events are: $BR(\mathrm{W} \to \ell\nu) \cdot BR(\mathrm{b} \to \mathrm{J}/\psi + X) \approx 32.57 \times 10^{-2} \cdot 2.5 \times 10^{-2} \approx 0.81\%$ and $BR(\mathrm{W} \to \ell\nu) \cdot BR(\mathrm{B}^{\pm} \to \mathrm{D}^0 + X) \cdot BR(\mathrm{D}^0 \to \mathrm{K}^{\pm}\pi^{\mp}) \approx 32.57 \times 10^{-2} \cdot 0.79 \cdot 0.039 \approx 1.0\%$. As the $\mathrm{J}/\psi \to \mathrm{e}^+\mathrm{e}^-$ and $\mathrm{W} \to \tau\nu$ are harder to cleanly reconstruct, they are not used in this analysis. This reduces the J/ψ branching ratio by

3.88×10^{-2}, and the D^0 meson branching ratio by 65.5×10^{-2}. The final branching ratio for J/ψ is $\approx 3.2 \times 10^{-4}$, while the final branching ratios for D^0 is $\approx 0.65 \times 10^{-2}$.

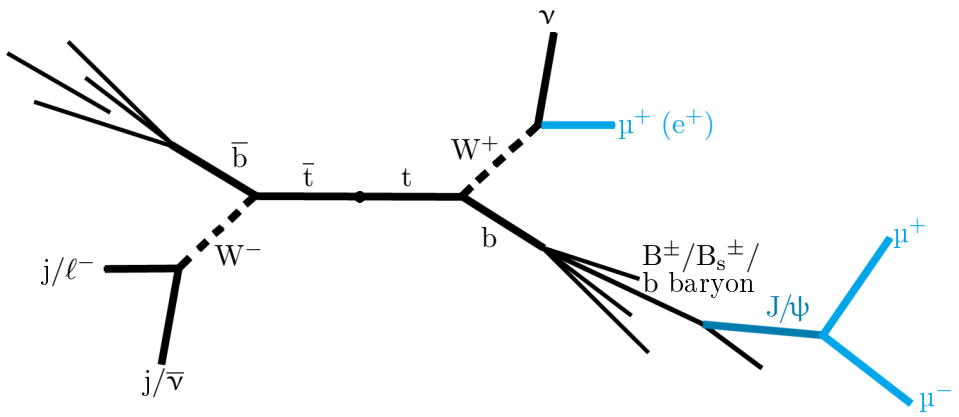

Figure 6.1: Pictorial view of an exclusive J/ψ production in a $t\bar{t}$ system.

6.1 Data and Monte Carlo simulation samples

6.1.1 Datasets

The triggers used to record the Single Muon (μ), Single Electron (e), Double Muon (μμ), Double Electron (ee), and Electron Muon (eμ) are listed in Table 6.1. The triggers with a DZ in the title have an additional constraint on the z position of the primary vertex. These are only used in epoch H because of the high instantaneous luminosity.

The complete 2016 dataset for BCDEFGH is used in this analysis. This has a total integrated luminosity of $35.9\,\text{fb}^{-1}$. An exhaustive list these samples can be found in Ta-

ble 6.2.

Table 6.1: HLT trigger paths for data epochs B–H

Channel	HLT	Epoch
μ	HLT_IsoMu24_v	BCDEFGH
	HLT_IsoTkMu24_v	
e	HLT_Ele32_eta2p1_WPTight_Gsf_v	BCDEFGH
μμ	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v	BCDEFG
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v	
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v	H
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v	
ee	HLT_DoubleEle24_22_eta2p1_WPLoose_Gsf_v	BCDEFGH
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	
eμ	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v	BCDEFG
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v	H
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v	

Table 6.2: Datasets and integrated luminosity

	Integrated luminosity ($\mathbf{fb^{-1}}$)				
Dataset	e	μ	ee	eμ	μμ
/Run2016B-23Sep2016-v3/MINIAOD	5.78	5.78	5.78	5.78	5.78
/Run2016C-23Sep2016-v1/MINIAOD	2.57	2.57	2.57	2.57	2.57
/Run2016D-23Sep2016-v1/MINIAOD	4.25	4.24	4.25	4.24	4.25
/Run2016E-23Sep2016-v1/MINIAOD	4.01	4.01	4.01	4.01	4.01
/Run2016F-23Sep2016-v1/MINIAOD	3.10	3.10	3.10	3.10	3.10
/Run2016G-23Sep2016-v1/MINIAOD	7.54	7.54	7.54	7.54	7.54
/Run2016H-PromptReco-v2/MINIAOD	8.39	8.39	8.39	8.39	8.39
/Run2016H-PromptReco-v3/MINIAOD	0.22	0.22	0.22	0.22	0.22
Total	35.86	35.86	35.86	35.86	35.86

6.1.2 Monte Carlo simulations

Table 6.3 lists the Monte Carlo (MC) simulations used for the signal and background. Signal for the $t\bar{t}$ samples is modeled using POWHEG V2 [20, 21, 22, 23] at next-to-leading order (NLO) with the mass of the t quark set to 172.5 GeV. A next-to-next-to-leading order (NNLO) cross section of 832^{+40}_{-46} pb [24, 25] is used to compute the number of $t\bar{t}$ events. The events are then processed in PYTHIA 8 (v. 8.219) [26] using tune CUETP8M1 [27] for the

underlying event (UE), and tune CUETP8M2T4 [28] for parton shower (PS) modeling. The parton distribution functions (PDF) used in POWHEG are the NLO set NNPDF30 [29]. The CMS detector is modeled using GEANT4 v9.4 [30].

The background processes are modeled using matrix element (ME) calculations from POWHEG and MADGRAPH5_aMC@NLO [31], while the parton shower and fragmentation are modeled using PYTHIA 8. The processes for W boson and Drell–Yan (DY) production are produced using MADGRAPH5_aMC@NLO and PYTHIA using the MLM [32] matching scheme. The simulation of single top quarks using the NLO cross section from HATHOR v2.1 [33] is modeled through the t-channel production using MADSPIN [34] and tW produciont is modeled using POWHEG. Multi-boson events are modeled using the NLO cross sections given by MCFM [31] where ZZ and WZ events are simulated using PYTHIA and WW events are simulated using POWHEG.

The lepton selection efficiencies and the trigger efficiencies in MC are corrected using scale factors to better reproduce the data. This includes μ and e identification, μ isolation, and e reconstruction. The statistical uncertainties are centrally provided. The recommendation for systematic uncertainties is an additional 1% uncertainty for the μ identification, 0.5% uncertainty for the μ isolation, 0.5% for the single μ triggers, and 1% for the e reconstruction (only for e $p_{\mathrm{T}} < 20\,\mathrm{GeV}$ and e $p_{\mathrm{T}} > 80\,\mathrm{GeV}$).

Table 6.3: MC samples

Process	Dataset	σ(pb)
$t\bar{t}$	/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	832
	/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	/TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
	/TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529
Di-Boson	/ZZ_TuneCUETP8M1_13TeV-pythia8	16.523
	/WWToLNuQQ_13TeV-powheg	49.997
	/WWTo2L2Nu_13TeV-powheg	12.178
	/WZ_TuneCUETP8M1_13TeV-pythia8	47.13
$W \rightarrow l\nu$	/W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9493.0
	/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3120.0
	/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	942.3
	/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	524.2
$Z \rightarrow ll$	/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	16270.0
	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4963.0
Single t	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	44.33
	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	26.38
	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.85
	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.85

6.2 Object selection

A clean $t\bar{t}$ environment is esseintial to this analaysis. Therefore, the data is split into five analysis channels: single muon with jets (μ+ jets), single electron with jets (e+ jets), double muon ($\mu\mu$), double electron (ee), and electron muon (eμ). A Kalman Vertex Fitter [35] is used to reconstruct charmed mesons (see Section 6.3). All jets containing at lest one charmed-meson candidate are flagged by the filter. The use of the Kalman Filter allows for a more relaxed jet criteria than is usually used for the CMS b-taggers, thus increasing the number of candidate jets. The momentum scale and resolution of all of the isolated muons have been corrected with the Rochester corrections method [15] as detailed in Section 5.3.3.

All particles are reconstrucitoned using the CMS Particle Flow (PF) algorithm [36]. Jets are clustered using the anti-k_{T} algorithm with a size peramteter of 0.4 [16, 37]. The PF tracks are limited to either the tracks used in the primary vertex (PV) fit, or the track that is the closest in the $z-$axis to the primary vertex which is not associated with another primary vertex. This helps remove pile-up (PU) tracks.

6.2.1 Isolated muons (μ+jets)

The muons that are not clustered in a jet must have a $p_{\mathrm{T}} > 26$ GeV, $|\eta| < 2.4$, and a relative isolation I_{rel} <0.15, where

$$I_{\mathrm{rel}} = \frac{\Sigma p_{\mathrm{T}}(\text{charged hadrons from PV})}{p_{\mathrm{T}}} + \frac{\max[0, \Sigma E_{\mathrm{T}}(\text{neutral hadrons}) + \Sigma E_{\mathrm{T}}(\gamma) - 0.5 * \Sigma p_{\mathrm{T}}(\text{charged hadrons from PU})]}{p_{\mathrm{T}}} \tag{6.2}$$

in a ΔR=0.4 cone around the muon candidate. All muon candidates must fulfill the following requirements: each must be identified by the PF algorithm as Global Muon, a global track normalized $\chi^2 < 10$, at least one muon chamber hit by definition (the muon system is required to tag a track as a muon), a muon segment in at least two muon stations, a transverse impact parameter $d_{xy} < 2\,\mathrm{mm}$ with respect to the primary vertex, a longitudinal impact parameter $d_z < 5\,\mathrm{mm}$ with respect to the primary vertex, at least one hit in the pixel detector, and must be identified by at least six tracker layers. Events are ignored if any additional muons have a $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, $|\eta| < 2.4$, and I_{rel} <0.24.

Single muon events must contain at least one jet with p_{T} >30 GeV and at least one additional jet flagged by the Kalman Filter.

6.2.2 Isolated electrons (e+jets)

The electrons not clustered in a jet must have a $p_{\mathrm{T}} > 35\,\mathrm{GeV}$ and $|\eta| < 2.1$ where superclusters located in the transition region $|\eta| = 1.479$ between the central and forward regions are excluded. Additional criteria for electron candidates are as follows: the distance between a match in the ECAL and the extrapolated electron track must have a $\Delta\eta < 3.08 \times 10^{-3}$ in central region and $\Delta\eta < 6.04 \times 10^{-3}$ in the forward region along with a $\Delta\phi < 8.16 \times 10^{-2}$ in the central region and $\Delta\phi < 3.94 \times 10^{-2}$ in the forward region, the ratio of the energy H associated with the electron in the HCAL divided by the energy E in the ECAL must be less than 4.14×10^{-2} in the central region and less than 6.41×10^{-2} in the forward region, the absolute difference between the inverse of the energy from the ECAL and the inverse of the momentum p of the electron track candidate is less than $12.9\,\mathrm{MeV}^{-1}$, up to one missi ng hit to suppress $\gamma \to \mathrm{e}^+\mathrm{e}^-$, and the electron shower in the ECAL is restricted to $\sigma_{\eta\eta} < 9.98 \times 10^{-3}$ in the central region and $\sigma_{\eta\eta} < 2.92 \times 10^{-2}$ in the endcap, where

$$\sigma_{\eta\eta} = \sqrt{\Sigma(\eta_i - \overline{\eta}^2 w_i / \Sigma w_i)}, \tag{6.3}$$

where the sum is taken over the five by five grid of crystals around the highest E_{T} crystal, η_i is measured in terms of unit crystals, $\overline{\eta}$ is the weighted average in η of the energy of the shower, and w_i are the weights which depend on $\log E$ in each crystal. Electrons must also have a relative isolation parameter $I^{\mathrm{e}}_{\mathrm{rel}} < 0.0588$ in the central region and $I^{\mathrm{e}}_{\mathrm{rel}} < 0.0571$ in the forward region where $I^{\mathrm{e}}_{\mathrm{rel}}$ is defined similar to the I_{rel} used for muons, with a cone of $\Delta R < 0.3$ around the electron candidate. The pile-up is estimated as $A_{\mathrm{eff}} \cdot \rho$ where A_{eff} is an η dependent area and ρ is the median of the transverse energy density in a $\delta\eta \times \delta\phi$ region which is found using the tracks of charged particles associated with pile-up. Events

are ignored if any additional electrons have: $\sigma_{\eta\eta} < 1.15 \times 10^{-2}$ in the central region and 3.7×10^{-2} in the forward region, $\Delta\eta < 7.49 \times 10^{-3}$ in the central region and 8.95×10^{-3} in the forward region, $\Delta\phi < 2.28 \times 10^{-1}$ in the central region and 2.13×10^{-1} in the forward region, $H/E < 35.6\%$ in the central region and 21.1% in the forward region, $I^{e}_{rel} < 0.175$ in the central region and 0.159 in the forward region, $|1/E - 1/p| < 299\,\text{MeV}^{-1}$ in the central region and $|1/E - 1/p| < 150\,\text{MeV}^{-1}$ in the forward region, and up to two missing hits in the central region and three in the forward region. These selection requirements were produced using the Toolkit for Multivariate Data Analysis (TMVA) [38] within the ROOT framework [39], producing an average selection efficiency of 70%. Electrons from Drell–Yan ($Z/\gamma^* \to e^+e^-$) (produced in MADGRAPH using MLM) and $t\bar{t}$ (produced in MADGRAPH5_aMC@NLO) samples were used in the MVA. Reconstructed electrons in the Drell–Yan which were kinematically matched to generator level PROMPT electrons were treated as signal, while unmatched electrons, electrons matched to NON-PROMPT generator level electrons, and electrons from $t\bar{t}$ decays were treated as background.

Single electron events must contain at least one jet with a p_T >30 GeV and at least one additional jet flagged by the Kalman Filter.

6.2.3 Dileptons (μμ, eμ, and ee)

Events with two leptons not clustered in jets have slightly looser criteria because the di-lepton channels are much cleaner. The muons must have a $p_T > 20$ GeV and $|\eta| < 2.4$, while the electrons must have a $p_T > 30$ GeV and $|\eta| < 2.1$ where superclusters located in the transition region $|\eta| = 1.479$ are still excluded. All other conditions listed in Section 6.2.1

and Section 6.2.2 are applied to the muons and electrons respectively. Resonance decays from the Z boson are suppressed by restricting $|M_{\mathrm{Z}} - M_{\mathrm{ll}}| > 15\,\mathrm{GeV}$. Low mass resonance backgrounds are suppressed by requiring $M_{\mathrm{ll}} > 20\,\mathrm{GeV}$. Drell–Yan processes involving same flavor di-lepton events are suppressed by requiring $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\mathrm{GeV}$.

Di-lepton events must have at least one jet flagged by the Kalman Filter. No additional jets are requited due to the di-lepton channels being much cleaner than the lepton + jets channels.

6.2.4 Jets

All jets must be defined by the anti-k_{T} algorithm [16, 37] with a $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta|$<2.4. Jets must pass the following criteria: the fraction of energy in the HCAL from neutral particles must be less than 99% and the fraction from charged particles must be greater than zero, the fraction of energy in the ECAL from neutral particles must be less than 99% and the fraction from charged particles must be greater than zero, jets must have at least two constitute particles, and jets must have a charge multiplicity greater than zero.

Jet cleaning is also preformed. This ensures no jets overlap with the isolated leptons within a cone of ΔR=0.4. Some control plots for the combined channels can be seen in Figure 6.2.

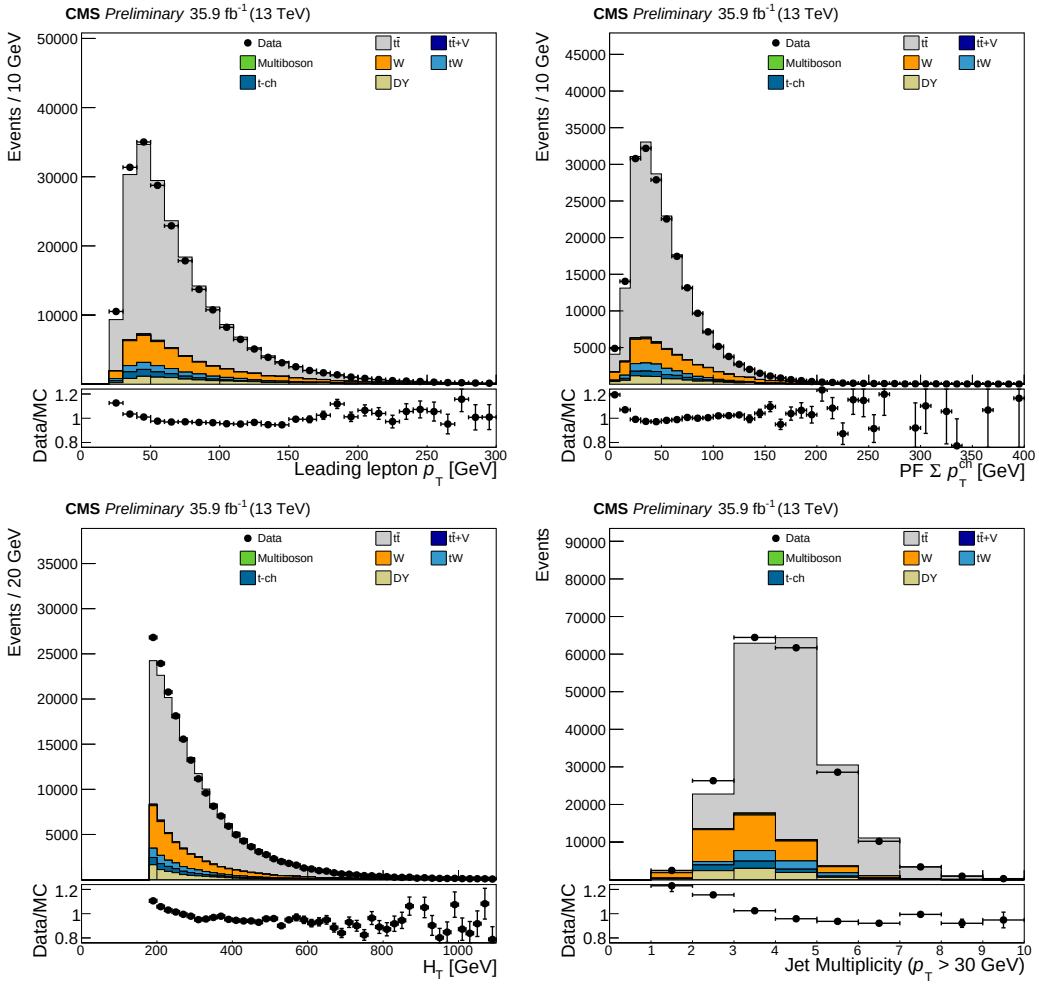

Figure 6.2: (D^0 selection) All channels isolated lepton p_T (top left) all channels jet $\sum p_T^{ch}$ (top right) all channels H_T (bottom left) and all channels jet multiplicity (bottom right)

6.3 Charmed mesons

The charmed mesons produced from the b quark fragmentation provide a relatively clean study sample. A Kalman Vertex Fit [35] is used to fit the vertices of the J/ψ and D^0 candidates to help reduce the combinatorial background. All tracks are required to have a large number of observed detector hits, and a small number of missed hits in the CTF. All muons are required to be GLOBAL MUONS.

A Kalman filter is an iterative algorithm used for combining experimental data points with theoretical calculations. A term known as the Kalman gain is calculated using the experimental uncertainty

$$K_k = P'_k H^T (H P'_k H^T + R)^{-1}, \tag{6.4}$$

where K is the Kalman gain, P' is the a posteriori error covariance matrix, H is the the observation model, and R is the the covariance of the observation noise. The algorithm produces a new state

$$\hat{x}_k = \hat{x}'_k + K_k(z_k - H\hat{x}'_k), \tag{6.5}$$

where $\hat{x}_k$ is the new estimate, $\hat{x}'_k$ is the prior estimate, and z_k is the experimental measurement. The a posteriori matrix is then updated using

$$P = (I - K_k H)P'_k, \tag{6.6}$$

where I is the unit matrix, and the next state in the iteration is calculated as

$$\hat{x}'_{k+1} = \Phi P_k \Phi^T + Q, \tag{6.7}$$

where Φ is the state transition model to go from state $\hat{x}_k$ to state $\hat{x}_{k+1}$ and Q is the the covariance of the process noise. When the Kalman gain is high, the algorithm will output results closer to the experimentally measured points, while if the gain is low the algorithm will output results closer to the theoretical predictions. A χ^2 merit can also be calculated using

$$\chi^2 = [z - h(a,x)]R^{-1}[z - h(a,x)]^T, \tag{6.8}$$

where h is the data model (assumed linear in the parameter a).

The Kalman filter can be used to correct the helical paths used fit detector hits in tracks. A high Kalman gain (low experimental uncertainty) results in the algorithm using the hits in the tracker, and a low Kalman gain (high experimental uncertainty) results in the algorithm using the theoretical helical path. Once a precise track is fit, the parameters such as d_{xy}, η, and ϕ are recomputed, and a goodness-of-fit for the vertex is calculated as a χ^2 value.

J/ψ reconstruction To help isolate J/ψ decays all events must have a scalar sum of the p_{T} of the jets (H_{T}) which is greater than 80 GeV. All soft muons used to reconstruct a J/ψ must have a $p_{\mathrm{T}} > 3$ GeV. High p_{T} isolated muons cannot be used in any events triggered with a muon (μ+jets, μμ, and eμ). The invariant mass of the J/ψ candidates is shown in the left panel of Figure 6.3. This plot shows how clean the J/ψ channel is in spite of the low statistics. Any disagreement between Data and MC is attributed to statistical fluctuations in the background samples.

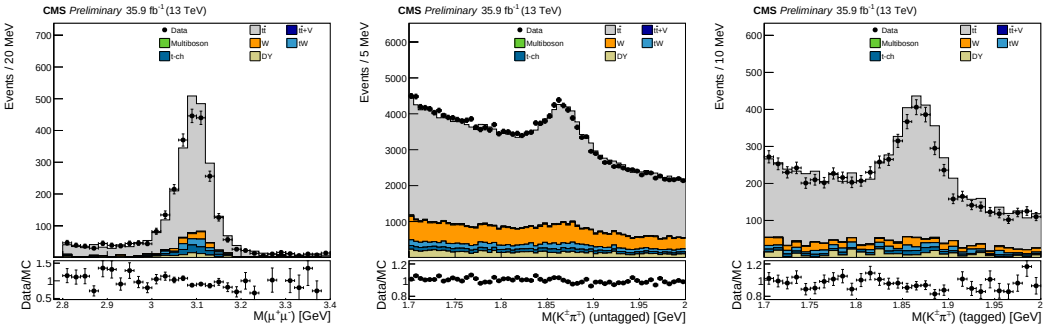

Figure 6.3: J/ψ mass (left) D^0 mass (middle), and $D^0{}_\mu$ mass (right)

D^0 reconstruction To help isolate D^0 decays all events must have an H_T >180 GeV. The silicon tracker cannot distinguish a track produced from a π from one produced from a K, so the mass hypothesis for both tracks is tested. Only track pairs with an invariant mass 1.7 GeV $< m_{K\pi} <$ 2.0 GeV are kept. All π and K used to reconstruct a D^0 must have a $p_T > 5$ GeV and a $p_T > 1$ GeV respectively. Due to the high combinatorics in this sample, a single mismodeled track can create a large discrepancy between Data and MC. Additional studies were done on the signal and background (see Section 6.4) of the D^0 events (Fig. 6.4). The quantities used to measure the fragmentation (proxies) are less sensitive to the b quark fragmentation when the scalar sum of the p_T of charged particles in the jet is large, and most of the D^0 disagreement is due to the background when this scalar sum is large. An optimal cut on the scalar sum of the p_T of charged particles is below 100 GeV (see Fig. 6.4). This was imposed on the un-tagged D^0 sample only. The π and K for the un-tagged D^0 must also have an $|\eta|$ <1.5 as the large η particles were not modeled well in MC. The invariant mass of the D^0 candidates is in the middle of Figure 6.3.

Flavor-tagging In addition to reconstructing the D^0, a SOFT MUON* can also be required. This allows for flavor-tagging, in which the charge of the parent meson is followed to ensure the correct mass assumptions for the $\pi^{\pm}$ and $K^{\mp}$ (see Table 6.4). The SOFT MUON must have a minimum of $p_T > 1$ GeV to reach the outer muon systems. The flavor-tagged D^0 is denoted as $D^0{}_{\mu}$. Flavor tagging significantly reduces the combinitorial background in D^0 events as can be seen in the right panel of Figure 6.3. This reduction allows the restrictions on the scalar sum of charged particle p_T, and on the η of the π and K to be removed.

Table 6.4: Flavor tagging

Decay chain						Final Charged Products
$B^{\pm}$	$\rightarrow$	D^0	$\rightarrow$	$K^{\pm}$		$K^{\pm}\pi^{\mp} + \mu^{\pm}$
$u/\overline{u}$	$\rightarrow$	$u/\overline{u}$	$\rightarrow$	$u/\overline{u}$		$K^{\pm}$
$\overline{b}/b$	$\rightarrow$	$\overline{c}/c$	$\rightarrow$	$\overline{s}/s$		
	$\downarrow$		$W^{\mp}$	$\rightarrow$	$\pi^{\mp}$	$\pi^{\mp}$
	$W^{\pm}$	$\rightarrow$	$\mu^{\pm} + \nu$			$\mu^{\pm}$

Additional criteria for the fitted vetices of all meson candidates are as follows: a vertex fit $\chi^2 < 5$ to remove combinatorial background, meson candidate time-of-flight significance $(c\tau/\sigma_{c\tau}) > 10$ to remove PROMPT mesons (mesons associated with the PV, not the b quark hadronization), and an uncertainty on the meson candidate time-of-flight of

*A muon with low p_T.

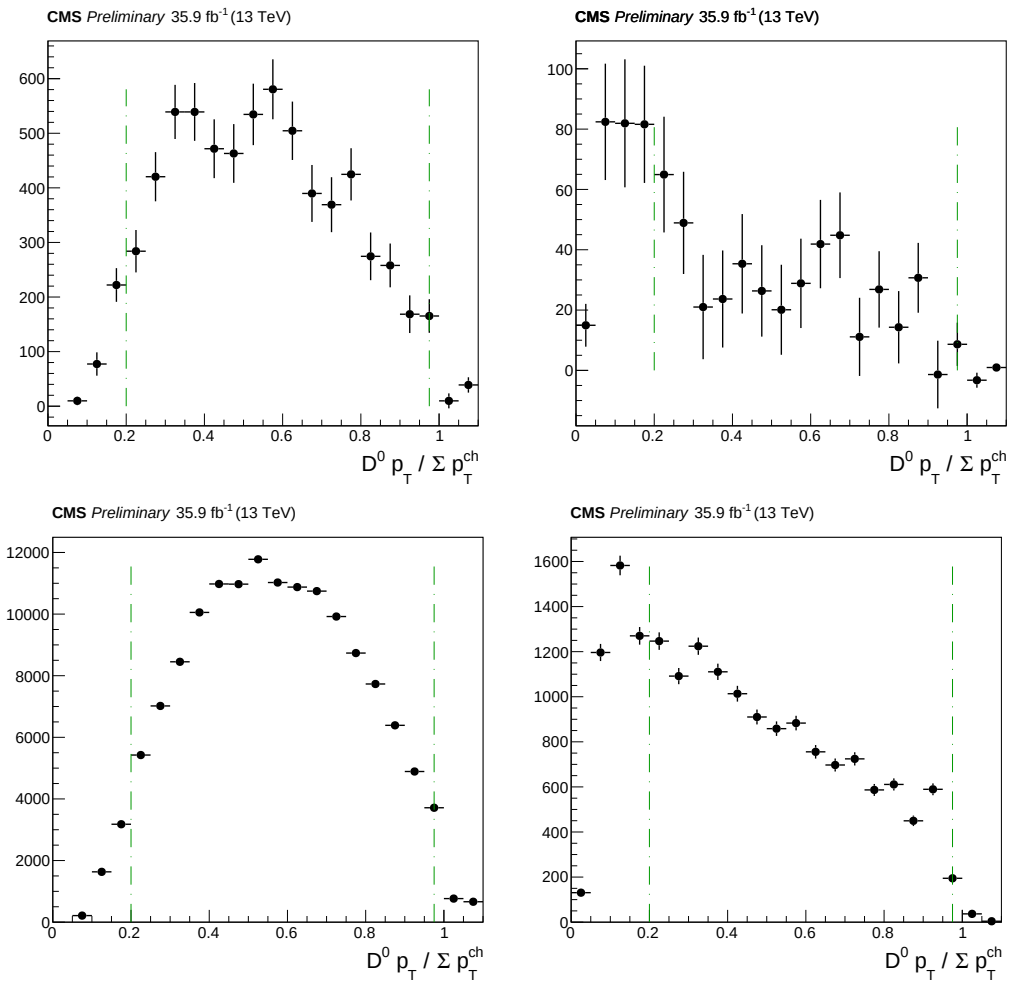

Figure 6.4: Signal (top) and background (bottom) $\sum p_{\mathrm{T}}^{\mathrm{ch}} < 100$ (left) and $\sum p_{\mathrm{T}}^{\mathrm{ch}} > 100$ (right)

$\sigma_{c\tau} > 2\times10^{-4}$ to remove residual PROMPT background and miss-reconstructed events. The number of meson candidates for each channel are listed in Table 6.5.

Table 6.5: Number of charmed meson candidates

Sample	J/ψ Events	D^0 Events	D^0_μ Events
Before Background subtraction			
Data	2768	199901	6521
MC	2772	199993	6847
Data/MC (%)	96	100	95
After Background subtraction			
Data	1948	6937	1308
MC	2036	6611	1403
Data/MC (%)	96	105	93

6.4 Study of b quark fragmentation function

The b quark fragmentation function is modeled by CMS using MC with a Lund–Bowler function [6] as show in Equation (6.1). The standard values in PYTHIA 8 are: $a = 0.68$, $b = 0.98\,\mathrm{GeV}^{-2}$, $r_\mathrm{b} = 0.855$, and $m_\mathrm{b} = 4.78\,\mathrm{GeV}$ (the b quark pole mass) [11]. In previous analyses the b quark fragmetnation function shape parameter r_b was measured by fully reconstructing the B mesons. These mesons were created in $\mathrm{e^+e^-}$ collision at the Z pole ($\mathrm{Z^0 \to b\bar{b}}$) [7, 8, 9, 10]. A best measurement of $r_\mathrm{b} = 0.8949^{+0.1841}_{-0.1968}$ in PYTHIA 8 was obtained. In order to measure the b-fragmentation in this analysis, the charmed mesons are used as a proxy for the parent b quark. The ratio of the p_T the charmed meson divided by the scalar sum of the p_T of all the charged tracks in the jet

$$\frac{\text{charmed meson } p_\mathrm{T}}{\sum p_\mathrm{T}^\mathrm{ch}} \tag{6.9}$$

is used to infer the fragmentation of the b quark producing the charmed mesons. Restricting the analysis to charged particles makes use of the inner tracker, and bypasses the systematics associated with the jet energy scale and jet energy resolution. The proxy method also provides increased statistics over fully reconstructing the B mesons. The samples are split into two separate data epochs, BCDEF and GH, due to issues with the silicon tracker in 2016 (see Appendix A), and are recombined after background subtraction is performed (see Background subtraction). The flavor-tagged $\mathrm{D^0}$ combined with the muon used for tagging gives the closest kinematics to the b quark (Fig. 6.5)

$$\frac{p_\mathrm{T}(\mathrm{D^0_\mu} + \mu)}{\sum p_\mathrm{T}^\mathrm{ch}}. \tag{6.10}$$

This is because only the ν and SOFT tracks from the fragmentation are missing. The final ratio of data divided by MC for all three samples can be found in Figures 6.8 and 6.9. The un-tagged D^0 sample has much larger statistics, making it just as valuable as a proxy for the b-fragmentation function. The J/ψ sample has the lowest statistics, but the highest purity, making it an important sample as a cross check.

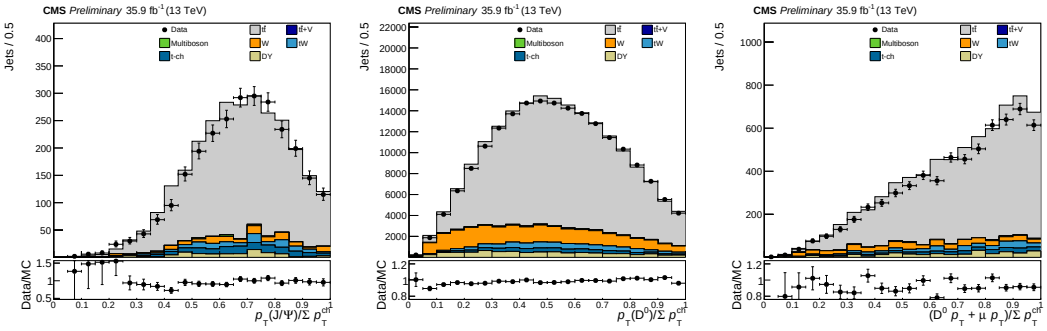

Figure 6.5: J/ψ p_{T} divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$, D^0 p_{T} divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$, and ($\mathrm{D}^0_{\mu}$ p_{T} + μ p_{T}) divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$

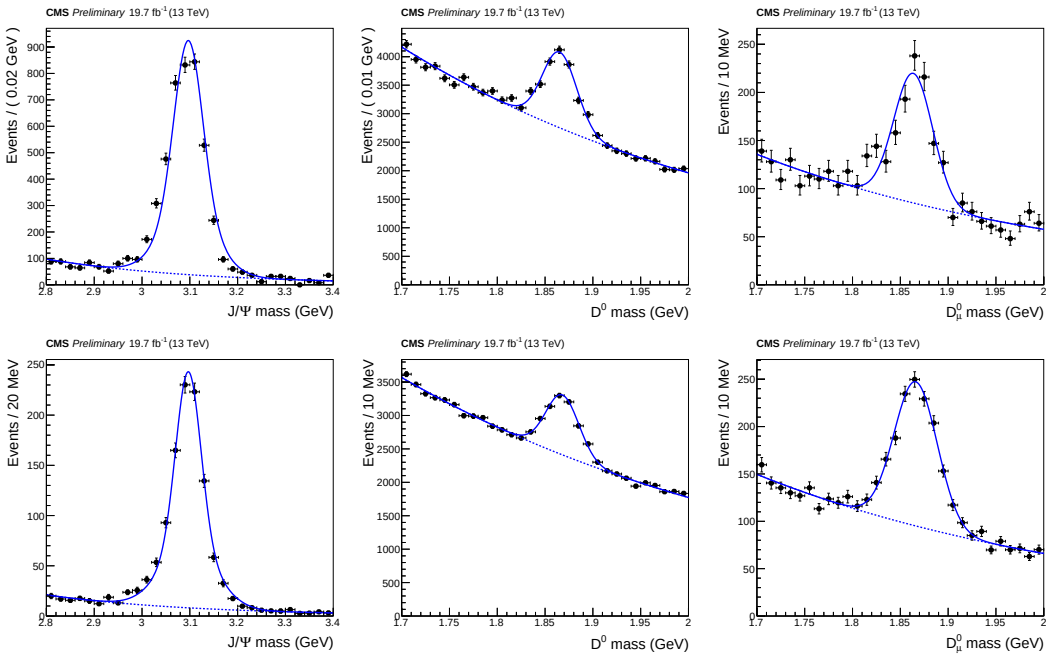

Figure 6.6: Mass fit for the J/ψ (left) D^0 (center) and D^0_μ samples (right) for epochs B–F in the Data (top) and MC (bottom)

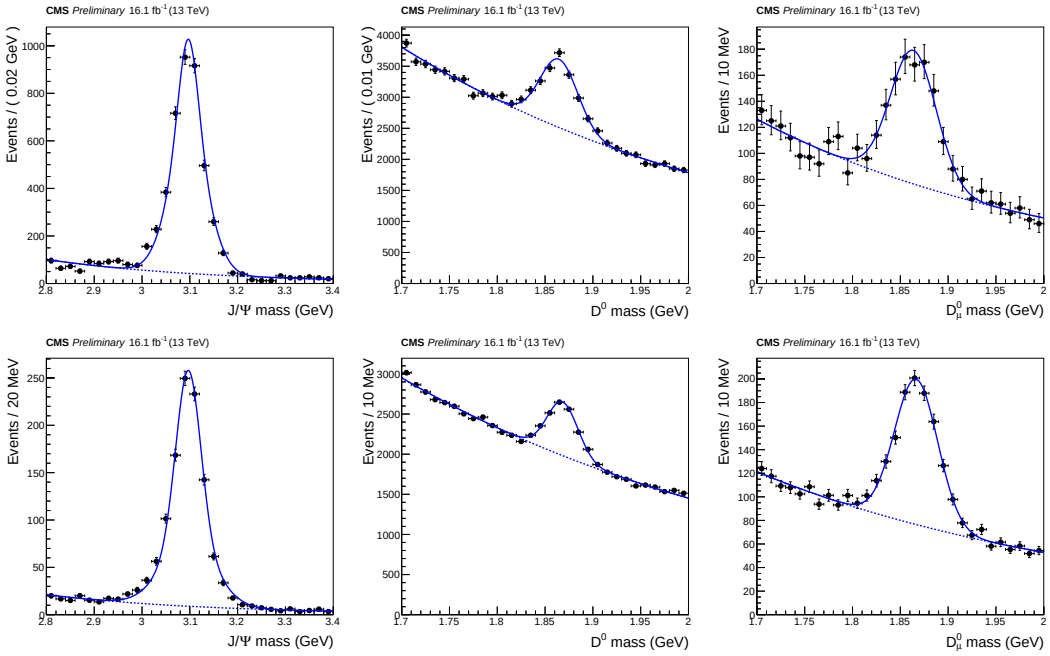

Figure 6.7: Mass fit for the J/ψ (left) D^0 (center) and D^0_μ samples (right) for epochs GH in the Data (top) and MC (bottom)

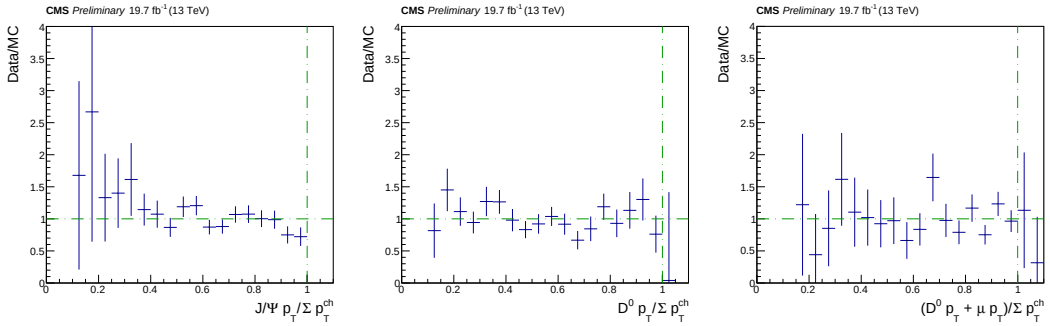

Figure 6.8: Ratio of J/ψ p_T divided by $\sum p_\mathrm{T}^\mathrm{ch}$ (left), D^0 p_T divided by $\sum p_\mathrm{T}^\mathrm{ch}$ (center), and ratio of ($D^0_\mu\, p_\mathrm{T} + \mu\, p_\mathrm{T}$) divided by $\sum p_\mathrm{T}^\mathrm{ch}$ (right) for data epochs BCDEF

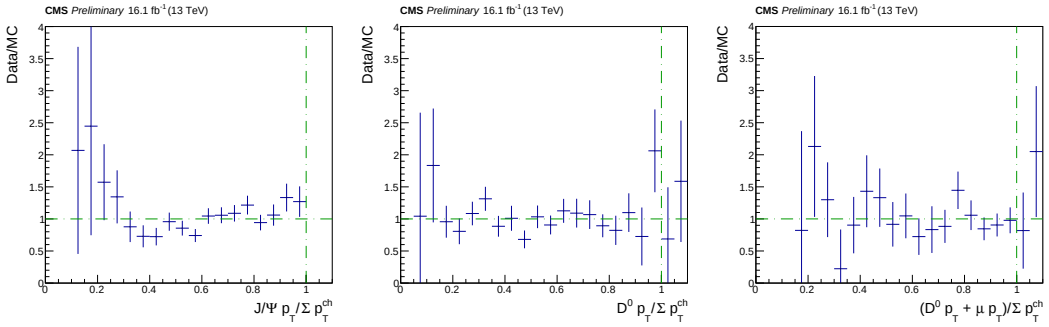

Figure 6.9: Ratio of J/ψ p_{T} divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ (left), D^0 p_{T} divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ (center), and ratio of $(\mathrm{D}^0_{\mu}\, p_{\mathrm{T}} + \mu\, p_{\mathrm{T}})$ divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ (right) for data epochs GH

Background subtraction In order to get a clean sample, we must remove as much background as possible. First a narrow window is selected around the mass peak. For the J/ψ sample a window of the PARTICLE DATA GROUP (PDG) [1] mass (3.097 GeV) ±110 MeV is selected around the mass peak. For the un-tagged D^0 sample a window of the PDG mass (1.864 GeV) ±40 MeV is selected. The D^0_μ has a window of ±45 MeV due to a small difference in resolution. The data and MC are background subtracted using SPLOT [40]. This method uses a minimized negative log-likelihood (NLL) fit ($-\mathcal{L}$) of the invariant mass to produce event weights known as sWeights. These weights are used to distinguish the signal and the background. The fit is obtained using a Gaussian for the peak, and an exponential decay for the background (Figs. 6.6 and 6.7). The J/ψ peak is the sum of two Gaussians. The SPLOT technique first computes the inverse of the variance matrix from the NLL fit

$$V_{nj}^{-1} = \frac{\partial(-\mathcal{L})}{\partial N_n \partial N_j} = \sum_{e=1}^{n} \frac{f_n(y_e) f_j(y_e)}{(\sum_{k=1}^{N_s} N_k f_k(y_e))^2}, \tag{6.11}$$

where N_s is the the event yield for each species i (e.g. signal or background), N_i is the number of events expected on the average for the i^{th} species, and $f_i(y_e)$ is the probability density function (PDF) evaluated for event e for the discriminating variable y_e. The so called sWeights ${}_s\mathcal{P}_n(y_e)$ for a particular event and estimator y_e are give by

$${}_s\mathcal{P}_n(y_e) = \frac{\sum_{j=1}^{N_s} V_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}. \tag{6.12}$$

The statistical uncertainty can be obtained from the relation

$$\sigma^2 = \sum_{e \subset \delta x} ({}_s\mathcal{P}_n(y_e))^2, \tag{6.13}$$

where the sum runs over all events e in the bin centered at $\overline{x}$ with a bin width of δx.

Fragmentation templates Performing a measurement of r_{b} requires templates for different tune parameters. Alternate MC samples are produced at generator level for various values of r_{b}: 0.655, 0.700, 0.725, 0.755, 0.775, 0.800, 0.825, 0.875, 0.900, 0.925, 0.955, 0.975, 1.00, and 1.055. The ratio of

$$x_{\mathrm{B}} = \frac{\text{B meson } p_{\mathrm{T}}}{\text{b-jet } p_{\mathrm{T}}} \tag{6.14}$$

is computed for each tune, and an event weight is produced by dividing this value by the nominal x_{B} value (Fig. 6.10). These weights are then used to alter the shape of the MC at reconstruction level. A subset of the re-weighted proxies after background subtraction are

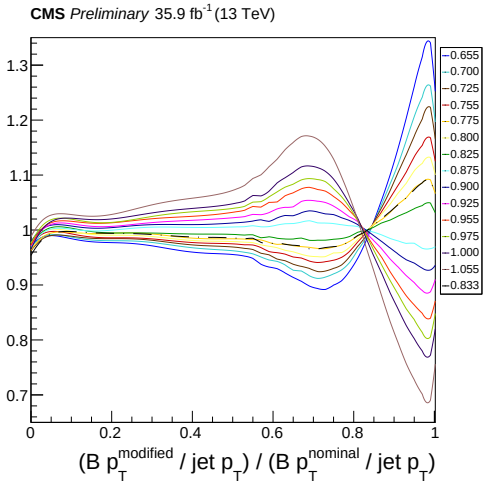

Figure 6.10: Event weights for various r_{b} values (combined measurement value in dash-dotted line)

shown in Figure 6.11. The weights are ignored in the region of $x_{\mathrm{B}} < 0.2$ as the sensitivity is very low, and in the region of $x_{\mathrm{B}} > 0.975$ as these are solely due to resolution effects.

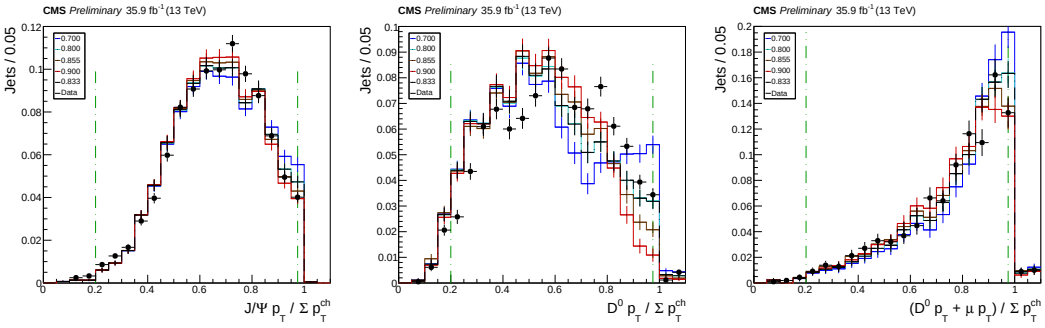

Figure 6.11: $J/\psi, D^0$, and D^0_μ proxies for $r_b = 0.700$, $r_b = 0.800$, $r_b = 0.855$, $r_b = 0.900$, and the combined measurement value (dash-dotted line) compared with data

χ^2 fit The data is used to find a best-fit value for r_b. A χ^2 goodness-of-fit is performed on each of the generated templates with respect to the data. The χ^2 definition for two weighted histograms is used. This is because sPlot weights events in both data and MC with a probability of being signal. The χ^2 test is defined as

$$\chi^2 = \sum_{i=0}^{N} \frac{(W_{\text{Data}} w_{\text{MC}_i} - W_{\text{MC}} w_{\text{Data}_i})^2}{W^2_{\text{Data}} \sigma^2_{\text{MC}_i} + W^2_{\text{MC}} \sigma^2_{\text{Data}_i}}, \tag{6.15}$$

where N is the number of bins, W_{Data} is the number of events in the data, W_{MC} is the number of events in MC, w_{Data_i} is the number of events in the data in bin i, w_{MC_i} is the number of events in MC in bin i, σ_{Data_i} is the error on bin i in the data, and σ_{MC_i} is the error on bin i in MC. The results of the χ^2 scan are plotted as a function of the fragmentation parameter r_b used to generate the templates. This plot is then fit with a 3rd order polynomial, and the minimum is extracted (Fig. 6.12). A 3rd order polynomial is needed due to the asymmetry of the χ^2 curves. The best measurement of the r_b values are $r_b = 0.874 \pm 0.045\,(\text{stat})$, $r_b = 0.836 \pm 0.017\,(\text{stat})$, and $r_b = 0.846 \pm 0.036\,(\text{stat})$for the

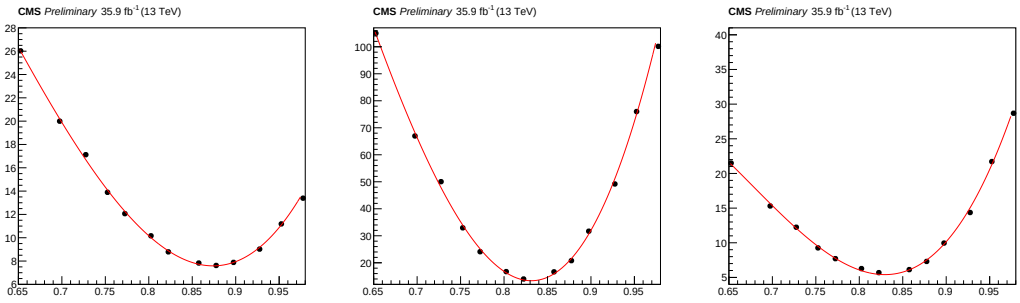

Figure 6.12: χ^2 goodness-of-fit vs. r_b for J/ψ (left), D^0 (center), and D^0_μ (right). The number of degrees of freedom (n.d.f) are 12, 16, and 14 for the J/ψ, D^0, and D^0_μ respectively.

J/ψ, D^0, and D^0_μ samples respectively. The statistical uncertainty is the difference between the minimum r_b value and the r_b value at $\chi^2_{min} + 1$ for each measurement.

6.4.1 Kinematics cross checks

The kinematic plots for the measured r_b values are compared to the nominal samples for each charmed meson (Figs. 6.13 to 6.15.). This shows the nominal MC is consistent with the measured fragmentation function parameter r_b. No artifacts were introduced by the re-weighting the nominal MC.

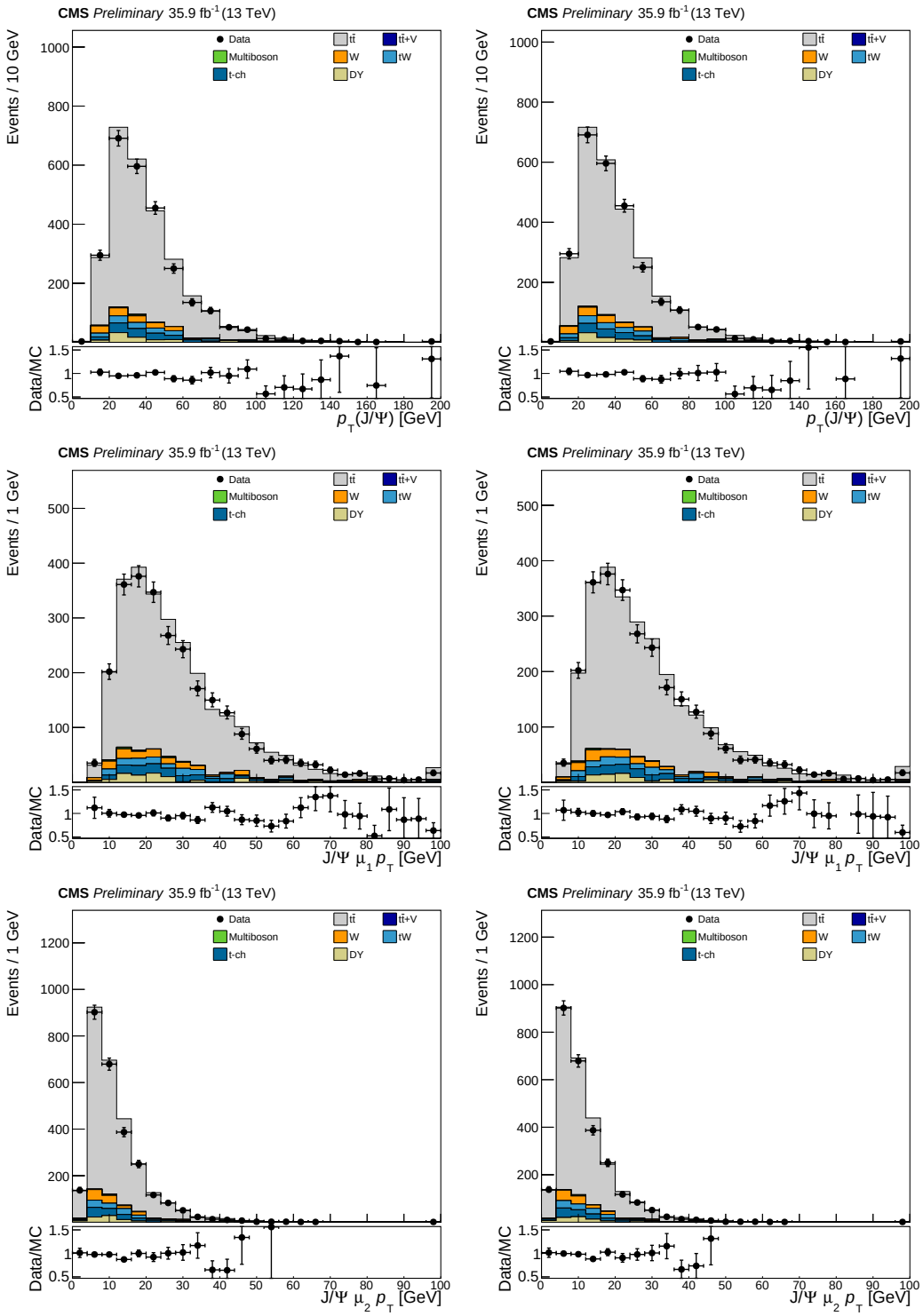

Figure 6.13: J/ψ nominal (left) and measured (right)

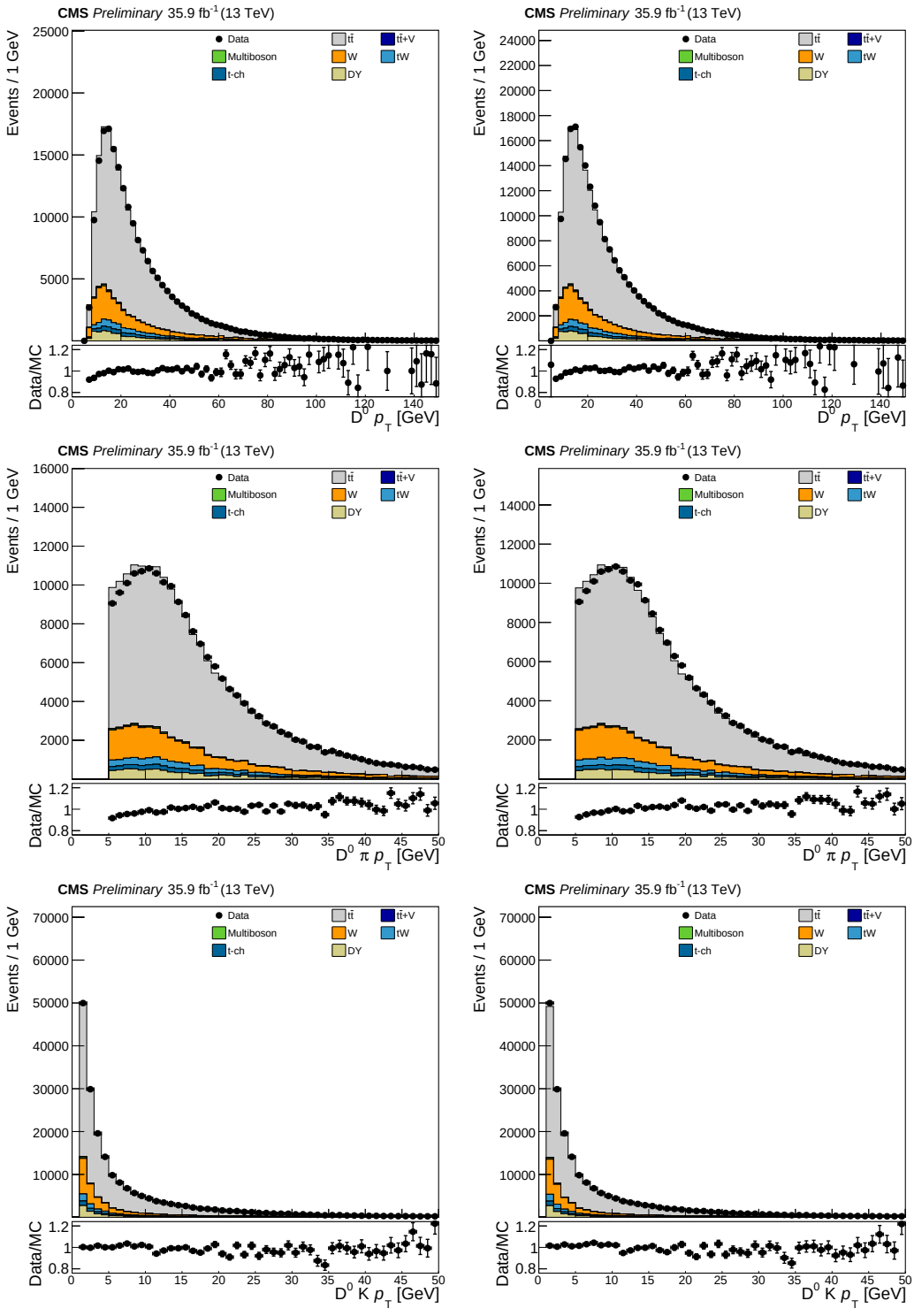

Figure 6.14: D^0 nominal (left) and fitted (right)

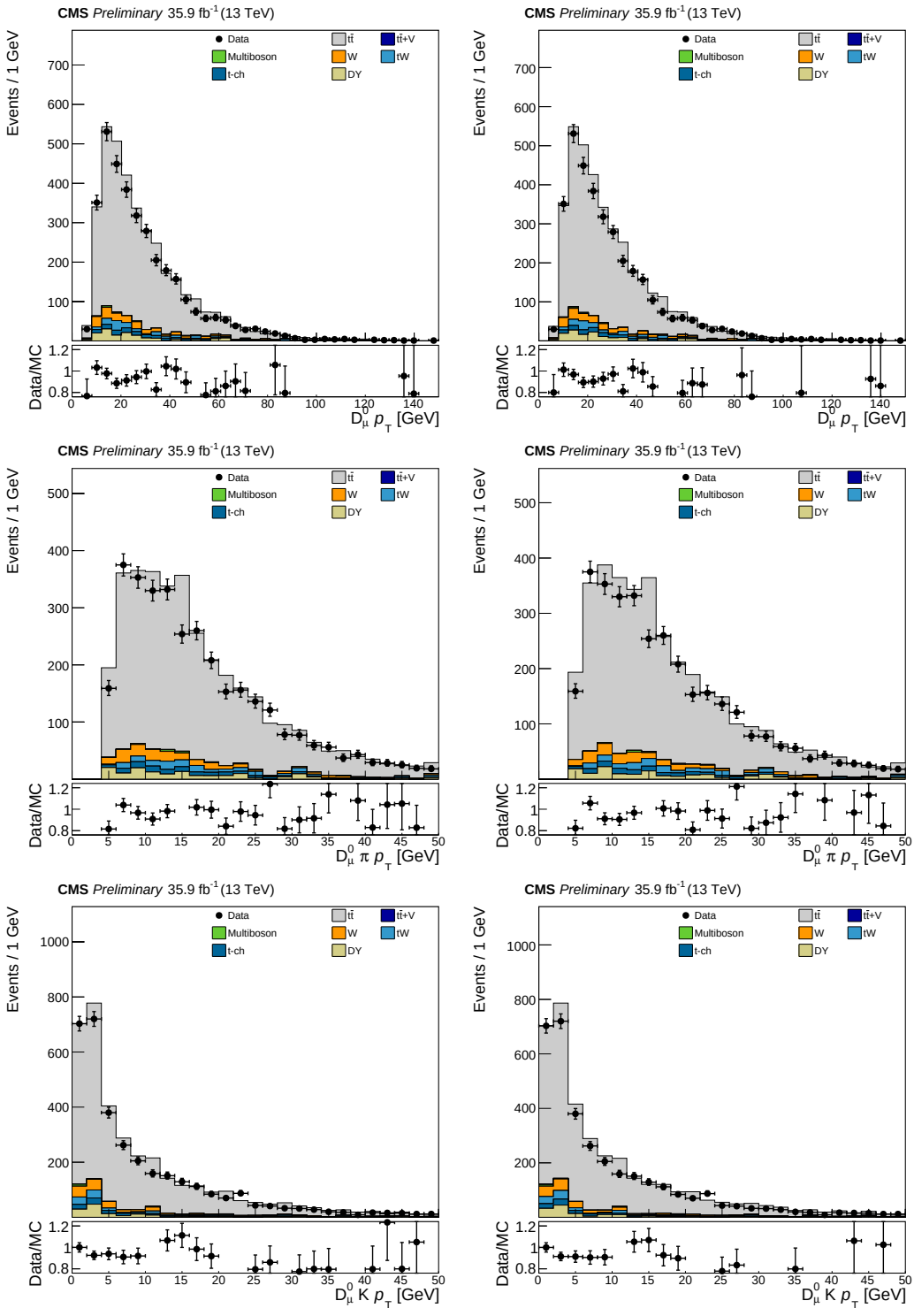

Figure 6.15: D^0_μ nominal (left) and measured (right)

Chapter 7

Systematic uncertainties

Systematic uncertainties are evaluated by performing the the χ^2 fit on each systematic sample, and subtracting the nominal measurement of r_{b} value from the result. Statistical uncertainties on these measurements are extracted only when the systematic sample has independent statistics from the nominal, and are computed as the quadrature sum of the statistical uncertainty of the systematic measurement of r_{b} ($\chi^2_{\mathrm{min}}+1$) and the statistical uncertainty of the nominal measurement of r_{b}. If the quadrature sum is larger than the systematic uncertainty, this number is used instead. A complete list of systematic uncertainties can be found in Table 7.1. Most of the systematic uncertainties are compatible with the statistical uncertainty of the measurements and are therefore symmetrized.

Pseudo-experiments are performed on the nominal MC sample to cross-check the statistical uncertainty obtained when fitting the data. A random subset of the MC is picked corresponding to the total number of events in the data for each pseudo-experiment, and this

subset is then treated as the toy data. The χ^2 measuement procedure is then performed. The mean of the measurement errors for the pseudo-experiments is extracted as the statistical uncertainty on the nominal MC sample.

Final state radiation The final state radiation (FSR) relies on a shape tune from LEP data for light-flavor quarks (u, d, s, c) [11]. The effective value of the strong coupling constant (α_s) is varied in PYTHIA by a factor of $\sqrt{2}$ up and $1/\sqrt{2}$ down to obtain the systematic uncertainties.

Matrix element parton shower matching The matching of ME from POWHEG to PS in PYTHIA is controlled by the model parameter h_{damp} [41] in PYTHIA 8. The ME/PS systematic uncertainty is obtained by varying the h_{damp} parameter up and down by 1 σ.

Initial state radiation The initial state radiation (ISR), similar to the FSR, relies on a shape tune from LEP data for light-flavor quarks (u, d, s, c) [11]. The effective value of the strong coupling constant (α_s) is varied in PYTHIA by a factor of 2 up and 1/2 down to obtain the systematic uncertainties.

Color reconnection The color reconnection (CR) affecting the resonance decays is compared to alternative models to obtain the systematic uncertainties. There are three models available: early resonance decay (ERD) [42] is enabled, a QCD inspired model [43] where the QCD color rules are also taken into account, and a Gluon move model [44] where gluons can be moved to another string. The model yielding the largest deviation in r_{b} from the nominal value is used as the systematic uncertainty for each sample.

Underlying event Remnants after initial parton scattering and multiple particle interactions (MPI) are know as the underlying events (UE). The UE is varried up and down by the uncertainties in PYTHIA [45].

Pile-up The inelastic proton-proton cross section (known as the mimum-bias cross section) is varied by $\pm 5\%$ to estimate the pile-up in the data [46]. The minimum-bias cross section is measured to be 69.2 mb.

Lepton selection efficiency Scale factors are used to correct the lepton selection efficiency in MC. This includes μ and e identification, μ isolation, and e reconstruction. The statistical uncertainties on these efficiencies are provided by the respective Physics Object Groups. The recommendation for systematic uncertainties on these efficiencies is an additional 1% uncertainty for the μ identification, 0.5% uncertainty for the μ isolation, and 1% for the e reconstruction (only for e $p_{\mathrm{T}} < 20\,\mathrm{GeV}$ and e $p_{\mathrm{T}} > 80\,\mathrm{GeV}$). The total uncertainties on these scale factors are used to shift the efficiencies up and down to obtain the systematic uncertainties.

Tracker efficiency Additional scale factors are derived for this analysis to account for issues seen in the silicon tracker during the 2016 data taking period (see Appendix A). The scale factors are used as a probability to drop reconstruction level particles in MC to help reproduce the Data. These scale factors are shifted up and down by 1 σ to obtain the systematic uncertainties. The systematic uncertainty due to these scale factors is expected to be anti-correlated; shifting the scale factors up or down results in smaller or larger probability of losing tracks respectively.

Top quark p_{T} re-weighting It was shown in [47, 48] that the p_{T} of the t quark in the Data was significantly softer than the MC. Scale factors are derived to correct this shape mismatch

$$SF(p_{\mathrm{T}}) = e^{0.0615 - 0.0005 \cdot p_{\mathrm{T}}} \tag{7.1}$$

$$w = \sqrt{SF\left(\mathrm{t}\right) SF\left(\bar{\mathrm{t}}\right)} \tag{7.2}$$

where SF is the scale factor and w is an event weight. The uncertainty is calculated by taking the difference of the r_{b} measured value with the re-weighting turned on and off.

Fit procedure A large portion of the systematic shifts are well within the statistical uncertainty of their respective samples. An uncertainty on the fitting procedure has been adopted to encompass this effect. The uncertainty is obtained by performing pseudo-experiments (see MC statistics for details) only on the MC samples which are obtained from re-weighting the nominal sample (such as ME/PS and UE). These samples by definition contain the same number of events as the nominal MC, so any differences in the up and down variations are attributed to the error of the measurement procedure. The sign of these shifts are dropped assuming the shift is a statistical effect. The up and down shifts are then averaged, resulting in the measurement procedure uncertainty.

MC statistics The statistical limitations on the MC samples are tested using pseudo-experiments. The experiments involve treating the nominal r_{b} template as toy data, and performing the χ^2 fitting procedure as normal. For each pseudo-experiment, each bin in the toy data is shifted up or down by a random Gaussian number with a width corresponding the uncertainty on each bin. The results for r_{b} for all pseudo-experiment are fit with a

Gaussian function, and the width of the Gaussian is extracted. This width is a convolution of the MC statistical uncertainty and the measurement procedure uncertainty. The quadrature difference of the width and the measurement procedure uncertainty is computed, and the result is taken as the MC statistical uncertainty.

Mass fit functional form The shape of the PDFs used to model the signal and background before the sPlot method is performed is another possible source of uncertainty. The background shape for both D^0 samples is modified from an exponential PDF to the product of an exponential and a Gaussian (with $\sigma = 1$). The signal for the J/ψ is modified from the sum of two Gaussians to the sum of three Gaussians. In all cased the visual change is negligible. The fitting procedure is then repeated as normal, staring with generating the r_b templates and performing the χ^2 scan.

Top quark mass The current uncertainties on the top quark mass measurements due to b quark fragmentation are restricted by the measurement of r_b from e^+e^- data. As result, the r_b response as a function of m_t should be checked. The sensitivity to the top quark mass is measured by varying m_t in PYTHIA 8 between 166.5 GeV and 178.5 GeV in steps of 1 GeV. The response in r_b is shown to be linear, and the shift from varying m_t by ± 0.5 GeV is interpolated. This corresponds to the current sensitivity of the top quark mass from CMS [3]. The response in r_b is essentially zero at the level of precision available in this analysis, and provides no additional uncertainty. This is because the measurement of r_b presented is compatible with the nominal value which entered into the m_t measurements. The previously

measured uncertainties on r_{b}, which is used to measure systematic uncertainties for m_{t}, are much larger than the measurement presented.

Trigger Efficiency The efficiencies of the triggers in MC were also corrected using scale factors. The uncertainties on these scale factors are used to to shift the efficiencies up and down. The p_{T} of the isolated µand e are selected beyond the trigger turn on curves. This results in the trigger efficiency only affecting the overall normalization, and not the shape of the fragmentation function. All distributions are normalized to unity before the r_{b} measurement is performed, so the trigger efficiency provides no additional uncertainty.

Jet energy resolution The jet energy resolution (JER) can be corrected to better reproduce the data [49]. This involves scaling the resolution of MC particles matched to generator level particles, or a stochastic smearing otherwise. The systematic uncertainty is obtained by varying these corrections by their statistical and systematic uncertainties. Most of the uncertainty in the JER is due to reconstructing neutral particles in the ECAL and HCAL. As the analysis only uses charged tracks, it is not sensitive to these corrections. The systematic uncertainty due to shifting the JER was checked, and the response in r_{b} is indeed negligible.

Table 7.1: Sources of systematic uncertainty

Source	J/ψ	D^0	D^0_μ	Combined
Fit procedure	± 0.012	± 0.008	± 0.004	± 0.006
MC stat	± 0.008	± 0.008	± 0.013	± 0.005
Functional form	± 0.008	± 0.001	± 0.008	± 0.006
FSR	$^{-0.129}_{+0.084}$	$^{-0.004}_{+0.019}$	$^{-0.057}_{+0.090}$	$^{-0.026}_{+0.033}$
Other systematics	± 0.016	± 0.011	± 0.016	± 0.010
Statistics	± 0.045	± 0.017	± 0.036	± 0.014

Chapter 8

Results

The r_b values obtained from the three separate channels are:

$$r_b = 0.874 \pm 0.045\,(\text{stat}) \pm 0.016\,(\text{syst})^{-0.129}_{+0.084}(\text{FSR}); \tag{8.1}$$

$$r_b = 0.836 \pm 0.017\,(\text{stat}) \pm 0.011\,(\text{syst})^{-0.004}_{+0.019}(\text{FSR}); \tag{8.2}$$

and

$$r_b = 0.846 \pm 0.036\,(\text{stat}) \pm 0.016\,(\text{syst})^{-0.057}_{+0.090}(\text{FSR}) \tag{8.3}$$

for the J/ψ, D^0, and D^0_μ samples respectively. The measurement results are compared to the data in Figure 8.1. Good agreement is found in all cases A combined measurement is performed by fitting each sample and summing the χ^2 values for each value of r_b. The χ^2 scan is then fit using the same procedure as the separate channels, producing a combined r_b result of

$$r_b = 0.841 \pm 0.014\,(\text{stat}) \pm 0.010\,(\text{syst})^{-0.026}_{+0.033}(\text{FSR}). \tag{8.4}$$

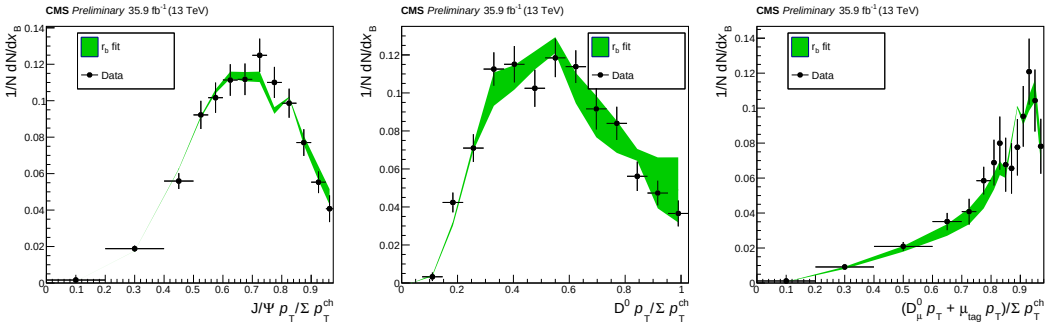

Figure 8.1: J/ψp_{T} divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$, $\mathrm{D}^0 p_{\mathrm{T}}$ divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$, and ($\mathrm{D}^0_{\mu}$ p_{T} + μ p_{T}) divided by $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ (bands are the total measurement uncertainties for each sample)

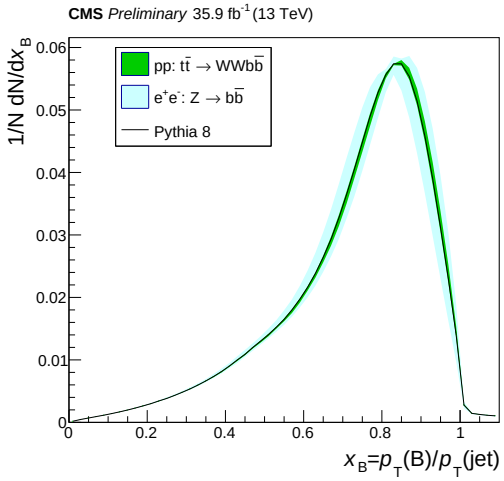

Figure 8.2: Fragmentation function (bands are the quadrature sum of the MC statistical and total measurement uncertainties)

This analysis gives almost a factor of five better precision than the analysis using $\mathrm{e}^+\mathrm{e}^-$ Z pole data [17] ($r_{\mathrm{b}} = 0.894^{+0.184}_{-0.197}$). No disagreement between the measurements is observed, indicating that the fragmentation process appears to be independent of the color

environment. A comparison between the generator level fragmentation fucntion for this analysis, the e^+e^- tune, and the default PYTHIA 8 tune is shown in Figure 8.2. The default PYTHIA 8 tune, $r_b = 0.855$ based on the e^+e^- data [11], is in good agreement with this analysis.

Chapter 9

Conclusion

A measurement of the shape of the b quark fragmentation function has been performed in a $t\bar{t}$ environmentusing $35.9\,\mathrm{fb}^{-1}$ of proton-proton collision data recorded by the CMS experiment during 2016. The analysis used b quark decays via the J/ψ meson and two orthogonal samples of D^0 meson decays (without and with flavor tagging) and is performed in samples of semi-leptonic and di-leptonic top quark pair events. The measurement of r_b for the three meson channels are:

$$r_b = 0.874 \pm 0.045\,(\mathrm{stat}) \pm 0.016\,(\mathrm{syst})^{-0.129}_{+0.084}(\mathrm{FSR}); \tag{9.1}$$

$$r_b = 0.836 \pm 0.017\,(\mathrm{stat}) \pm 0.011\,(\mathrm{syst})^{-0.004}_{+0.019}(\mathrm{FSR}); \tag{9.2}$$

and

$$r_b = 0.846 \pm 0.036\,(\mathrm{stat}) \pm 0.016\,(\mathrm{syst})^{-0.057}_{+0.090}(\mathrm{FSR}) \tag{9.3}$$

for the J/ψ , D^0, and D^0_μ samples respectively. The combined analysis of the three channels gives:

$$r_b = 0.841 \pm 0.014\,(\mathrm{stat}) \pm 0.010\,(\mathrm{syst})^{-0.026}_{+0.033}(\mathrm{FSR}). \tag{9.4}$$

The result is in consistent with the nominal value $r_b = 0.855$, and improves the statistical significance of the measurement by roughly a factor of four or five over the e^+e^- Z pole measurements. This indicates that there is no apparent dependence of the fragmentation on the color environment.

Bibliography

[1] M. Tanabashi *et al.*, "Review of Particle Physics," *Phys. Rev.*, vol. D98, no. 3, p. 030001, 2018.

[2] S. Abachi *et al.*, "Observation of the top quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, 1995.

[3] CMS Collaboration, "Measurement of the top quark mass using proton-proton data at $\sqrt{(s)} = 7$ and 8 tev," *Phys. Rev. D*, vol. 93, p. 072004, Apr 2016.

[4] J. R. Espinosa, "Implications of the top (and Higgs) mass for vacuum stability," *PoS*, vol. TOP2015, p. 043, 2016.

[5] V. Khachatryan *et al.*, "Measurement of the ratio $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV," *Phys. Lett.*, vol. B736, pp. 33–57, 2014.

[6] M. G. Bowler, "e^+e^- Production of Heavy Quarks in the String Model," *Z. Phys.*, vol. C11, p. 169, 1981.

[7] 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.*, vol. C71, p. 1557, 2011.

[8] A. Heister *et al.*, "Study of the fragmentation of b quarks into B mesons at the Z peak," *Phys. Lett.*, vol. B512, pp. 30–48, 2001.

[9] G. Abbiendi *et al.*, "Inclusive analysis of the b quark fragmentation function in Z decays at LEP," *Eur. Phys. J.*, vol. C29, pp. 463–478, 2003.

[10] K. Abe *et al.*, "Measurement of the b quark fragmentation function in Z^0 decays," *Phys. Rev.*, vol. D65, p. 092006, 2002. [Erratum: Phys. Rev.D66,079905(2002)].

[11] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune," *Eur. Phys. J.*, vol. C74, no. 8, p. 3024, 2014.

[12] F. Beaudette, "The CMS Particle Flow Algorithm," in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*, pp. 295–304, 2013.

[13] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting," *Nucl. Instrum. Meth.*, vol. A262, pp. 444–450, 1987.

[14] E. Chabanat and N. Estre, "Deterministic annealing for vertex finding at CMS," in *Computing in high energy physics and nuclear physics. Proceedings, Conference, CHEP'04, Interlaken, Switzerland, September 27-October 1, 2004*, pp. 287–290, 2005.

[15] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov, "Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments," *Eur. Phys. J.*, vol. C72, p. 2194, 2012.

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

[17] M. Seidel, *Precise measurement of the top-quark mass at the CMS experiment using the ideogram method.* PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2015.

[18] L. Evans and P. Bryant, "LHC Machine," *JINST*, vol. 3, p. S08001, 2008.

[19] S. Chatrchyan *et al.*, "The CMS Experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[20] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004.

[21] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP*, vol. 11, p. 070, 2007.

[22] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP*, vol. 06, p. 043, 2010.

[23] S. Alioli, S.-O. Moch, and P. Uwer, "Hadronic top-quark pair-production with one jet and parton showering," *JHEP*, vol. 01, p. 137, 2012.

[24] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders," *Comput. Phys. Commun.*, vol. 185, p. 2930, 2014.

[25] M. Czakon, P. Fielder, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$," *Phys. Rev. Lett.*, vol. 110, p. 252004, 2013.

[26] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[27] V. Khachatryan *et al.*, "Event generator tunes obtained from underlying event and multiparton scattering measurements," *Eur. Phys. J.*, vol. C76, no. 3, p. 155, 2016.

[28] "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV," Tech. Rep. CMS-PAS-TOP-16-021, CERN, Geneva, 2016.

[29] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions," *JHEP*, vol. 01, p. 013, 2007.

[30] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.

[31] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *JHEP*, vol. 07, p. 079, 2014.

[32] J. Alwall *et al.*, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions," *Eur. Phys. J. C*, vol. 53, p. 473, 2008.

[33] P. Kant *et al.*, "Hathor for single top quark production: updated predictions and uncertainty estimates for single top quark production in hadron collisions," *Comp. Phys. Comm.*, vol. 191, p. 74, 2015.

[34] R. Artoisenet, R. Frederix, O. Mattelaser, and R. Rietkerk, "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations," *JHEP*, vol. 03, p. 015, 2013.

[35] T. Speer, K. Prokofiev, Physik-Institut, U. Zurich, Zurich, S. R. Fruehwirth, I.-K. fur, HEPHY, Wien, and Austria, "Vertex Fitting with the Kalman Filter Formalism in the ORCA Reconstruction Program," Feb 2003.

[36] A. M. Sirunyan *et al.*, "Particle-flow reconstruction and global event description with the cms detector," *JINST*, vol. 12, p. P10003, 2017.

[37] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J.*, vol. C72, p. 1896, 2012.

[38] A. Hocker *et al.*, "TMVA - Toolkit for Multivariate Data Analysis," 2007.

[39] I. Antcheva *et al.*, "ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization," *Comput. Phys. Commun.*, vol. 180, pp. 2499–2512, 2009.

[40] M. Pivk and F. R. L. Diberder, "sPlot: a statistical tool to unfold data distributions," 2004.

[41] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV," 2016.

[42] T. Sjostrand and M. van Zijl, "A Multiple Interaction Model for the Event Structure in Hadron Collisions," *Phys. Rev.*, vol. D36, p. 2019, 1987.

[43] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour," *JHEP*, vol. 08, p. 003, 2015.

[44] J. R. Christiansen and T. Sjöstrand, "Color reconnection at future e^+e^- colliders," *Eur. Phys. J.*, vol. C75, no. 9, p. 441, 2015.

[45] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements," 2018.

[46] M. Aaboud *et al.*, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC," *Phys. Rev. Lett.*, vol. 117, no. 18, p. 182002, 2016.

[47] V. Khachatryan *et al.*, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV," *Phys. Rev.*, vol. D95, no. 9, p. 092001, 2017.

[48] A. M. Sirunyan *et al.*, "Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons," *JHEP*, vol. 02, p. 149, 2019.

[49] S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, vol. 6, p. P11002, 2011.

[50] E. Butz, "Operation and Performance of the CMS outer tracker," *PoS*, vol. Vertex 2017, p. 013, 2018.

Appendix A

Corrections for the issues in the 2016 data

During first half of the 2016 data taking period the CMS silicon tracker had known issues with the APV configuration [50]. The instantaneous luminosity from the LHC was steadily increased over this period; and the APV25 readout per-amplifier was saturated due to the amount of charge deposited by tracks in the silicon tracker. This was corrected halfway through 2016 by changing the drain speed, and we see that the second half of the data is unaffected by this problem.

From comparison of the unaffected data and the MC simulation a normalization correction factor of $11\% \pm 0.42\%$ was derived as shown in Fig. A.2. This is associated with the multiplicity mismodeling inside jets using PYTHIA 8. The correction is needed for the data from the epochs B–F and GH. To correct for the kinematic dependent effects in the

early data, a data-driven method is used. After the sub-samples of the data are normalized to their respective luminosity, the ratio of the data for epochs B–F divided by epochs GH is compared on an (η, p_{T}) grid. The results are shown in Table A.1. These values are used as a probability to drop particles from the simulation. A comparison of the corrected and uncorrected MC can be found in Fig. A.1.

Table A.1: Data B–F divided by data GH ratios binned by p_{T} and η for π and K associated with a D^0

		η					
		-1.5	-0.8	-0.4	0	0.4	0.8
p_{T} (GeV)	300	0.90 ± 0.19	1.01 ± 0.20	1.11 ± 0.24	0.99 ± 0.21	1.07 ± 0.23	1.41 ± 0.30
	100	1.00 ± 0.06	0.98 ± 0.06	1.08 ± 0.07	0.97 ± 0.06	1.07 ± 0.07	1.02 ± 0.06
	50	0.96 ± 0.03	0.91 ± 0.03	0.96 ± 0.03	0.94 ± 0.03	0.92 ± 0.03	1.03 ± 0.03
	30	0.92 ± 0.02	0.92 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.93 ± 0.03	0.98 ± 0.03
	20	0.94 ± 0.02	0.87 ± 0.02	0.88 ± 0.02	0.88 ± 0.02	0.91 ± 0.02	0.95 ± 0.02
	10	0.91 ± 0.02	0.82 ± 0.02	0.86 ± 0.02	0.85 ± 0.02	0.86 ± 0.02	0.93 ± 0.02
	6	0.88 ± 0.03	0.83 ± 0.03	0.87 ± 0.03	0.85 ± 0.03	0.81 ± 0.03	0.98 ± 0.03
	4	0.88 ± 0.02	0.81 ± 0.02	0.79 ± 0.02	0.83 ± 0.02	0.83 ± 0.02	0.91 ± 0.02
	2	0.87 ± 0.02	0.84 ± 0.02	0.79 ± 0.02	0.86 ± 0.02	0.81 ± 0.02	0.90 ± 0.02

It was observed that the $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ comparison between data from both run periods show a p_{T} and η dependence beyond the 11% normalization correction and that the results in Table A.1 are not quite sufficient to correct for this. A similar data to data comparison was performed for this quantity, resulting in the ratios observed in Table A.2. After correcting the MC using these ratios, following the same method as above, good agreement between the data and simulation is observed (see Fig. A.3). We believe that the differences between the corrections the result of a bias in the track selection coming from the Kalman filter.

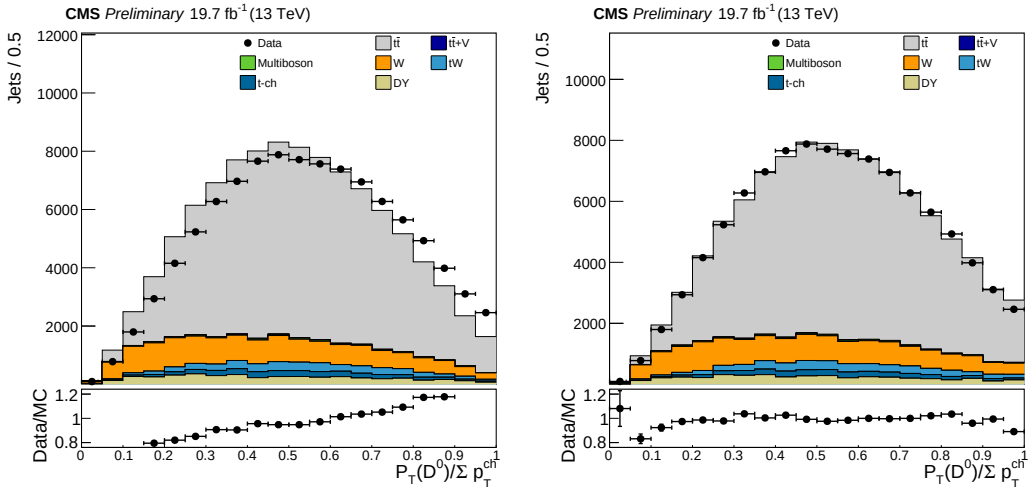

Figure A.1: Effects of the tracker corrections (uncorrected left and corrected right) on the fragmentation proxies D^0 p_T divided by $\sum p_\mathrm{T}^\mathrm{ch}$ for the first half of the data

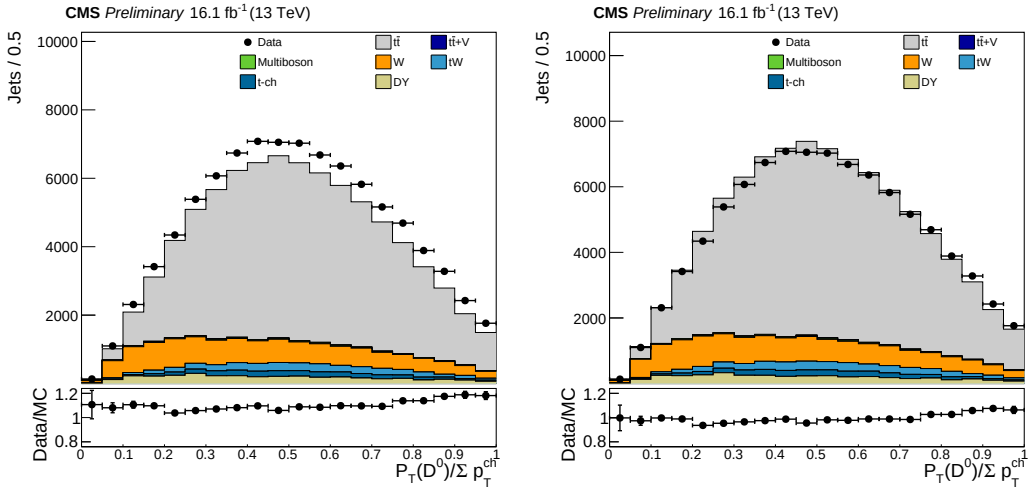

Figure A.2: Effects of the normalization correction (uncorrected left and corrected right) on the fragmentation proxies D^0 p_T divided by $\sum p_\mathrm{T}^\mathrm{ch}$ for the second half of the data

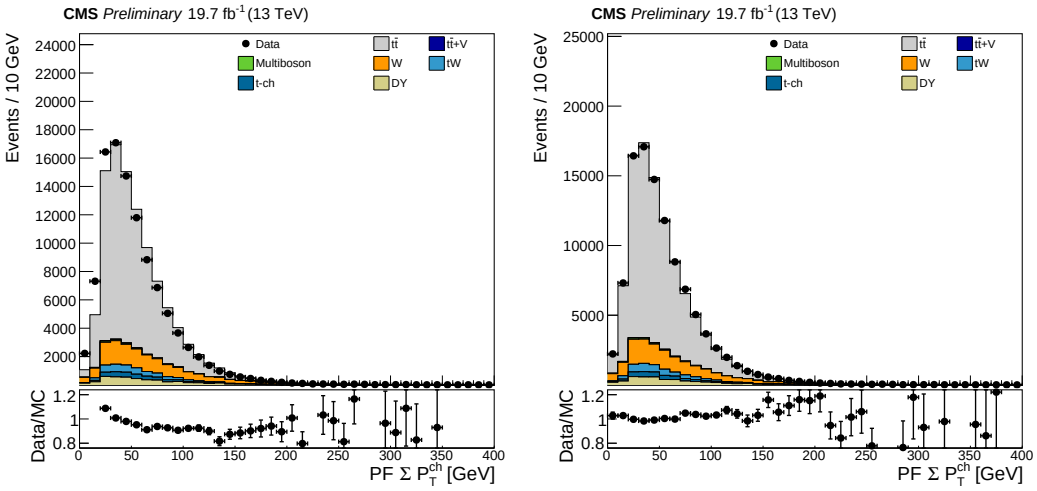

Figure A.3: Effects of the tracker corrections (uncorrected left and corrected right) on the $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ for the first half of the data

Table A.2: Data B–F dvided by data GH ratios binned by p_{T} and η for all charged tracks within b-jets containing a D^0

p_{T} (GeV) \ η	-2.4	-1.5	-0.8	-0.4	0	0.4	0.8	1.5
300	1.17 ± 0.11	1.10 ± 0.08	0.76 ± 0.05	0.89 ± 0.05	0.80 ± 0.05	0.72 ± 0.05	1.03 ± 0.06	0.88 ± 0.94
100	0.88 ± 0.03	0.94 ± 0.02	0.93 ± 0.02	0.91 ± 0.02	0.86 ± 0.02	0.92 ± 0.02	0.96 ± 0.02	0.94 ± 0.97
50	0.87 ± 0.02	0.87 ± 0.01	0.91 ± 0.01	0.86 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.94 ± 0.01	0.94 ± 0.97
30	0.91 ± 0.02	0.89 ± 0.01	0.89 ± 0.01	0.86 ± 0.01	0.88 ± 0.01	0.88 ± 0.01	0.93 ± 0.01	0.85 ± 0.92
20	0.87 ± 0.01	0.91 ± 0.01	0.84 ± 0.01	0.83 ± 0.01	0.85 ± 0.01	0.87 ± 0.01	0.90 ± 0.01	0.90 ± 0.95
10	0.86 ± 0.01	0.87 ± 0.01	0.82 ± 0.01	0.80 ± 0.01	0.84 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.85 ± 0.92
6	0.85 ± 0.01	0.85 ± 0.01	0.83 ± 0.01	0.80 ± 0.01	0.84 ± 0.01	0.85 ± 0.01	0.92 ± 0.01	0.84 ± 0.91
4	0.82 ± 0.01	0.86 ± 0.01	0.81 ± 0.01	0.79 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.90 ± 0.01	0.84 ± 0.92
2	0.86 ± 0.01	0.91 ± 0.01	0.87 ± 0.01	0.82 ± 0.00	0.93 ± 0.00	0.88 ± 0.00	0.96 ± 0.00	0.92 ± 0.96

Appendix B

Cross-check of the result

The result for the un-tagged D^0 measurement was cross-checked using the full jet p_T. This turns the fragmentation proxy into

$$x_B = \frac{D^0\ p_T}{\text{jet}\ p_T}. \tag{B.1}$$

The proxy for the D^0 sample is in Figure B.1. The proxy peaks at a much lower value of x_B which this analysis is less sensitive to (see Fig. 6.10). As a result, the statistical and systematic uncertainties increase. By including the neutral particles in the jet from the ECAL and HCAL, the JSF and JER corrections must be included. A complete list of the full jet systematics are listed in Table B.1. The final measurement for the D^0 sample is $r_b = 0.873 \pm 0.079\,(\text{stat}) \pm 0.109\,(\text{syst})$.

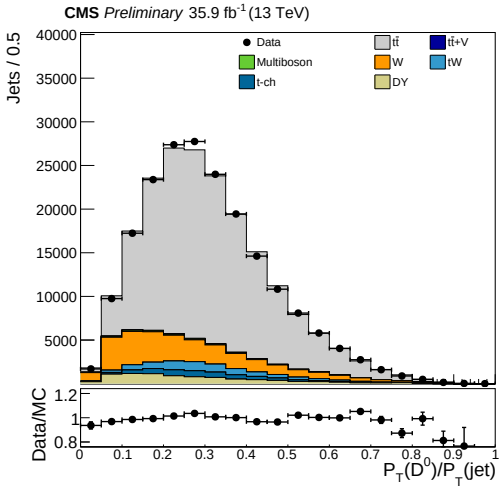

Figure B.1: D^0 p_T divided by full jet p_T

Table B.1: Sources of systematic uncertainty for D^0 full jet p_T proxy

ME/PS	±0.071
FSR	±0.041
ISR	±0.036
Underlying event	±0.035
JSF	±0.033
JER	±0.028
Pile-up	±0.025
Color reconnection	±0.024
Top p_T	±0.014
Lepton selection	±0.012

Appendix C

Monitoring the Cathode Strip Chamber Low Voltage

The low voltage (LV) system for the CSCs provides power to all the essential components except for the high voltage (HV) gas chambers themselves. Currently, the best way to identify a problem with the LV is to wait until a component breaks, or if a user is lucky enough to check a voltage value shortly before an issue occurs. A more preventative system is desirable. The LV values for each CSC chamber are stored in a STRUCTURED QUERY LANGUAGE (SQL) database. I have developed software to parse these databases, and look for trends in the LV system over time. It is hoped this software will help identify LV issues quicker, and may indicate potential future issues.

C.1 Chamber measurements

The simplest way to check a chamber for issues is to plot the various voltages over time. Each chamber is labeled by its location in the CSC system (Figure 4.3). The convention is ME (for MUON ENDCAP) followed by $\pm$ for chambers in the positive* z (near side) or negative z (far side) respectively, then the station number, ring number, and finally the chamber number. The station number increase with z for the near (or $-z$ for far side), and the ring number increases with r. For example ME $+1/1/01$ is the chamber located on the $+x$ side in the first station, the innermost ring, and the first chamber in that ring. An example of the ME $+1/1/01$ analog/digital seven volt (A/D 7 V) channel is in Figure C.1. Other than a few fluctuations, this figure indicates chamber ME $+1/1/01$ was operating within the accepted voltage levels for all of 2016. While this plot appears useful, there are 540 chambers in the CSC system. Each chamber also receives several different voltages for various components. This results in thousands of plots, and checking each one is inefficient.

The first step in generalizing is to look at the mean voltage for a group of chambers at a time. The ring structure of the CSCs provides a natural grouping. Mean voltage values are plotted for the ME $+1/1$ ring (Figure C.2) ring 1 excluding ME $+1/1$ (Figure C.3) and ring 2 (Figure C.4).

*See Section 4.1 for coordinate definitions.

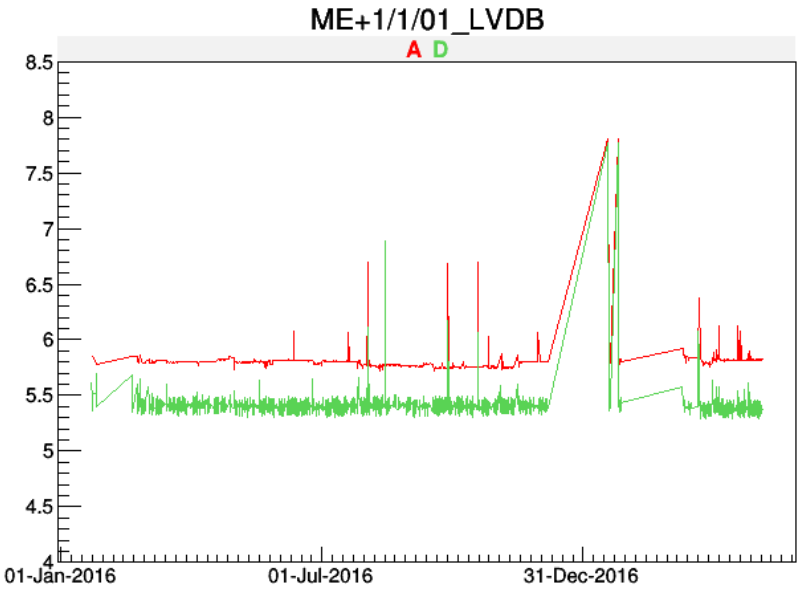

Figure C.1: The A/D 7 V values plotted over time for CSC chamber ME +1/1/01 in 2016 and part of 2017

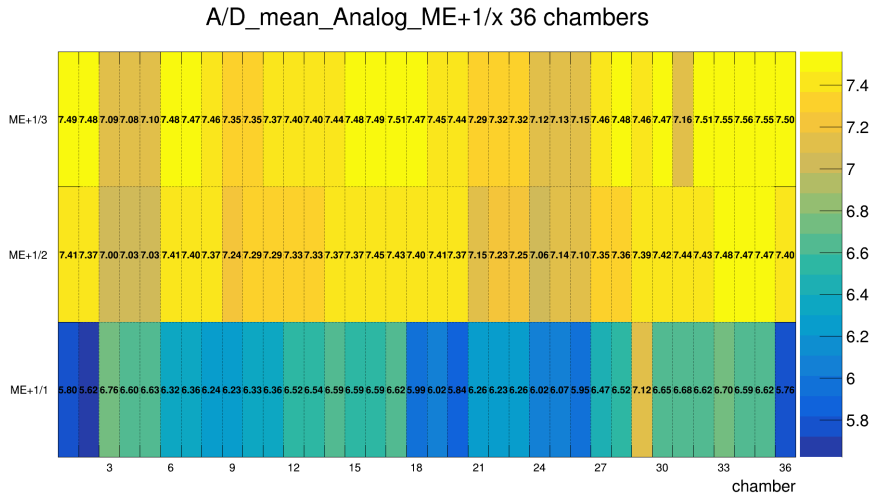

Figure C.2: Mean A/D 7 V values plotted for CSC chambers in the ME +1/x rings for 2016 and part of 2017

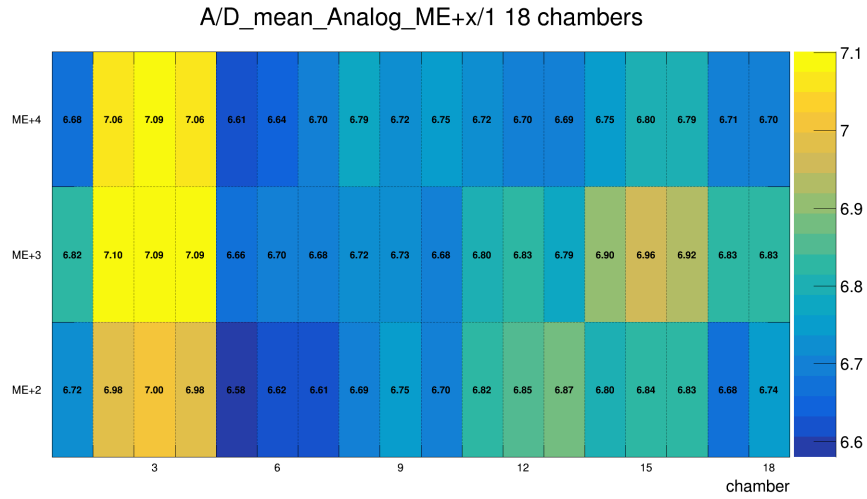

Figure C.3: Mean A/D 7 V values plotted CSC chambers in the ME +x/1 rings for 2016 and part of 2017

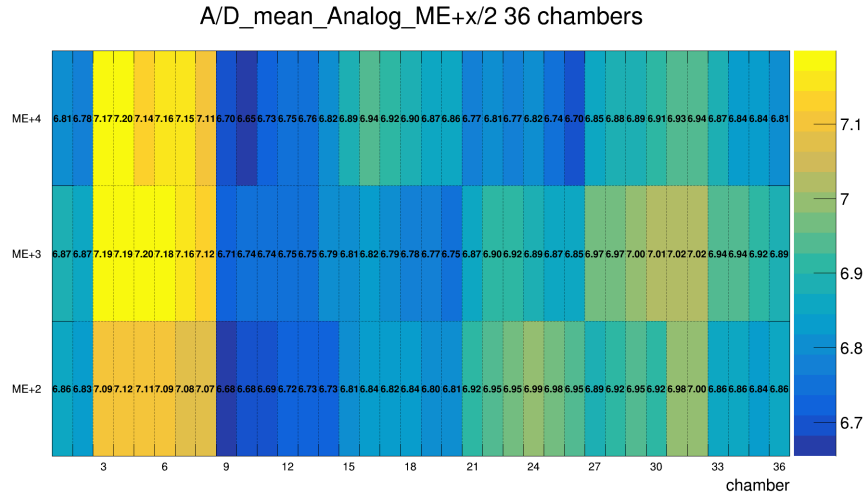

Figure C.4: Mean A/D 7 V values plotted CSC chambers in the ME +2/x rings for 2016 and part of 2017

C.2 Root-Mean-Square of chambers

The Root-Mean-Square (RMS) of each chamber may also be plotted. The mean and RMS differ slightly. An ideal mean will average around zero, whereas RMS will average

around $1/\sqrt{2}$ the amplitude. Figure C.5 shows ME +1/1/29 analog 7 V RMS was large compared to other chambers. Further investigation (Figure C.6) shows this chamber was off for an extended period of time. Chambers which are off store a value of -2 in the database. Values below zero are ignored by this software. The large dip in the early part of 2017 is a residual artifact of the chamber turning on. Otherwise, the chamber shows no issues.

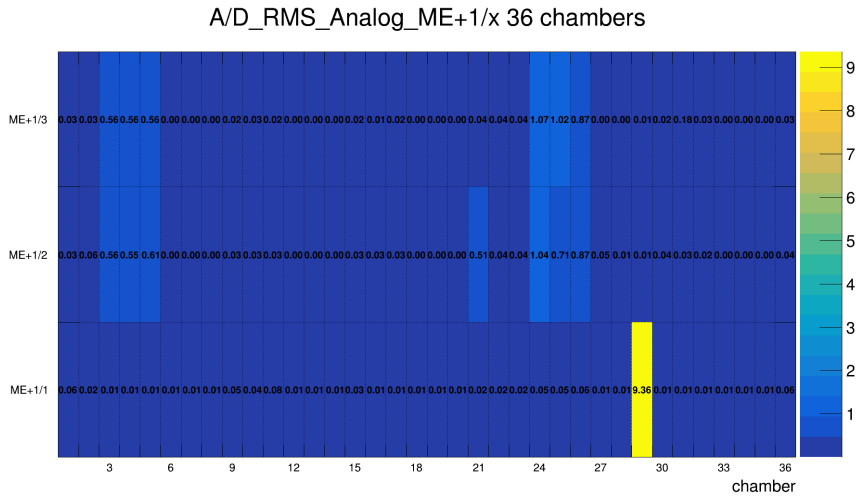

Figure C.5: RMS of the chambers in the ME +1 rings A/D 7 V plotted for 2016

The chambers may also be plotted with respect to the RMS of the ring. The ring RMS is computed, and the voltage of each chamber is compared to the ring RMS. Figure C.7 shows the percentage of times each chamber in the ME+1 rings was bigger than the ring RMS. An empty chamber in the plot signifies the chamber was within the RMS of its ring for the entire time period. The chambers which do show value are within a few percent, and are acceptable.

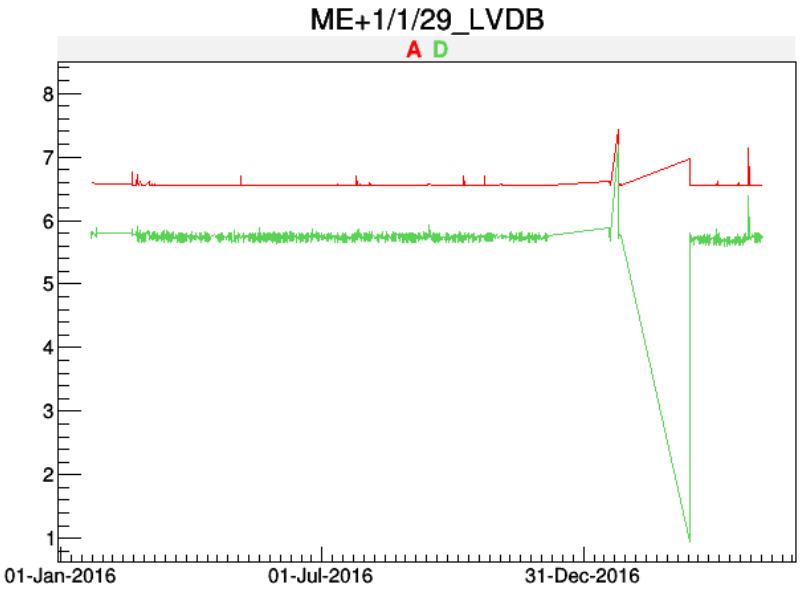

Figure C.6: The A/D 7 V values plotted over time for CSC chamber ME +1/1/29 in 2016 and part of 2017

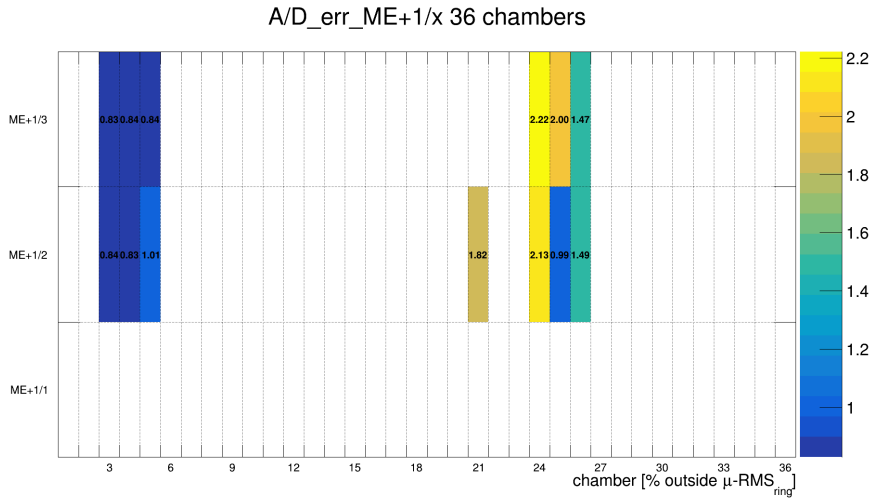

Figure C.7: Percentage of each chamber outside the ME +1 rings A/D 7 V RMS plotted for 2016