Measurement of Total and Differential Charm Cross Sections at 7 TeV with the CMS Detector

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

Nur Zulaiha Binti Jomhari

aus Kuala Lumpur, Malaysia

Hamburg

2022

Gutachter der Dissertation:	PD. Dr. Achim Geiser
	Prof. Dr. Elisabetta Gallo
Zusammensetzung der Prüfungskommission:	PD. Dr. Achim Geiser
	Prof. Dr. Elisabetta Gallo
	Prof. Dr. Daniela Pfannkuche
	Dr. Katarzyna Wichman
	Prof. Dr. Dieter Horns
Vorsitzender des Prüfungskommission:	Prof. Dr. Daniela Pfannkuche
Datum der Disputation:	16 March 2022
Vorsitzender des Fach-Promotionsauschusses Physik:	Prof. Dr. Wolfgang Hansen
Leiter des Fachbereichs Physik:	Prof. Dr. Günter H. W. Sigl
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Eidesstattliche Versicherung / Declaration on oath

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den 16.01.2022 Nur Zulaiha Binti Jomhari

삶에게 지는 날들도 있겠지 또다시 헤매일지라도 돌아오는 길을 알아

– IU

ACKNOWLEDGMENT

I would like to express my gratitude to people who I encountered during my Ph.D. For those who I mention the names are the one who contributed significantly to my life.

First is Achim Geiser, an interesting man whom I've never met before. It is a fact that he is the one who gave the opportunity to do a Ph.D. in Hamburg when I am a nobody and a newbie in particle physics. He has many intriguing physics projects benefiting the science and do things out of the box compared to other common people. For that, I truly respect him. I will definitely not forget his advice in life. Second is Daniela Pfannkuche. She is like my mother and sister who supported me emotionally when my family is far away from here. I love the positive vibe and energy that I got from her every time I went to see her. Third is Jordan Martin which I really enjoyed working under him and the team itself. I felt motivated doing the tasks, and he inspired me on what kind of leader I want to be in the future.

I am thankful to Kati and DPOA group for the interaction and support regarding CMS Open Data. Special thanks to the examination committee members for their time and effort. I had a good time during my defense and our fruitful session. Also, thanks to UHH scholarship, my group members, NanoAODplus team, office mates, German teachers and class members, friends in Malaysia, Hamburg and around the world, for our nice chatting and sharing knowledge. For people that know/help me directly or indirectly and wishes me all the best, good luck and happiness, you are included in this section and thank you so much.

Not forgetting my family in Malaysia for their encouragement especially my beloved mom, Halimah. I love you ma. To my late dad, Jomhari, I did it! I hope you are proud of me. To my nieces and nephews, Maksu will come back home at one point in time so wait for me∼

I also want to dedicate a section for those who I lost during the journey and gave an impact in my life. Jihyun, Azrul, abah Osman, and abang K. In addition, Fadzlin Adibah, my best friend during high school. May you rest in peace.

Last but not least, Afiq Aizuddin Anuar, a guy who gave me a lot of headache but stay by my side when I am at my worst. All the time that we spent together and memories that we made are precious and priceless. Sayang sayang sayang.

ABSTRACT

This thesis presents the total charm cross section measurement in the full kinematic phase space of pp collisions at the LHC. It was performed using open data (2010) from the CMS experiment at a center-of-mass energy of 7 TeV due to its special low p_T tracking. The integrated luminosity is determined in a data-driven way using unprescaled triggers as a basis and information from the luminosity database of the CMS collaboration to be 3.00 nb^{-1}. This is the first and so far only measurement of charm at 7 TeV in CMS.

The charm reconstruction was done through the $D^{*\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ final state. Since this analysis is statistically limited, one of the main strategies of this analysis is to use pileup vertices from muon and electron datasets as a physics resource. The $D^{*\pm}$ measured cross section is 1096 $\pm$ 133 (stat.) μb in the phase space of $0 < |y| < 2$ and $p_T > 1$ GeV, and also $2 < |y| < 2.5$ and $p_T > 8$ GeV, in which the double-differentially measured cross sections were integrated. This analysis has covered the largest possible kinematic phase space at the LHC from a single experiment with p_T down to 1 GeV. This includes three new phase space regions where no D^* cross section measurement has been done before at the LHC for p_T below 3.5 GeV. The total charm cross section is then extracted from this analysis. By combining with the LHCb measurements, which covered most of the region outside the CMS detector coverage, and some extrapolation from `PYTHIA` and FONLL, the total charm cross section for the full kinematic phase space is measured to be 9.40 $\pm$ 0.45 (statistical) $^{+2.50}_{-0.95}$ (FONLL/`PYTHIA`) mb with an extrapolation factor of 1.4 throughout all phase space. This is the smallest extrapolation achieved for the total charm cross section at the LHC so far.

Apart from the main result of this thesis, the validation of CMS Open Data with the Higgs to four leptons example at 7 (2011) and 8 (2012) TeV is also shown. This validation reproduces approximately part of the CMS Higgs discovery publication at 7 and 8 TeV with partial datasets in CMS Open Data. The purpose of the validation is to give an example of using CMS Open Data for educational purposes and show its potential for research applications. The Higgs peak is extracted at a significance of two standard deviations, compared to the original publication, which is 3.2 standard deviations in this channel alone. The corresponding example code was publicly released together with the CMS primary dataset for 2012. It has been used as reference by many ever since.

ZUSAMMENFASSUNG

Diese Dissertation präsentiert die Messung des totalen Charm-Wirkungsquerschnitts im gesamten kinematischen Phasenraum von pp-Kollisionen am LHC. Die Analyse wurde durchgeführt unter Verwendung öffentlicher Daten von 2010 aus dem CMS-Experiment bei einer Schwerpunktsenergie von 7 TeV, da für diese Daten ein spezielles Tracking für niedrige Transversalimpulse verfügbar ist. Die integrierte Luminosität wird datengetrieben unter Verwendung von nicht vorskalierten Triggern als Basis und Informationen aus der Luminositätsdatenbank der CMS-Kollaboration zu 3,00 nb^{-1} bestimmt. Dies ist die erste und bislang einzige Charm-Messung bei 7 TeV in CMS.

Die Charm-Rekonstruktion erfolgte über den Endzustand $\mathrm{D}^{*\pm} \to \mathrm{K}^{\mp}\pi^{\pm}\pi^{\pm}$. Da diese Analyse statistisch begrenzt ist, besteht eine der Hauptstrategien dieser Analyse darin, Pileup-Vertices aus Myon- und Elektronendatensätzen als physikalische Ressource zu verwenden. Der gemessene D^*-Wirkungsquerschnitt beträgt 1096 ± 133 (stat.) μb im Phasenraum von $0 < |\mathrm{y}| < 2$ und $\mathrm{p_T} > 1$ GeV, sowie $2 < |\mathrm{y}| < 2,5$ und $\mathrm{p_T} > 8$ GeV, in dem die doppelt differenziell bestimmten Wirkungsquerschnitte integriert werden. Diese Analyse hat den größtmöglichen kinematischen Phasenraum am LHC von einem einzigen Experiment mit $\mathrm{p_T}$ bis hinunter zu 1 GeV abgedeckt. Dazu gehören drei neue Phasenraumregionen, in denen noch keine D^*-Querschnittsmessung am LHC für $\mathrm{p_T}$ unter 3,5 GeV durchgeführt wurde. Der total Charm-Querschnitt wird dann aus dieser Analyse extrahiert. Durch Kombination mit den LHCb-Messungen, die einen Großteil des Bereichs außerhalb der CMS-Detektorakzeptanz abdecken, und Extrapolation von `PYTHIA` und FONLL für die verbleibenden Bereiche, wird ein totaler Charm-Wirkungsquerschnitt für den gesamten Kinematischen Phasenraum bestimmt. Dieser ergibt sich zu 9,40 ± 0,45 (statistisch) $\pm\ ^{+2,50}_{-0,95}$ (FONLL/`PYTHIA`) mb mit einem Extrapolationsfaktor von 1,4 im gesamten Phasenraum. Dies ist die bisher kleinste Extrapolation, die für den gesamten Charm-Querschnitt am LHC erreicht wurde.

Neben dem Hauptergebnis dieser Arbeit wurde auch eine Validierung der CMS Open Data bei 7 (2011) und 8 (2012) TeV durchgeführt anhand eines Analysebeispiels für das Massenspektrum für Higgs nach vier Leptonen. Diese Validierung reproduziert approximativ einen Teil der CMS-Veröffentlichung zur Higgs-Entdeckung bei 7 und 8 TeV mit teilweisen

Datensätzen in CMS Open Data. Der Zweck der Validierung besteht darin, ein Beispiel für die Verwendung von CMS Open Data für Bildungszwecke zu geben und sein Potenzial für Forschungsanwendungen aufzuzeigen. Der Higgs-Peak wird mit einer Signifikanz von zwei Standardabweichungen extrahiert, verglichen mit der ursprünglichen Veröffentlichung, die allein in diesem Kanal 3,2 Standardabweichungen beträgt. Der entsprechende Beispielcode wurde zusammen mit den 2012-Primärdatensatzen eröffentlicht. Seitdem wird er von vielen als Referenz verwendet.

CONTENTS

1 Introduction **12**
- 1.1 Motivation of This Work . . . 12
- 1.2 Bibliography . . . 13

2 Theoretical Overview **15**
- 2.1 The Standard Model of Particle Physics . . . 15
  - 2.1.1 Electroweak interaction . . . 16
- 2.2 Strong Interaction . . . 19
- 2.3 Proton-Proton Collisions . . . 21
- 2.4 Charm Quarks at the LHC . . . 22
  - 2.4.1 Charm quark production . . . 23
  - 2.4.2 Charm quark fragmentation . . . 23
  - 2.4.3 D^* meson decay . . . 24
  - 2.4.4 FONLL . . . 24
- 2.5 Bibliography . . . 25

3 Experimental Setup **29**
- 3.1 The Large Hadron Collider . . . 29
- 3.2 The CMS Experiment . . . 32
  - 3.2.1 Coordinate system . . . 32
  - 3.2.2 Superconducting magnet . . . 33
  - 3.2.3 Tracker . . . 33
  - 3.2.4 Electromagnetic calorimeter . . . 35
  - 3.2.5 Hadron calorimeter . . . 36
  - 3.2.6 Muon system . . . 36
- 3.3 Data Acquisition and Processing . . . 37
  - 3.3.1 Trigger system . . . 37
  - 3.3.2 Offline analysis . . . 37
- 3.4 Bibliography . . . 40

4 Object Reconstruction and Event Simulation **42**
4.1 Track reconstruction . . . 42
4.2 Primary vertex reconstruction . . . 45
4.3 Particle Flow . . . 45
4.3.1 Muon reconstruction . . . 46
4.3.2 Electron reconstruction . . . 47
4.4 Event Generator and Simulation . . . 48
4.4.1 Treatment of charm in simulation . . . 49
4.5 Bibliography . . . 49

5 Higgs to Four Leptons with CMS Open Data **53**
5.1 Introduction . . . 53
5.2 CMS Open Data . . . 54
5.3 Data and MC Samples . . . 54
5.4 Event/Object Selection . . . 55
5.5 Comparison with Higgs discovery . . . 57
5.6 Bibliography . . . 60

6 Measurement of $D^{*\pm}$ Cross Section at 7 TeV **62**
6.1 Introduction . . . 62
6.2 Analysis Strategy . . . 63
6.2.1 Pileup as a physics resource . . . 63
6.2.2 $D^{*\pm}$ meson . . . 64
6.3 Data and MC Samples . . . 66
6.3.1 Effective luminosity . . . 66
6.3.2 MC samples . . . 67
6.4 $D^{*\pm}$ Meson Reconstruction and Selection . . . 71
6.4.1 $D^{*\pm}$ preselection cut . . . 74
6.4.2 $D^{*\pm}$ analysis cuts . . . 75
6.5 $D^{*\pm}$ Cross Section Measurement . . . 76
6.5.1 $D^{*\pm}$ Signal extraction method . . . 76
6.5.2 Efficiency calculation . . . 78
6.5.3 Combined prompt and non-prompt $D^{*\pm}$ cross section . . . 84
6.5.4 Beauty and charm separation . . . 86
6.5.5 Prompt $D^{*\pm}$ cross section . . . 86
6.6 Systematic Uncertainties . . . 86
6.7 Bibliography . . . 94

7 Total Charm Cross Section Measurement **97**
7.1 Total Charm Production Cross Section . . . 97
7.2 Bibliography . . . 99

8 Summary and Outlook **102**

Appendices **105**

A Δm performance with and without beamspot constraint **106**

B Mass cut of D^0 and D^* at low p_T **110**

C Signal extraction for different p_T and y phase space **111**

D $N_{reco\&true}$ and N_{true} for efficiency calculation **122**

E D^{*+} cross section comparison with ATLAS **125**

CHAPTER

1

INTRODUCTION

1.1 Motivation of This Work

Humanity is full of curiosity to acquire knowledge. This knowledge includes the question what are the smallest particles in the world that make up matter. Particle physics is one of the physics fields that allow the study of matter at its fundamental level. The theory that describes the elementary particles and their interaction is referred to as the Standard Model. Exploring the particles within the Standard Model is interesting because there are still many aspects that are not fully understood. In addition, experimental measurements are important as they may reveal deviations that point to physics beyond the Standard Model. In this work, a study on a particle called charm quark is conducted. The interaction of charm quarks, like any other quarks, is mainly governed by a part of the Standard Model called Quantum Chromodynamics (QCD). The measurement of charm production thus provides a way to test QCD predictions. These can then be used to further constrain the charm quark mass, the proton parton density functions, and other QCD parameters.

Figure 1.1 shows the phase space in transverse momentum vs rapidity in which different LHC experiments performed their measurements of charm quark production so far [1–10]. As in the present analysis, these analyses measured D mesons or other charm hadrons. Some of these measurements were extrapolated to the full kinematic phase space with a large extrapolation factor in order to extract the total charm cross section. The corresponding total cross section values are thus strongly theory/model dependent. This analysis aims to cover the largest possible fraction of the total phase space in order to extract the total charm cross section with the smallest possible extrapolation, and thus smaller theoretical uncertainty. The analysis is conducted with 2010 7 TeV data taken by the CMS experiment due to its special low transverse momentum tracking.

The thesis is structured as follows. Chapter 2 gives a brief overview of the theoretical aspects of the Standard Model for Higgs and charm analyses in proton-proton collisions. The experimental aspects which provided the data to perform the analyses are presented

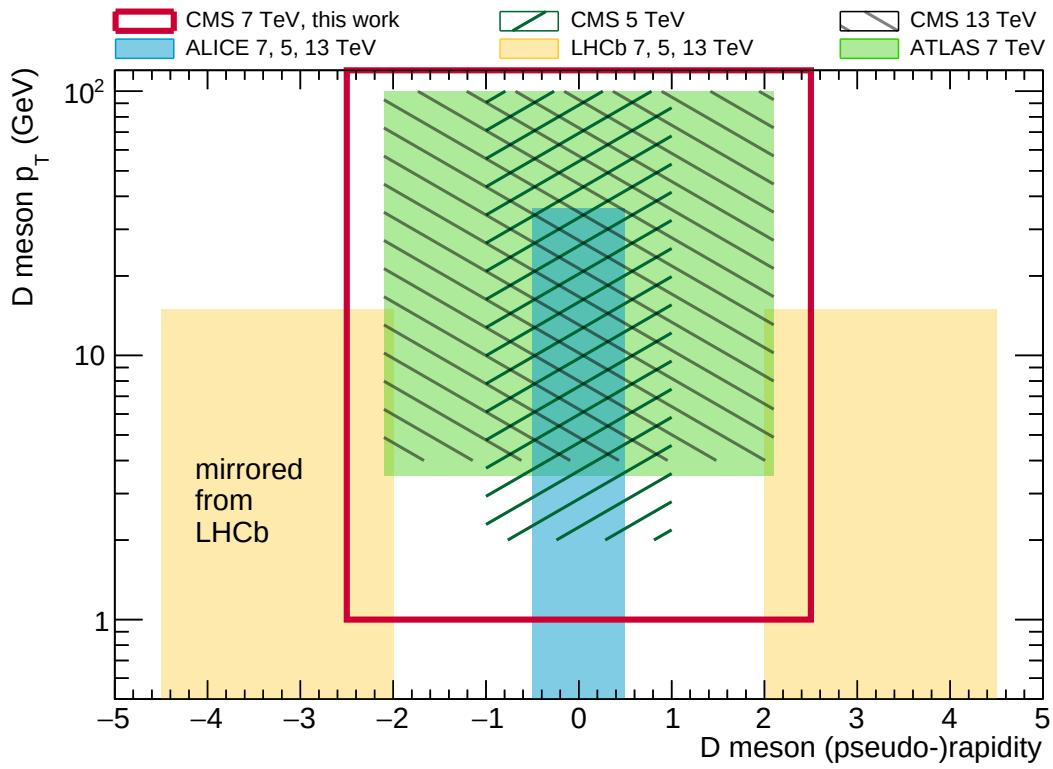


Figure 1.1: Phase space covered by the ALICE [1–3], LHCb [5–7], ATLAS [8], CMS 5 TeV HIN-16-001 [9] and CMS 13 TeV BPH-18-003 [10] experiments for charm meson measurements. The main analysis of this thesis is the CMS 7 TeV analysis shown in the red box.

in Chapter 3 and Chapter 4. The machinery starting from the collider, the detectors and how data is recorded is explained in Chapter 3. The reconstruction of events in terms of physics objects related to this work is described in Chapter 4. In addition, this chapter also includes an overview of event generators and simulation that is needed to compare with the data and correct for the detector effects. Chapter 5 manifests one of the author's service works that represent an important milestone for the CMS Open Data project. Chapter 6 covers the analysis aspects of the core topic of this thesis. It focuses on the measurement of differential charm meson cross sections in the early data taking of CMS and on the comparison to other measurements. Chapter 7 discusses the total charm cross section that covers the full kinematic phase space. This thesis ends with a summary and outlook in Chapter 8. The author's main work is covered in Chapter 5, Chapter 6, and Chapter 7.

1.2 Bibliography

[1] ALICE Collaboration, "Measurement of charm production at central rapidity in proton-proton collisions at $\sqrt{s} = 2.76$ TeV", *JHEP* **07** (2012) 191, doi:10.1007/JHEP07(2012)191, arXiv:1205.4007.

[2] ALICE Collaboration, "Measurement of D^0 , D^+ , D^{*+} and D_s^+ production in pp collisions at $\sqrt{s}$ = 5.02 TeV with ALICE", *Eur. Phys. J. C* **79** (2019), no. 5, 388, doi:10.1140/epjc/s10052-019-6873-6, arXiv:1901.07979.

[3] ALICE Collaboration, "Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **77** (2017), no. 8, 550, doi:10.1140/epjc/s10052-017-5090-4, arXiv:1702.00766.

[4] ALICE Collaboration, "Measurement of prompt D^0, Λ_c^+, and $\Sigma_c^{0,++}(2455)$ production in pp collisions at $\sqrt{s} = 13$ TeV", arXiv:2106.08278.

[5] LHCb Collaboration, "Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 5$ TeV", *JHEP* **06** (2017) 147, doi:10.1007/JHEP06(2017)147, arXiv:1610.02230.

[6] LHCb Collaboration, "Prompt charm production in pp collisions at sqrt(s)=7 TeV", *Nucl. Phys. B* **871** (2013) 1–20, doi:10.1016/j.nuclphysb.2013.02.010, arXiv:1302.2864.

[7] LHCb Collaboration, "Measurements of prompt charm production cross-sections in *pp* collisions at $\sqrt{s} = 13$ TeV", *JHEP* **03** (2016) 159, doi:10.1007/JHEP03(2016)159, arXiv:1510.01707. [Erratum: JHEP 09, 013 (2016), Erratum: JHEP 05, 074 (2017)].

[8] ATLAS Collaboration, "Measurement of $D^{*\pm}$, $D^{\pm}$ and $D_s^{\pm}$ meson production cross sections in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Nucl. Phys. B* **907** (2016) 717–763, doi:10.1016/j.nuclphysb.2016.04.032, arXiv:1512.02913.

[9] CMS Collaboration, "Nuclear modification factor of D^0 mesons in PbPb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *Phys. Lett. B* **782** (2018) 474–496, doi:10.1016/j.physletb.2018.05.074, arXiv:1708.04962.

[10] CMS Collaboration, "Measurement of prompt open-charm production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2021) 225, doi:10.1007/JHEP11(2021)225, arXiv:2107.01476.

CHAPTER

2

THEORETICAL OVERVIEW

This chapter will give an overview of the theoretical foundation of this work, starting with Section 2.1 where all the particles of the Standard Model (SM), and electroweak interaction including the Higgs mechanism are introduced. The next sections are the ones directly relevant to the main focus of this thesis. The section on strong interactions (Section 2.2) describes the theory of Quantum Chromodynamics (QCD), followed by a description of the interaction of partons in proton-proton (pp) collisions in Section 2.3. Lastly, the chapter is closed with the description of the production of charm quarks at the LHC in Section 2.4.

2.1 The Standard Model of Particle Physics

The Standard Model is a well-known and experimentally validated physics theory that describes elementary particles and their interactions. According to the SM [1–3], the smallest particles that make up matter are fermions, and their interactions are mediated by gauge bosons. The SM consists of 12 types of fermions, 4 types of gauge bosons, and 1 scalar boson as shown in Figure 2.1.

The fermions are spin 1/2 particles and consist of 6 quarks and 6 leptons. They are grouped into 3 generations where the only difference between each generation is their mass. The higher the generation, the heavier the particle's mass and the more unstable it is. The first generation is the everyday matter (made up of electrons, protons and neutrons)[1], and contains the lightest and the most stable particles. The second and third generation particles are heavier and quickly decay to the more stable particles.

The quarks have fractional electric charge of +2/3 (up (u), charm (c), and top (t)) and -1/3 (down (d), strange (s) and bottom (b)) and these particles are called up-type and down-type quarks, respectively. The quarks also possess color charge which causes them to engage in strong interactions (see Section 2.2). The leptons have integer electric charge of -1 (electron

[1] a proton consists of u, u, d valence quarks while a neutron consists of u, d, d valence quarks. Together with the electron, they form atoms. See Section 2.3 for details of proton structure.

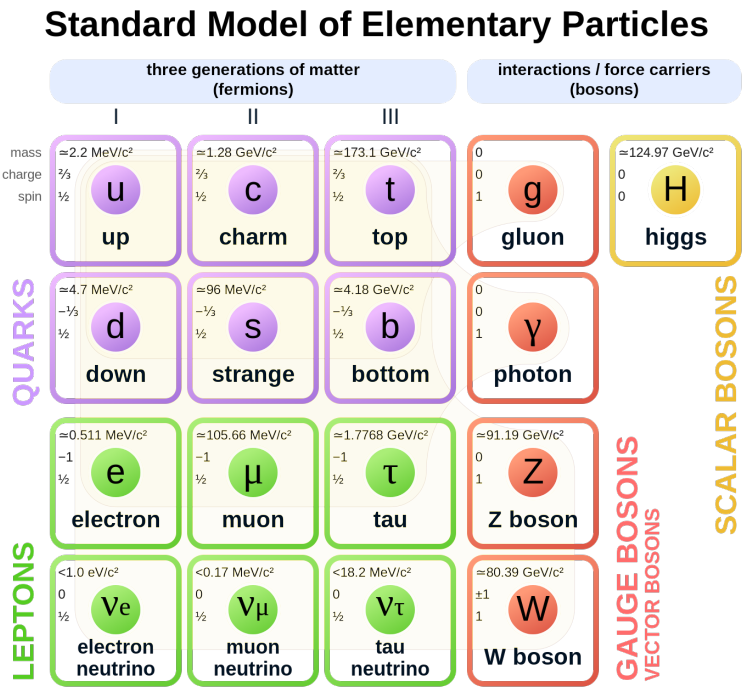


Figure 2.1: The Standard Model of particle physics. The picture was taken from [4].

(e), muon (μ) and tau (τ)) and neutral charge (electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ)). The masses of the neutrinos are very small and often assumed to vanish. Each fermion has its anti-particle[2] (anti-quarks and anti-leptons) that has the same mass but opposite charges.

The gauge bosons are spin 1 particles and consist of gluon (g), photon (γ), Z boson, W^+ boson, and W^- boson (both are denoted as W bosons). The first two mediators are massless, while the rest are massive. Their masses are given by the Higgs mechanism, which results in an additional boson with spin 0 called the Higgs boson. The interaction between elementary particles in the SM mediated by the gauge bosons and the Higgs mechanism is described in Subsection 2.1.1 below.

2.1.1 Electroweak interaction

The particle's interactions mediated by the gauge bosons in the SM are the strong, electromagnetic and weak interactions[3]. The former will be described in detail in Section 2.2 while the last two can be combined into 1 electroweak (EWK) interaction.

The electromagnetic interaction is mediated by photons, and they act on non-zero electric charged particles. γ is a massless and chargeless particle. Due to these properties, γ has an

[2]For anti-quarks, their symbols are $\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$, $\bar{t}$, $\bar{b}$ while for anti-lepton they are e^+, $\bar{\nu_e}$, μ^+, $\bar{\nu_\mu}$, τ^+, $\bar{\nu_\tau}$

[3]Gravity is also a force/ an interaction that occurs between objects in the universe but is not included in the Standard Model.

unlimited range of interaction (even though the force gets weaker when the distance between particles increases) and cannot couple to itself. The theory that describes the electromagnetic interaction is Quantum Electrodynamics (QED) [1].

The weak interaction [5, 6] is mediated by W and Z bosons, acting on all fermions. W and Z bosons participate in charge- and neutral-current weak interactions, respectively. One property of weak interactions is that they can change a quark or a lepton to a different flavor through W bosons. For example, a charm quark in a D meson can change to a strange quark in a Kaon (see Figure 2.8) by emitting a W^+. However, the quark flavor can only be changed from up-type quark to down-type quark and vice versa[4]. Plus, only left-handed-chiral particles can participate in the interaction. The flavor mixing between quarks in such interactions is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix (denoted as V) shown below [7]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.1}$$

where d' s' and b' are the weak eigenstate and d, s, b are the mass eigenstate. For each element in the CKM matrix, its experimentally measured magnitudes are [8]:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix} \tag{2.2}$$

where V_{ij} is the probability of quark flavor j to transform into quark flavor i. The CKM matrix is a unitary matrix that expresses the degree of mixing between the quark flavor and mass eigenstates. As one can see in Equation 2.2, the values in the diagonal elements are close to 1, meaning that the probability of particle j transforming to particle i within one generation is more likely to happen compared to other elements in the matrix.

The combination of electromagnetic and weak interaction formed the electroweak unification by Glashow-Salam-Weinberg [5, 6, 9]. The group structure of the weak interaction is $SU(2)_L \times U(1)$, where the subscript L refers to the fact that W bosons couple only to left-handed fermions. Due to gauge invariance, the gauge bosons are required to be massless. However, W and Z bosons have been measured to be massive [8] which breaks the electroweak symmetry.

The Higgs field was postulated by Brout-Englert-Higgs [10, 11] where any fermions or gauge bosons that interact with the Higgs field will be given a mass without violating gauge invariance. This mechanism is called the BEH mechanism. The Higgs field ϕ is a complex

[4]For the case of lepton, the flavor can change from charged lepton to neutrino and vice versa.

scalar doublet with four degrees of freedom:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix} \tag{2.3}$$

where ϕ^+ and ϕ^0 have weak isospin of 1/2 and -1/2, respectively. The energy potential of this field is given as:

$$V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4 \tag{2.4}$$

where $V(\phi)$ is symmetric under changes of complex components and rotation in ϕ space, therefore it is invariant under gauge transformations. μ^2 and λ are real parameters of the theory. λ must be > 0 so that there is at least 1 minimum otherwise the electroweak vacuum would not be stable. Spontaneous symmetry breaking happens when the value of μ^2 is negative as shown in Figure 2.2. The left figure is obtained when $\mu^2 > 0$ and $\lambda > 0$ where the minimum remains at $|\phi| = 0$. When $\mu^2 < 0$ and $\lambda > 0$, the minimum is at $|\phi| = v$ where v is the vacuum expectation value:

$$v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{2.5}$$

which is measured to be around 246 GeV.

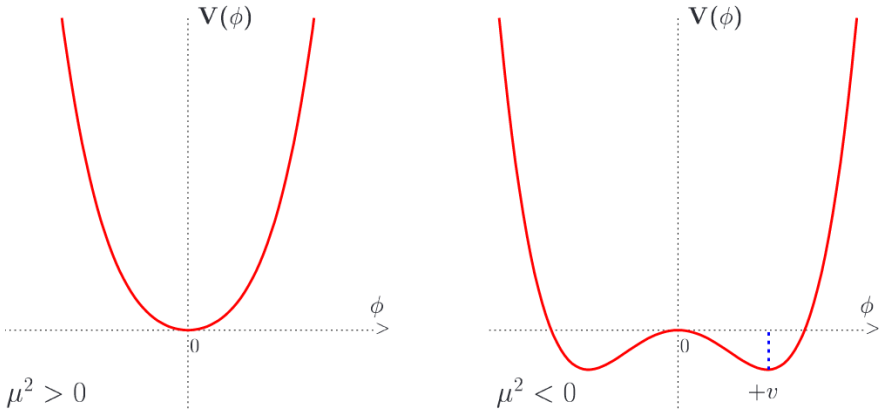


Figure 2.2: The Higgs potential before (left) and after (right) spontaneous symmetry breaking. The picture was taken from [12].

The interaction between the Higgs and weak fields gives mass to the W and Z bosons, in accordance with experimental constraints. Fermion masses are obtained through a Yukawa interaction between the Higgs and fermion fields [13]. The fermion masses are proportional to the vacuum expectation value and the fermion Yukawa couplings, which are free parameters of the theory. Another consequence of the existence of the Higgs field is the manifestation of a new particle called the Higgs boson. This particle is a scalar boson, whose mass is not predicted by the theory. It was discovered in 2012 by the CMS [14] and ATLAS [15] experiments with a mass of 125.09 ± 0.24 GeV [16].

2.2 Strong Interaction

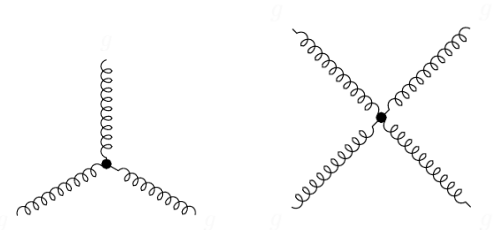
The strong interaction [1] is mediated by gluons and acts on particles with color charge, which are quarks and gluons. The color charge for a quark can be either red, blue or green. Unlike photons, which are electrically neutral, gluons have color charge which allow them to interact among themselves in addition to mediating interactions between quarks. The theory behind the strong interaction is Quantum Chromodynamics (QCD) which is a gauge theory based on the non-Abelian SU(3) symmetry group. The theory has 8 linearly independent generators called Gell-Mann matrices [17], leading to the existence of 8 gluons. The Lagrangian density of QCD, $\mathcal{L}_{QCD}$, is given by:

$$\mathcal{L}_{QCD} = \sum_n \overline{\psi}_n (i\gamma^\mu D_\mu - m_n)\psi_n - \frac{1}{4} G^a_{\mu v} G^{a\mu v} \tag{2.6}$$

where n is an index that runs over the six quark flavors, ψ is the Dirac spinor of the quark field, γ^μ refers to the Dirac matrices. m_n refers to quark masses which are free parameters in the Lagrangian, $G^a_{\mu v}$ is the gluon field strength tensor, and D_μ is the covariant derivative defined as:

$$D_\mu = \partial_\mu + i g_s T^a G^a_\mu \tag{2.7}$$

The D_μ contain the gluon fields, G^a_μ (a = 1, ..., 8), which are responsible for the interaction of colored particles. T^a are the linearly independent hermitian 3 x 3 matrices (Gell-Mann matrices) in color space, and g_s is the strong coupling strength. In Equation 2.6, the first term represents the kinetic and mass term of quarks, including the quark-gluon coupling. The second term corresponds to the kinetic energy of the gluons and their self-interaction in the form of 3 and 4 gluon couplings as shown in Figure 2.3.



(a) Triple gluon vertex (b) Quartic gluon vertex

Figure 2.3: Feynman diagrams of gluon-gluon interaction that enter the $\mathcal{L}_{QCD}$.

QCD has some unique properties, which are quark confinement and asymptotic freedom. This is best explained by referring to Figure 2.4. The strength between a particle and a field in strong interaction is represented by a coupling constant ($\alpha_s = \frac{g_s^2}{4\pi}$). This constant is depending on the energy scale (Q); therefore, it is called running α_s. At a low energy scale, quarks cannot be observed as isolated particles, as α_s increases as the distance between quarks increases. Such phenomenon is called quark confinement [18]. At some point, the quarks eventually form a colorless state (hadrons) by forming a quark and anti-quark state (meson) or a combination of three quarks (baryon). The process of forming hadrons is

called hadronization. It involves small momentum transfer and a long-distance process which continues until no more color particles are available to interact. At a high energy scale, an asymptotic freedom [19–24] occurs as α_s is decreasing in value and asymptotically weaker. At this stage, the quarks behave as quasi-free particles. The property of asymptotic freedom allows for perturbative QCD (pQCD) calculations to be performed where a cross section calculation is done via expansion in powers of α_s:

$$\sigma_{pert} = \sum_{i=0}^{n} C_i \alpha_s^i \tag{2.8}$$

where n is the order of the calculation and C_i is the coefficient calculated from the Feynman diagrams. To perform pQCD calculations, α_s has to be smaller than one for the series to converge. The higher the order of the calculation, the more loops/number of Feynman diagrams are produced, as shown in Figure 2.5. The lowest non-zero order of a calculation is called leading order (LO), and the following order is referred to as next-to-leading order (NLO).

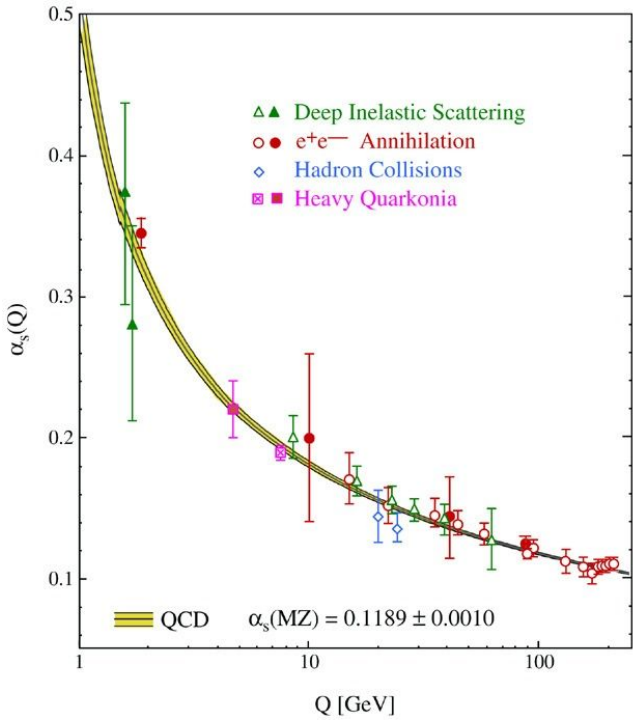


Figure 2.4: Measurement of the coupling constant (α_s) as a function of momentum transfer (Q). Picture taken from [25].

Higher-order Feynman diagrams can lead to a certain problem. For example, in Figure 2.5 (e) and (f), the four-momentum is required to be conserved at each vertex, but the momentum in the loops can acquire arbitrary values. Integration over all possible momenta of the loop when calculating the matrix element, $\mathcal{M}$, will be infinite and introduces divergences. To fix this, a renormalization scheme [26] which introduces a renormalization scale, μ_R, has to be

applied. The formula for the renormalized value of α_s at LO is shown in Equation 2.9:

$$\alpha_s(\mu_R^2) = \frac{12\pi}{(33 - 2n_f)\ln(\mu_R^2/\Lambda_{QCD}^2)} \tag{2.9}$$

where μ_R is the renormalization scale (see next section), n_f is the number of quark flavors with mass less than μ_R and Λ_{QCD} is the cutoff parameter for the use of pQCD where the value is experimentally determined to be ∼250 MeV.

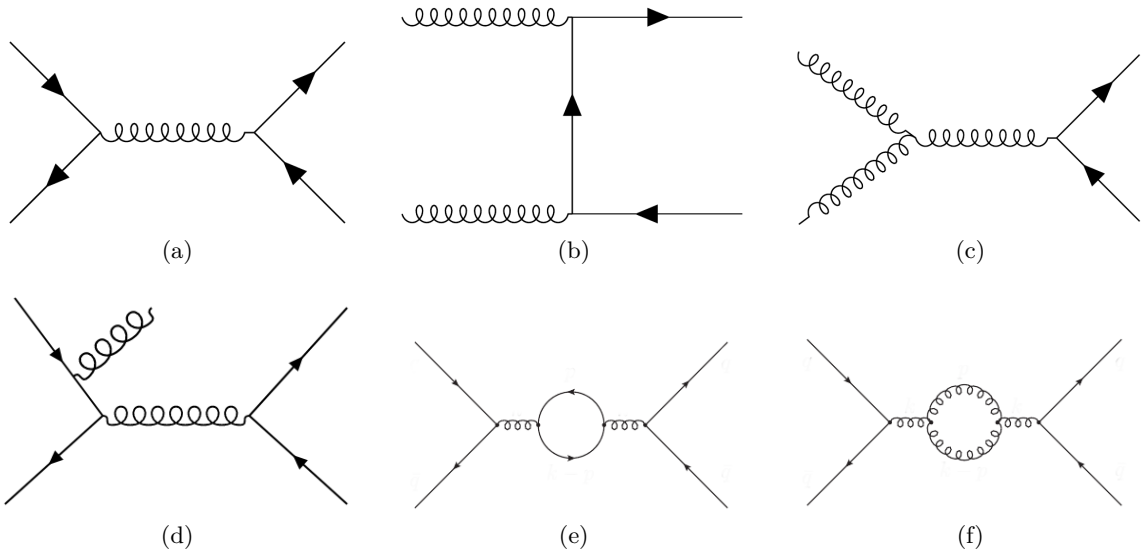


Figure 2.5: Examples of Feynman diagrams at a LO (a) - (c) and with a higher order real gluon radiation (d), (e) and (f) are higher-order diagrams with one quark loop correction and a gluon loop correction, respectively.

2.3 Proton-Proton Collisions

The proton is a hadron consisting of three valence quarks (uud), is color neutral and has electric charge of +1. Naively, one would expect that each valence quark carry a third of the total proton momentum but that is not the case. The quarks are bound by gluons and the gluons in the proton are interacting and splitting into quark-anti-quark pairs known as sea quarks [27]. Together, the valence quarks, gluons and sea quarks in the proton are referred to as partons. Perturbative pp cross sections can be calculated using the factorization theorem [28]. This theorem allows the cross section calculation by separating between soft and hard interactions as shown below:

$$\sigma_{pp\to X} = \sum_{i,j} \int dx_i dx_j f_i(x_i, \mu_F) f_j(x_j, \mu_F) \cdot \hat{\sigma}_{i,j\to X}(x_i x_j s, \mu_F, \mu_R) \tag{2.10}$$

where i, j are the partons that interact with each other, x_i is the parton momentum fraction with respect to the proton momentum, and $f_i(x_i, \mu_F)$ are the parton distribution functions (PDF) that give the probability to find a parton i with the parton momentum fraction x_i at

some scale of μ_F. The PDF is extracted from fits to experimental measurements [29, 30]. Several tools are available within the literature for this purpose, one example being `XFITTER` [31]. μ_F is called factorisation scale, a scale that defines the boundary between soft and hard interactions in the initial state. μ_R is the renormalization scale, the scale at which α_s is evaluated (Equation 2.9) for the expansion of the cross section in Equation 2.8 (see Section 2.2). Both are arbitrary scale choices and serve as parameters to avoid divergences. $\hat{\sigma}$ is the partonic cross section for the final state X that depends on $x_i x_j s$ (where s is the pp center of mass-energy), μ_F and μ_R. $\hat{\sigma}$ corresponds to the partonic hard interaction cross section that can be calculated in pQCD, which is proportional to the square of the matrix element of the process: $\hat{\sigma} \sim |\mathcal{M}|^2$.

In addition to proton-dissociative interactions, there can also be interactions in which one or both of the protons stay intact. They are categorized as elastic and diffractive processes, where elastic collision refers to the case when both protons stay intact. Diffractive processes [32] are further split into single and double dissociation. In the former, only one proton remain intact while in the latter, both protons are destroyed in the collision. Figure 2.6 shows the schematic diagram of these processes. Both diffractive and non-diffractive processes are included in the analysis within |y| of 2.5.

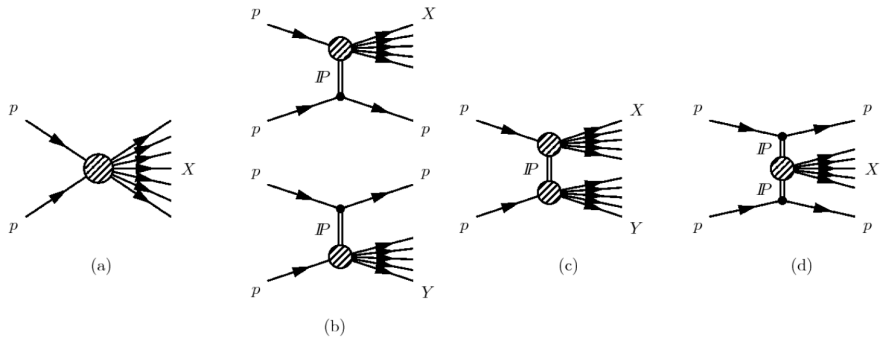


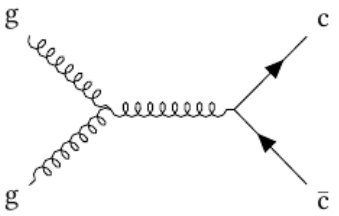
Figure 2.6: Schematic diagrams of non-diffractive (a), single dissociation (b) and double dissociation (c) of diffractive processes with an exchange of a pomeron and (d) central diffraction (pomeron-pomeron scattering) . Picture taken from [32].

2.4 Charm Quarks at the LHC

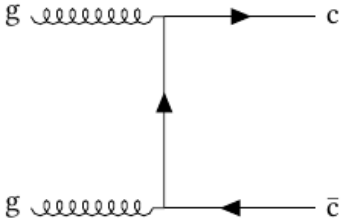
The existence of the charm quark was postulated in 1964 [33] and the J/ψ was the first discovery of a charm quark bound state in 1974 [34, 35] from which this breakthrough is named the November Revolution. The charm quark is part of the second generation of the SM (see Section 2.1) with a mass of 1.27 ± 0.02 GeV [8]. The total cross section of charm production at the LHC is expected to be ∼10 mb [36] (∼10% of the total proton-proton cross section). Since the charm mass provides a (semi)hard scale, this cross section is perturbatively calculable, currently up to Next-to-Next-to-Leading Order (NNLO, terms up to α_s^4) [36] for the total cross section, and Next-to-Leading Order (NLO, up to α_s^3), with [37] or without [38] Next-to-Leading Log (NLL, all order resummation up to $\log^2$) contributions, for differential cross sections.

2.4.1 Charm quark production

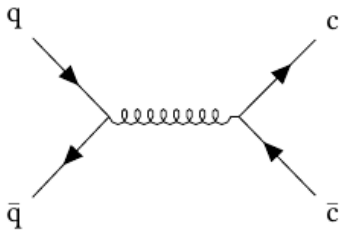
Diagrams for the production of a charm quark pair ($c\bar{c}$) in pp collisions at leading order are shown in Figure 2.7. The main production is coming from gluon-gluon fusion (Figure 2.7 a-b) with a small contribution from quark-anti-quark (Figure 2.7 c) fusion. The former allows the $c\bar{c}$ to have color-octet and color-singlet states while the latter allows only $c\bar{c}$ color-octet states. At higher order, many other configurations are possible [38].



(a) feynman diagram gg - cc in s-channel



(b) feynman diagram gg - cc in t-channel



(c) feynman diagram qq - cc

Figure 2.7: Possible Feynman diagrams for charm production at LO. The main contribution to charm production at the LHC is gluon-gloun fusion.

In addition to the above processes, charm quarks can also be produced through the decay of heavier particles. One common source of charm quarks produced in this way is bottom quark production followed by $b \to cX$ decay. In this work, such processes are referred to as non-prompt charm production, in contrast to the above case which is referred to as prompt charm production (see Subsection 2.4.3 for the definition of prompt and non-prompt). The treatment of non-prompt charm production can be signal or background or both. More details are explained in Subsection 6.3.2 and Subsection 6.5.4.

2.4.2 Charm quark fragmentation

Even though at the charm production level ($p\bar{p} \to c\bar{c}X$), the interactions between the partons are treated perturbatively, the hadronization of a charm quark to a charm hadron involves non-perturbative QCD and can be described through charm fragmentation fractions and charm fragmentation functions.

The former represents the probability of a charm quark to hadronize into a particular charm hadron $f(c \to h)$ and the latter describes the energy transfer of a charm quark to

a given charm hadron. There are several parameterizations in the literature used for the fragmentation function for heavy quarks, such as Bowler [39], Kartvelishvili [40] and Peterson [41].

2.4.3 D^* meson decay

Since quarks are confined (see Section 2.2), the study of charm quarks can be made through a measurement of charmed hadrons, for example, D mesons. One of the available D mesons is the D^* meson. There are several charged and neutral D^* mesons but only $D^*(2010)^{\pm}$ will be discussed here due to its good signal to background ratio for a broad phase space coverage (see Section 6.1). It is important to take a note that from here onwards, any particles mentioned always include their charge conjugate (for charged particles) or anti-particle (for neutral particles) unless stated otherwise.

The D^{*+} has three different strong or electromagnetic interaction decay modes with their respective branching ratio (BR) [8]

- $D^0\pi^+$, BR = 67.7 ± 0.5 %
- $D^+\pi^0$, BR = 30.7 ± 0.5 %
- $D^+\gamma$, BR = 1.6 ± 0.4 %

The $D^0\pi^+$ decay channel is chosen for the analysis of this thesis. The D^0 is further decaying through weak interactions in hadronic mode to a Kaon (K^-) and a pion (π^+) with a BR of 3.95 ± 0.03 % which sets the total BR of D^{*+} to $K^-\pi^+\pi^+$ to be 2.67 ± 0.03 %. Table 2.1 shows the properties of the D^{*+} and its decayed particles. Based on Table 2.1, the mass difference between D^{*+} and D^0 is small (just above the pion mass), which gives a limited momentum to the pion from the D^{*+} decay. Hence, this pion is usually referred to as "slow" pion (π_s), and this abbreviation will be used from now onwards. Figure 2.8 illustrates the D^{*+} decay including the quark content of the states involved. The advantages of choosing $D^{*\pm} \rightarrow K^{\mp}\pi^{\pm}\pi_s^{\pm}$ are explained further in Subsection 6.2.2.

In this work, a prompt D^{*+} is referred to as prompt charm production where the D^{*+} is produced at the collision point. The D^0 from the prompt D^{*+} decay is called prompt D^0. Since the D^{*+} can also be produced in decays of beauty hadrons (non-prompt charm production), by default the CMS data contains both beauty and charm contributions.

2.4.4 FONLL

Fixed-order plus next-to-leading-logs (FONLL) [43, 44] is a calculation tool that calculates NLO plus NLL QCD predictions. For heavy quark production, FONLL is used as the benchmark prediction to calculate total inclusive cross sections (within cuts) of charm production. The calculation is done at pp collider energy 7 TeV with bins of 1 GeV in p_T and 0.5 in |y| ($D^{*\pm}$) [37]. The hadronic final state selected is D^* and the PDF used in the calculation is CTEQ6.6. The uncertainty in the FONLL prediction is the sum in quadrature

Table 2.1: Properties of D^{*+} and its decay particles. I is isospin, P is parity of spin J. The information is taken from the latest PDG [8].

Particles	Quark content	Mass (MeV)	Mean lifetime (s)	I (J^P)
D^{*+}	$c\bar{d}$	2010.26 ± 0.05	-	$\frac{1}{2}(1^-)$
D^0	$c\bar{u}$	1864.84 ± 0.05	410.1 ± 1.5 (10^{-15})	$\frac{1}{2}(0^-)$
K^-	$s\bar{u}$	493.68 ± 0.02	1.2379 ± 0.002 (10^{-8})	$\frac{1}{2}(0^-)$
π^+	$u\bar{d}$	139.57 ± 0.0002	2.6033 ± 0.0005 (10^{-8})	$1^-(0^-)$

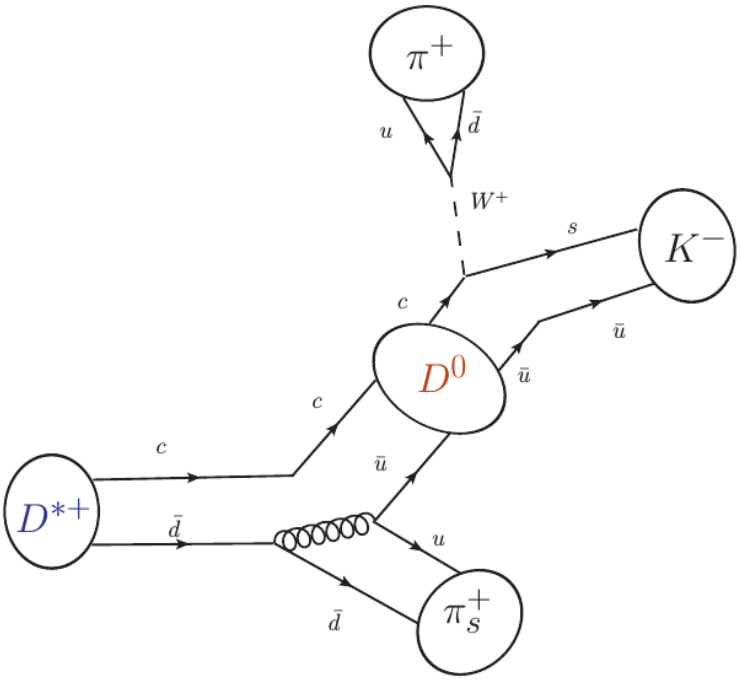


Figure 2.8: Quark content of D^{*+} decays to π_s^+ and $D^0 \to K^-\pi^+$. Picture taken from [42].

of uncertainties due to scale variation, charm quark mass and PDF. The scale variation uncertainty is estimated by varying μ_R and μ_F by a factor of two while keeping $1/2 < \mu_R/\mu_F < 2$. The nominal scale choice used in FONLL calculation is $\sqrt{m_c^2 + p_T^2}$. The uncertainty due to the charm quark mass is estimated by varying m_c by 0.2 GeV from a default choice of 1.5 GeV. The PDF uncertainty is estimated according to the recipe provided with the CTEQ 6.6 PDF [45]. The fragmentation fraction of the charm quark to D^{*+} used in FONLL is 0.236 [46].

2.5 Bibliography

[1] D. J. Griffiths, "Introduction to elementary particles; 2nd rev. version". Physics textbook. Wiley, New York, NY, 2008.

[2] M. Thomson, "Modern particle physics". Cambridge University Press, New York, 2013.

[3] D. H. Perkins, "Introduction to High Energy Physics". Cambridge University Press, 4th edition, 2000.

[4] "Standard Model", (2021). https://en.wikipedia.org/wiki/Standard_Model.

[5] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367–377, doi:10.1142/9789812795915_0034.

[6] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[7] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.

[8] Particle Data Group Collaboration, "Review of Particle Physics", *PTEP* **2020** (2020), no. 8, 083C01, doi:10.1093/ptep/ptaa104.

[9] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[10] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[11] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[12] A. Djouadi, "The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model", *Phys. Rept.* **457** (2008) 1–216, doi:10.1016/j.physrep.2007.10.004, arXiv:hep-ph/0503172.

[13] H. Yukawa, "On the Interaction of Elementary Particles I", *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48–57, doi:10.1143/PTPS.1.1.

[14] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[15] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[16] ATLAS, CMS Collaboration, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.

[17] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215, doi:10.1016/S0031-9163(64)92001-3.

[18] K. G. Wilson, "Confinement of Quarks", *Phys. Rev. D* **10** (1974) 2445–2459, doi:10.1103/PhysRevD.10.2445.

[19] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (Jun, 1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[20] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[21] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. I", *Phys. Rev. D* **8** (Nov, 1973) 3633–3652, doi:10.1103/PhysRevD.8.3633.

[22] H. D. Politzer, "Asymptotic Freedom: An Approach to Strong Interactions", *Phys. Rept.* **14** (1974) 129–180, doi:10.1016/0370-1573(74)90014-3.

[23] D. J. Gross, "The discovery of asymptotic freedom and the emergence of QCD", *Proc. Nat. Acad. Sci.* **102** (2005) 9099–9108, doi:10.1073/pnas.0503831102.

[24] F. Wilczek, "Asymptotic freedom: From paradox to paradigm", *Proc. Nat. Acad. Sci.* **102** (2005) 8403–8413, doi:10.1103/RevModPhys.77.857, arXiv:hep-ph/0502113.

[25] S. Bethke, "Experimental tests of asymptotic freedom", *Prog. Part. Nucl. Phys.* **58** (2007) 351–386, doi:10.1016/j.ppnp.2006.06.001, arXiv:hep-ex/0606035.

[26] G. 't Hooft, "Renormalization of Massless Yang-Mills Fields", *Nucl. Phys. B* **33** (1971) 173–199, doi:10.1016/0550-3213(71)90395-6.

[27] R. Vogt, "Physics of the nucleon sea quark distributions", *Prog. Part. Nucl. Phys.* **45** (2000) S105–S169, doi:10.1016/S0146-6410(00)90012-7, arXiv:hep-ph/0011298.

[28] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, doi:10.1142/9789814503266_0001, arXiv:hep-ph/0409313.

[29] S. Forte and G. Watt, "Progress in the Determination of the Partonic Structure of the Proton", *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 291–328, doi:10.1146/annurev-nucl-102212-170607, arXiv:1301.6754.

[30] J. M. Campbell, J. W. Huston, and W. J. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics", *Rept. Prog. Phys.* **70** (2007) 89, doi:10.1088/0034-4885/70/1/R02, arXiv:hep-ph/0611148.

[31] S. Alekhin et al., "HERAFitter", *Eur. Phys. J. C* **75** (2015), no. 7, 304, doi:10.1140/epjc/s10052-015-3480-z, arXiv:1410.4412.

[32] CMS Collaboration, "Measurement of diffraction dissociation cross sections in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. D* **92** (2015), no. 1, 012003, doi:10.1103/PhysRevD.92.012003, arXiv:1503.08689.

[33] J. D. Bjorken and S. L. Glashow, "Elementary Particles and SU(4)", *Phys. Lett.* **11** (1964) 255–257, doi:10.1016/0031-9163(64)90433-0.

[34] J. J. Aubert et al., "Experimental Observation of a Heavy Particle J", *Phys. Rev. Lett.* **33** (Dec, 1974) 1404–1406, doi:10.1103/PhysRevLett.33.1404.

[35] J. E. Augustin et al., "Discovery of a Narrow Resonance in e^+e^- Annihilation", *Phys. Rev. Lett.* **33** (Dec, 1974) 1406–1408, doi:10.1103/PhysRevLett.33.1406.

[36] A. Accardi et al., "A Critical Appraisal and Evaluation of Modern PDFs", *Eur. Phys. J. C* **76** (2016), no. 8, 471, doi:10.1140/epjc/s10052-016-4285-4, arXiv:1603.08906.

[37] M. Cacciari et al., "Theoretical predictions for charm and bottom production at the LHC", *JHEP* **10** (2012) 137, doi:10.1007/JHEP10(2012)137, arXiv:1205.6344.

[38] M. L. Mangano, P. Nason, and G. Ridolfi, "Heavy quark correlations in hadron collisions at next-to-leading order", *Nucl. Phys. B* **373** (1992) 295–345, doi:10.1016/0550-3213(92)90435-E.

[39] M. G. Bowler, "e+ e- Production of Heavy Quarks in the String Model", *Z. Phys. C* **11** (1981) 169, doi:10.1007/BF01574001.

[40] V. Kartvelishvili, A. Likhoded, and V. Petrov, "On the fragmentation functions of heavy quarks into hadrons", *Physics Letters B* **78** (1978), no. 5, 615–617, doi:https://doi.org/10.1016/0370-2693(78)90653-6.

[41] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, "Scaling violations in inclusive e^+e^- annihilation spectra", *Phys. Rev. D* **27** (Jan, 1983) 105–111, doi:10.1103/PhysRevD.27.105.

[42] O. Bachynska, "Measurement of $D^{*\pm}$ meson production in deep-inelastic scattering at HERA". Dissertation, Universität Hamburg, Hamburg, 2012. Dissertation, Universität Hamburg, 2012.

[43] M. Cacciari, M. Greco, and P. Nason, "The P(T) spectrum in heavy flavor hadroproduction", *JHEP* **05** (1998) 007, doi:10.1088/1126-6708/1998/05/007, arXiv:hep-ph/9803400.

[44] M. Cacciari, S. Frixione, and P. Nason, "The p(T) spectrum in heavy flavor photoproduction", *JHEP* **03** (2001) 006, doi:10.1088/1126-6708/2001/03/006, arXiv:hep-ph/0102134.

[45] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev. D* **78** (2008) 013004, doi:10.1103/PhysRevD.78.013004, arXiv:0802.0007.

[46] L. Gladilin, "Fragmentation fractions of c and b quarks into charmed hadrons at LEP", *Eur. Phys. J. C* **75** (2015), no. 1, 19, doi:10.1140/epjc/s10052-014-3250-3, arXiv:1404.3888.

CHAPTER

3

EXPERIMENTAL SETUP

This chapter briefly introduces the Large Hadron Collider (LHC), the collider that smashes proton bunches together. The final state particles produced in the collisions are detected by the CMS experiment, which is then described starting with the coordinate system, the individual sub-detectors, and finally, the data acquisition and event processing.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [1] is a circular particle accelerator and collider and the world's largest. It has a 26.7 km circumference and is built underground beneath the France and Switzerland border and is hosted by the European Organization of Nuclear Research (CERN, 'Conseil Européen pour la Recherche Nucléaire'). It accelerates and collides proton beams[1], but before proton beams are injected into the LHC, they undergo several steps as described in Figure 3.1.

Inside the LHC, two proton beams circulate along the ring in opposite directions. The beams are focused and can travel along the ring due to the 8 Tesla magnetic field from superconducting electromagnets. The beams then collide at four interaction points, and there are four main detectors/experiments. They are:

- ALICE (A Large Ion Collider Experiment) is 26 m long, 16 m high, and 16 m wide with a weight of 10,000 tons. It is a detector designed to investigate mainly heavy ion collisions. The physics goal is to study strongly interacting matter in an extreme quark-gluon plasma state [2].

- LHCb (Large Hadron Collider beauty) is 21 m long, 10 m high, and 13 m wide with a weight of 5,600 tons. It is a single-arm detector with a forward coverage of $\eta < 4.9$ for heavy flavor physics. The physics goal is to study CP violation and rare decays of beauty and charm hadrons [3].

[1] also heavy-ion beams, but these are not covered in this thesis

- ATLAS (A Toroidal LHC Apparatus) is 46 m long, with a diameter of 25 m, and weighs 7,000 tons. It is one of the general-purpose detectors at LHC. It serves various physics goals from Standard Model measurements to the search for new physics [4].

- CMS (Compact Muon Solenoid) is 21.6 m long, with a diameter of 14.6 m, and weighs 12,500 tons. It is the second general-purpose detector at the LHC and the detector involved in this work [5]. More details are described in Section 3.2.

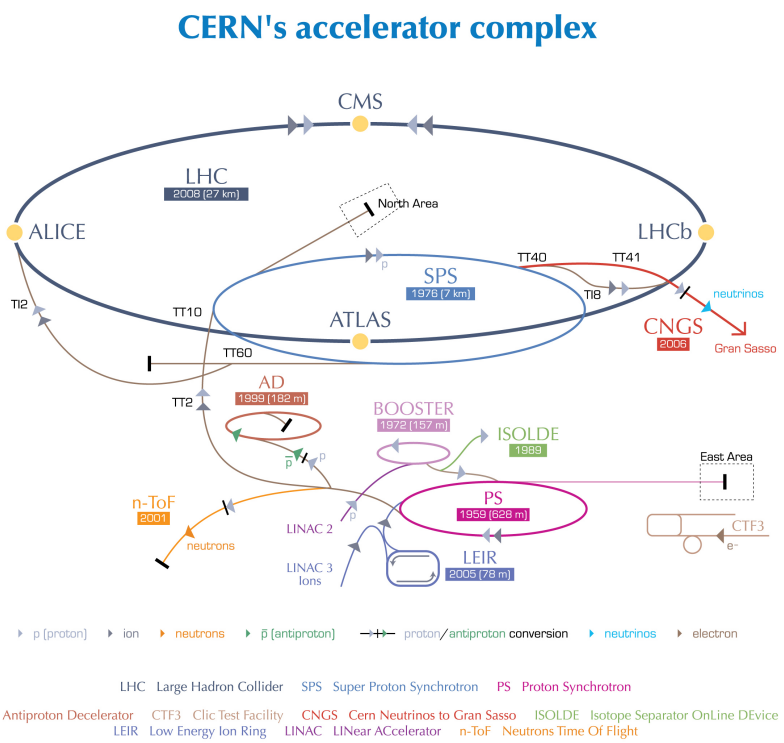


Figure 3.1: A diagram of CERN's accelerator complex showing several proton synchrotrons that accelerate protons before they are injected into the LHC. Starting from protons extracted from a bottle of hydrogen, they enter LINAC 2, are accelerated up to 50 MeV and are sent to BOOSTER, which increases the energy up to 1.4 GeV to inject the protons into the PS. The PS further raises the energy up to 26 GeV and sends the protons to the SPS, which provides 450 GeV to the LHC. The protons are then accelerated to different center of mass energies such as 7 TeV [6].

A physics quantity related to an accelerator is the integrated luminosity (L):

$$\mathrm{L} = \frac{\mathrm{N}}{\sigma} \tag{3.1}$$

where N is the total number of events, σ is the total cross section of a given process, and L is an integral of instantaneous luminosity ($\mathcal{L}$) over time, $\mathrm{L} = \int \mathcal{L} dt$. The dimension for L is usually expressed in the inverse unit of cross section, e.g. /pb where $1/\mathrm{pb} = 1 \times 10^{-36}\mathrm{cm}^{-2}$.

The luminosity is directly proportional to the number of collisions at the interaction point.

The LHC was designed to achieve a peak of instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ at a center of mass-energy, $\sqrt{s} = 14$ TeV. The LHC so far completed two data taking periods, namely, Run 1 with $\mathcal{L}$ of $1.2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ and Run 2 with $\mathcal{L}$ more than twice the design value of the LHC [7]. Data were taken at several $\sqrt{\text{s}}$ values during these periods. Run 1 has a $\sqrt{\text{s}}$ of 7(8) TeV for the period of 2010-2011 (2012). Run 2 has $\sqrt{\text{s}}$ of 13 TeV for 2015-2018. There are also special low luminosity runs at $\sqrt{\text{s}}$ of 0.9 (2010 and 2021), 2.76 (2011 and 2013) and 5.02 TeV (2015 and 2017). A timeline on these data taking is shown in Figure 3.2.

The luminosity which LHC has delivered for each year is shown in Figure 3.3. Data taking at 13.6 TeV to almost reach the design luminosity is planned between 2022 and 2024. In this work, we mainly focus on the analysis of $\sqrt{s} = 7$ TeV (2010), which has $\mathcal{L}$ of $2{\times}10^{32}\text{cm}^{-2}\text{s}^{-1}$ and L of 45 pb^{-1}.

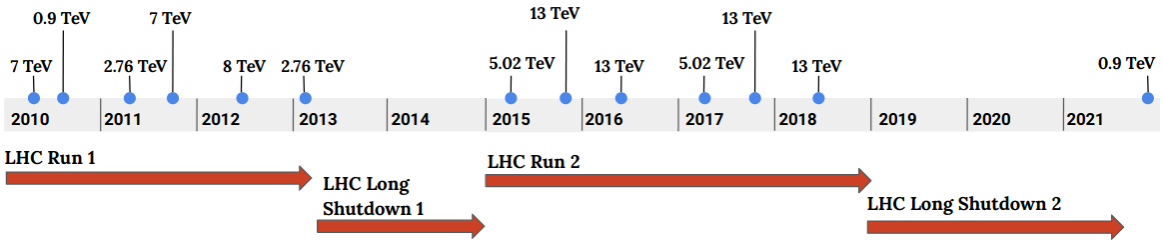


Figure 3.2: CMS data taking timeline throughout all years. The analysis that the author worked on is at 7 TeV 2010. This measurement is part of a bigger project aiming to measure charm cross section for all center of mass energies.

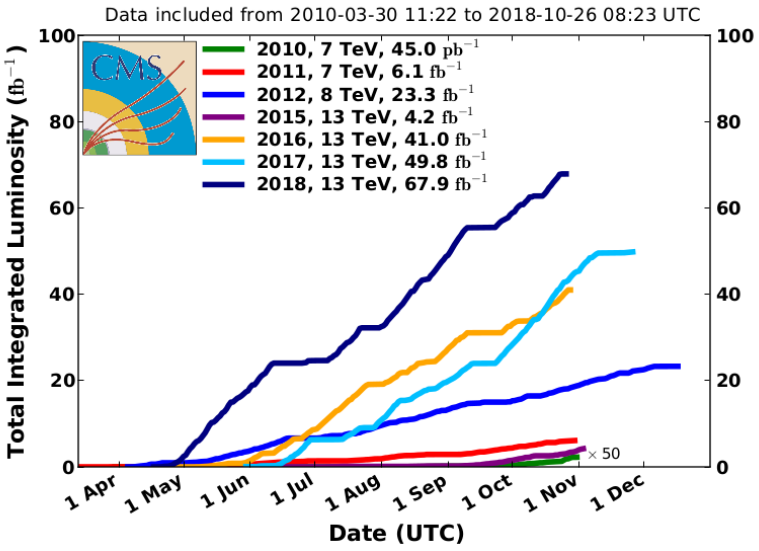


Figure 3.3: Integrated luminosity delivered by the LHC for each year data taking [8].

3.2 The CMS Experiment

The CMS detector [5] is a cylindrical detector that is the heaviest detector at the LHC. It looks like an onion-like structure in the x-y plane that consists of several sub-detectors starting from inside out: tracker, electromagnetic calorimeter, hadronic calorimeter, superconducting solenoid, and lastly, muon chamber that enclose the detector. Each part of the detector has a role in identifying particles.

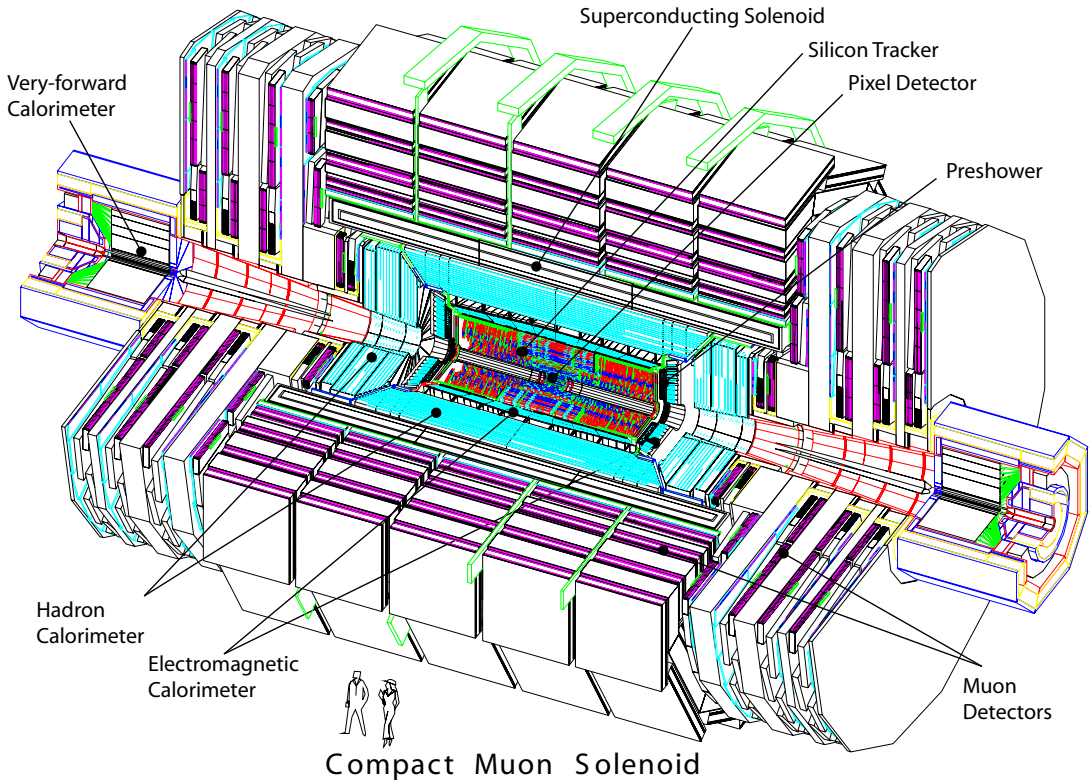


Figure 3.4: Schematic view of the CMS detector showing the different components. Picture taken from [5].

3.2.1 Coordinate system

CMS uses a right-handed coordinate system where the origin is at the nominal interaction point. The x-axis points to the LHC center, the y-axis points upwards, and the z-axis lies along the proton beam direction. The radial coordinate, r, is the distance from the interaction point within the transverse plane. The azimuthal angle, ϕ, is measured from the x-axis, and the polar angle, θ, is measured relative to the $+z$-axis, referred to as the forward direction. These quantities are illustrated in Figure 3.5. Since physics analysis is mainly dealing with momenta, the coordinate system $\{x, y, z\}$ will be in terms of momentum space, $\{\mathrm{p}_x, \mathrm{p}_y, \mathrm{p}_z\}$, and $\{r, \theta, \phi\}$ will be in terms of $\{\mathrm{p_T}, \eta, \phi\}$. Transverse momentum, $\mathrm{p_T}$, is defined as

$$\mathrm{p_T^2} = \mathrm{p}_x^2 + \mathrm{p}_y^2 \tag{3.2}$$

and the pseudorapidity, η, is given by:

$$\eta = -\ln(\tan\frac{\theta}{2}) \tag{3.3}$$

η is chosen instead of θ as $\Delta\eta$ stays invariant under Lorentz boosts along the z-axis for massless particles while $\Delta\theta$ is not. Another interesting quantity is rapidity, y, which is defined as:

$$\mathrm{y} = \frac{1}{2}\ln(\frac{\mathrm{E} + \mathrm{p}_z}{\mathrm{E} - \mathrm{p}_z}) \tag{3.4}$$

where E is the energy of the particle. Δy is not only invariant under the Lorentz boost along the z-axis for massless particles but also for massive particles. Furthermore, if the mass of the particle is zero or extremely small compared to its momentum, $\eta \simeq$ y.

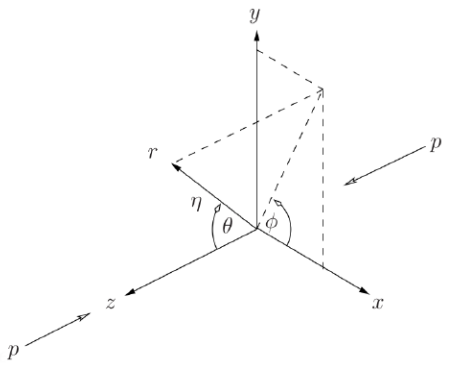


Figure 3.5: Illustration of the coordinate system in a proton-proton collision.

3.2.2 Superconducting magnet

The basic principle of magnets in collider experiments is to bend charged particles to determine the charged particle's momentum and electric charge type. A particle with higher momentum gives less curvature in the trajectory, and the direction of the particle's bend gives information on the particle's positive or negative charge.

The CMS magnet has a length of 12.5 m, and a 6 m bore diameter with 12,000 tonnes of mass. It surrounds the tracker and calorimeter detectors (hadronic and electromagnetic). The muon detector is interleaved in 3 layers with the 12-sided iron structure to guide the magnetic field outside of the solenoid [9]. The magnet generates a magnetic field of 3.8 T, to measure the momenta of particles. Figure 3.6 shows the general view of the CMS magnet.

3.2.3 Tracker

The finely segmented CMS tracker is located closest to the interaction point [11]. It is designed to achieve two things: precise and efficient measurement of charged particle trajectories from the collisions and precise reconstruction of secondary vertices. The tracker consists of a pixel detector and a strip detector, and both are made of silicon sensors, covering $|\eta| <$ 2.5 in total. The pixel detector is the innermost part of the tracker. It consisted in Run 1

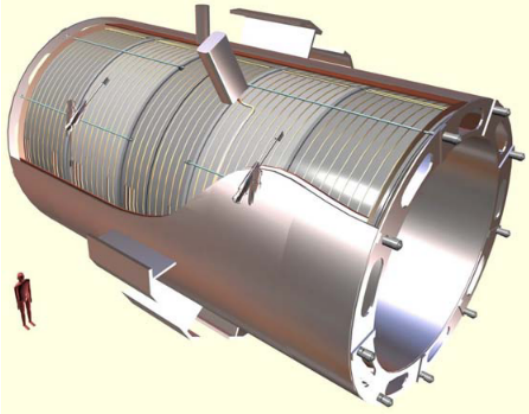


Figure 3.6: General view of CMS magnet compared to human size. Picture taken from [10].

and 2015-2016 of 3 barrel layers (BPix) and two endcap disks (FPix) in the forward region. The strip detector consists of several subparts starting with the Tracker Inner Barrel (TIB) and the Tracker Inner Disk (TID), composed of 4 barrel layers and 3 disks at each end. It is enclosed by the Tracker Outer Barrel (TOB) with 6 layers and the Tracker Endcap (TEC) with 9 layers on each side.

An upgraded pixel detector [12] was installed at the end of 2016. A new pixel read-out chip was implemented to minimize a significant amount of data loss beyond 25 ns per bunch crossing and degradation due to radiation damage. Second, an additional lightweight pixel layer was added on each barrel and endcap to reduce the negative effect of high pile-up and to avoid degradation in performance due to detector material.

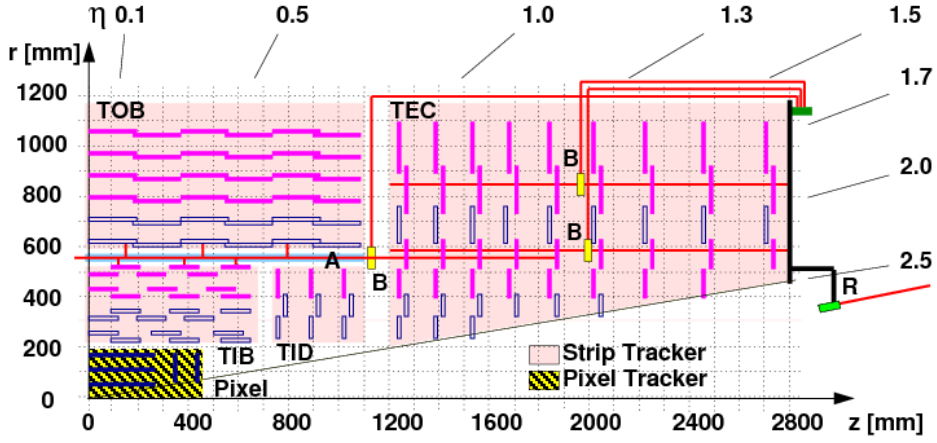


Figure 3.7: A quarter of the tracker schematic diagram in the r-z plane. The figure also illustrates the paths of the laser rays (R), the alignment tubes (A) and the beam splitter (B) of the laser alignment system [11].

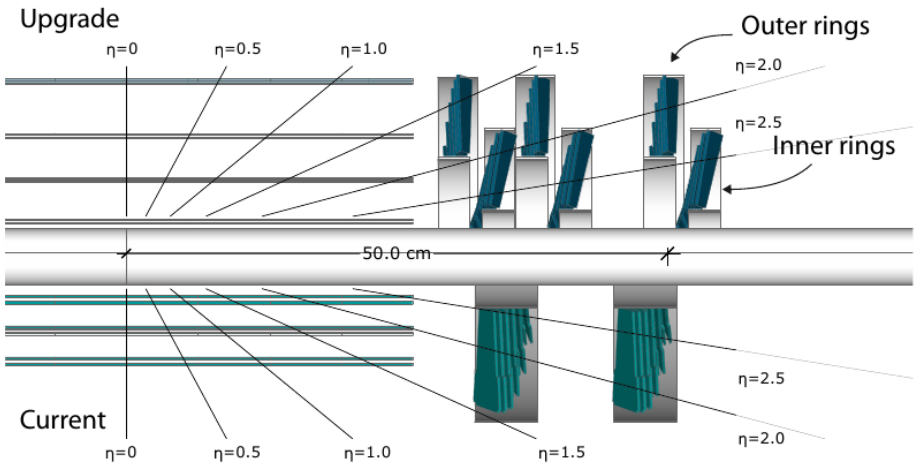


Figure 3.8: Upgraded pixel detector layout in barrel and endcap (top) compared to the old one (bottom), relevant for this thesis.

3.2.4 Electromagnetic calorimeter

The Electromagnetic CALorimeter (ECAL) [5] is a hermetic and homogeneous calorimeter to detect electrons and photons. It is located between the silicon tracker and the hadronic calorimeter. There are two parts of ECAL, namely the Barrel ECAL (EB) and the Endcap ECAL (EE). Both parts are made of lead tungstate ($PbWO_4$) crystal. In front of the endcap crystals, a Preshower detector (ES) is placed. A schematic diagram of the ECAL is shown in Figure 3.9.

Lead tungstate crystals with fine granularity were chosen as the material for ECAL due to its characteristics: high density, short radiation length, and radiation-hard crystals, which suit the demand during LHC operation conditions: high magnetic field, high radiation, and only 25 ns per bunch crossing. The EB covers $|\eta| < 1.479$, and the EE provides $|\eta|$ coverage between 1.479 and 3.0. The ES is a silicon-based sensor that aims to detect neutral pions in the endcaps between $1.653 < |\eta| < 2.6$. It also allows CMS to distinguish between high and low-energy photons.

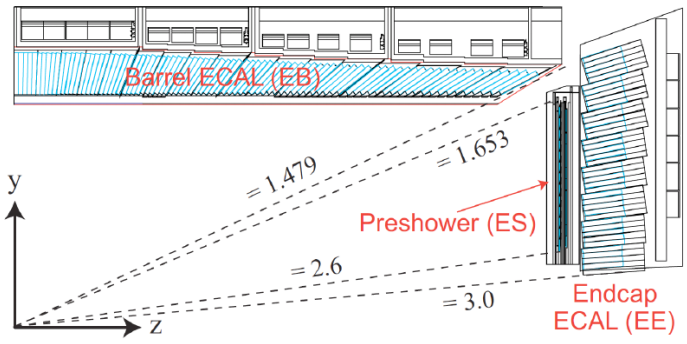


Figure 3.9: Schematic diagram of the CMS ECAL

3.2.5 Hadron calorimeter

The Hadronic CALorimeter (HCAL) [5] is a hermetic sampling calorimeter designed to detect charged and neutral hadrons. There are different subdetectors in HCAL, namely Hadron Barrel (HB), Hadron Endcap (HE), Hadron Outer (HO), and Hadron Forward (HF) shown in Figure 3.10. HB and HE are located behind the ECAL, covering $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$ respectively. The material is made with alternating layers of brass absorber and plastic scintillator. Brass has been chosen as absorber material because it has a short interaction length, is easy to machine, has reasonable cost, and is non-magnetic. The plastic scintillator is chosen for its long-term stability and moderate radiation hardness. HO is placed outside the solenoid to complement the HB as a tail catcher. They are used to identify late starting showers and to measure the shower energy deposited after the HB. Lastly, HF is sitting at 11.2 m from the interaction point and covering $3 < |\eta| < 5.2$. It is made of steel absorbers and embedded radiation hard quartz fibers, which provide a fast collection of Cherenkov light. This is because HF captures bulk of particles energies at a shallow angles relative to the beam line.

3.2.6 Muon system

As the name suggests, the muon system [5] is used to detect muons by measuring their momenta. There are three different chambers located in different detector regions, as shown in Figure 3.10. Drift Tube (DT) chambers are located in the barrel region, covering $|\eta| < 1.2$. This is the region where the neutron-induced background, the muon rate, and the residual magnetic field are low. The chambers are organized in 4 layers of Muon Barrel (MB). Cathode Strip Chambers (CSC) are located at the endcap region, covering $|\eta| < 2.4$. In addition to this, Resistive Plate Chambers (RPC) are used in both the barrel and the endcap regions.

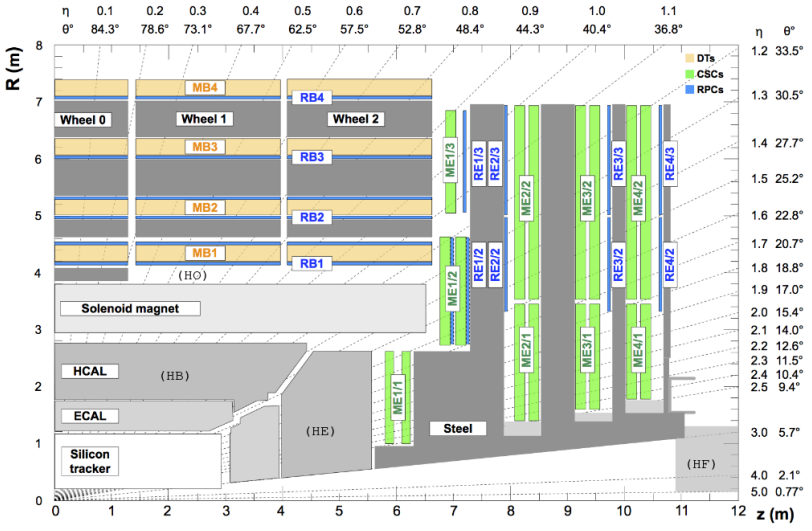


Figure 3.10: A quadrant of the CMS detector showing HCAL's subdetectors and muon system. DT stations are labeled as MB 1-4, starting from the innermost layer. CSC is labeled as ME (Muon Endcap), and RPC is labeled as RB and RE in both regions

3.3 Data Acquisition and Processing

3.3.1 Trigger system

The proton-proton collision rate at the LHC is exceptionally high, and corresponds to a nominal frequency of 40 MHz, i.e. a proton bunch crossing every 25ns. However, it is impossible to record and store all of these events with the current technology; thus, only a small fraction is kept for further offline analysis. It is the job of a trigger system to reduce the rate by selecting potentially interesting physics events. In CMS, the event rate was reduced at two levels; Level 1 (L1) trigger and High Level Trigger (HLT) [13].

The L1 trigger is implemented in custom hardware that takes information from the calorimeters and the muon system. To do this, it identifies and uses leptons, photons, jets, and missing transverse energy. The event rate is reduced to 100 kHz before transfer to HLT. HLT is implemented in a custom C++ software to further refine the physics objects reconstruction and selection with an average rate of 400 Hz for offline storage [5].

The overall output rate of the L1 trigger and HLT can be adjusted by prescaling the number of events that pass the selection criteria of specific algorithms. For example; Zero Bias and Minimum Bias triggers used in this analysis. The Zero Bias triggers require only the LHC machine beam crossing signal and are completely unbiased by anything going on in the detector while Minimum Bias triggers require some very generic detector signals, mostly forward instrumentation inclusively sensitive to the proton remnant(s) of inelastic events. Both triggers contribute to very high rates, therefore they are prescaled. A prescale of N, where N is a positive integer, means that the triggers record only 1/N events passing them. If the prescale is 1, the triggers record all of the events, and if 0, it means that the trigger is turned off.

3.3.2 Offline analysis

The events recorded by the HLT are transferred to the Tier-0 computing center at CERN to be analyzed [14]. Tier-0 is responsible to secure the raw data, perform the prompt offline reconstruction and distribute the reconstructed data to the Tier-1 centers. These Tier-1 centers are long-term data storage sites, hosted and operated by large collaborating national laboratories and computing centers around the world. CMS physicists perform selection and further reconstruction on this data and provide them to the Tier-2 centers hosted at associated CMS institutes. Tier-2 centers are designed mainly to support the final-stage analyses of reduced datasets and for the production of Monte Carlo samples (see Chapter 4). Lastly, Tier-3 centers are modest facilities at institutes for local use. One of the usage of these centers is to act as a medium for the Open Data transfer within End of Service (EOS) public areas. EOS public area is the place where the datasets used in Chapter 5 and Chapter 6 are stored. Data reprocessing and computing centers can be summarized in Figure 3.11.

CMS introduced several data formats with different level of detail, precision and therefore size per event. Starting from the `RAW` data format containing the full recorded information from the detector after online formatting, the L1 trigger result and the HLT selections. `RAW`

data is used as an input to the Tier-0 reconstruction. The next data format is RECO. This format is produced by applying several levels of pattern recognition and algorithms to the RAW data. It contains high-level physics objects such as tracks, vertices, electrons and muons in addition to low level information such as detector hits and supports re-fitting of tracks as discussed in Chapter 4. However, the RECO format is still large in size, therefore a derived data format is provided for physics analysis. The Analysis Object Data (AOD) format is a compact version of the RECO format, chosen to satisfy the needs of a large fraction of analyses. It contains the parameters of high-level physics objects plus sufficient additional information to allow kinematic refitting.

From 2015, the colliding protons' energy has increased to 13 TeV, and ten times more data were collected; therefore, they needed vastly more CMS storage resources. This challenge was tackled by developing a condensed data format called MINIAOD [15]. As the name implies, the size of this data format is smaller (30 - 50 kb/event) than the AOD. Some of the main contents of MINIAOD are high-level physics objects with some preselection requirement applied on the objects, and all particle candidates reconstructed by the CMS Particle Flow (PF) algorithm [16, 17] that store only basic kinematic information for each objets. Up to MINIAOD, the data is fully organized in the so-called Event Data Model (EDM) which is part of the CMS framework and requires the use of CMSSW(CMS SoftWare).

A further reduced format called NANOAOD [18] is introduced in the same spirit as MINIAOD, to enable the analysis of much larger datasets throughout each year 2016-2018 (see Figure 3.2). NANOAOD is smaller ($\sim$1 kB per event) than MINIAOD, contains only top level information typically used in the last steps of physics analysis, but at the same time maintains the physics analysis capability to be studied. Since NANOAOD is only available starting from the 2016 Run 2 data, an ongoing effort in producing a NANOAOD-like format called NANOAODplus for Run 1 data is conducted under the Data Preservation and Open Access (DPOA) group [19]. Unlike NANOAOD which is created from MINIAOD, NANOAODplus is defined directly from AOD (see Figure 3.12). The aim of NANOAODplus is to function consistently with NANOAOD; plus including additional AOD variables and extensions (hence the named NANOAODplus). The advantage of using a NANO data format compared to the other data formats mentioned above is its independence from any particular CMSSW version and most importantly it is a flat ntuple format readable with plain ROOT [20]. Contributions to the NANOAODplus development were one of the author's service tasks.

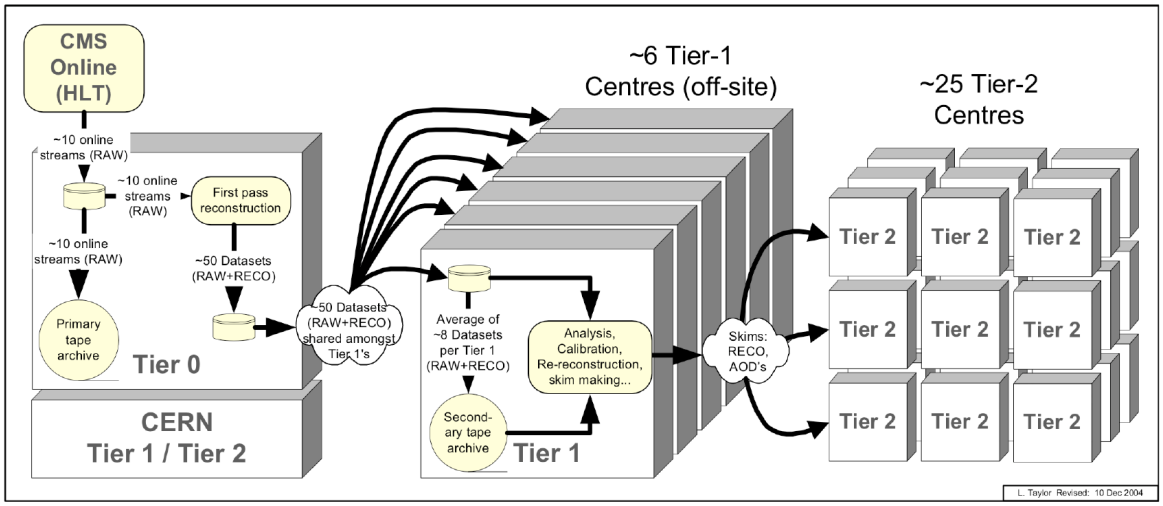


Figure 3.11: Schematic flows of CMS events in different computing centers with varying degree of detail, size and refinement [14].

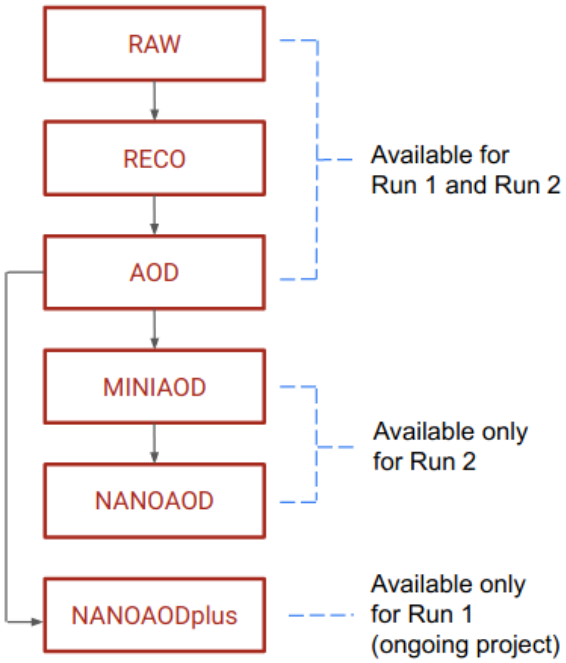


Figure 3.12: Overview of data tiers in CMS data format. The arrow represent the flow of data format creation. NanoAODplus is an ongoing project under the Data Preservation and Open Access (DPOA) group.

3.4 Bibliography

[1] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (Aug, 2008) S08001–S08001, doi:10.1088/1748-0221/3/08/s08001.

[2] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[3] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[4] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[5] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[6] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN", (Dec, 2008). https://cds.cern.ch/record/1260465.

[7] C. Pralavorio, "Record luminosity: well done LHC", (2017). https://home.cern/news/news/accelerators/record-luminosity-well-done-lhc.

[8] CMS Collaboration, "Public CMS luminosity information", (2019). https://twiki.cern.ch/twiki/bin/view/CMSPublicLumiPublicResults.

[9] CMS Collaboration, "Public CMS superconducting magnet information", (2011). http://cms.web.cern.ch/news/superconducting-magnet.

[10] F. Kircher et al., "Final design of the CMS solenoid cold mass", *IEEE Transactions on Applied Superconductivity* **10** (2000) 407–410, doi:10.1109/77.828259.

[11] CMS Collaboration, "Alignment of the CMS tracker with LHC and cosmic ray data", *JINST* **9** (2014) P06009, doi:10.1088/1748-0221/9/06/P06009, arXiv:1403.2286.

[12] A. Dominguez et al., "CMS Technical Design Report for the Pixel Detector Upgrade", Technical Report CERN-LHCC-2012-016, Sep, 2012.

[13] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[14] C. Grandi et al., "CMS Computing Model: The "CMS Computing Model RTAG"", Technical Report CERN-NOTE-2004-031, CERN, Geneva, Dec, 2004.

[15] CMS Collaboration, "Mini-AOD: A New Analysis Data Format for CMS", *J. Phys. Conf. Ser.* **664** (2015), no. 7, 7, doi:10.1088/1742-6596/664/7/072052, arXiv:1702.04685.

[16] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", Technical Report CMS-PAS-PFT-10-001, 2010.

[17] CMS Collaboration Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr, 2009.

[18] CMS Collaboration, "A further reduction in CMS event data for analysis: the NANOAOD format", *EPJ Web Conf.* **214** (2019) 06021, doi:10.1051/epjconf/201921406021.

[19] CMS Collaboration, "Simplified legacy data formats for Run1", (2021). https://twiki.cern.ch/twiki/bin/view/CMS/DataPreservationOpenAccess.

[20] CMS Collaboration, "NanoAODplus documentation", (2021). https://twiki.cern.ch/twiki/bin/viewauth/CMS/DPOANanoAODlike.

CHAPTER

4

OBJECT RECONSTRUCTION AND EVENT SIMULATION

Physics object reconstruction is done to identify the particles produced from the LHC proton-proton collisions and measure their momentum, direction, charge etc.. The object reconstruction is crucial as it determines whether analyses can be conducted or not. Furthermore, it also applies implicit cuts in the reconstruction to have better acceptance and efficiency for the majority of the analyses. This chapter covers only the reconstruction of objects involved in this thesis.

Starting from the proton-proton beam interaction, particles first enter the tracker, in which charged-particle trajectories (tracks) and origin (vertices) are reconstructed from signals (hits) in the sensitive layers. The details of track and vertex reconstruction are described in Section 4.1 and Section 4.2, respectively. The algorithms for muon and electron reconstruction used for Particle-Flow (PF) and in Chapter 5 are briefly explained in Section 4.3.

To ensure the implementation of object reconstruction is working properly and the result behaves as expected, physics event generators and detector simulation are needed. The Monte Carlo (MC) simulation uses the MC method to model physics processes, leading to a sample of simulated events. Often, the MC simulation is used to estimate detector acceptance, resolution effects, and background contributions. In addition, it is used in the optimization of event selection criteria and to compare the consistency of different predictions against measurements. Section 4.4 will discuss MC production and some details of the treatement of charm quarks in MC.

4.1 Track reconstruction

The track is one of the basic objects that is reconstructed in the CMS detector. It is reconstructed by combining the information of hits that formed trajectories to estimate the track momentum and direction. Hits are clusters of signals above certain thresholds in the

tracker's sensor. The information about the hits is used in the track reconstruction.

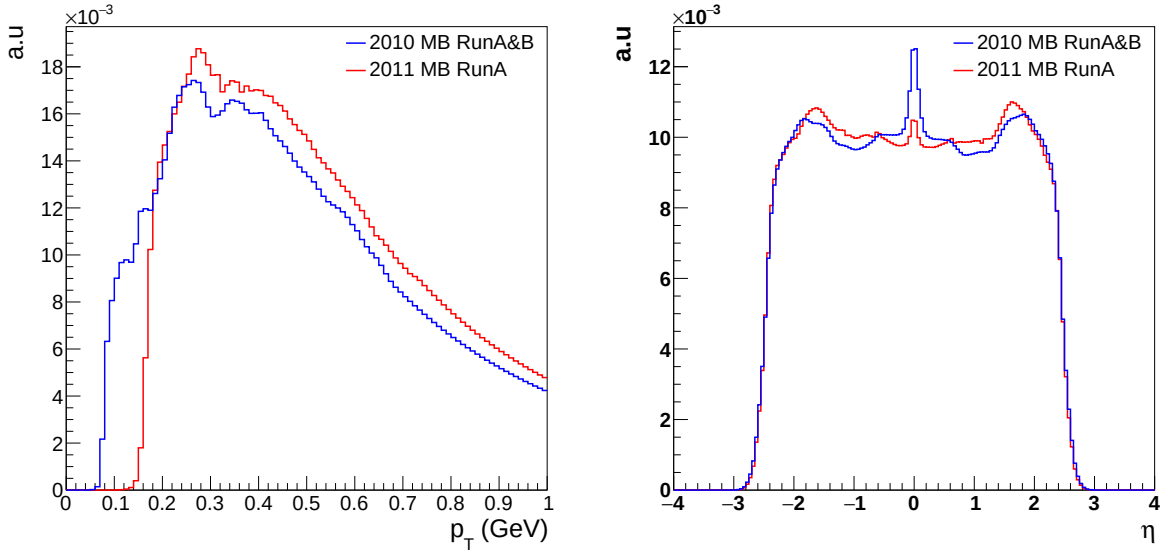
In CMS, a track is reconstructed using a tracking algorithm called Combinatorial Track Finder (CTF) [1]; an adaptation of the combinatorial Kalman Filter [2]. The algorithm allows pattern recognition and track fitting to occur in the same framework. The reconstructed tracks are produced iteratively considering the pileup condition (higher pileup, more iterations). The initial iterations start with a search for tracks that are easiest to find, such as tracks with relatively large p_T and produced near the interaction region. After each iteration, hits associated with the tracks are removed, thereby reducing the combinatorial complexity and simplifying subsequent iterations in a search for more difficult track classes such as low-p_T tracks.

Each iteration has four main steps. The first is the seed generation that provides initial track candidates found using only 2-3 hits. The seeds are constructed using the inside-out approach, starting from the inner region to the tracker's outermost region. This approach is chosen because of the pixel detector's high granularity which ensures that the fraction of channels that are hit in the pixel layers is lower than in the strip layers. A pixel layer produces a 3-D spatial measurement, which provides constraints and estimates of trajectory parameters. High efficiency for reconstructing tracks is gained, especially for the low-momentum tracks that are deflected by the strong magnetic field before reaching the outer tracker layer and for electrons that lose a significant fraction of their energy due to bremsstrahlung radiation in the tracker.

The second step is the track finding based on the Kalman Filter technique [2]. It extrapolates the seed trajectories along the expected flight path of a charged particle and searches for additional hits that can be assigned to the track candidates. A search for additional hits uses an outside-in approach, starting by taking all the hits assigned to the track, excluding those which already belong to the track seed, and using them to fit the track trajectory. The trajectory is propagated inwards through the seeding layer until the tracker's inner edge is reached or too many ghost hits[1] are found.

The third step is track fitting to smooth out the trajectories, thus providing the best possible estimates of the trajectory parameters. The result is used to extract the track kinematics in the last step named track selection. In this step, only tracks that satisfy good quality tracks are kept, and the rest is discarded. The selection is based on the number of layers that have hits, whether their fit yielded a good χ^2/dof, and how compatible they are with originating from a primary interaction vertex. If several primary vertices are present in the event, all tracks are considered for reconstruction. The track information is stored in `reco :: TrackCollection` for offline analysis, and this analysis used the `generalTracks` input tag from this collection. Additionally, the analysis also used `TransientTracks` objects as they have access to the magnetic field and detector geometry allowing their parameters to be refitted. Figure 4.1 and Figure 4.2 show the track p_T and η distributions for 2010 – 2017.

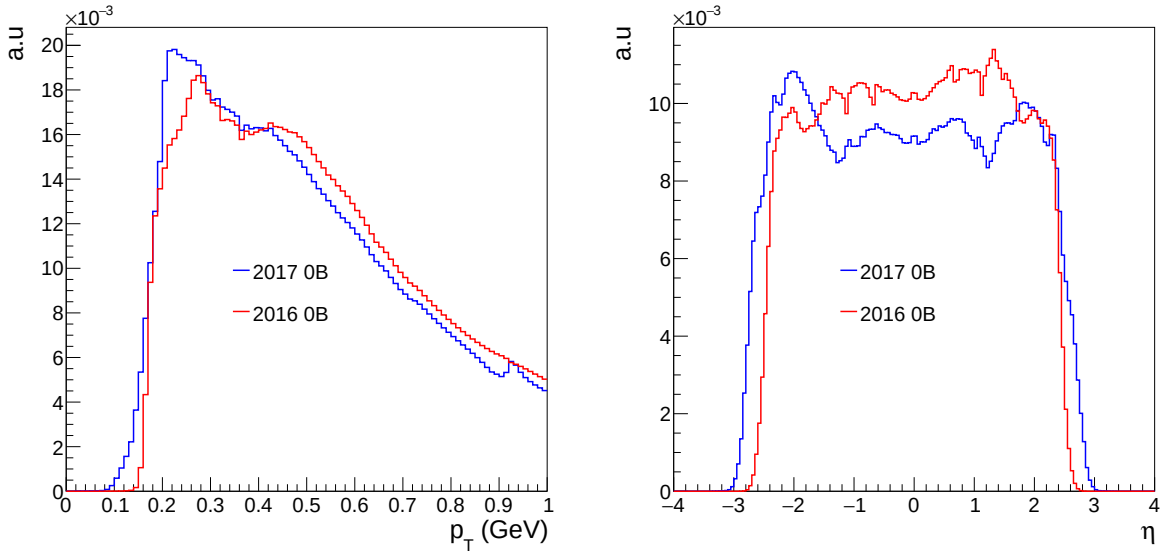
[1] Ghost hits are defined as no hits found in some of the tracker's layer.



(a) Track p_T distribution

(b) Track η distribution

Figure 4.1: Track p_T and η distributions comparison between 2010 and 2011 Minimum Bias (MB) data. Both distributions are normalized to one for shape comparison. In (a), it can be seen that the low track p_T acceptance is better in 2010. In (b), the peak at $\eta = 0$ might be caused by low momentum loopers.



(a) Track p_T distribution

(b) Track η distribution

Figure 4.2: Track p_T and η distributions comparison between 2016 and 2017 Zero Bias (0B) data. Both distrbutions are normalized to one for shape comparison. In (a) and (b), it can be seen that the track p_T and η acceptance is slightly better at low p_T and large pseudorapidity in 2017 data due to the new pixel tracker that was installed before the 2017 data taking [3].

4.2 Primary vertex reconstruction

The primary vertex (PV) is defined as the point where protons collide with each other leading to particles from the collision being scattered or decayed. The PV reconstruction aims to measure the location and uncertainty of all proton-proton interaction vertices in each event using the available reconstructed tracks. At the LHC, multiple primary vertices are usually produced per event in the same bunch crossing due to multiple proton-proton interactions. Since most of the CMS analyses are interested in hard scattering events, a main PV is defined to be the PV with the largest sum of p_T^2 of tracks associated to it. Pileup vertices are the remaining primary vertices reconstructed in the event.

PV reconstruction consists of three steps; selection of the tracks, clustering of the tracks that appear to originate from the same interaction vertex, and fitting for the position of each vertex using its associated tracks [1]. Track selection involves choosing prompt tracks originating from the primary interaction region. This is done by imposing requirements on the significance of the transverse impact parameter[2] with respect to the beam spot[3], the number of hits in the strips and pixel tracker, and the normalized track χ^2 from a fit to the trajectory. The selected tracks are then clustered based on their distance in the z coordinate using a Deterministic Annealing (DA) algorithm [4] starting from 2011 onwards. During 2010, a simple 'gap clustering' algorithm [5] was used. Vertex candidates containing at least two tracks are fitted using an adaptive vertex fitter [6] to compute the best estimate of the vertex parameters, e.g., the position of the vertex. The vertex information is stored in `reco` :: `VertexCollection`, e.g. offline primary vertices without and with beamspot (with input tag `offlinePrimaryVertices` and `offlinePrimaryVerticesWithBS` respectively). Both vertex types were reconstructed using the tracks taken from the `generalTracks` collection; the latter imposed the beam spot as a constraint in the fit of the vertex position [7]. However, the PV reconstruction does not differ between hard and soft scattering events. Figure 4.3 shows the average number of pileup for each year in pp collisions.

4.3 Particle Flow

Since the CMS detector is a general-purpose detector(see Section 3.1), it is intended to identify and reconstruct individually each particle produced from the LHC proton-proton collisions. Such reconstruction can be achieved by using a Particle-Flow (PF) event reconstruction algorithm [9]. The PF algorithm aims to identify all stable particles in the event, i.e., muons, electrons, charged hadrons, neutral hadrons, and photons, by combining information from all CMS sub-detectors, thus optimizing the determination of particle types, directions, and energies. A muon is the only visible particle that traverses the entire detector, producing hits in the muon system before escaping the CMS detector. Electrons and charged hadrons also deposit their energies in ECAL and HCAL, respectively. Photons and neutral hadrons deposit energy in ECAL and HCAL without producing hits in the tracker. A sketch for the detection of these stable particles is shown in Figure 4.4. The neutrino, which is not in this figure, is one of the invisible particles that does not interact with any detector material. However, missing momentum can indicate its presence in the event. Some

[2]The impact parameter is the shortest distance between a track and some object, e.g., beam spot

[3]The beam spot is the region where the most pp collisions take place in the detector, detail in Ref. [1]

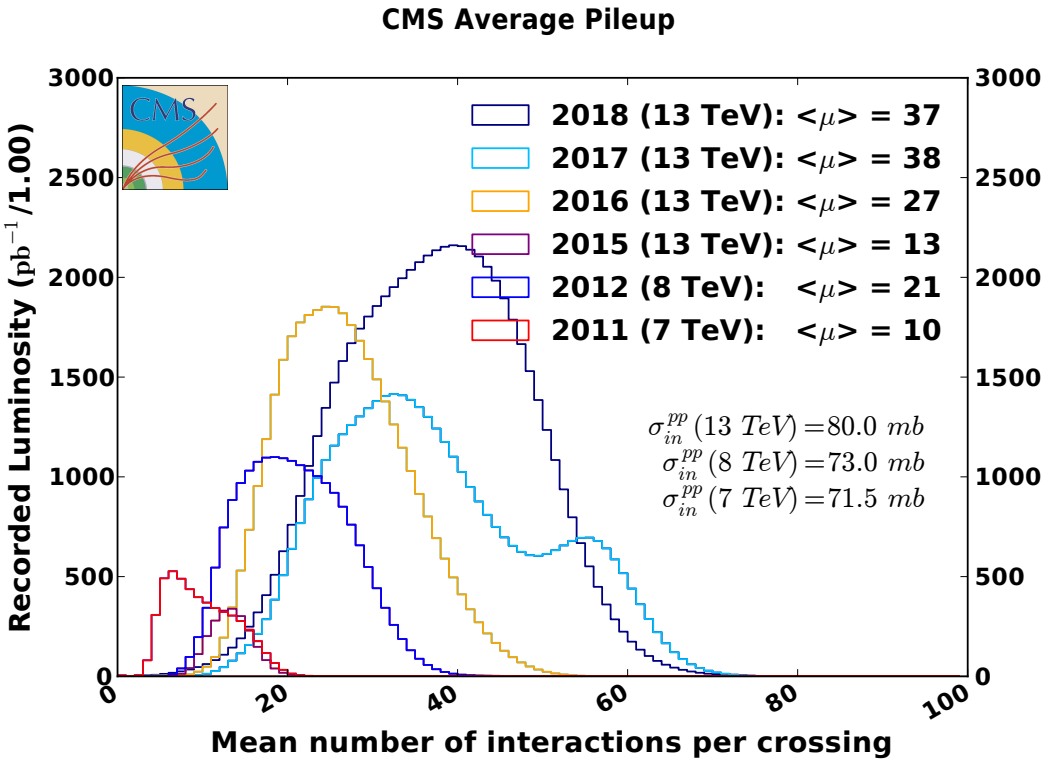


Figure 4.3: Distribution of average number of interaction per crossing (pileup) in 2011 – 2018. The overall mean value is denoted as $<\mu>$ and the cross section is the Minimum Bias cross section which is taken from theoritical predictions. Picture taken from [8].

muons and electrons for example, those from W and Z boson decays, are isolated. Isolated here means the muons and electrons happen away from other activities in the event. One particular isolation[4] variable is defined as:

$$RelPFiso = \frac{\Sigma \text{chargedhadron}_{p_T} + \Sigma \text{neutralhadron}_{p_T} + \Sigma \text{photon}_{p_T}}{p_T^{lepton}} \tag{4.1}$$

4.3.1 Muon reconstruction

Muon reconstruction in CMS uses input from the tracks in the inner tracker and standalone-muon tracks[5] in the muon system. Two approaches are used to reconstruct the muon [10]:

- **Global Muon reconstruction** (outside-in): The muons are reconstructed combining information from the standalone and tracker muon track and fitted using the Kalman filter technique [11]. The global-muon fit can improve the momentum resolution at large transverse momenta.

[4]The variable isolation is defined by performing the scalar sum of the transverse momentum of the particle flow candidates reconstructed in a ΔR cone of 0.4 (CMS Note AN-12-141).

[5]Standalone-muon tracks are tracks reconstructed using the information from the muon system only.

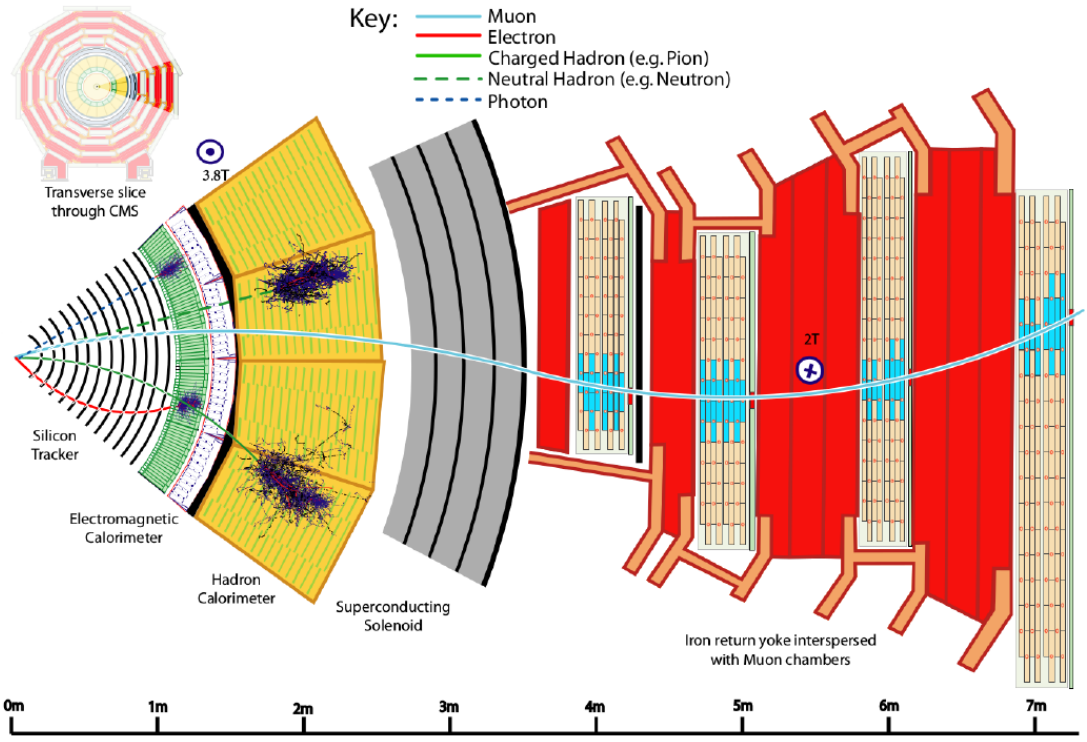


Figure 4.4: Transverse slice of the CMS detector showing how each particle types interact with each subdetector. Bending of charged particle trajectories is due to the magnetic field. In this figure, the muon and charged hadron are positively charged, and the electron is negatively charged.

- **Tracker Muon reconstruction** (inside-out): All tracker tracks with $p_T > 0.5$ GeV and total momentum, $p > 2.5$ GeV are considered as possible muon candidates and are extrapolated to the muon system. If more than one muon segment matches the extrapolated track, this track qualifies as a Tracker Muon. Tracker Muon reconstruction is more efficient at lower momenta, $p < 5$ GeV, because it requires only a single muon segment in the muon system.

Muon identification in the PF algorithm is based on a set of selections of global and tracker muon properties. It is used for the Higgs to four lepton analysis in Chapter 5.

4.3.2 Electron reconstruction

Electrons are reconstructed by associating a cluster of energy deposited in ECAL with a track. The first step of electron reconstruction is called seeding, which consists of finding and selecting the two or three first hits in the tracker from which the track can be initiated. There are two complementary algorithms used to generate the electron track seeds [12]

- **ECAL-based seeding**: Starts from the supercluster[6] energy and position, used to extrapolate the electron trajectory towards the collision vertex.

- **Tracker-based seeding**: Starts from the hits in pixel tracks, used to extrapolate the electron trajectory towards the ECAL and match to the supercluster.

The selected electron seeds are used for electron track reconstruction and fitted using the Gaussian-Sum Filter (GSF) [14] algorithm. This algorithm is chosen as it models the energy losses due to bremsstrahlung[7] in the tracker layers.

4.4 Event Generator and Simulation

The study of partons in pp collisions is complex. In order to ensure the measurement from the data collected in the collision is reliable, a comparison with Monte Carlo (MC) simulation is needed. Often, the MC simulation is used to estimate detector acceptance, resolution effects, and background contribution. In addition it is used in the optimization of event selection criteria and to compare the consistency of different predictions against measurements. The MC simulation uses the Monte Carlo method in the simulation programs to model physics processes, leading to a sample of simulated events.

The event is generated starting from a matrix element (see Section 2.2 and Section 2.3). It is calculated perturbatively up to some fixed order. It is followed by parton showering to approximately account for higher order effects, including initial and final state radiation[8]. Next, hadronization takes place, and the final state partons are transformed into hadrons. Hadronization cannot be computed perturbatively therefore it is performed by using some models such as the Lund string model [15] which are tuned to the data. These hadrons can further decay to other particles if they are unstable. Figure 4.5 shows a pictorial representation behind pp collisions. There are several examples of MC generators such as POWHEG [16], MADGRAPH [17] and PYTHIA [15, 18]. POWHEG and MADGRAPH generate events at matrix element level and usually interface with PYTHIA for parton shower and hadronization. However, PYTHIA is also a multipurpose tool, that is able to generate events in stand alone mode when matrix elements are calculated at LO.

Event generation is followed by detector simulation. This is done using GEANT4 [19] to simulate the interactions of particles in the event with detector elements, leading to a set of simulated detector hits. The hits are then passed through the object reconstruction chain in a similar manner as with real data, in order to ensure that the two are as similar as possible.

[6] A supercluster electron is an electron that is built up by merging a group of energy clusters in the ECAL of similar η extended in ϕ [13]

[7] When electrons propagate in the detector, they emit a sizeable fraction of their energy in the form of bremsstrahlung due to the thickness of the tracker layer, and the spectrum of this radiation is highly non-Gaussian; thus the KF algorithm is not optimal for electron track reconstruction.

[8] Final state radiation is where the scattered particles from the partons emit radiation causing their energy to decrease after the splitting. Initial state radiation is where the partons emit radiation before it interacts with other particles.

4.4.1 Treatment of charm in simulation

The main project of this thesis uses PYTHIA6 to simulate charm production (see Subsection 2.4.1) in pp collisions. Since the previously available MC samples for 7 TeV 2010 had low statistics, a study on the charm MC generation was conducted to provide the MC samples for this project. PYTHIA6 (version 424) is used to generate inelastic parton scattering events. The events are selected such that they receive contributions mainly from QCD subprocesses that include quark or gluon scattering and also single and double diffractive (see Section 2.3) Minimum Bias subprocesses. Both $c\bar{c}$ and $b\bar{b}$ events are produced. In the $b\bar{b}$ events, a $c\bar{c}$ pair can also be produced in addition from parton showering or gluon splitting (which are described by the DGLAP evolution [20–22]), so a clean conceptual separation of charm and beauty is not possible. Therefore, the events were generated treating inclusive charm and beauty samples together, and charm or beauty filters were partially applied only at the parton level. Due to the larger beauty mass which gives a smaller cross section, the fraction of charm and beauty event is 90% and 10%, respectively. The events also are not required to have minimum and maximum p_T as the analysis is mostly interested in the production near threshold. At the hadron level, several tunes are available to simulate underlying events [23] and TuneZ2star is commonly used for PYTHIA6. One component of this tune for heavy flavor is the Bowler fragmentation function (see Subsection 2.4.2). Figure 4.6 shows the comparison of different tunes in MC at 7 TeV for a validation study. The generated MC with charm filter has a higher fraction of D^0 and D^* in all events and their p_T and y at truth level is comparable with other available MCs [24]. A MC sample of explicit D^0 meson decay to K and π is also generated to compare with the data and theoretical predictions. A brief description on charm fragmentation function is mention in Subsection 2.4.2. In order to produce the samples needed for this analysis, the author has directly worked with the CMS generator group, performed validation studies, and prepared all the necessary CMSSW configuration files.

4.5 Bibliography

[1] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[2] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the Kalman filtering method", *Nucl. Instrum. Meth. A* **294** (1990) 219–228, doi:10.1016/0168-9002(90)91835-Y.

[3] A. Dominguez et al., "CMS Technical Design Report for the Pixel Detector Upgrade", Technical Report CMS-TDR-11, Sep, 2012.

[4] E. Chabanat and N. Estre, "Deterministic annealing for vertex finding at CMS", in *14th International Conference on Computing in High-Energy and Nuclear Physics*, pp. 287–290. 2005.

[5] CMS Collaboration Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", Technical Report CMS-PAS-TRK-10-005, CERN, Geneva, 2010.

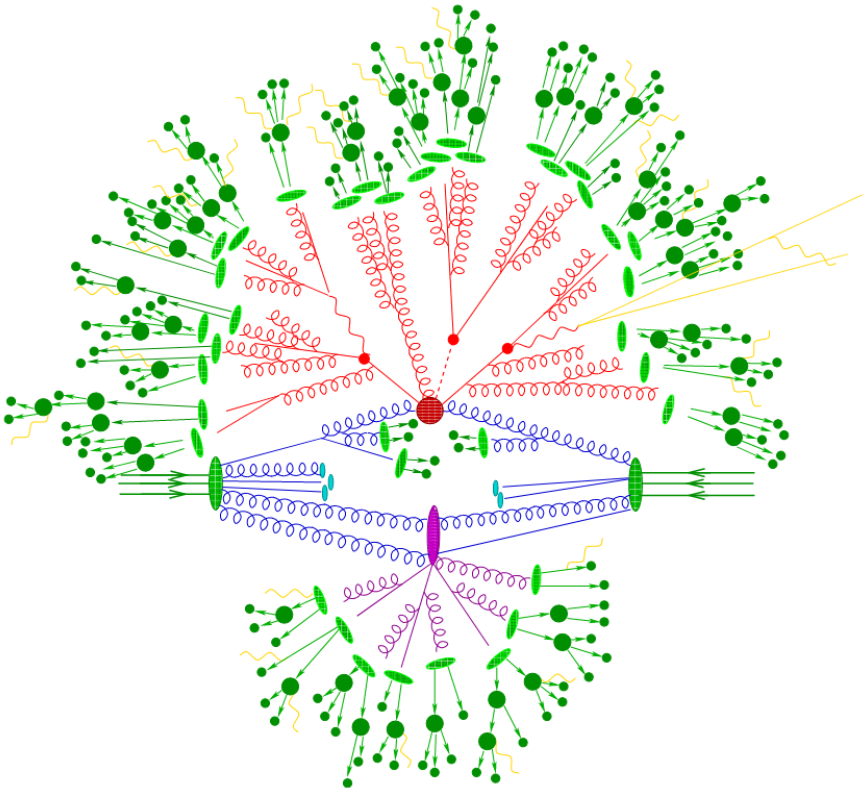


Figure 4.5: Schematic of pp collisions produced by an event generator. The big red blob is the hard interaction, and the small red blobs are the parton showers. The hadronization occurs in the light green blobs, and hadron decay is shown in the dark green blobs. Yellow represents photon radiation. The purple blob is the underlying event produced from the interaction of proton remnants. Picture taken from [25].

[6] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys. G* **34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.

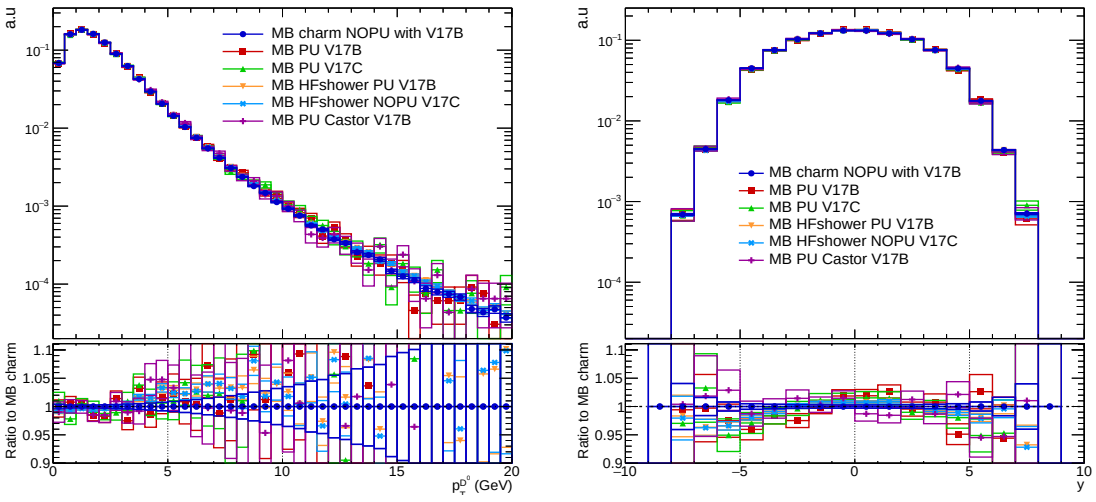
[7] CMS Collaboration, "Public TWiki page on RECO Data Format Table description", (2011). https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideRecoDataTable.

[8] CMS Collaboration, "Public CMS luminosity information", (2019). https://twiki.cern.ch/twiki/bin/view/CMSPublicLumiPublicResults.

[9] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", Technical Report CMS-PAS-PFT-10-001, 2010.

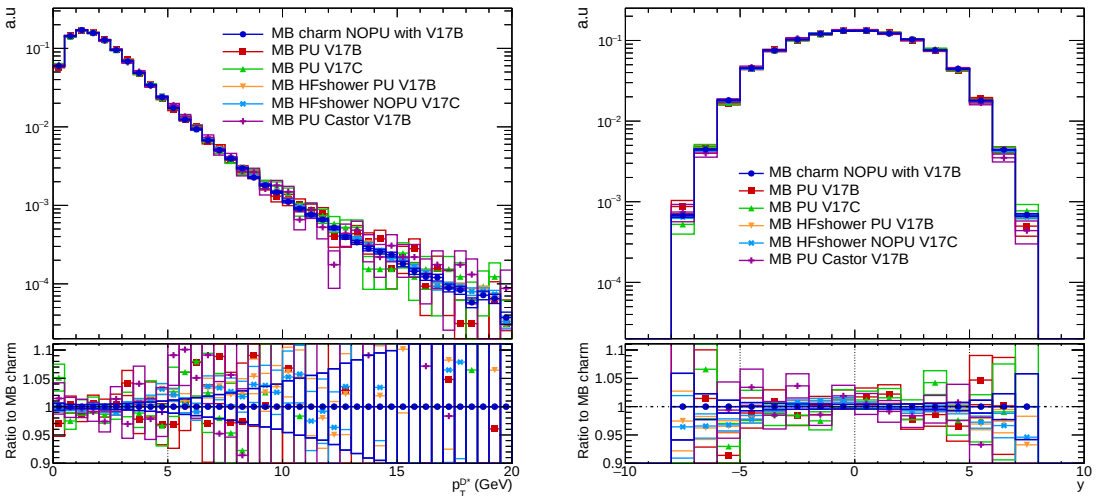
[10] CMS Collaboration, "Performance of CMS Muon Reconstruction in *pp* Collision Events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[11] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth. A* **262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.



(a) True D^0 p_T

(b) True D^0 y



(c) True D^* p_T

(d) True D^* y

Figure 4.6: Comparison of MC with charm filter and other available MCs at 7 TeV 2010.

[12] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[13] CMS Collaboration, "Electron reconstruction and identification at sqrt(s) = 7 TeV", Technical Report CMS-PAS-EGM-10-004, 2010.

[14] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC", *eConf* **C0303241** (2003) TULT009, doi:10.1088/0954-3899/31/9/N01, arXiv:physics/0306087.

[15] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[16] 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* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[17] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph", *JHEP* **02** (2003) 027, doi:10.1088/1126-6708/2003/02/027, arXiv:hep-ph/0208156.

[18] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[19] GEANT4 Collaboration, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[20] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641–653.

[21] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[22] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298–318, doi:10.1016/0550-3213(77)90384-4.

[23] CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV", *JHEP* **09** (2011) 109, doi:10.1007/JHEP09(2011)109, arXiv:1107.0330.

[24] CMS Collaboration, "Measurement of prompt open-charm production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2021) 225, doi:10.1007/JHEP11(2021)225, arXiv:2107.01476.

[25] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.

CHAPTER

5

HIGGS TO FOUR LEPTONS WITH CMS OPEN DATA

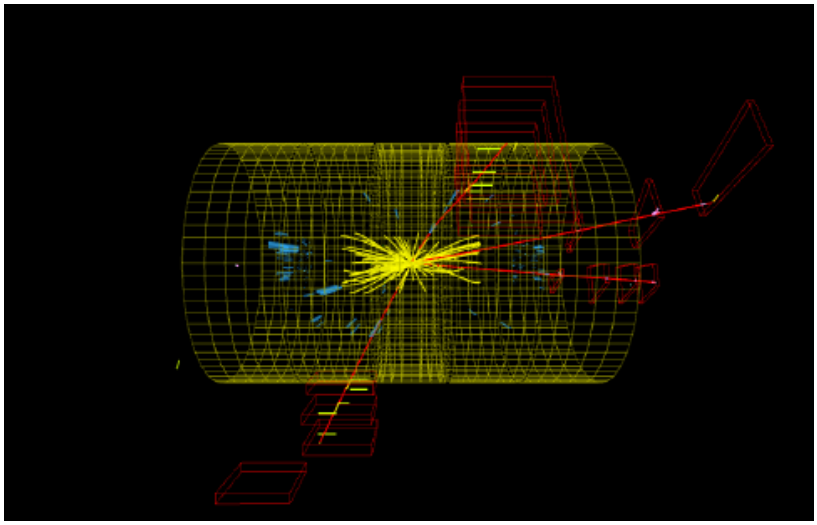


Figure 5.1: CMS event display that shows Higgs to 4μ final state (red lines)

5.1 Introduction

This chapter will cover the reconstruction of the Standard Model Higgs to four lepton mass spectrum using CMS Open Data. This validation analysis of CMS Open Data reproduces approximately part of the published CMS Higgs discovery paper at 7 and 8 TeV [1]. The available datasets are only partially overlapping with this publication, and the legacy software versions and corresponding calibrations are different from the original paper. This validation aims to give an example of using CMS Open Data for educational applications and show the potential of its usage for research applications. The Higgs boson can decay to

four charged leptons through intermediate Z bosons, and this decay channel is the cleanest experimentally; therefore, it is chosen for this analysis. Other Higgs final states (e.g., Higgs to two photons), which were also part of the same CMS paper and strongly contributed to the Higgs discovery, are not covered. Most of the cuts in this analysis are the same as the original cuts in the CMS paper [2]. Still, some of the more advanced analysis methods and systematic uncertainties of the discovery paper are skipped for simplicity reasons.
Nevertheless, it provides a qualitative insight into how the original result was obtained. The work shows that the Higgs peak is two standard deviations, compared to the original publication, which is 3.2 standard deviations in this channel alone. The corresponding example code is publicly released [3] together with the CMS primary dataset for 2012. Many people have used it as a reference or as a starting point on the further usage of CMS Open Data [4, 5].

5.2 CMS Open Data

CMS Open Data [6] are original data used by CMS members that are released to the public via the CERN Open Data portal [7]. In total, CMS made public more than 2 Petabyte of the Run 1 data collected by the CMS detector at the LHC and plans to release more data in the near future, as shown in Figure 5.2. A downloadable Virtual Machine (VM) tool with the CMS software environment is provided and maintained by CERN to use these legacy data. The purpose of these public data is to encourage external researchers and the general public to conduct their analysis using original preserved data. CMS Open Data can not only be used for research but also for educational purposes. Educational applications [8] usually target school pupils and university students to provide knowledge about high energy physics in a fun and educative way without any prior knowledge. Research applications [9] focus on people who want to conduct and publish their analysis using original preserved data. One example is a paper related to jet substructure studies by a theorist group, Jesse Thaler, and his team [10]. To achieve both purposes, several example codes of different analyses are available in the portal for users to analyze. The Higgs to four lepton example is in an educational and research potential category with four different levels of difficulties. The next sub-chapter will describe the details of this example.

5.3 Data and MC Samples

This example is based on the legacy version of both MC simulations and the original CMS primary datasets. This version slightly differs from the one used in the publication due to improved calibrations. These legacy data and MC sets were used in practice, exactly as they are, in many CMS publications. The data that was used is 7 and 8 TeV data, recorded in 2011 and 2012, respectively. This analysis used DoubleMuon and DoubleElectron datasets for the signal to prevent trigger overlap, and the events in these datasets were selected by the presence of at least two high-energy muon/electrons in the event. The total luminosity is 13.9 fb^{-1}, a sample corresponding to L = 2.3 fb^{-1} in 2011 at 7 TeV (Run A) and L = 11.6 fb^{-1} in 2012 at 8 TeV (Run B/C).

For the signal MC sample, we used SM Higgs bosons that are forced to decay to two Z-bosons, where these bosons are allowed to be off-shell; then each Z-boson is forced to decay

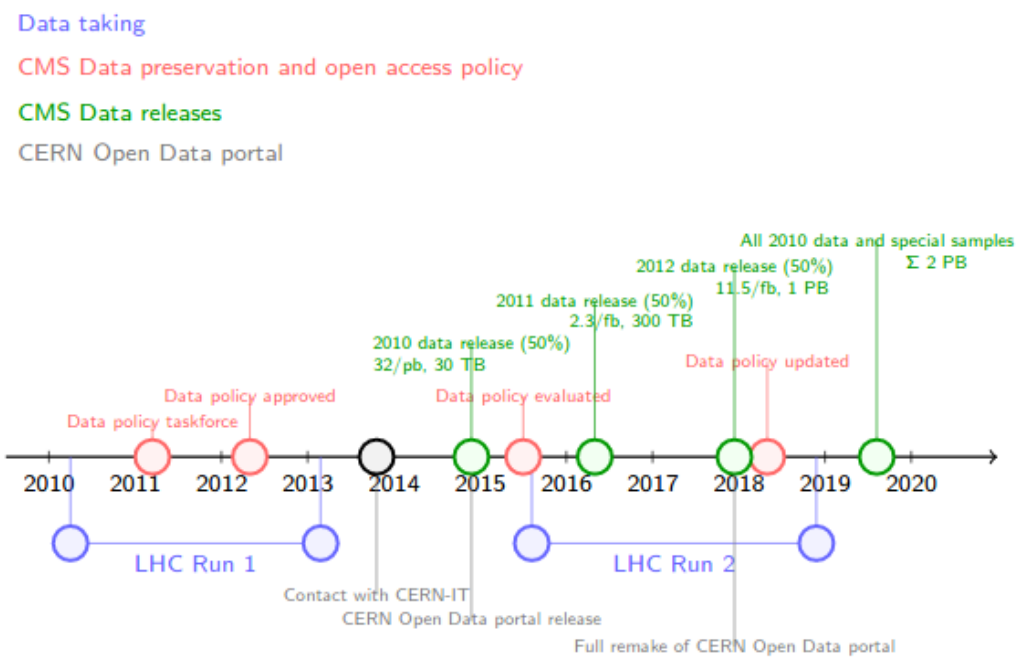


Figure 5.2: Timeline of CMS data releases starting 2014. Picture courtesy from Kati Lassila-Perini.

to 2 leptons. Contributions from di-boson, top, and Drell-Yan production that describe the background processes were used as background samples in this analysis. The first two background samples use POWHEG [11, 12] generators except for the Drell-Yan sample that used the MADGRAPH [13] event generator. Showering, hadronization, decay, and underlying events of the particles are then simulated by the PYTHIA [14] event generator. All of these datasets and MC samples are available in CMS Open Data, and the user can find a full list of datasets in [3].

5.4 Event/Object Selection

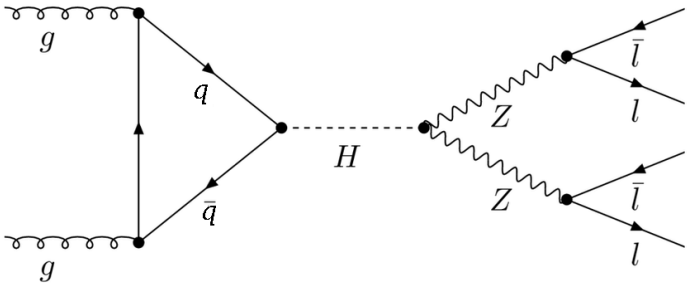


Figure 5.3: Feynman diagram of Higgs decays to Z boson pair, reconstructed in the 4 lepton final state (4μ, 4e and 2μ 2e)

This analysis focuses on the search of a Standard Model Higgs boson that decays into a Z boson pair, which then decays to 4 charged leptons in the final state as shown in Figure 5.3. One or both of the Z bosons can be off-shell, and the 4 leptons can be 4μ, 4e, and 2μ 2e. The leptons are reconstructed using the CMS PF event reconstruction algorithm as described in Chapter 4. First, good lepton candidates are selected by fulfilling the criteria below.

Both used PF relative isolation < 0.4 with ΔR cone of 0.4. The leptons should be originating from a "common primary vertex," which is defined as each lepton having an associated track with a small impact parameter with respect to the event primary vertex. This is achieved by putting cuts on the impact parameters to reject combinatorial background from other vertices and heavy flavor background. A cut on the transverse ($d_{xy} < 0.5$ cm) and longitudinal ($d_z < 1$ cm) impact parameter is applied, as well as a cut on its significance. The latter is defined as the significance of the impact parameter to the event vertex, $|SIP_{3D} = \frac{IP}{\sigma_{IP}}|$ where IP is the lepton impact parameter in the three dimensions at the point of closest approach with respect to the primary interaction vertex, and σ_{IP} is the associated uncertainty. This analysis applies a cut $|SIP_{3D}| < 4$. The muon(electron) candidate is required to have $p_T > 3(5)$ GeV and $|\eta| < 2.4(2.5)$. The electron candidate should also have expected missing inner hit <= 1 in order to reject conversions of radiated photons in the tracker material. After the good lepton selection, the muon and electron candidates are sorted from higher to lower p_T.

Next, a specific requirement for the $H \to ZZ \to 4\ell$ event selection is implemented.

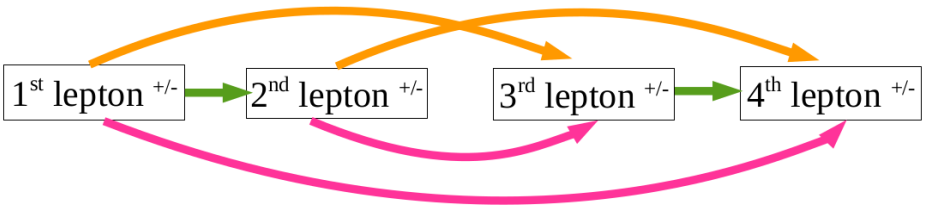


Figure 5.4: The chart above shows the first 4 leptons selected with the highest p_T. There are 3 possible combinations colored green, orange and pink that make 4 leptons where each combination has 2 lepton pairs with the same lepton flavor and opposite charge ($\mu^+\mu^-$ or e^+e^-). This is only valid for the 4μ and 4e case as for the 2μ 2e case, there is only 1 possible combination that satisfies the lepton pair requirement.

The event is required to have four or more leptons that passed the good lepton selection, and only the first four highest p_T leptons are chosen for further treatment. The selected four leptons are formed in groups of two that have the same flavor and opposite charge, as shown in Figure 5.4. Throughout this process, the total charge of each lepton pair is required to vanish.

Any two leptons from the selected four should have $p_T^i > 20$ GeV and $p_T^j > 10$ GeV. The p_T thresholds ensure that the leptons are on the high-efficiency plateau for the trigger. The lepton pair with invariant mass closest to the nominal Z boson mass is required to have a mass between 40 - 120 GeV, and the other lepton pair should have a mass between 12 - 120 GeV. Lastly, the mass of 4ℓ should be more than 70 GeV to provide better sensitivity for

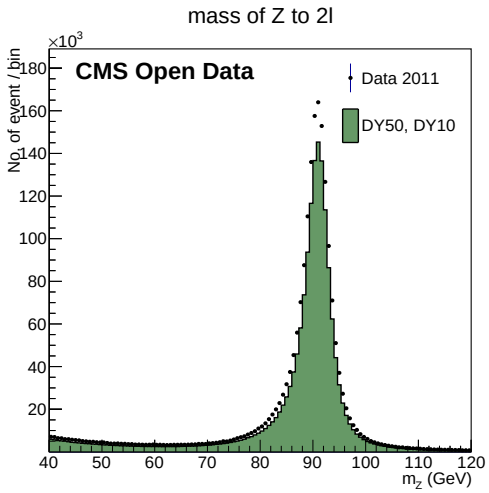
Higgs masses < 130 GeV.

5.5 Comparison with Higgs discovery

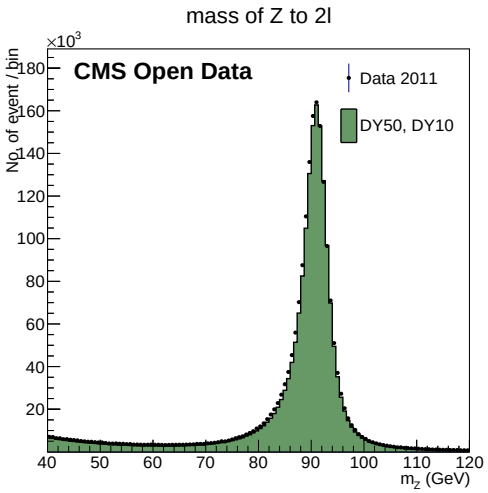
The background is estimated in a data driven way. In order to account for missing higher order contributions from the MC, scale factors are derived within control regions such that the normalization of the MC samples matches the data. The control region for the Drell-Yan process is taken to be the 40 - 120 GeV interval in the dilepton mass distribution, considering events that contain two or more same flavor and opposite charge leptons. For the di-boson process, the control region is the interval above 180 GeV in the 4ℓ mass distribution, considering events that contain four leptons. Both examples are shown in Figure 5.5. For the Higgs signal sample, the MC is normalized to the predicted cross section.

Based on Figure 5.6 (b), the Higgs mass peak is clearly visible. This is compatible with the original publication [2], where it has 3.2 standard deviations in this channel alone. The main differences are due to the less sophisticated analysis strategy and no systematics is applied. This example provides four levels of increasing complexity adjusted with the user's need [3].

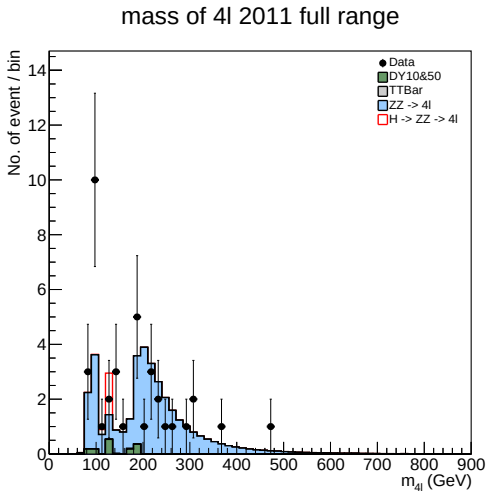
The first level compares only the final Higgs mass plot to the published one to see the difference with the caveat mentioned above. The second level is reproducing the mass plot from the predefined histogram files using a root macro. The third level produces a root data input file from original data and MC files for one Higgs signal candidate and the simulated Higgs signal with reduced statistics (for speed reasons). This level will allow the user to reproduce the final output plot containing their input using a root macro. Lastly, the fourth level, the highest level of difficulties, reproduces the full example analysis, which can take up to a few months depending on internet connection speed, computer performance, and user familiarity with CMS software [15]. This Higgs example also provides the resulting ROOT files containing many undocumented plots as a side product and the documented core results. The minimum requirement to conduct this example code is a minimal acquaintance with the Linux operating system and the ROOT analysis tool, which can be obtained in the CMS Open Data portal.



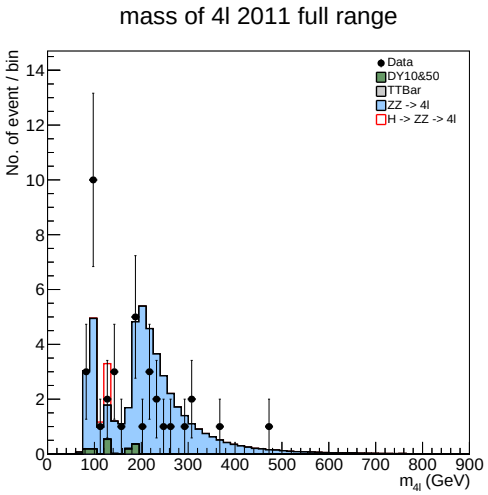
(a) Mass of Z $\rightarrow 2\ell$ before MC normalization



(b) Mass of Z $\rightarrow 2\ell$ after MC normalization

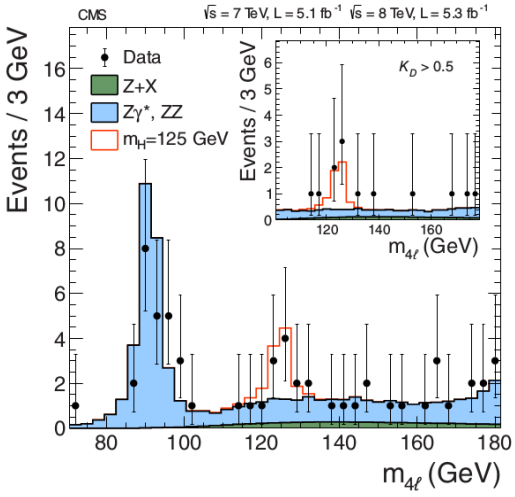


(c) Mass of 4ℓ in full mass range before MC normalization

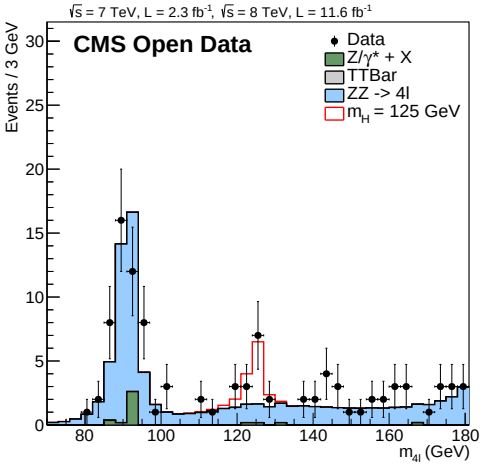


(d) Mass of 4ℓ in full mass range after MC normalization at region $>$ 180 GeV

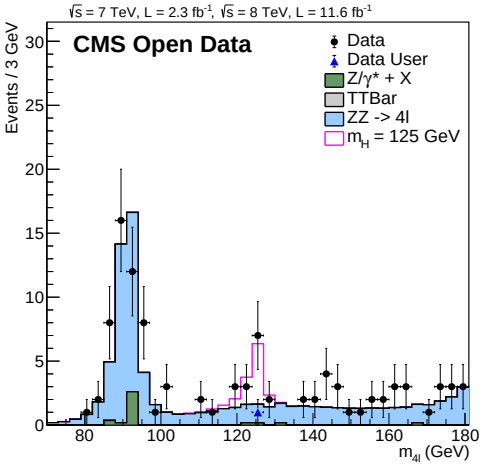
Figure 5.5: Control plots of H $\rightarrow$ ZZ $\rightarrow 4\ell$ for different MC processes to estimate the background.



(a) Mass of 4ℓ from paper [2]



(b) Mass of 4ℓ from CMS Open Data



(c) Mass of 4ℓ from CMS Open Data of the third level Higgs example

Figure 5.6: Mass of 4ℓ comparison

5.6 Bibliography

[1] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[2] S. Chatrchyan et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** (Sep, 2012) 30–61, doi:10.1016/j.physletb.2012.08.021.

[3] N. Jomhari, A. Geiser, and A. A. Bin Anuar, "Higgs-to-four-lepton analysis example using 2011-2012 data", *CERN Open Data portal* (2017) doi:10.7483/opendata.cms.jkb8.rr42.

[4] X. Chen, S. Dallmeier-Tiessen, R. Dasler, and et al., "Open is not enough", *Nature Physics* **15** (November, 2019) 113–119, doi:10.1038/s41567-018-0342-2.

[5] K. Lassila-Perini, C. Lange, E. Carrera Jarrin, and M. Bellis, "Using CMS Open Data in research – challenges and directions", *EPJ Web of Conferences* **251** (August, 2021) doi:10.1051/epjconf/202125101004.

[6] CMS Collaboration, "Public CMS Open Data information", (2014). http://opendata.cern.ch/docs/about-cms.

[7] CERN, "Public CERN Open Data Portal information", (2014). http://opendata.cern.ch/docs/about.

[8] CMS Collaboration, "CMS Guide to education use of CMS Open Data", (2014). http://opendata.cern.ch/docs/cms-guide-for-education.

[9] CMS Collaboration, "CMS Guide to education use of CMS Open Data", (2014). http://opendata.cern.ch/docs/cms-guide-for-research.

[10] A. Tripathee et al., "Jet Substructure Studies with CMS Open Data", *Phys. Rev. D* **96** (2017), no. 7, 074003, doi:10.1103/PhysRevD.96.074003, arXiv:1704.05842.

[11] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG", *Journal of High Energy Physics* **2009** (apr, 2009) 002–002, doi:10.1088/1126-6708/2009/04/002.

[12] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG", *Journal of High Energy Physics* **2010** (Feb, 2010) doi:10.1007/jhep02(2010)037.

[13] J. Alwall et al., "MadGraph/MadEvent v4: the new web generation", *Journal of High Energy Physics* **2007** (sep, 2007) 028–028, doi:10.1088/1126-6708/2007/09/028.

[14] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *Journal of High Energy Physics* **2006** (may, 2006) 026–026, doi:10.1088/1126-6708/2006/05/026.

[15] Simko, Tibor et al., "Open data provenance and reproducibility: a case study from publishing CMS open data", *EPJ Web Conf.* **245** (2020) 08014, doi:10.1051/epjconf/202024508014.

CHAPTER

6

MEASUREMENT OF $D^{*\pm}$ CROSS SECTION AT 7 TEV

6.1 Introduction

So far, there is no measurement of D meson cross sections in the CMS experiment at 7 TeV. This analysis focuses on charm reconstruction in the $D^{*\pm}$ final state, which offers the highest reconstruction redundancy and hence the best signal to background ratio for a broad phase space coverage. Since the reconstruction of the π_s from D^* decays (see Subsection 2.4.3) in the full phase space needs access to the lowest possible transverse momentum tracks, this analysis performs the measurement on a subset of data at 7 TeV for which special low p_T tracking was available.

Figure 6.1 shows the phase space in p_T vs y covered by the ALICE [1], LHCb [2] and CMS (this analysis) experiments for D^* measurements at 7 TeV. Due to detector constraints, only the rapidity range up to about 2.5 is experimentally accessible in CMS. However, the combination of the CMS measurements presented here with those from LHCb essentially covers the full phase space for charm production. For the uncovered regions, the measurement will be extrapolated. The measurements in the overlapping region between CMS and ALICE will be compared for a consistency check. This thesis covers the total and differential charm cross section at 7 TeV from the 2010 CMS data.

In this chapter, the steps toward the $D^{*\pm}$ meson cross section measurement are described. Starting with Section 6.2, where one of this analysis's uniquenesses is the strategy used to increase the statistics. Section 6.3 describes the data and MC samples as well as the luminosity calculation. The reconstruction of $D^{*\pm}$ down to p_T 1 GeV and their selection cuts are explained in Section 6.4. This chapter continues with the signal extraction of the $D^{*\pm}$ followed by the efficiency calculation and the result of the $D^{*\pm}$ cross section in Section 6.5. Lastly, the systematic uncertainties are listed in Section 6.6.

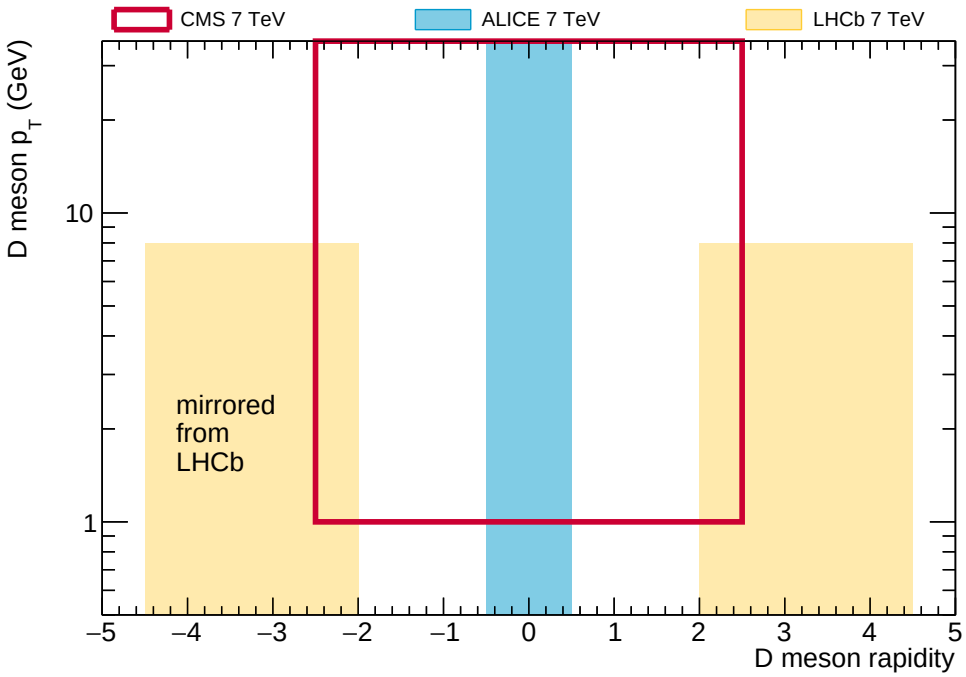


Figure 6.1: Phase space covering D^* meson measurements at 7 TeV by different LHC experiments; ALICE [1] in the blue region, LHCb [2] covered the forward region (right side) and its mirror (left side) in the yellow region, and finally CMS (this analysis) in the red box. For the phase space at different center of mass energies, see Figure 1.1.

6.2 Analysis Strategy

6.2.1 Pileup as a physics resource

In a classical approach, a low p_T charm analysis can be performed using Zero Bias and Minimum Bias triggers only, since no dedicated triggers are available for charm production. However, in this analysis, statistics is a limiting factor. Even though the cross section of charm in pp collision is high ($\sim$10 mb^{-1} [3]), these triggers, unfortunately, were heavily prescaled (see Section 3.3 for prescale definition). Therefore, the strategy is to use also pileup vertices as they can recover untriggered charm events.

The source of pileup vertices comes from the events triggered by non-Zero Bias and non-Minimum Bias triggers on any suitable dataset. In this case, muon and electron datasets are used since the muon and electron detector deposits which fire the trigger can be uniquely associated to a particular vertex which can then be excluded. Only purely leptonic triggers in these datasets are used.

To cope with this strategy, in this work, instead of the number of events, the number of collisions with good PV selection is used. In an event, there can be several pp collisions and sometimes no collision at all (empty event), which can happen e.g. for Zero Bias events. However, when there are collisions, not all of the collisions lead to reconstructed PVs. There

can be some cases where a PV cannot be reconstructed. Therefore, a good selection of PVs in an event is needed. The criteria for a good PV are defined as [4]:

- the vertex is from the primary vertex collection with beamspot constraint (see Section 4.2)
- the vertex is valid i.e. its fit converged reasonably
- the vertex is not fake (not an empty vertex)
- number of degrees of freedom, ndof > 4
- absolute distance in the z-direction between the vertex and the beamspot, |vertex z - beamspot z| < 15 cm
- distance in the xy-direction between the vertex and the beamspot < 2 cm.

A collision fulfilling these criteria is called a Minimum Bias (MB) collision if the events it occurs in is triggered by a Zero Bias or Minimum Bias trigger. In addition, collisions within events that are triggered by leptonic triggers are considered as Next-to-Minimum Bias (NMB) collisions, after the exclusion of the triggering lepton PVs (and those within $\sim$2 cm in z around these PVs). More technical details can be found in [5].

6.2.2 $D^{*\pm}$ meson

As mentioned in Subsection 2.4.3, the reconstruction of a $D^{*\pm}$ with a final state of $K^{\mp}\pi^{\pm}\pi_s^{\pm}$ has some advantages. First, the kinematics of the π_s from the D^* decay is limited by the small difference between the D^* and D^0 masses, giving D^* events a distinctive signature, thus reducing the background contribution, as shown in Figure 6.2 and Figure 6.3. Second, by having three charged particles in the $D^{*\pm}$ final state, it is the cleanest decay channel compared to others (see Subsection 2.4.3) that makes the $D^{*\pm}$ signal extraction simpler. The K charge from the D^0 always has an opposite sign to the π_s and a correct charge combination of D^0 daughters can differentiate the $D^{*\pm}$ resonance and combinatorial background (see Figure 2.8). More details about the $D^{*\pm}$ reconstruction will be described in Section 6.4.

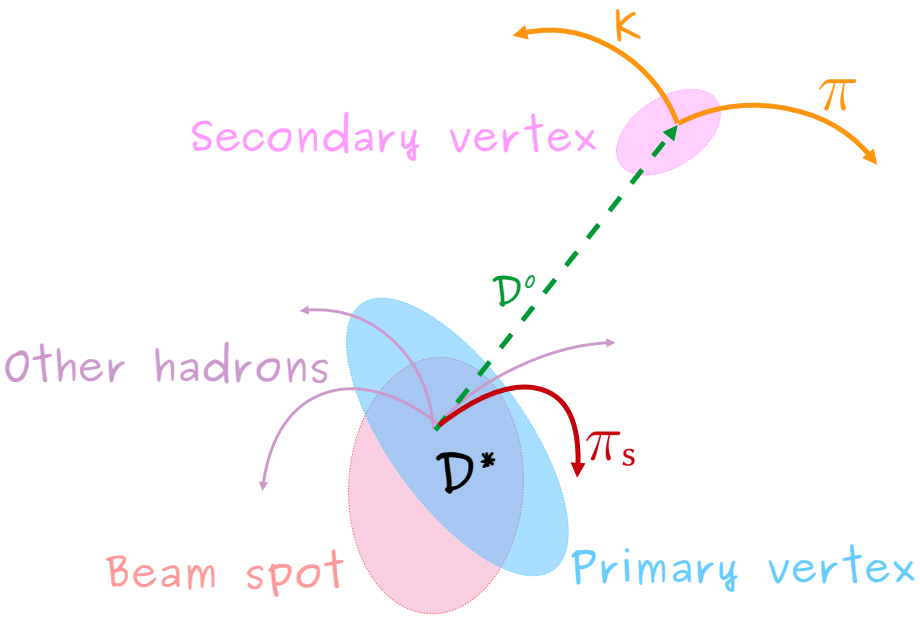


Figure 6.2: Sketch of primary $D^* \to D^0\pi_s$ where $D^0 \to K^-\pi^+$ in the xy plane. In reality, the curve of π_s is more like a helix curve as shown in Figure 6.3.

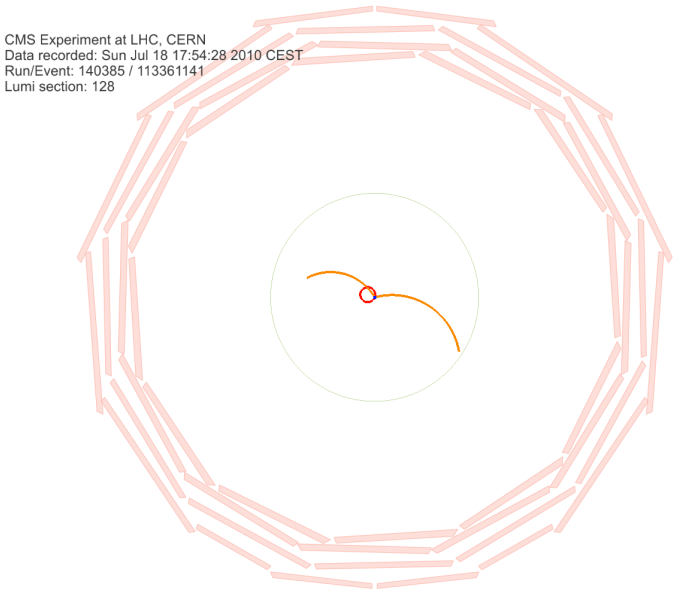


Figure 6.3: Real event of $D^* \to K\pi\pi$ at 7 TeV data in the CMS event display in the xy plane. The K and π tracks are shown in orange and the π_s track in red. It can be seen that the π_s track has a very low momentum such that it is curled up.

6.3 Data and MC Samples

The analysis is performed with 2010 pp data at $\sqrt{s} = 7$ TeV using the CMSSW_4_2_8 software version. The data used are public in `AOD` or `RECO` format published in the CMS Open Data portal. A dedicated virtual machine needs to be set up [6] in order to use these public data. The list of datasets used in the analysis is shown in Table 6.1. These datasets were recorded during the Commissioning10 and Run2010A-B data taking periods. The table is ordered such that events that are selected based on the triggers from a dataset early in the list are not considered in any other dataset later in the list, in order to prevent double counting of events.

Typically, one should also include a corresponding official CMS JSON file when running on these data. The file contains some quality selection of events certified by CMS that can be used for most CMS analyses and saved in JSON format. However, this analysis is statistically limited, and by applying this file, the statistic losses are significant for Zero Bias and Minimum Bias datasets. Since only the tracker is needed to deliver quality data, the official JSON file is not applied when running on Zero Bias and Minimum Bias datasets. Instead, dedicated tracking quality checks are made (see Subsection 6.3.1). For the rest of the datasets, the JSON file is applied. The JSON file for 7 TeV 2010 for Run A and Run B is shown in Table 6.2.

The second column of Table 6.1 is the raw number of events in the datasets before applying the JSON file. The third column is the number of (N)MB collisions used in the analysis satisfying the criteria of the trigger and vertex selection in these datasets. For Zero Bias and Minimum Bias collisions, if more than one PV occurs in the same event, each PV will be used as a separate unbiased collision if they are separated well enough between each other. For (N)MB collisions, see Subsection 6.2.1 for the definition and the treatment of using this collision. As one can see in the third column, some of the (N)MB collision numbers are significantly lower than the #Events due to empty events (see Subsection 6.2.1), since they do not pass the JSON data quality check, or the trigger quality checks or the events are already selected in one of the previous datasets.

6.3.1 Effective luminosity

In the CMS experiment, the luminosity is calculated using a command-line tool called Brilcalc provided by the Beam Radiation Instrumentation and Luminosity (BRIL) group [7]. For the 2010 data, the Brilcalc tool initially did not seem to return useful numbers yet for triggers which are prescaled, i.e. part of the prescales seemed to be ignored. The original lumicalc tool is no longer available. This was kindly fixed by the BRIL groups. The luminosity of all data for Run A and Run B returned by Brilcalc is 3.13 pb^{-1} and 31.88 pb^{-1}, respectively. This number contributes to the effective luminosity ($\mathcal{L}_{\text{eff}}$) calculation. The $\mathcal{L}_{\text{eff}}$ was determined in a different, data driven way, using unprescaled triggers as a basis and information from Brilcalc. The strategy is the following:

- Identify some high rate reference triggers which are active and unprescaled throughout a relevant data taking (sub)period (e.g. `HLT_Mu9` for 2010 Run A, `HLT_DoubleMu3` for 2010 Runs A and B, on the Mu dataset).

- Get the luminosity from Brilcalc for Run A and Run B for these triggers.
- For inclusive Zero/Minimum Bias triggers, measure the fraction of events with the reference trigger which also fire the prescaled inclusive trigger. This yields a measurement of the effective prescale factor.
- The efficiency for Minimum Bias (98-99%) and 100% for Zero Bias are applied, the product of the reference luminosity with this effective prescale factor then yields the $\mathcal{L}_{\text{eff}}$ for the respective Zero/Minimum Bias trigger for the relevant data (sub)set.
- Dividing the number of good primary vertices (see Subsection 6.2.1 for definition of good PV) obtained with this trigger by this luminosity then yields an "effective good vertex cross section", σ_{eff}, akin to the total inelastic cross section (σ_{eff} = total cross section × good vertex efficiency). This in turn can be used to calculate the $\mathcal{L}_{\text{eff}}$ of any Zero/Minimum Bias-like data subset, including pileup vertices, through good vertex counting.

The value obtained for σ_{eff} is 52.1 ± 2.1 mb. A detailed calculation of σ_{eff} has been done in Ref. [5]. This value is then used to get all the $\mathcal{L}_{\text{eff}}$ for each dataset using the formula below:

$$\mathcal{L}_{\text{eff}} = (\text{N})\text{MB}/\sigma_{\text{eff}} \tag{6.1}$$

To ensure the trigger/pileup and quality selections used on these datasets are fine, a tracking quality check is done by looking at the multiplicity and vertex stability for each trigger through all 2010 running eras. Figure 6.4 and Figure 6.5 shows the trigger and vertex stability of Zero Bias/Minimum Bias triggers and pileup from Muon/Electron triggers, respectively. A clear gap between groups of run numbers indicates different data taking periods (e.g. Commisioning10) as shown in Figure 6.4 (a). Stability is achieved when the average PV track multiplicity of each trigger is consistent throughout different run numbers.

6.3.2 MC samples

There are several Minimum Bias MC samples for 7 TeV 2010 with different tune settings as shown in Table 6.3. The MC samples have low statistics of charm events. Nevertheless, they are useful for cross checks and studies of pileup effects (see Subsection 4.4.1). There are three MC samples generated specially for the use in this analysis (see Table 6.4). The first is the MC sample with charm filter, thus the events contain prompt charm contribution. The second is the MC sample with beauty filter where the events contain the beauty contribution (plus a bit of additional charm in beauty events). Both the charm and beauty filters are applied at parton level. The filter efficiency for the inclusive charm and beauty filter MC is 0.09 and 0.0066, respectively. They are needed for the systematics.

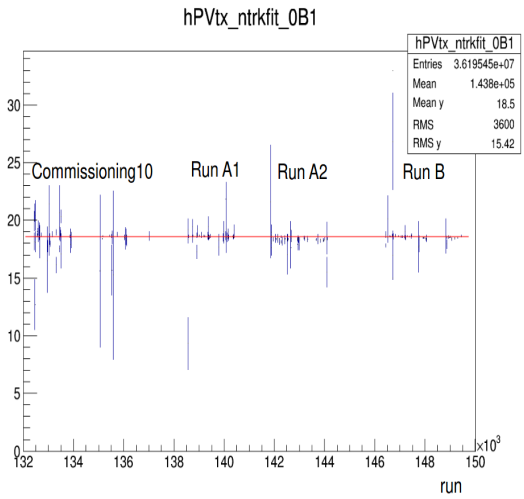
The third sample is the MC sample with D^0 decays to K and π. It is the main sample used in this analysis i.e. for the detector efficiency determination. The MC contains both contributions from charm and beauty. The $D^0 \to K^-\pi^+$ sample has a filter efficiency of 0.004. Since these three MCs are based on Minimum Bias events plus a filter (charm filter, beauty filter or $D^0 \to K^-\pi^+$ filter), a cross section from the Minimum Bias sample (71.21

Table 6.1: The list of datasets used in the analysis. The second column is the raw number of input events, while the third column gives the number of collisions actually used in the analysis. The last column is the effective luminosity calculated from the formula in Equation 6.1.

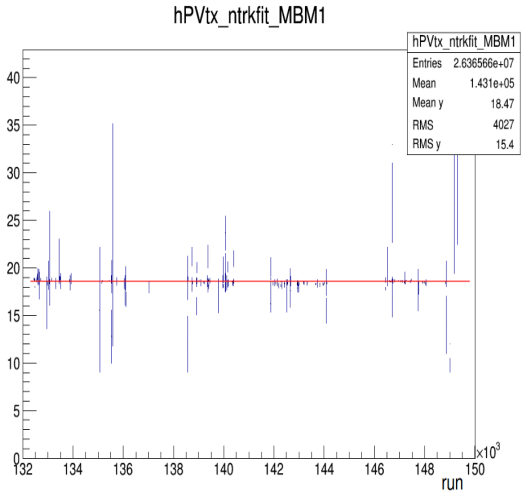
Data	#Events	(N)MB	$\mathcal{L}_{\text{eff}}$ (nb^{-1})
/ZeroBias/Commissioning10-May19ReReco-v1/RECO	129,186,198	646,080	0.0124
/ZeroBias/Run2010A-Apr21ReReco-v1/AOD	34,923,622	9,884,247	0.190
/MinimumBias/Commissioning10-May19ReReco-v1/RECO	46,553,963	32,246,050	0.619
/MinimumBias/Run2010A-Apr21ReReco-v1/AOD	103,848,957	25,950,980	0.498
/MinimumBias/Run2010B-Apr21ReReco-v1/AOD	40,785,403	16,092,377	0.309
/MuOnia/Run2010A-Apr21ReReco-v1/AOD	33,021,472	4,258,204	0.0817
/MuOnia/Run2010B-Apr21ReReco-v1/AOD	26,685,576	20,388,790	0.391
/Mu/Run2010A-Apr21ReReco-v1/AOD	51,802,592	6,039,449	0.116
/Mu/Run2010B-Apr21ReReco-v1/AOD	32,376,291	15,094,680	0.290
/MuMonitor/Run2010A-Apr21ReReco-v1/AOD	55,740,719	717,184	0.0138
/MuMonitor/Run2010B-Apr21ReReco-v1/AOD	12,728,741	1,799,123	0.0345
/EG/Run2010A-Apr21ReReco-v1/AOD	53,163,466	5,034,667	0.097
/Electron/Run2010B-Apr21ReReco-v1/AOD	32,772,061	16,598,760	0.319
/EGMonitor/Run2010A-Apr21ReReco-v1/AOD	67,929,392	3,315	0.000
/EGMonitor/Run2010B-Apr21ReReco-v1/AOD	11,826,859	1,457,837	0.028
Total	733,345,312	156,211,743	3.00

Table 6.2: JSON file 7 TeV 2010

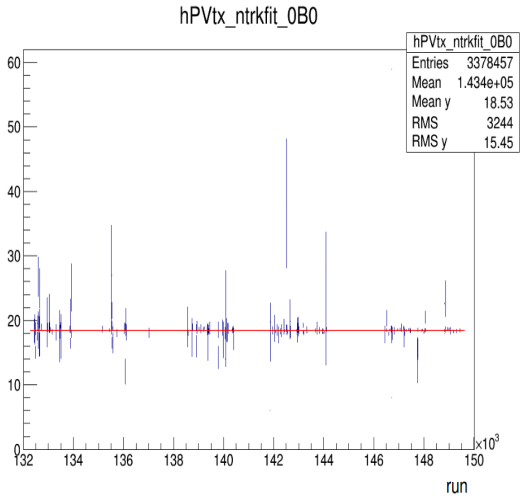
JSON file
Run2010A, Run2010B:
Cert_136033-149442_7TeV_Apr21ReReco_Collisions10_JSON_v2.txt



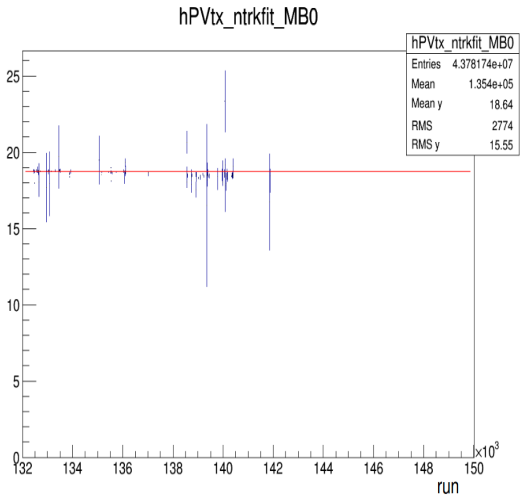
(a) HLT_ZerobiasPixel_SingleTrack



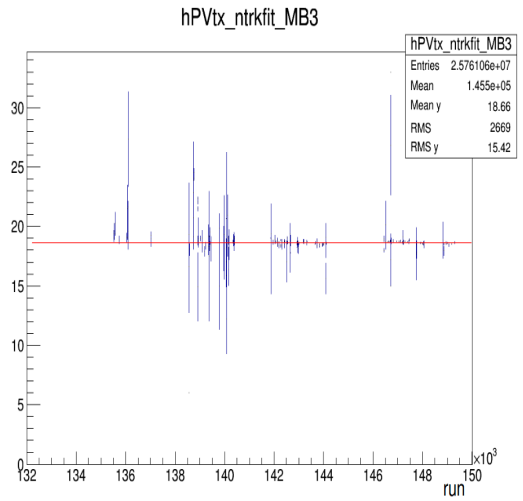
(b) HLT_Zerobias



(c) HLT_L1_BPTX

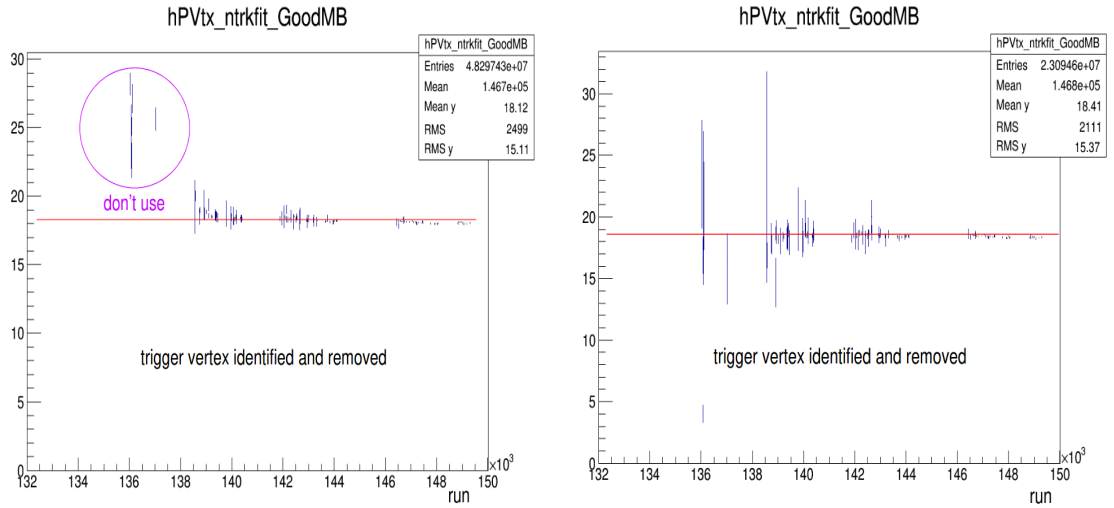


(d) HLT_L1_BscMinBiasOR_BptxPlusORMinus



(e) HLT_L1Tech_HCAL_HF

Figure 6.4: The average PV track multiplicity of Zero Bias/Minimum Bias trigger selection is stable throughout run numbers.



(a) Pileup from (di)muon triggers

(b) Pileup from electron triggers

Figure 6.5: The average PV track multiplicity of pileup from the di(muon) and electron trigger selection is mostly stable throughout run numbers after the trigger vertex is identified and removed (see Subsection 6.2.1).

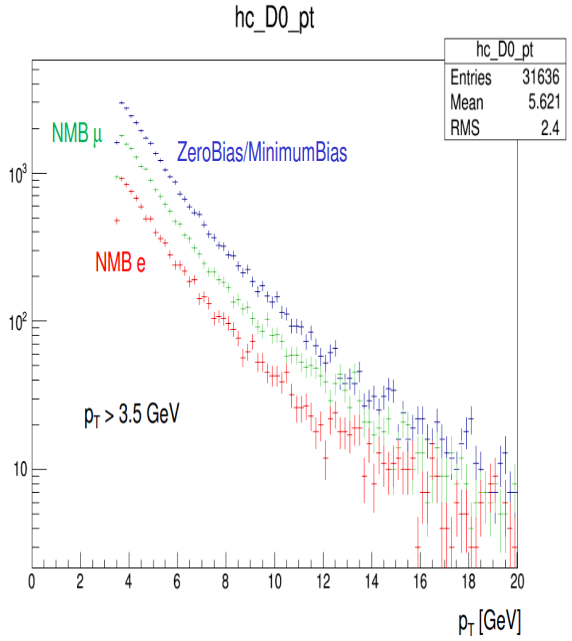


Figure 6.6: The distribution of p_T D^0 candidates in pileup (N)MB from muon and electron dataset and Zero/Minimum Bias dataset shows that the shape of the distribution is similar. This means the pileup vertices treatment is done properly and there is no trigger bias.

mb) is used for the $\mathcal{L}_{\text{eff}}$ calculation of the MC samples. The formula is defined as:

$$\mathcal{L}_{\text{eff}}^{MC} = \frac{N}{\sigma \cdot f_{\text{eff}}} \tag{6.2}$$

where N is the generated events, σ is the Minimum Bias σ and f_{eff} is the filter efficiency. The effective luminosities for these MC samples are shown in Table 6.4.

Table 6.3: List of Minimum Bias 7 TeV 2010 pp MC used for various studies of this analysis.

Sample	#Events
/MinBias_Tune4C_7TeV_pythia8/ Summer12-LowPU2010_DR42-PU_S0_START42_V17B-v1/AODSIM	500,000
/MinBias_Tune4C_Castor_7TeV_pythia8/ Summer12-LowPU2010_DR42-PU_S0_START42_V17B-v1/AODSIM	500,000
/MinBias_Tune4C_HFshowerLibrary_7TeV_pythia8/ Summer12-LowPU2010_DR42_NoPileUp_START42_V17C-v1/AODSIM	19,976,800
/MinBias_TuneD6T_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_NoPileUp_START42_V17C-v1/AODSIM	9,995,200
/MinBias_TuneZ1_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_PU_S0_START42_V17B-v1/AODSIM	9,994,400
/MinBias_TuneZ1_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_NoPileUp_START42_V17C-v1/AODSIM	9,994,400
/MinBias_TuneZ2star_7TeV_pythia6/ Summer12-LowPU2010_DR42-PU_S0_START42_V17B-v1/AODSIM	500,000
/MinBias_TuneZ2star_7TeV_pythia6/ Summer12-LowPU2010_DR42-PU_S0_START42_V17C-v1/AODSIM	500,000
/MinBias_TuneZ2star_Castor_7TeV_pythia6/ Summer12-LowPU2010_DR42-PU_S0_START42_V17B-v1/AODSIM	500,000
/MinBias_TuneZ2star_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_PU_S0_START42_V17B-v2/AODSIM	19,868,000
/MinBias_TuneZ2star_HFshowerLibrary_7TeV_pythia6/ Summer12-LowPU2010_DR42_NoPileUp_START42_V17C-v1/AODSIM	19,866,400

6.4 $D^{*\pm}$ Meson Reconstruction and Selection

The reconstruction of $D^{*\pm}$ starts from its final state particles. It requires at least three tracks where two tracks are the decayed particles of the D^0 and the third track is the π_s track. The two tracks are assigned as K or π with an opposite charge. The vertex of these tracks is refitted using the KalmanVertexFitter but to limit the revertexing, some events, for example, the ones that have low $p_T^{D^0}$ and low p_T^{frac} are rejected (see Subsection 6.4.1 for the $D^{*\pm}$ pres-

Table 6.4: MC samples with charm filter, beauty filter and D^0 decays to K and π.

MC Sample	#Events	eff.lumi (nb^{-1})
`/MinBias_charmfilter_TuneZ2star_7TeV-pythia6-evtgen/` `LowPU2010DR42-NoPU2010_DR42_START42_V17B-v2/AODSIM`	19,444,773	3.03
`/MinBias_beautyfilter_TuneZ2star_7TeV-pythia6-evtgen/` `LowPU2010DR42-NoPU2010_DR42_START42_V17B-v2/AODSIM`	1,827,530	3.89
`/D0Kpi_pT0toInf_TuneZ2star_7TeV-pythia6-evtgen/` `LowPU2010DR42-NoPU2010_DR42_START42_V17B-v2/AODSIM`	5,801,549	20.37

election cuts). The refitted vertex allows the reevaluation and to give a better estimation of the track parameters. If the fit quality is acceptable (p-value >1%), the position of the D^0 vertex and its covariance matrix is determined. The four-momenta of these two tracks are updated and combined to be a D^0 candidate and this vertex is called a secondary vertex (SV).

Since the D^0 has non-negligible mean lifetime, $\tau = (410.1 \pm 1.5) \times 10^{-15}$ s [8], it travels a bit before its decay. The length of the path is called decay length of the D^0. The 3D decay length of D^0, dl_{3D}, is calculated as:

$$dl_{3D} = \frac{\vec{p}^{D^0} \cdot \vec{\Delta}^{D^0}}{|\vec{p}^{D^0}|} \tag{6.3}$$

where $\vec{p}^{D^0}$ is the D^0 momentum and $\vec{\Delta}$ is the distance between the PV and SV. The cosine of the angle between these two vectors, $\cos\phi$, is defined as:

$$\cos\phi = \frac{dl_{3D}}{|\vec{\Delta}^{D^0}|} \tag{6.4}$$

Additionally, to evaluate how significantly the secondary vertex is away from the primary vertex, the decay length significance is defined as:

$$dl_{sig} = \frac{dl_{3D}}{dl_{err}^{D^0}} \tag{6.5}$$

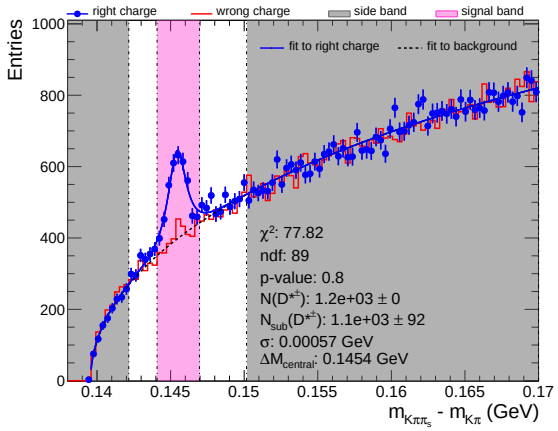
where $dl_{err}^{D^0}$ is the uncertainty of dl_{3D}.

After the D^0 candidates for D^* are successfully reconstructed, the π_s must originate from the same PV as the D^0 for the prompt contribution. The PV used in this analysis is the PV with beam spot constraint with input tag `offlinePrimaryVerticesWithBS` (see Section 4.2). A study on using PV with and without beamspot constraint is given in the Appendix A. Since the contributions to the total charm cross section measurement are mostly at low p_T, using `offlinePrimaryVerticesWithBS` helps to recover $D^{*\pm}$ efficiency, reduces the vertex uncertainty in xy and limits potential biases from the inclusion of secondary tracks in the PV fit. The π_s track is refitted to the PV to improve its momentum resolution. Figure 6.7 shows the study on the π_s refit in the lower ($p_T < 3.5$ GeV) and higher ($p_T > 3.5$ GeV) p_T regions and

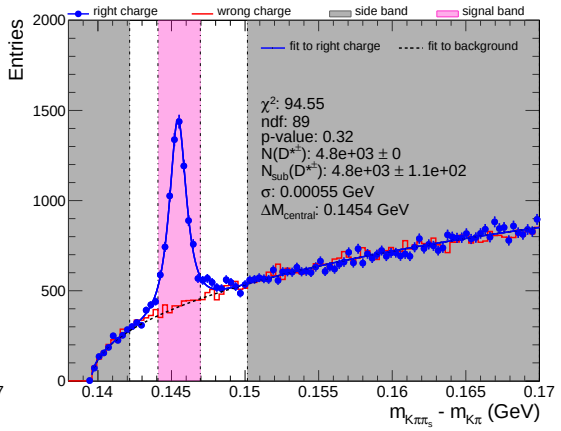
the variables in the plots are defined below.

The right charge combinations of $K^{\mp}\pi^{\pm}\pi_s^{\pm}$ receive contributions from the $D^{*\pm}$ resonance and combinatorial background, while the wrong charge combinations of $K^{\mp}\pi^{\mp}\pi_s^{\pm}$ receive contributions only from the latter, since doubly charged D mesons do not exist. For this reason, the wrong charge combinations can be used to estimate the background directly from the data. The $D^{*\pm}$ signal extraction is usually done using the Δm distribution instead of the $D^{*\pm}$ mass spectrum due to its superior resolution. Δm is defined as:

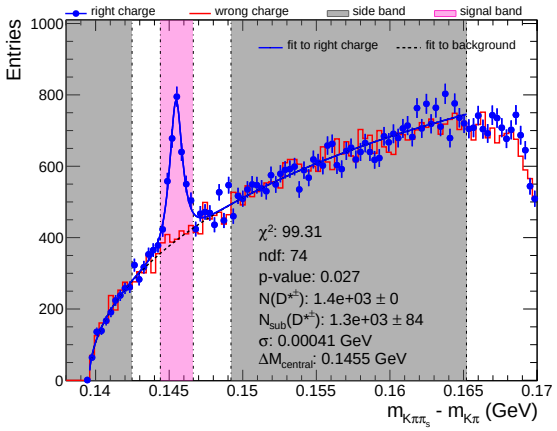
$$\Delta m = m(K^{\mp}\pi^{\pm}\pi_s^{\pm}) - m(K^{\mp}\pi^{\pm}) \tag{6.6}$$



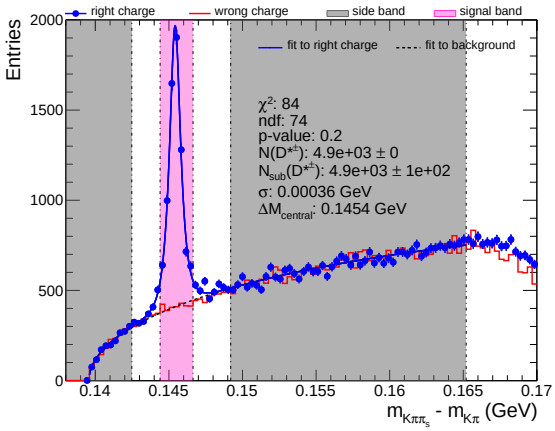
(a) Δm before π_s refit at $p_T < 3.5$ GeV



(b) Δm before π_s refit at $p_T > 3.5$ GeV



(c) Δm after π_s refit at $p_T < 3.5$ GeV



(d) Δm after π_s refit at $p_T > 3.5$ GeV

Figure 6.7: The figures above and below show the D*- D^0 mass difference before and after the π_s refit, respectively. The refit gives better signal to background ratio. The Δm width also becomes narrower. The side band region and the fit will be explained in Subsection 6.5.1

6.4.1 $D^{*\pm}$ preselection cut

A list of $D^{*\pm}$ preselection cuts is shown in Table 6.5. It was implemented at the AOD level[1] to reconstruct reasonable $D^{*\pm}$ candidates with loose cuts. The cut on the transverse momentum of the K and π is used to reduce the combinatorial background while keeping reasonable detector acceptance at low p_T. The variables d_{xy} and d_z are defined as the distance in the transverse plane and along the beam direction, respectively. They represent the distance between two tracks or vertices and the purpose is to ensure the combinatorial background is reduced as much as possible. p_T^{frac} is defined as ratio between the scalar sum of the p_T of the K^- and π^+ tracks over the scalar sum of p_T of all tracks within some d_{xy} and d_z ranges from the two tracks (< 0.5 cm). p_T^{frac} is given by:

$$p_T^{frac} = \frac{p_T \text{ of } D^0}{\sum p_T \text{ of all tracks}} \tag{6.7}$$

A cut on this quantity is imposed to ensure that the p_T of the D^0 candidate contributes significantly to the total p_T of all tracks associated to the vertex. Another useful preselection cut is the mass 'cross cut' between D^0 from $D^{*\pm}$ and Δm as shown in Figure 6.10 (a). A cross cut (see also Table 6.5) is applied on either of these masses to reduce the storage of background contributions in the Ntuple. Most of the cuts implemented at the AOD level are loose cuts. A further selection is implemented at the analysis level[2] on some of the variables and described in Subsection 6.4.2.

Table 6.5: $D^{*\pm}$ preselection at AOD level

Variable	Cut
Transverse momentum of K and π, $p_T^{K,\pi}$	> 0.3 GeV
d_{xy} and d_z between K and π track origins	< 0.5 cm
p_T^{frac} for $p_T^{D^0} < 0.9$ GeV	> 0.15
d_z distance between D^0 vertex and PV	< 2 cm
$p_T^{D^0}$	> 0.9 GeV
loose m_{D^0} cut	$\in [1.5, 2.3]$ GeV
ratio of $p_T^{\pi_s}$ and $p_T^{D^0}$	$\in [0.03, 0.20]$
d_{xy} and d_z distances between π_s and D^0 vertex	< 2 cm
mass difference between $D^{*\pm}$ and D^0, Δm	< 0.17 GeV
mass 'cross' cut on D^0 candidate from $D^{*\pm}$ and Δm	$m^{D^0} \in [1.828, 1.908]$ GeV or $\Delta m \in [0.1434, 0.1474]$ GeV

[1] A stage where AOD data (see Subsection 3.3.2) is read and written out to an Ntuple.
[2] A stage where an Ntuple is read and are written out histograms.

6.4.2 $D^{*\pm}$ analysis cuts

At analysis level, tighter cuts are implemented and shown in Table 6.6 and Table 6.7. In Table 6.6, the cuts are applied on the D^* daughters. The d_{xy} and d_z cuts of K and π are treated separately at this level. A sketch that shows the d_z from the PV is given in Figure 6.8. Another important cut at this level is the dE/dx cut which allows the separation of K and π at very low momenta. The dE/dx is the rate of energy loss by the particle as it travels through the detector. This takes advantage of the fact that different particle types lose energy at different rates, thus, allowing the particle species to be identified, by cutting on this variable. An example of this phenomenon can be seen in Figure 6.9.

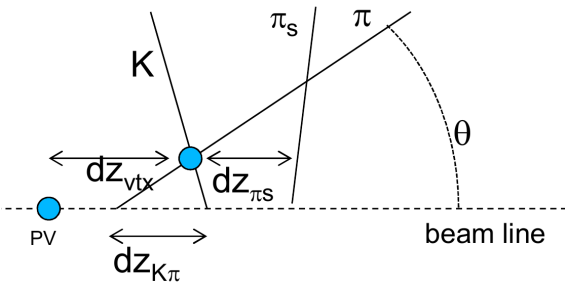
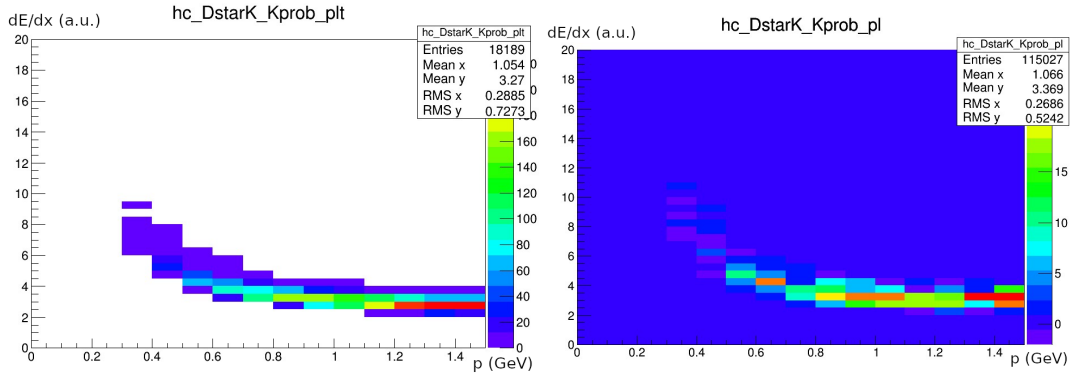


Figure 6.8: A sketch that shows the distance of K, π, and π_s from primary vertex along the beam direction. Picture taken from [9].

Table 6.6: Cuts on the D^* daughters at analysis level.

Variable	Cut
p_T^{π}	> 0.5 GeV
p_T^{K}	> 0.3 GeV
d_{xy} of K or π from PV	< 0.15 cm
d_{xy} of π_s from PV	< 0.3 cm
d_z of K or π from PV	$< 0.1/\sin\theta$ cm
d_z of π_s from PV	$< 0.2/\sin\theta$ cm
dE/dx of K for $p^K < 1.5$ GeV	$\in \left[0.6/\lvert p^K\rvert + 2, 1.0/\lvert p^K\rvert + 3.5\right]$

A final selection of $D^{*\pm}$ candidate is listed in Table 6.7. At the lower p_T region ($p_T < 3.5$ GeV), the $D^{*\pm}$ candidates are cut tighter to reduce more background. In Figure 6.10 (c), a tight signal region (shown in pink) of Δm is applied to the D^0 mass entries in Figure 6.10 (b) and vice versa. It is worth to mention that in Figure 6.10 (b), one observes a bump which is coming from partial reconstruction reflection, for example $D^0 \to \mu\nu X$ at the lower D^0 mass spectrum.



(a) dE/dx cut on D^0 MC

(b) dE/dx cut on data

Figure 6.9: dE/dx cut of $K^{\pm}$ (K band) from true $D^{*\pm}$ in D^0 MC is shown in (a) and from data after wrong charge subtraction (see Section 6.4) is shown in (b) for $p_T^{D^*} < 3.5$ GeV. Picture taken from [9].

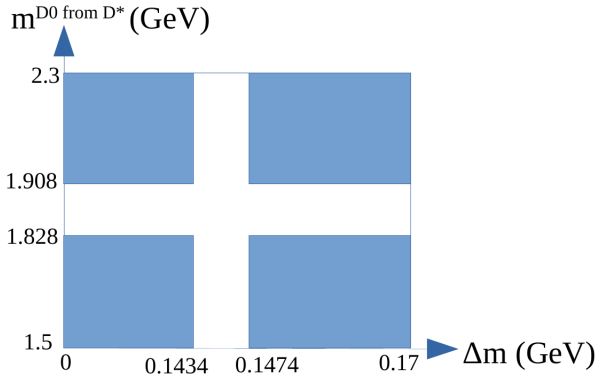
Table 6.7: $D^{*\pm}$ candidate selection at analysis level for each p_T range.

$p_T^{D^*}$	$> \in [1.0, 3.5]$ GeV	> 3.5 GeV
m^{D^0}	$\in [1.84, 1.89]$ GeV	$\in [1.85, 1.88]$ GeV
dl_{sig}	($dl_{sig} > 1.5$ and $p_T^{frac} > 0.15$) or [($dl_{sig} > 3$ or ($dl_{sig} > 2$ and $\cos\phi > 0.99$)) and $p_T^{frac} > 0.1$]	($dl_{sig} > -1$ and $p_T^{frac} > 0.15$) or $dl_{sig} > 2$
$\cos\phi$	> 0.8	
Δm	$\in [0.14440, 0.14664]$ GeV	

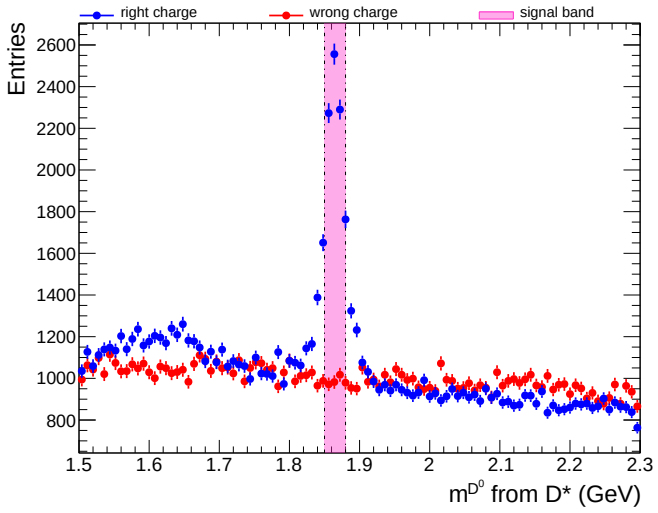
6.5 $D^{*\pm}$ Cross Section Measurement

6.5.1 $D^{*\pm}$ Signal extraction method

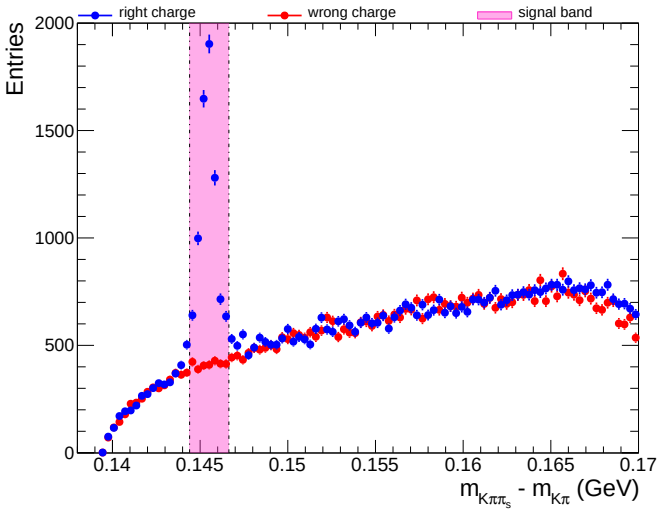
In this analysis, a method called wrong charge background subtraction is used to extract the $D^{*\pm}$ signal. It is done by subtracting the wrong charge Δm distribution from the right charge Δm distribution as shown in Figure 6.11. Before doing the subtraction, the wrong sign distributions are normalized by taking the ratio of the right and wrong charge distribution excluding the region between 0.14248 to 0.14920 GeV. The number of $D^{*\pm}$ signal events is taken from the integral of the background-subtracted right charge distribution within the signal region from 0.1444 to 0.14664 GeV. This is compared with a fitting method where the wrong charge distribution, consisting purely of background, is fitted with a threshold function, which is given by:



(a) Loose signal region for m^{D^0} and Δm.



(b) m^{D^0} spectrum ($p_T > 3.5$ GeV) in data indicate Δm cut.



(c) Δm spectrum ($p_T > 3.5$ GeV) in data indicate m_{D^0} cut.

Figure 6.10: Figure (a) shows a loose mass cut is implemented at AOD level on either D^* or D^0 mass which created a 'cross'. Figure (b) and (c) show the D^0 mass and Δm distributions, respectively, where the Δm cut in the signal region (pink color) is applied to (b) and the D^0 mass cut in the signal region is applied to (c). Both cuts in (b) and (c) are implemented at analysis level. For the distribution for $p_T < 3.5$ GeV, please see Appendix B.

$$A \cdot (\Delta m - m_\pi)^B \cdot \exp[C \cdot (\Delta m - m_\pi)] \tag{6.8}$$

where m_π is the π mass and A, B and C are free parameters. The signal contribution on the other hand is fitted with a modified Gaussian function [10] which is given by:

$$\text{Gauss}^{\text{mod}} = N \cdot \exp[-0.5 \cdot x^{1+1/(1+0.5 \cdot x)}] \tag{6.9}$$

where N is a normalisation paramater and $x = |(\Delta m - m_0)/\sigma|$. The signal position m_0, width σ and N are free parameters. This function is used to describe the non-Gaussian tails of a resonant signal. The fitting of the Δm distribution is only for illustration and not used in the charm cross section calculation.

The $D^{*\pm}$ signal extraction is divided into a finer binning instead of only $p_T < 3.5$ GeV and $p_T > 3.5$ GeV. There are several criteria that were considered when choosing the range of each bin. First, they must have enough statistics where the D^* signal can be extracted. The bins are chosen such that the bin width is bigger than the resolution. In this analysis, the binning of the p_T is chosen to be in steps of 1 GeV up to 10 GeV (> 10 GeV is the overflow bin) and for |y| in steps of 0.5. The $D^{*\pm}$ distributions in each bin are shown in Figure 6.12.

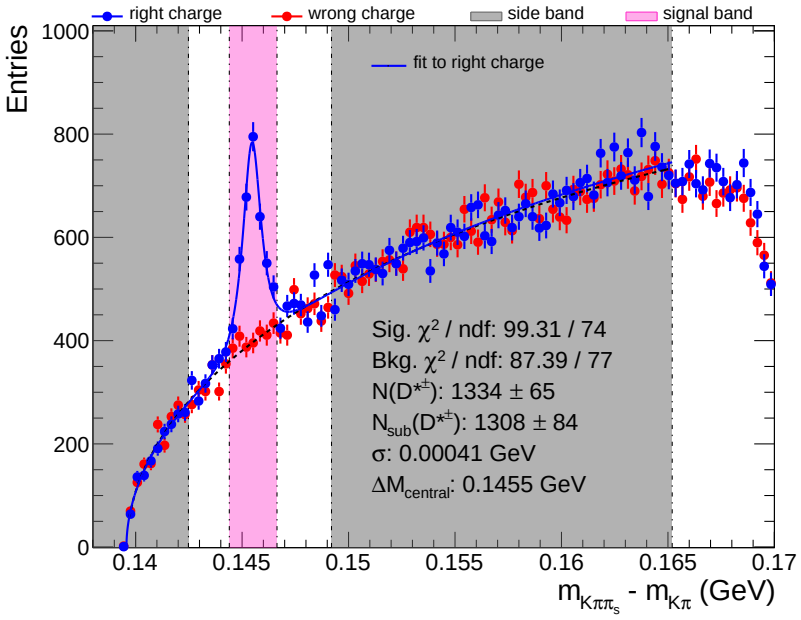
The reconstructed $D^{*\pm}$ may not be as good as the generated $D^{*\pm}$ due to the detector performance, underlying events etc. For example, some of the events at lower p_T might migrate to high p_T or vice versa (see Figure 6.13). To check the quality of the reconstruction and to ensure that the chosen binning for the $D^{*\pm}$ reconstruction is fine, a migration plot is shown in Figure 6.14. It is produced by applying generator matching where the reconstructed $D^{*\pm}$ candidates are matched with the generated $D^{*\pm}$ based on their kinematics. The figure shows that the correlation plot is roughly diagonal which indicates that the migrations between the reconstructed and generated event bins are small for both $D^{*\pm}$ p_T and y. Due to this, the number of extracted signal events can be used directly to extract the cross section without using any complex unfolding techique. The small migrations are accounted for in the efficiency correction (bin-by-bin unfolding).

6.5.2 Efficiency calculation

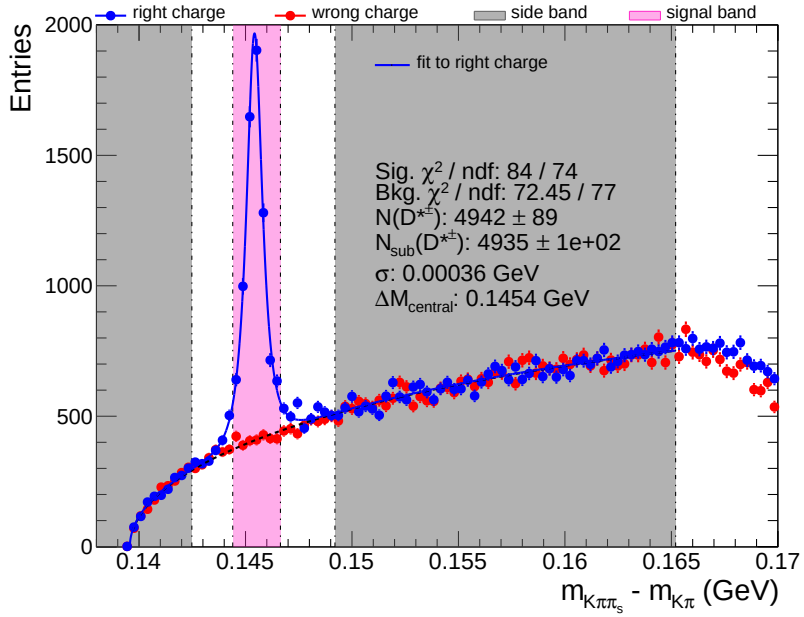
Another important ingredient for the cross section measurement is the $D^{*\pm}$ efficiency. It is calculated from the generated events of D^0 MC to determine how well the $D^{*\pm}$ can be reconstructed for each p_T and y bin by applying the formula below:

$$\text{efficiency} = \frac{N_{\text{reco\&true}}}{N_{\text{true}}} \tag{6.10}$$

where $N_{\text{reco\&true}}$ is the number of reconstructed $D^{*\pm}$ in each bin that match a generated $D^{*\pm}$ inside the bin of Figure D.1 and N_{true} is the number of generated $D^{*\pm}$ in the bin of Figure D.2. As such, the calculated efficiency encodes the detector efficiency. Figure 6.15 shows the detector efficiency of both prompt and non-prompt $D^{*\pm}$ decays to $K^{\mp}\pi^{\pm}\pi_s^{\pm}$ as a function of $D^{*\pm}$ p_T and |y|. The MC sample used in the calculation is mentioned in the figure's caption. The efficiency reaches up to 59% at higher p_T and lower |y|.



(a) Δm at $p_T < 3.5$ GeV



(b) Δm at $p_T > 3.5$ GeV

Figure 6.11: Δm in the lower(above) and higher(below) p_T regions.

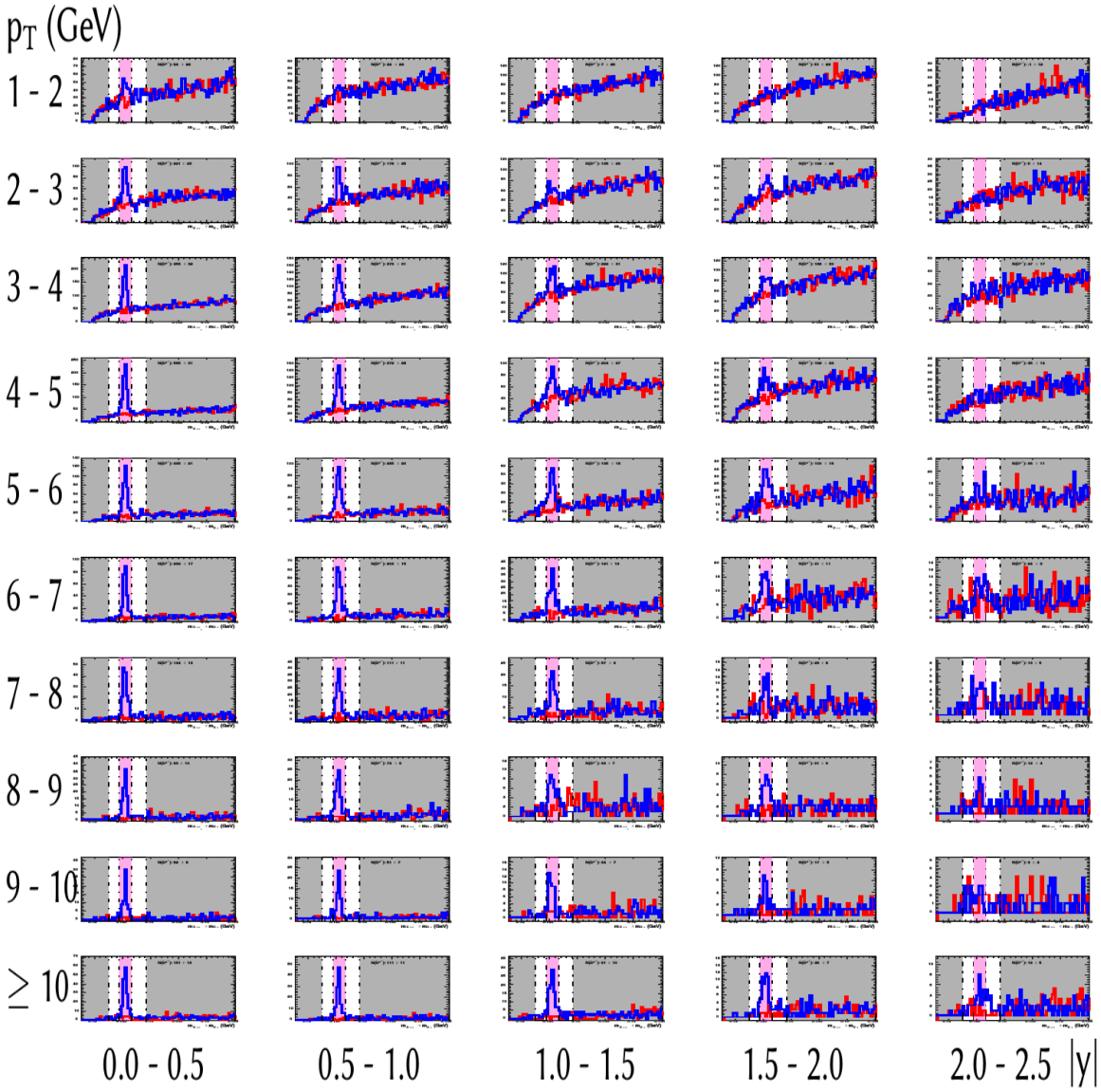


Figure 6.12: The Δm distribution in different bins of $K^{\mp}\pi^{\pm}\pi_s^{\pm}$ p_T and |y|. In most of the p_T and |y| bins, the $D^{*\pm}$ contribution can be seen as a peak in the Δm distribution. Individual Δm plots in each bin are shown in Appendix C.

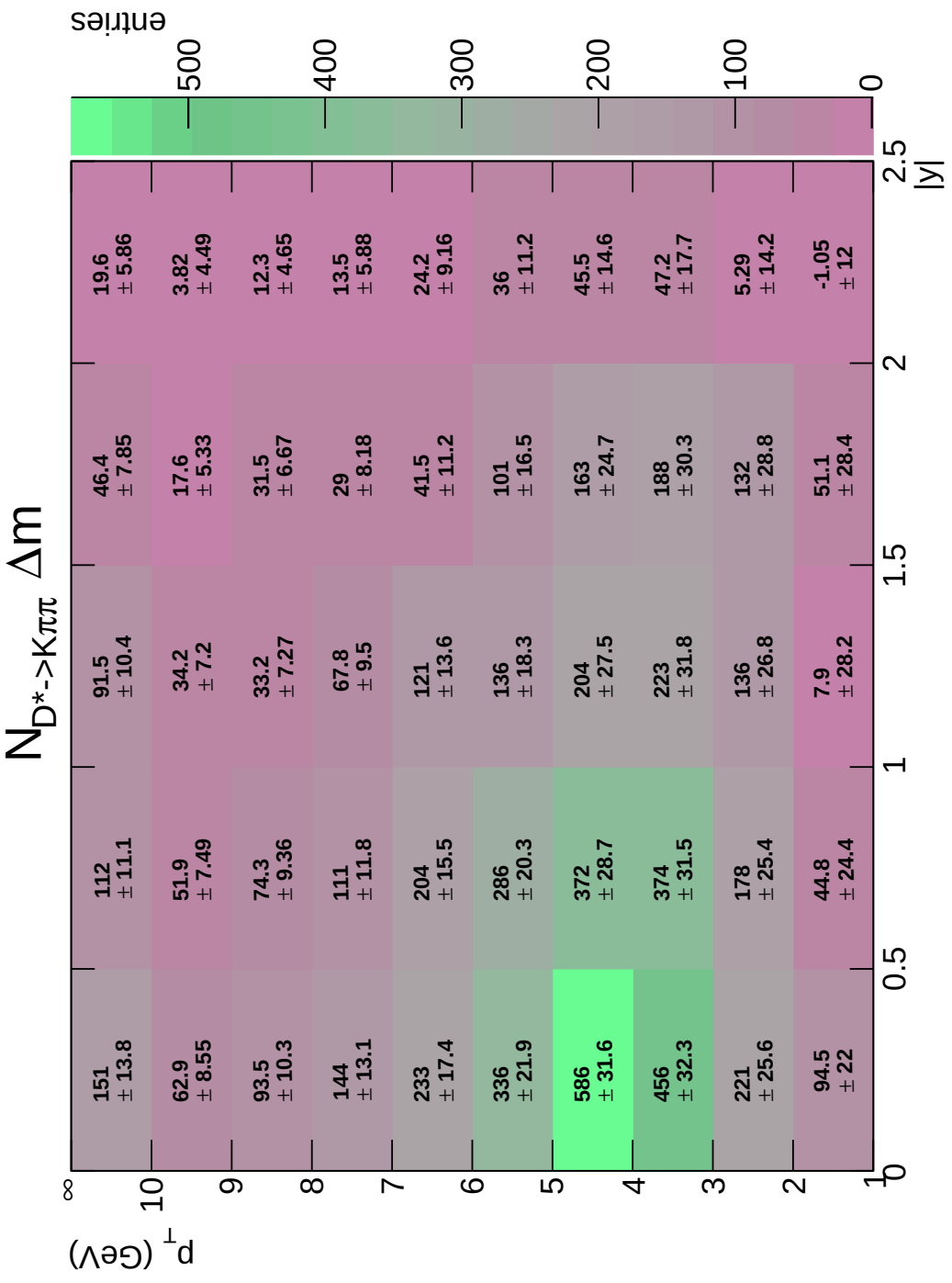
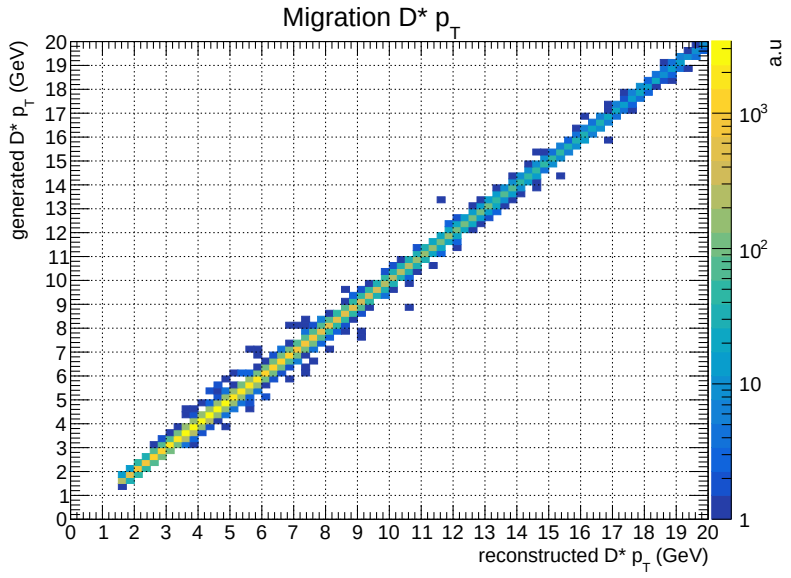
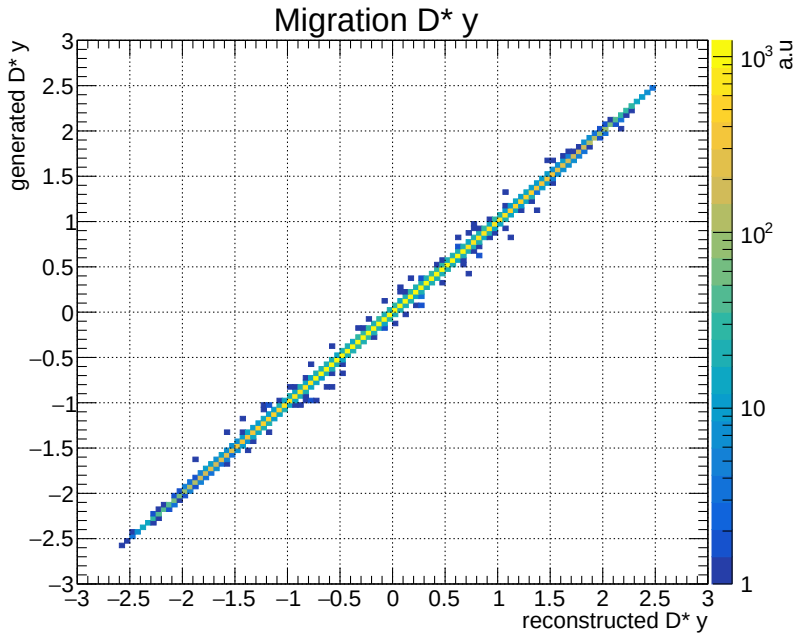


Figure 6.13: Number of extracted $D^{*\pm}$ signal events for each phase space bin. The majority of the phase space has a D^{*} signal except for a few regions in the higher |y| bins.



(a)



(b)

Figure 6.14: Scatter plots showing the migration between generated and reconstructed $D^{*\pm}$ p_T (a) and y (b). The plots are highly diagonal, indicating that there is little migration between the generated and reconstructed quantities.

Figure 6.15: Efficiency of $D^{*\pm}$ reconstruction in the D^0 MC.

6.5.3 Combined prompt and non-prompt $D^{*\pm}$ cross section

The cross sections of the $D^{*\pm}$ are determined and compared with the result of other LHC experiments and theoretical prediction. At this stage, the result of this analysis has both prompt and non-prompt contributions to the cross section and no separation between them was made. The cross section is defined as:

$$\sigma = \frac{N_{sub}(D^{*\pm})}{\mathcal{L} \cdot \text{efficiency} \cdot \text{BR}} \tag{6.11}$$

where σ is the cross section of $D^{*\pm}$ after including its decay channel BR of 0.0267 [8], N_{sub} is the number of $D^{*\pm}$ signal events from the background subtraction method (see Figure 6.13), $\mathcal{L}$ is the luminosity of the data shown in Table 6.1 and the efficiency is given in Figure 6.15 (see Subsection 2.4.3). The measured cross section of $D^{*\pm}$ for each p_T and |y| phase space bin is shown in Figure 6.16.

Figure 6.17 shows the differential cross sections of $D^{*\pm}$ as a function of p_T in bins of |y|. Only statistical uncertainties are included in Figure 6.17 for qualitative comparison. In all cases, the last bin in each distribution is the overflow bin. The circle bullet is the measurement from CMS (this analysis), the rectangle bullet is the ALICE measurement and the cross represents LHCb. As shown in Figure 6.1, the CMS region is overlapping at the central rapidity ($|y| < 0.5$) with the ALICE [11] measurements, while the |y| region 2-2.5 is overlapping with the LHCb [12] measurements. The CMS measurement contains both the charm and beauty contributions while ALICE and LHCb contain only the charm contribution. A factor of two is needed and taken into account for the ALICE and LHCb measurements to include the charge conjugate and negative rapidity, respectively.

The result is also compared with the theoretical predictions such as Pythia (LO) and FONLL (see Subsection 2.4.4). The Pythia prediction is coming from the D^0 MC sample (see Table 6.4). The sample is split into the charm (c) and beauty (b) contribution, based on whether or not the D^* meson is produced promptly at the generator level. Pythia is used for shape comparison only and an arbitrary number of 0.7 is used to normalize it to the data. From Figure 6.17, it can be seen that the b contribution is ∼10%. In FONLL, the CTEQ 6.6 PDF is used based on the study in Ref. [13]. The uncertainty of FONLL is calculated from the combination of scale variation, charm quark mass and the PDF (see Subsection 2.4.4). A factor of four is taken into account for FONLL to include the D^* charge conjugate and |y|. Qualitatively, the measurement of this analysis is comparable with the ALICE and LHCb measurements and is within the uncertainty of the theoretical prediction as shown in the ratio plot of each distribution.

The result of this analysis in Chapter 6 onwards is not compared with the ATLAS measurements [14] for several reasons. The kinematic phase space of ATLAS doesn't include the lower p_T region ($p_T < 3.5$ GeV), the bins in p_T and |y| are wider, and instead of using rapidity, pseudorapidity was used in the ATLAS measurement. However, the ATLAS result was used as a reference point during on early stage of this analysis. That is the reason why our Δm distributions are divided into lower p_T (< 3.5 GeV) and higher p_T (> 3.5 GeV) regions (see Figure 6.11) before it is divided into smaller binning. A cross check and comparison of

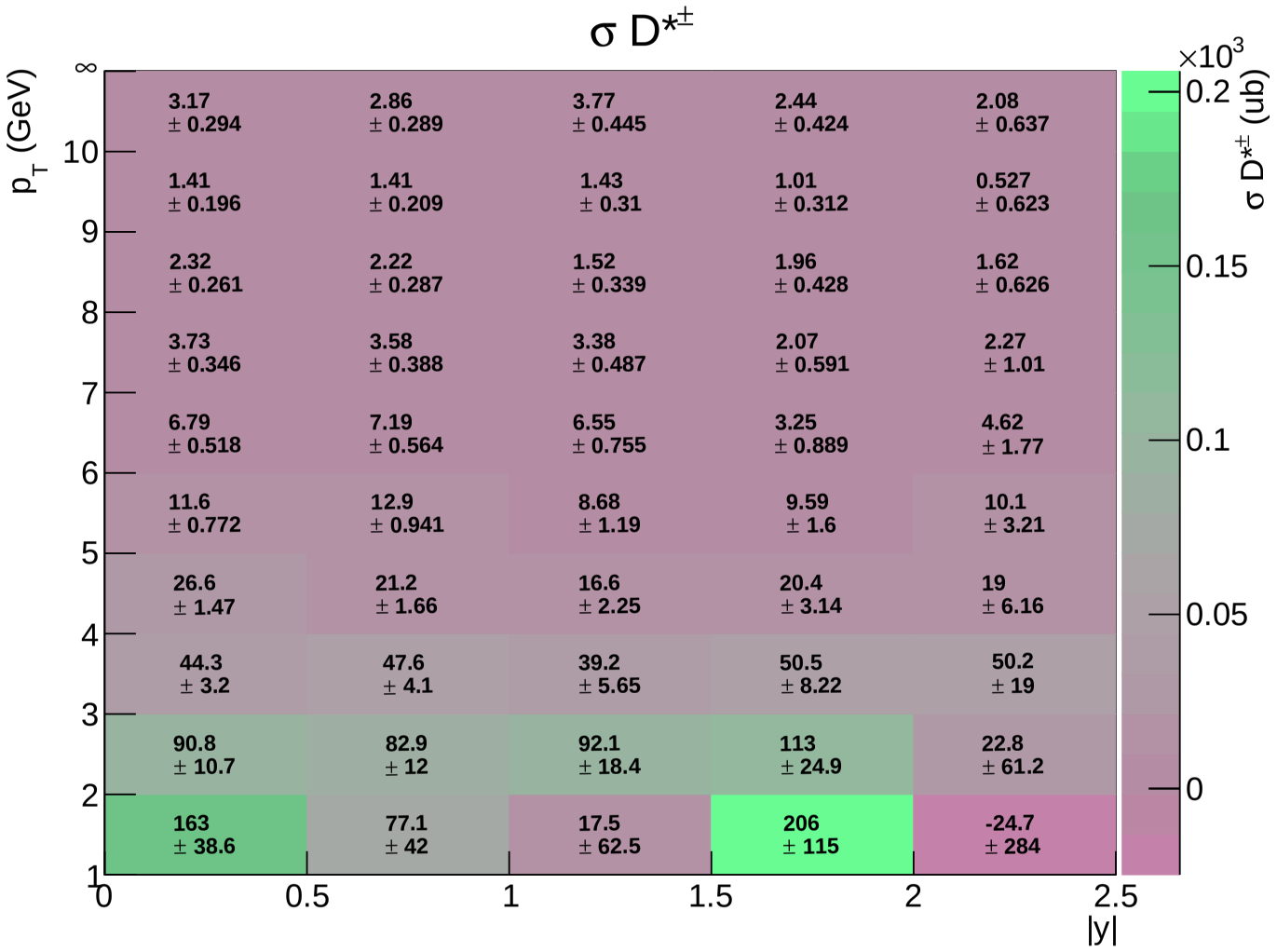


Figure 6.16: The measured value of the $D^{*\pm}$ cross section (μb) for each p_T and |y| bin before beauty and charm separation.

the D^* cross section was obtained during this study and is shown in Appendix E.

6.5.4 Beauty and charm separation

Based on Figure 6.17, the $D^{*\pm}$ cross section in the CMS data contains both the beauty and charm contributions. To measure the total charm cross section, these contributions need to be separated. Since the lifetime of b hadrons is long enough to produce the $D^{*\pm}$ away from the PV, a variable called D^0 distance of closest approach (DCA) is calculated to distinguish them. This variable is sensitive to the lifetime of the $D^{*\pm}$ mother particles. D^0 DCA is defined as:

$$D^0 \text{ DCA} = \vec{\Delta}^{D^0} \cdot \sin\phi \tag{6.12}$$

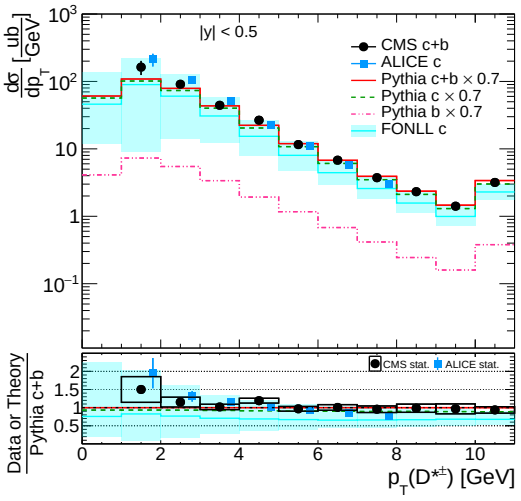
where D^0 can be a prompt D^0 (coming from charm/PV) or a non-prompt D^0 (coming from beauty)and $\vec{\Delta}^{D^0}$ is defined the same as Equation 6.3. Figure 6.18 shows a study from Ref. [15] where the contributions from prompt and non-prompt D^0 as a function of D^0 DCA are determined. The smaller the D^0 DCA, the higher the probability that the D^0 is prompt. In this analysis, the fraction of the charm contribution is assumed to be 0.90 ± 0.05, following input from the studies reported in Ref. [16] which is consistent with the value of b contribution mentioned in Subsection 6.5.3.

6.5.5 Prompt $D^{*\pm}$ cross section

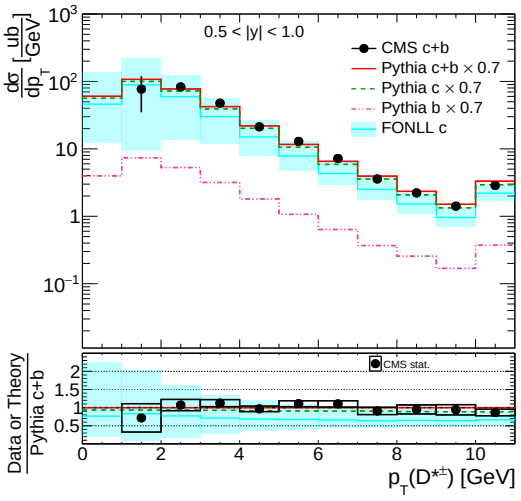
After the charm and beauty contributions have been separated and the charm fraction is implemented at the cross section level, the prompt $D^{*\pm}$ cross section as a function of p_T in |y| bins and vice versa are plotted. This cross section without the beauty contribution in the CMS data is shown in double differential plots for quantitative comparison. Due to time constraints, our result includes only statistical uncertainty, while the ALICE and LHCb measurements include statistical and systematic uncertainties. Figure 6.19 shows the $D^{*\pm}$ cross sections as a function of p_T in log scale. At $|y| < 0.5$, our result is compared with the ALICE measurement and they are in agreement. At |y| 0.5-1.0, 1.0-1.5, and 1.5-2.0 in Figure 6.19, there are three phase space regions where no D^* cross section measurement has been done at the LHC for p_T below 3.5 GeV. This analysis is the first measurement at these regions and the result is comparable with Pythia c and FONLL. Lastly, at |y| 2-2.5, our result is compared with the LHCb measurement. In this region, even though our result is in agreement with LHCb, their uncertainties are smaller, therefore, we use their result for the total charm cross section in that region. Overall, it can be stated that our results are in agreement with the results from ALICE, LHCb, FONLL and Pythia c within uncertainty. Figure 6.20 and Figure 6.21 show the $D^{*\pm}$ cross section as a function of |y|. These figures are the same cross sections as in Figure 6.19 but in linear scale, and also include $|y| > 2.5$ which covers the full phase space of LHCb measurement. Figure 6.22 and Figure 6.23 are the double differential plots that show the overview of the measurements at the LHC with theoretical predictions in all phase space.

6.6 Systematic Uncertainties

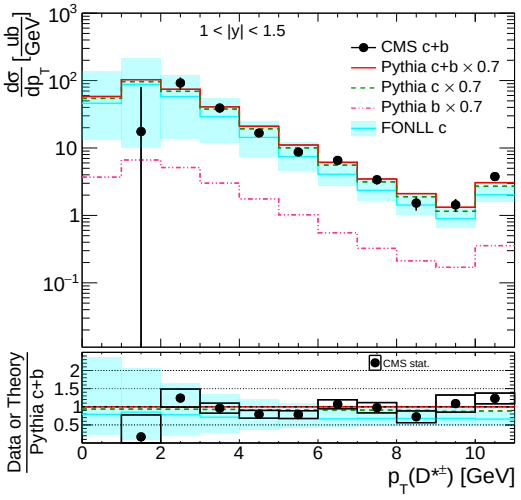
The systematic uncertainties of the cross section measurement are partially described in this section. Systematic uncertainties are important to quantify the limitations of the mea-



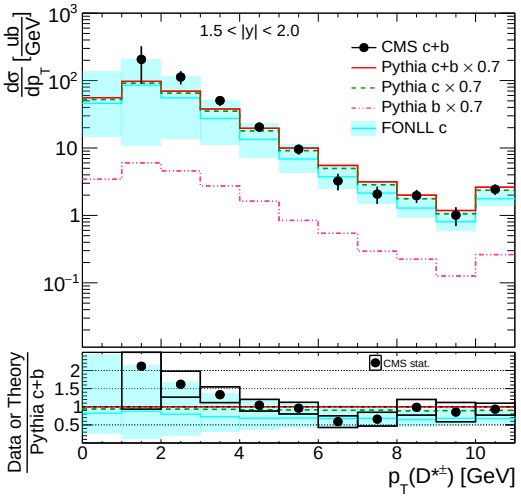
(a) $|y| < 0.5$



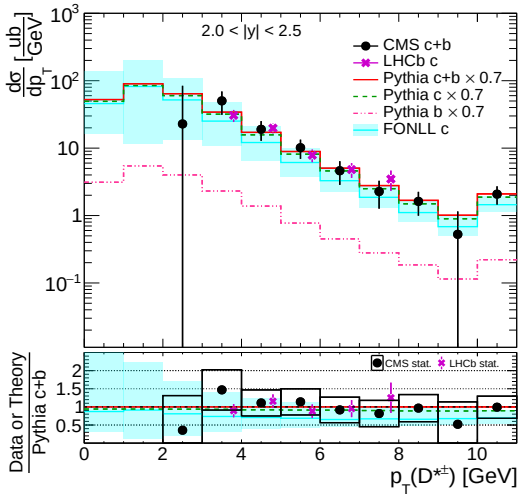
(b) $0.5 < |y| < 1.0$



(c) $1.0 < |y| < 1.5$



(d) $1.5 < |y| < 2.0$



(e) $2.0 < |y| < 2.5$

Figure 6.17: $D^{*\pm}$ cross section as a function of p_T in $|y|$ bins. The last bin is the overflow bin. See Subsection 6.5.3 for the meaning of Pythia c and b.

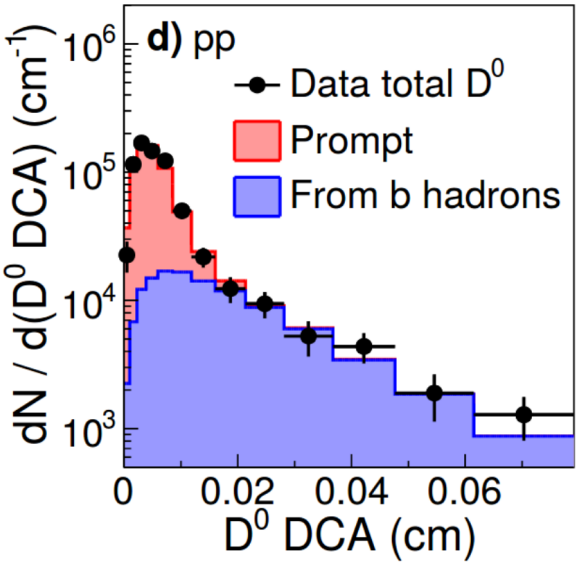
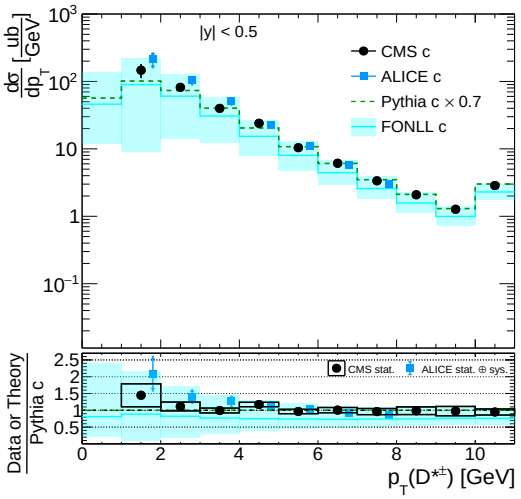


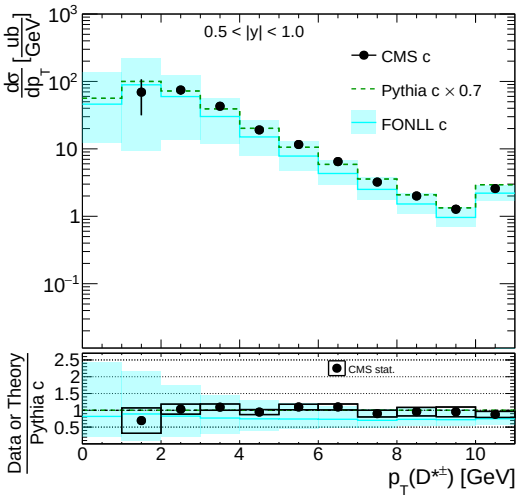
Figure 6.18: The stack plot shows the signal DCA distribution obtained with the sideband subtraction for D^0 from prompt (red) and beauty (blue) hadrons. Plot taken from Ref. [15].

surements and also to take into account the uncertainties of the performance of the detector efficiencies, resolutions as well as the assumptions used in the MC. The systematic uncertainties are estimated using the Minimum Bias MC; e.g. the last two datasets in Table 6.3 or based on previous study, e.g. in Ref. [17]. Even though the systematics uncertainties are not fully included in this analysis due to time constraints, several sources of the systematic uncertainties that are relevant to this analysis are already discussed below.

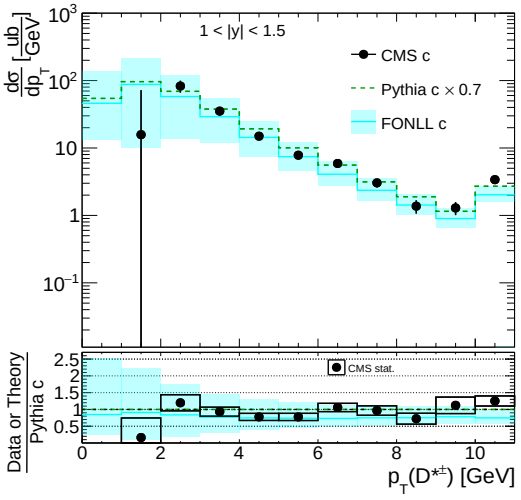
- **Trigger efficiencies**: The uncertainty of the trigger efficiency is considered to be negligible ($< 1\%$, see Subsection 6.5.2). This is because most of the events are coming from the Zero Bias or Minimum Bias triggers and untriggered Next-to-Minimum Bias events.
- **Tracking efficiencies**: The uncertainty for kaon and pion tracks is taken from the study in Ref. [17] which is 3.9% each. The uncertainty is estimated by taking the ratio of four-body ($K\pi\pi\pi$) and two-body ($K\pi$) D^0 decays. If the tracks are reconstructed with perfect efficiency, the ratio would simply be the ratio between the branching ratios of the two D^0 decay modes. However, as the tracking efficiency is not perfect, this ratio is the ratio between two branching ratios multiplied by the efficiencies to reconstruct the two-track and four-track final states. Assuming that the efficiency to reconstruct tracks within an event is independent from each other, and that the efficiency to reconstruct π and K is the same, this is simply the ratio between two branching ratios multiplied by the square of the track efficiency. For the π_s track, the efficiency uncertainty is estimated to be 5% following Ref. [18].
- **Luminosity**: The systematic uncertainty of the luminosity for 7 TeV 2010 data given by the CMS collaboration is 3.6% [19]. This enters through the reference luminosities used as described in Subsection 6.3.1.



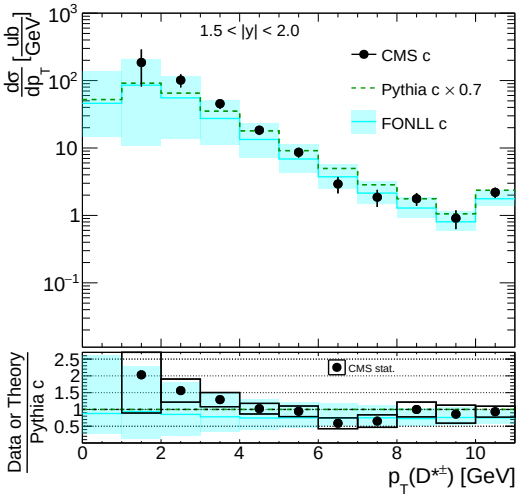
(a) $|y| < 0.5$



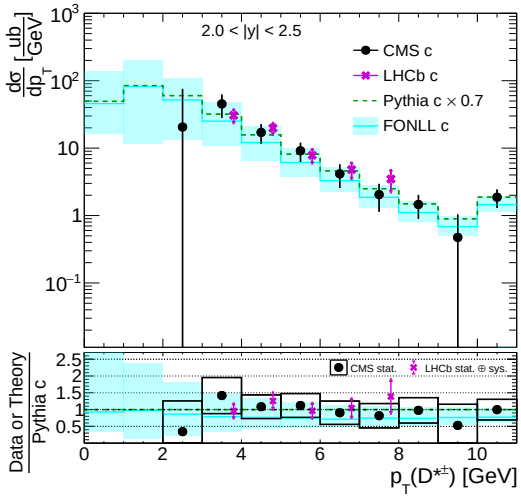
(b) $0.5 < |y| < 1.0$



(c) $1.0 < |y| < 1.5$

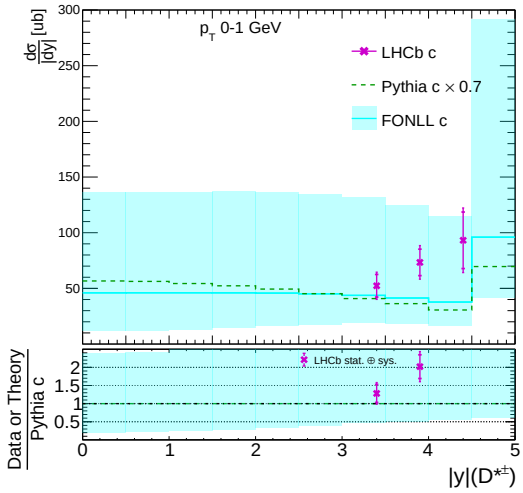


(d) $1.5 < |y| < 2.0$

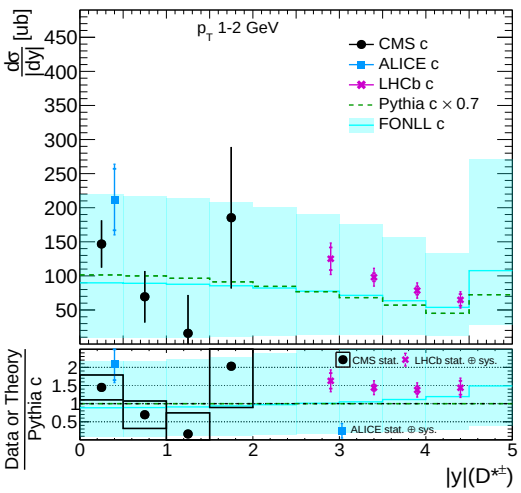


(e) $2.0 < |y| < 2.5$

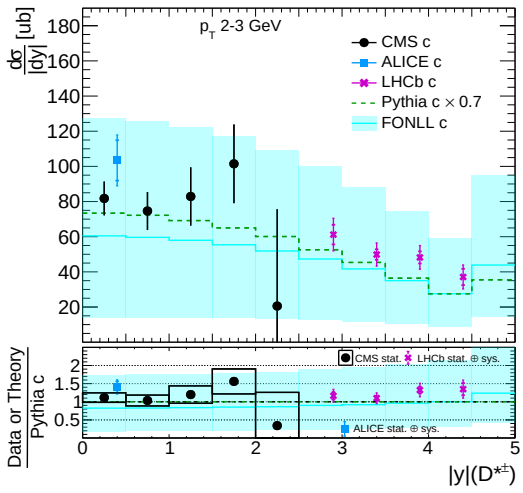
Figure 6.19: Prompt $D^{*\pm}$ cross section as a function of p_T in |y| bins. Last bin is the overflow bin. For |y| above 2.5, the charm cross section will use LHCb measurement.



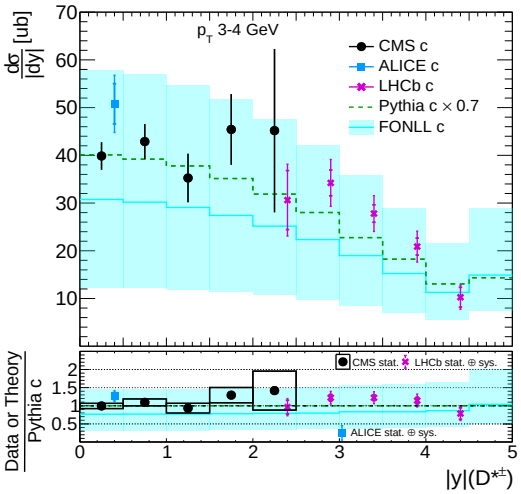
(a) p_T 0-1 GeV



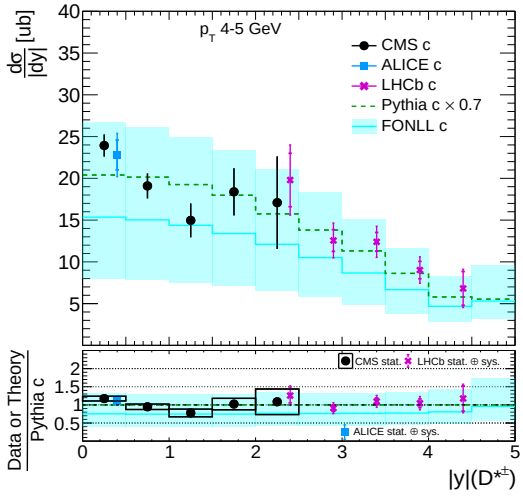
(b) p_T 1-2 GeV



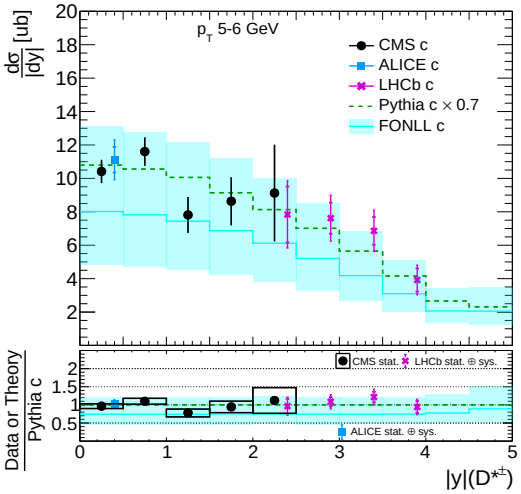
(c) p_T 2-3 GeV



(d) p_T 3-4 GeV

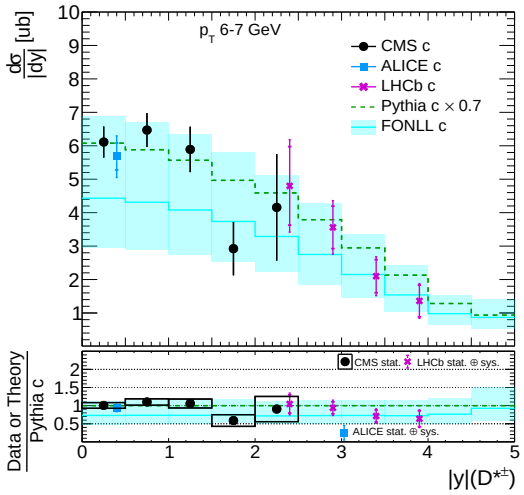


(e) p_T 4-5 GeV

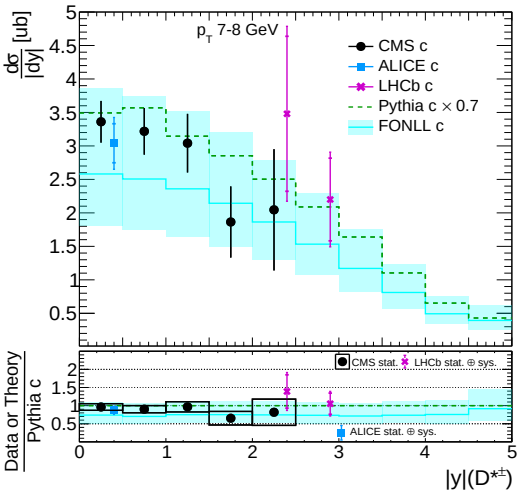


(f) p_T 5-6 GeV

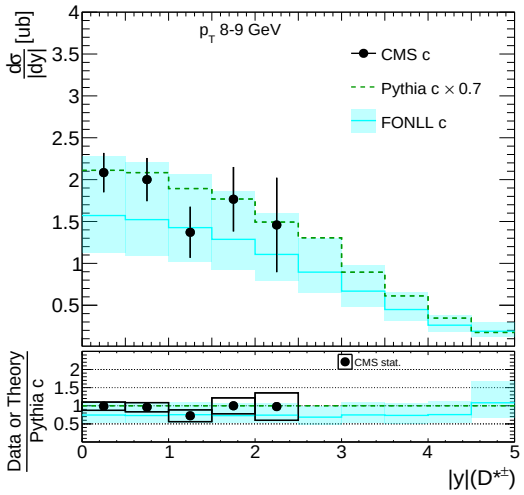
Figure 6.20: Prompt $D^{*\pm}$ cross section as a function of |y| in p_T bins 0 to 6 GeV. The last bin is the overflow bin.



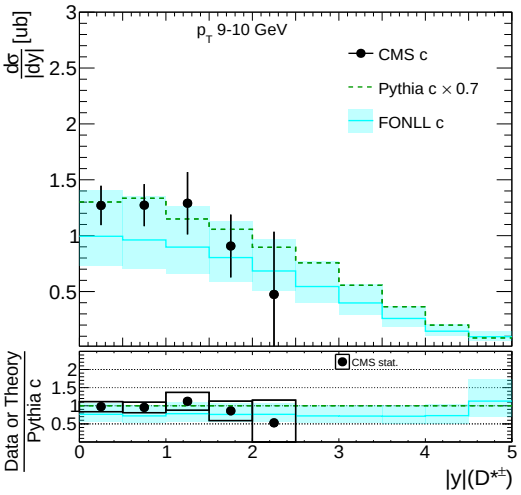
(a) p_T 6-7 GeV



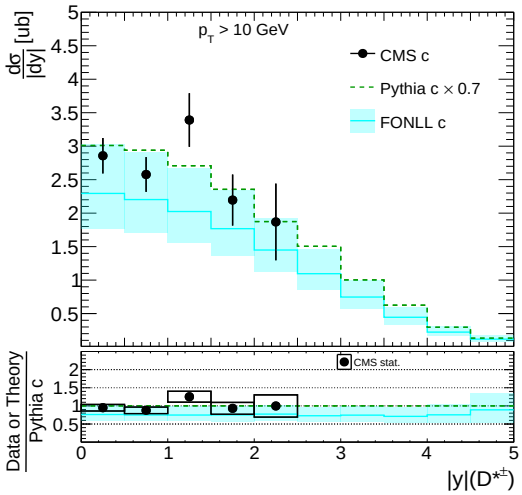
(b) p_T 7-8 GeV



(c) pT 8-9 GeV



(d) p_T 9-10 GeV



(e) $p_T > 10$ GeV

Figure 6.21: Prompt $D^{*\pm}$ cross section as a function of |y| in p_T bins > 7 GeV. The last bin is the overflow bin.

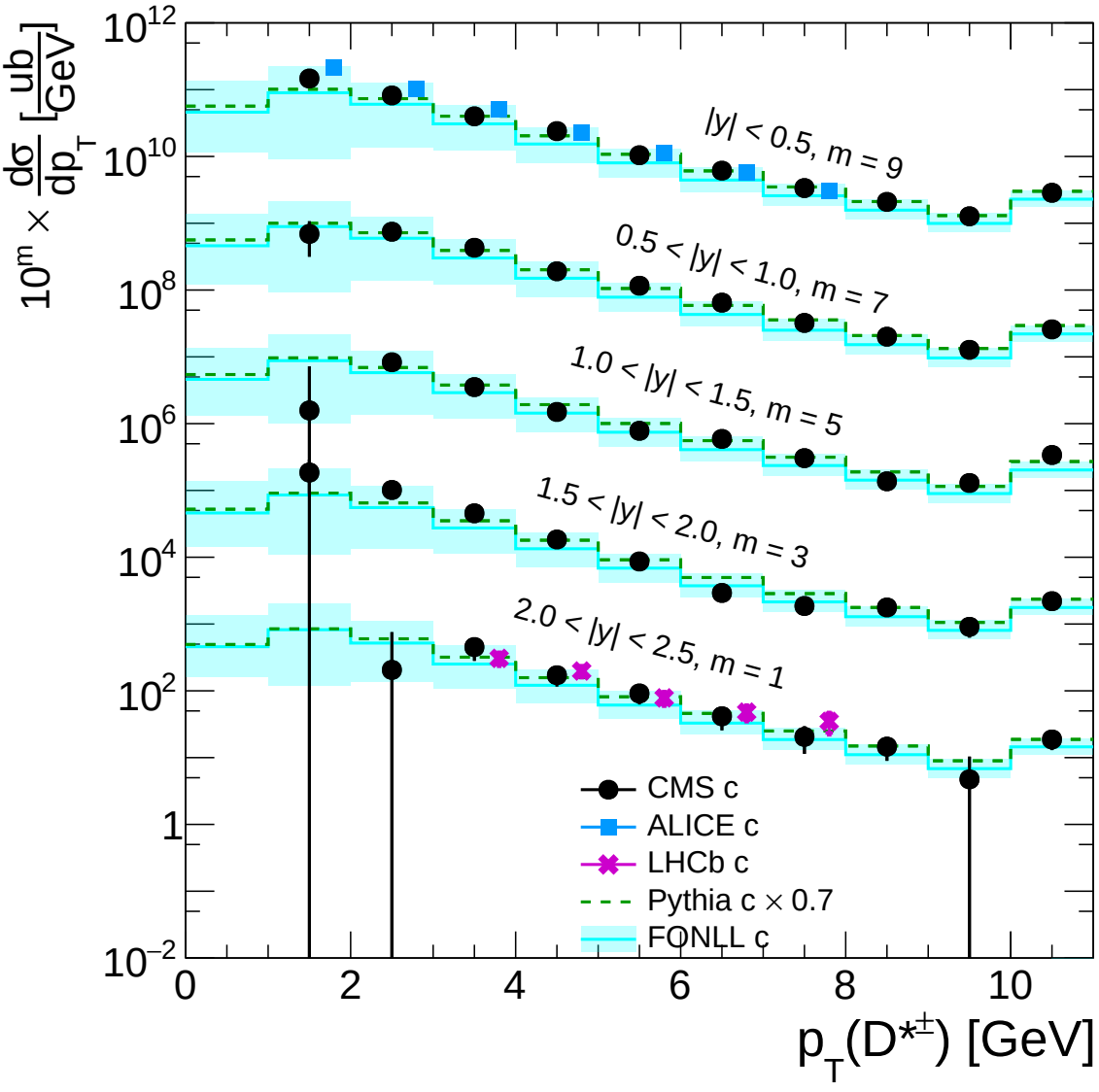


Figure 6.22: Prompt $D^{*\pm}$ double differential cross section as a function of p_T in |y| bins. The last bin is the overflow bin.

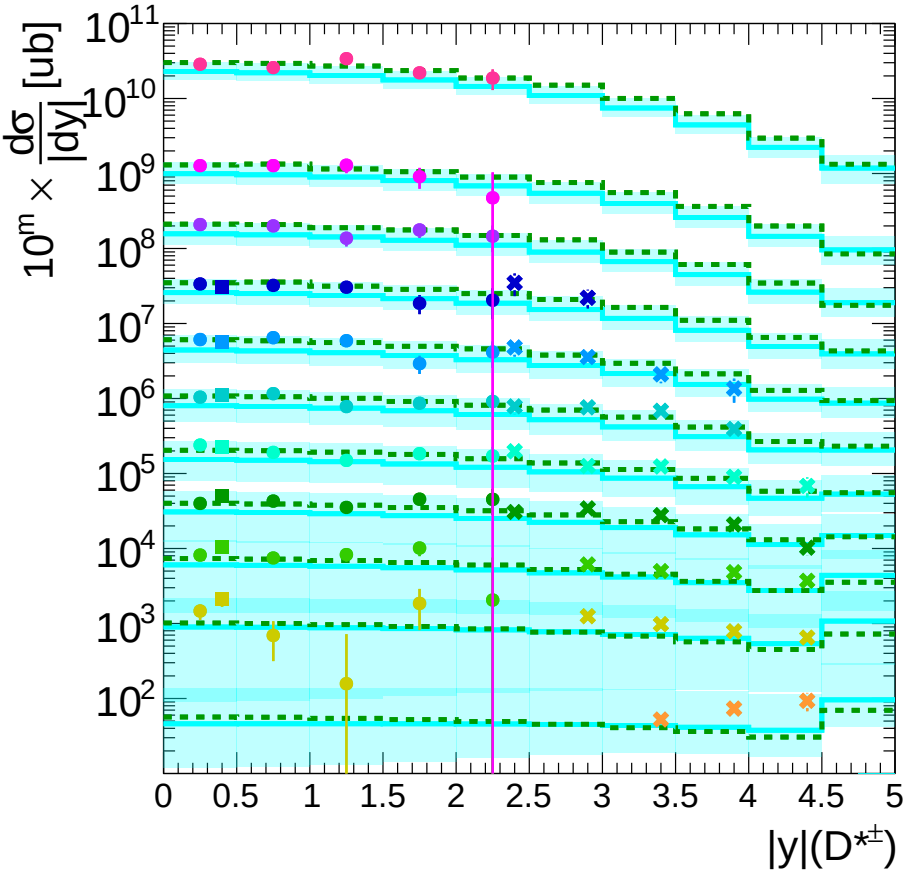


Figure 6.23: Prompt $D^{*\pm}$ double differential cross section as a function of |y| in p_T bins. Refer to Figure 6.24 for the legend. The last bin is the overflow bin.

- **Branching fractions**: The uncertainty related to the branching fractions is a combination of the uncertainties of the $D^{*\pm} \to D^0 \pi^\pm$ and the $D^0 \to K^-\pi^+$ branching ratios from the PDG [8]. The value of the BR is $2.67 \pm 0.03\%$ as described in Subsection 2.4.3.

- **Pileup**: The average pileup systematics in the MC is taken into account with respect to the data. It was done by checking the efficiency on the pileup and no-pileup MC samples of Minimum Bias TuneZ2Star (the last two samples in Table 6.3). The uncertainty is 5% using an older version of D^* kinematic reconstruction. It can potentially be improved by using the latest version of the D^* reconstruction in this analysis.

- D^* **Signal extraction**: One of the uncertainties related to the D^* signal extraction is due to the uncertainty in the normalization factor used in estimating the background contribution (see Subsection 6.5.1). The uncertainty estimated is 0.8% for $p_T < 3.5$ GeV and 0.8% for $p_T > 3.5$ GeV. The uncertainty should be calculated separately for each bin in the future. This estimation is a simpler approach than the one employed by Ref. [18], where the background contribution is estimated by performing a functional fit and therefore, the systematic uncertainties are due to the uncertainties in the function parameters.

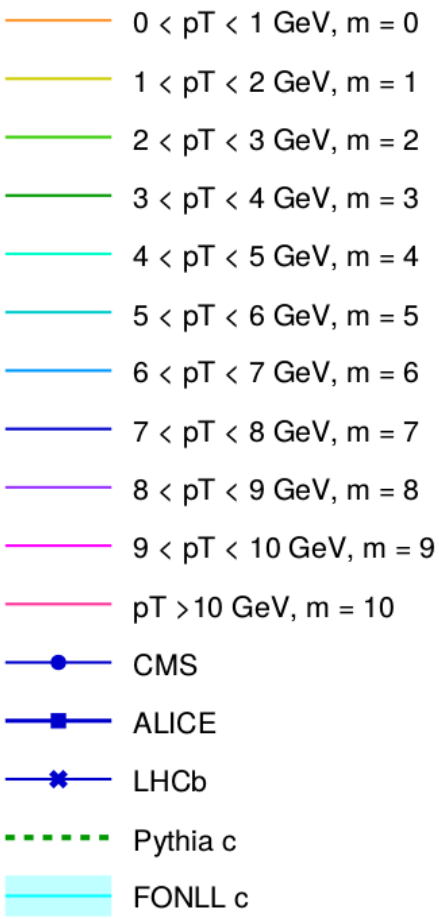


Figure 6.24: Legend for double differential cross section in Figure 6.23 and Figure 7.1.

- **Charm contribution in the** D^* **sample**: As discussed in Subsection 6.5.4, the charm contribution in the D^* sample is found to be 0.90 ± 0.05. The uncertainty on this number can be taken into account as a systematic uncertainty.

In addition to the sources of the systematics listed above, other systematics sources can also affect the results of this analysis. For example, the time dependence of the tracker may introduce a systematic effect through its impact on the selection cut efficiencies, in particular the D^0 significance cut. This can be evaluated by splitting the data into different time bins and comparing the extracted cross sections in each. Since the systematics are not yet complete, they will not yet be applied in the cross section measurement.

6.7 Bibliography

[1] ALICE Collaboration, "Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **77** (2017), no. 8, 550, doi:10.1140/epjc/s10052-017-5090-4, arXiv:1702.00766.

[2] LHCb Collaboration, "Prompt charm production in pp collisions at sqrt(s)=7 TeV", *Nucl. Phys. B* **871** (2013) 1–20, doi:10.1016/j.nuclphysb.2013.02.010, arXiv:1302.2864.

[3] A. Accardi et al., "A Critical Appraisal and Evaluation of Modern PDFs", *Eur. Phys. J. C* **76** (2016), no. 8, 471, doi:10.1140/epjc/s10052-016-4285-4, arXiv:1603.08906.

[4] CMS Collaboration Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", Technical Report CMS-PAS-TRK-10-005, CERN, Geneva, 2010.

[5] A. Geiser, N. Z. Jomhari, and J. Metwally, "Measurement of total and associated charm and beauty cross sections in the CMS experiment - technical information", CMS Analysis Note CMS AN-2018/284, 2018.

[6] CMS Collaboration, "Public CMS Open Data information", (2014). http://opendata.cern.ch/docs/about-cms.

[7] CERN, "CMS luminosity information, for 2010 CMS open data", (2014). https://opendata.cern.ch/record/1050.

[8] Particle Data Group Collaboration, "Review of Particle Physics", *PTEP* **2020** (2020), no. 8, 083C01, doi:10.1093/ptep/ptaa104.

[9] A. Geiser, N. Z. Jomhari, and J. Metwally, "Update of preselection cuts for D*", (2020). https://indico.desy.de/event/27610/contributions/92272/attachments/62171/75399/Precuts_-and_dEdx_QCD.pdf.

[10] ZEUS Collaboration, "Measurement of charm fragmentation fractions in photoproduction at HERA", *JHEP* **09** (2013) 058, doi:10.1007/JHEP09(2013)058, arXiv:1306.4862.

[11] ALICE Collaboration, "Measurement of charm production at central rapidity in proton-proton collisions at $\sqrt{s} = 2.76$ TeV", *JHEP* **07** (2012) 191, doi:10.1007/JHEP07(2012)191, arXiv:1205.4007.

[12] LHCb Collaboration, "Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 5$ TeV", *JHEP* **06** (2017) 147, doi:10.1007/JHEP06(2017)147, arXiv:1610.02230.

[13] Y. Yang, A. Geiser, N. Jomhari, and J. Metwally, "FONLL official code usage development and PROSA PDF", (2021). https://indico.desy.de/event/29314/contributions/99471/attachments/64362/78938/FONLL official code usage develpment and PROSA PDF.pdf.

[14] ATLAS Collaboration, "Measurement of $D^{*\pm}$, $D^{\pm}$ and $D_s^{\pm}$ meson production cross sections in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Nucl. Phys. B* **907** (2016) 717–763, doi:10.1016/j.nuclphysb.2016.04.032, arXiv:1512.02913.

[15] CMS Collaboration, "Studies of Beauty Suppression via Nonprompt D^0 Mesons in Pb-Pb Collisions at $Q^2 = 4$ GeV^2", *Phys. Rev. Lett.* **123** (2019), no. 2, 022001, doi:10.1103/PhysRevLett.123.022001, arXiv:1810.11102.

[16] J. Metwally, A. Geiser, N. Jomhari, and Y. Yang, "D cross sections at 5 TeV and charm beauty separation", (2021). https://indico.cern.ch/event/1039238/contributions/4366122/attachments/2245665/3808289/BPH_-talk.pdf.

[17] CMS Collaboration Collaboration, "Measurement of Tracking Efficiency", Technical Report CMS-PAS-TRK-10-002, CERN, Geneva, 2010.

[18] CMS Collaboration, "Measurement of prompt open-charm production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2021) 225, doi:10.1007/JHEP11(2021)225, arXiv:2107.01476.

[19] CMS Collaboration, "Luminosity Physics Object Group (Lumi POG)", (2021). https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM.

CHAPTER

7

TOTAL CHARM CROSS SECTION MEASUREMENT

7.1 Total Charm Production Cross Section

In this chapter, a preliminary result for the total charm cross section at 7 TeV for the whole phase space is calculated. The result is from the combination of CMS measurements (this analysis), LHCb measurements, and extrapolations from `PYTHIA` (LHCb's extrapolation region) and using FONLL calculations. Figure 7.1 shows the double differential $D^{*\pm}$ cross section in p_T and |y| bins. It is the same plot as Figure 6.23 but with boxes to indicate the different phase space regions.

Box 1 is the $D^{*\pm}$ measurement of this analysis that was explained in Chapter 6. The region covers |y| 0-2 and $p_T > 1$ GeV, and for |y| 2-2.5, only p_T above 8 GeV is covered. Below that p_T, the measurement from the LHCb is used as it has smaller uncertainties. The measured $D^{*\pm}$ cross section in this region is 1096 ± 133 (stat.) μb. The phase space covered by CMS is the largest possible kinematic phase space at the LHC from a single experiment with p_T down to 1 GeV.

Box 2 is the FONLL prediction that is calculated for |y| 0-2 and p_T 0-1 GeV. FONLL is used as there is no measurement performed in this region. The uncertainty in the FONLL calculation is the sum in quadrature of uncertainties due to variation of renormalization and factorization scales, charm quark mass with central value of 1.5 GeV and CTEQ 6.6 PDF (see Subsection 2.4.4). The calculated cross section in this region is 184^{+361}_{-133} (FONLL) μb.

Box 3 is based on the LHCb measurement from Ref. [1]. A factor two is always taken into account for |y| as mentioned in Subsection 6.5.3 for this analysis. Table 7.1 shows the phase space regions covered by LHCb for the $D^{*\pm}$ cross section. The total measured $D^{*\pm}$ cross section is 1010 ± 39 (stat.) ± 115 (sys.) μb.

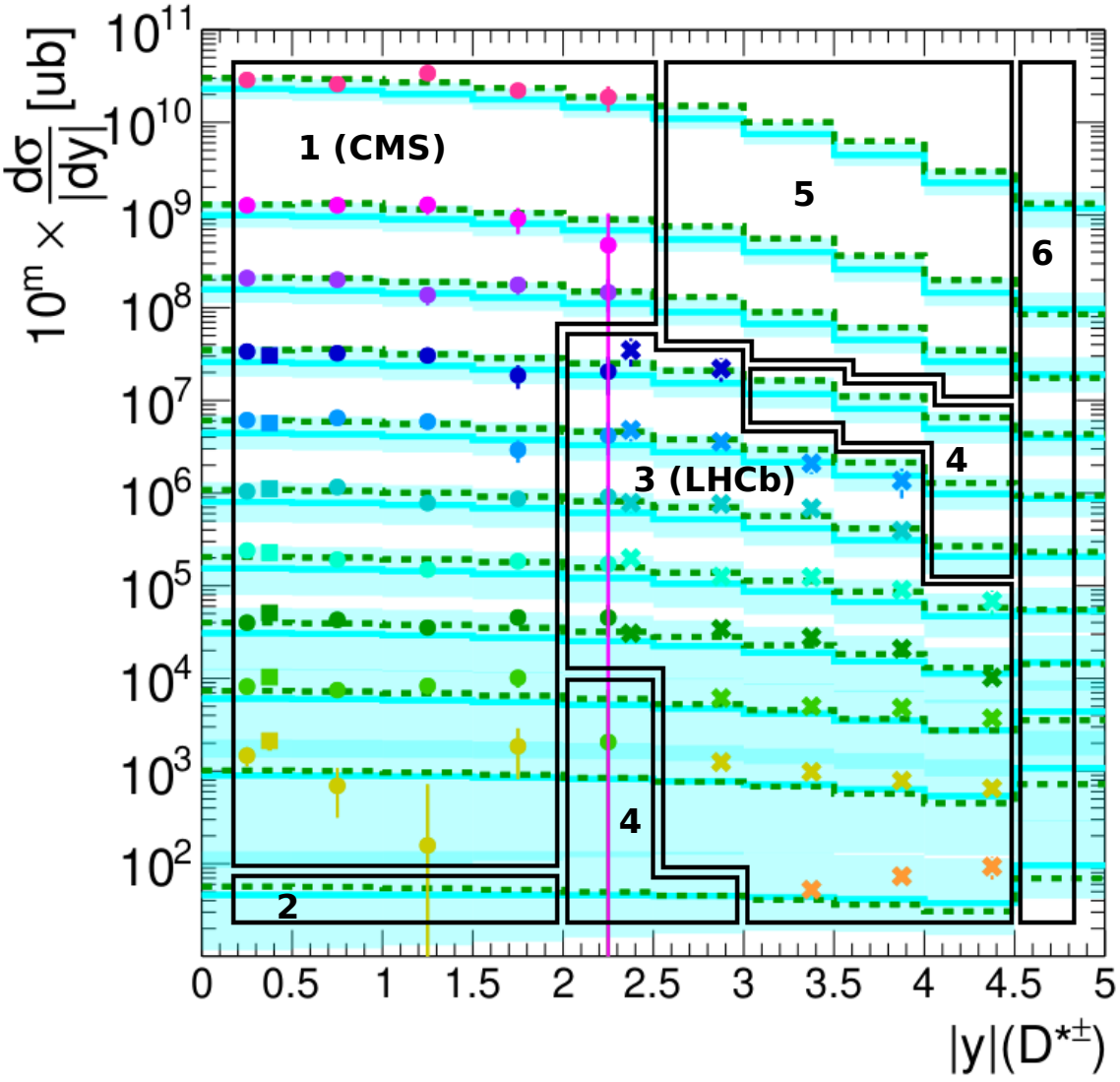


Figure 7.1: Boxes correspond to the phase space in calculating total charm cross section from the CMS and LHCb measurements, FONLL and LHCb's extrapolation. Please refer to Figure 6.24 for the legend of this plot and the text for details.

Box 4 is the extrapolated region from the LHCb measurement. In Table 7.1, cells without a reported measurement are included in the LHCb extrapolation region. The extrapolations are computed from `PYTHIA6.4` [2] predictions as described in Ref. [1]. The $D^{*\pm}$ cross section is calculated by taking the extrapolation factor minus 1 and multiply with the measured cross section. The statistical and systematic uncertainties are calculated in the same way but for `PYTHIA` extrapolation's uncertainty, the value is taken from extrapolation of the total $D^{*\pm}$ cross section in all phase space shown in Table 7.1. Thus, the $D^{*\pm}$ cross section in LHCb extrapolation region is 344 ± 6.6 (stat.) ± 20 (sys.) ± 38 (`PYTHIA`) μb.

Lastly, the $D^{*\pm}$ cross section in box 5 and 6 are calculated using FONLL predictions. The former covers a region of $2.5 < |y| < 4.5$ and $p_T > 8$ GeV, and the latter covers $|y| > 4.5$ and $p_T > 0$ GeV beyond LHCb, with a $D^{*\pm}$ cross section of $6.0^{+0.6}_{-0.4}$ (FONLL) μb and 272^{+428}_{-160} (FONLL) μb, respectively. The total $D^{*\pm}$ cross section for the full kinematic phase space is shown in Table 7.2. Only statistical uncertainties, and uncertainties from the `PYTHIA` and FONLL are included since systematics in this analysis are not yet completed. The latter two are combined to obtain the total FONLL/`PYTHIA` uncertainty. The FONLL calculations in box 2, 5 and 6 are considered correlated, therefore the uncertainty is added linearly. The same method is applied for LHCb in box 3 and 4 which are assumed to be correlated. Other than these, the uncertainty is computed by summing in quadrature.

In this analysis, the charm fragmentation fraction to D^* was asummed to be universal during the calculation of the predictions for the differential charm cross sections even though recently there is some hint that it might be not universal [3] based on the latest study by ALICE [4]. The measured charm fragmentation fraction in the ALICE paper is $0.155^{+0.04}_{-0.02}$. This value is chosen to calculate the total charm cross section because the measurement is done using pp collisions, the same as the data use in this analysis. The extrapolation factor obtained in this analysis is 1.4 calculated from the ratio of total $D^{*\pm}$ cross section in all phase space over total $D^{*\pm}$ cross section from the CMS and LHCb measurements. This is the smallest extrapolation achieved for the total charm cross section at the LHC so far.

The total charm cross section is measured to be $9.4 \pm 0.45\,(\text{stat.})\,^{+2.50}_{-0.95}$ (FONLL/`PYTHIA`) mb. Since the systematic uncertainties have not yet been estimated in our measurement, it is also not evaluated for the total charm cross section, even though some of the measurements that are used include systematic uncertainties. The ALICE experiment also calculated a total charm cross section [5] for the full phase space interpolating between the ALICE and LHCb measurements, with a value of 7.44 ± 0.14 (stat.) ± 0.46 (sys.) $^{+0.13}_{-0.07}$ (ext.) mb. Our result is consistent with ALICE's result and also with the theory (∼10 mb) [6] stated in the Section 2.4.

7.2 Bibliography

[1] LHCb Collaboration, "Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 5$ TeV", *JHEP* **06** (2017) 147, `doi:10.1007/JHEP06(2017)147`, `arXiv:1610.02230`.

Table 7.1: $D^{*\pm}$ cross section measurement covered by LHCb for p_T and y bins (box 3). The first uncertainty is statistical, and the second is the total systematic. The regions not measured by LHCb are the extrapolation regions from LHCb. The table, extrapolation factor and the total $D^{*\pm}$ σ in all phase space below are taken from Ref. [1].

p_T	y				
(GeV/c)	(2.0, 2.5)	(2.5, 3.0)	(3.0, 3.5)	(3.5, 4.0)	(4.0, 4.5)
(0, 1)			26.17 ± 5.17 ± 3.25	36.67 ± 6.02 ± 4.53	46.60 ± 12.77 ± 6.88
(1, 2)		62.56 ± 8.42 ± 7.91	49.02 ± 3.13 ± 5.73	39.27 ± 3.15 ± 4.62	32.40 ± 4.41 ± 4.06
(2, 3)		30.60 ± 2.85 ± 3.66	24.93 ± 1.54 ± 2.91	24.11 ± 1.77 ± 2.86	18.55 ± 2.37 ± 2.45
(3, 4)	15.31 ± 3.11 ± 2.12	17.11 ± 1.37 ± 2.04	13.90 ± 0.93 ± 1.63	10.44 ± 0.91 ± 1.34	5.13 ± 1.06 ± 0.70
(4, 5)	9.90 ± 1.61 ± 1.35	6.28 ± 0.66 ± 0.81	6.20 ± 0.57 ± 0.74	4.51 ± 0.53 ± 0.59	3.41 ± 1.02 ± 0.52
(5, 6)	3.92 ± 0.84 ± 0.55	3.81 ± 0.47 ± 0.50	3.43 ± 0.42 ± 0.45	1.96 ± 0.35 ± 0.27	
(6, 7)	2.40 ± 0.59 ± 0.36	1.78 ± 0.32 ± 0.24	1.05 ± 0.25 ± 0.15	0.68 ± 0.24 ± 0.10	
(7, 8)	1.74 ± 0.58 ± 0.30	1.10 ± 0.31 ± 0.17			

Total measured σ	505 ± 19.5 (stat.) ± 57.5 (sys.) μb
The extrapolation factor in all phase space above	1.340 ± 0.037
Total $D^{*\pm}$ σ in all phase space above	677 ± 26 (stat.) ± 77 (sys.) ± 19 (ext.) μb

Table 7.2: Table showing the total integrated D* cross section (μb), extrapolation factor, and total charm cross section (mb) covering full kinematic phase space of p_T and |y|. See text for explaination of each box.

Phase space	Integrated D* cross section (μb)
1. CMS region measured:	1096 ± 133 (stat.)
2. FONLL:	184^{+361}_{-133} (FONLL)
3. LHCb region measured:	1010 ± 39 (stat.) ± 115 (sys.)
4. Extrapolated LHCb:	344 ± 6.6 (stat.) ± 20 (sys.) ± 38 (PYTHIA)
5. FONLL:	$6.0^{+0.6}_{-0.4}$ (FONLL)
6. Beyond LHCb (FONLL):	272^{+428}_{-160} (FONLL)
Total $D^{*\pm}$ cross section	2912 ± 140 (stat.) $^{+790}_{-295}$ (FONLL/PYTHIA) μb

Extrapolation factor in full phase space	1.4
Total $c\bar{c}$ cross section	9.4 ± 0.45 (stat.) $^{+2.5}_{-0.95}$ (FONLL/PYTHIA) mb

[2] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[3] A. Metz and A. Vossen, "Parton Fragmentation Functions", *Prog. Part. Nucl. Phys.* **91** (2016) 136–202, doi:10.1016/j.ppnp.2016.08.003, arXiv:1607.02521.

[4] ALICE Collaboration, "Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC", arXiv:2105.06335.

[5] ALICE Collaboration, "Measurement of charm production at central rapidity in proton-proton collisions at $\sqrt{s} = 2.76$ TeV", *JHEP* **07** (2012) 191, doi:10.1007/JHEP07(2012)191, arXiv:1205.4007.

[6] A. Accardi et al., "A Critical Appraisal and Evaluation of Modern PDFs", *Eur. Phys. J. C* **76** (2016), no. 8, 471, doi:10.1140/epjc/s10052-016-4285-4, arXiv:1603.08906.

CHAPTER

8

SUMMARY AND OUTLOOK

The total charm cross section measurement in the full kinematic phase space of pp collisions at 7 TeV has been presented. The goal is to cover the largest possible fraction of the full kinematic phase space so that the cross section is less model-dependent. In CMS data, the charm reconstruction was done through the $D^{*\pm}$ final state of $K^-\pi^+\pi^+$ with the decay mode $D^{*\pm} \rightarrow D^0\pi^\pm$ where the D^0 decays further to $K^\mp\pi^\pm$. This final state gave the best signal to background ratio for a broad phase space coverage. Since the π from the D^* decay needs access to the lowest possible p_T, the analysis is performed at 7 TeV 2010 data, taken by the CMS experiment, due to its special low p_T tracking. It is the first and so far only measurement of charm at 7 TeV in CMS.

One of the main strategies of this analysis is to use pileup vertices as a physics resource since this analysis is statistically limited due to the fact that Zero Bias and Minimum Bias triggers are heavily prescaled. The pileup vertices in this analysis are from the muon and electron datasets as reconstructed muons and electrons can be uniquely associated with a particular vertex. One can remove this vertex and use the remaining (pileup) vertices. Thus, using pileup vertices from muon and electron datasets and the Zero Bias and Minimum Bias datasets helps to get sizeable additional statistics.

To reconstruct D^*, several techniques and methods were studied and implemented to benefit the analysis. A study on the primary vertex with and without beamspot constraints was conducted. The former is chosen as it recovers $D^{*\pm}$ efficiency at low p_T regions. It also reduces the vertex uncertainty in the xy plane and limits potential biases from including secondary tracks in the primary vertex fit. Furthermore, refitting the π track from D^* decay to the D^* vertex gives a better signal to background ratio. An important variable for the D^* reconstruction is the D^0 decay length significance. Since D^0 has a long mean lifetime, it can produce a secondary vertex before it decays. A cut on this variable selects candidates with their secondary vertex away from the D^* vertex, thus reducing the combinatorial background. Moreover, a cut on the p_T^{frac} variable is introduced. It is a cut on the ratio between the scalar sum of the p_T of the K and π tracks from the D^0 over the scalar sum of p_T pf all tracks to

ensure that the p_T of the D^0 contributes significantly to the total p_T of all tracks associated to the vertex. Another important variable for the lowest p_T bins is the dE/dx, which is the rate of energy loss by the particle as it travels through the detector. Since different particle types lose energy at different rates, one can take advantage of this feature to identify the K more accurately. The optimization of the event selection for the D^* candidate in this analysis is mainly focusing on low p_T as this is the region where the large combinatorial background is present but has a high charm cross section.

After the D^* selection cuts, the ingredients toward the D^* cross section measurement were prepared. First, the D^* signals were extracted. It was done by using a data-driven method called the wrong charge background subtraction method. This method can be used for D^* since there is a clear distinction between right charge (where the π from D^* decay has an electric charge opposite to that of K) and wrong charge (where the π from D^* decay and the K have the same electric charge) combinations. The former receives contributions from the $D^{*\pm}$ resonance and combinatorial background while the latter only receive contributions from the combinatorial background since doubly charged D mesons do not exit. Thus, this method can be used to estimate the background directly from the data. The $D^{*\pm}$ signal extraction is usually done using the Δm distribution (difference between D^* and D^0 masses) due to its superior resolution. Second, the D^* efficiency was calculated which reaches up to 59% efficiency at higher p_T and lower |y|. Third, the effective luminosity ($\mathcal{L}_{\text{eff}}$) was determined in a data-driven way using unprescaled triggers as a basis and information from the luminosity database of the CMS collaboration. The $\mathcal{L}_{\text{eff}}$ obtained from this analysis is 3.00 nb^{-1}.

Finally, the double differential $D^{*\pm}$ cross section measurement was presented. Since, by default, the CMS data contains charm (c) and beauty (b) contributions, the result first shows the $D^{*\pm}$ cross section from these contributions for qualitative comparison with the ALICE and LHCb measurements. The beauty contribution from Pythia is $\sim$10% to the total of c and b contributions. A separation of charm and beauty is then implemented at cross section level, and a study on this separation was reported consistent with the Pythia prediction. Thus, the fraction of the charm contribution is assumed to be 0.90 ± 0.05. After the beauty separation and charm fraction were implemented at the cross section level, the D^* cross section with only the charm contribution is presented. The results show agreement with both the ALICE and LHCb measurements, and with FONLL and Pythia predictions within their uncertainties. The $D^{*\pm}$ measured cross section is 1096 ± 133 (stat.) μb in the phase space of $0 < |y| < 2$ and $p_T > 1$ GeV, and also $2 < |y| < 2.5$ and $p_T > 8$ GeV. This analysis has covered the largest possible kinematic phase space at the LHC from a single experiment with p_T down to 1 GeV. This includes three new phase space regions where no D^* cross section measurement has been done before at the LHC for p_T below 3.5 GeV.

The total charm cross section is then extracted from this analysis. By combining with the LHCb measurements, which covered most the region outside the CMS detector coverage, and some extrapolation from FONLL and `PYTHIA`, a total charm cross section for the full kinematic phase space is achieved. A fragmentation fraction of $0.155^{+0.04}_{-0.02}$ is used based on the latest measurement by the ALICE experiment. The total charm cross section obtained in this analysis is 9.4 ± 0.45 (statistical) $^{+2.50}_{-0.95}$ (FONLL/`PYTHIA`) mb with an extrapolation factor of 1.4 throughout all phase space. This is the smallest extrapolation achieved for the

total charm cross section at the LHC so far.

Apart from the main result of this thesis, the validation of CMS Open Data with the Higgs to four leptons example at 7 (2011) and 8 (2012) TeV is also included. This validation is one of the service tasks that was done during the early stage of the Ph.D., yet an important achievement. This validation reproduces approximately part of the published CMS Higgs discovery paper at 7 and 8 TeV. The available datasets are only partially overlapping with this publication, and the legacy software versions and corresponding calibrations are different from the original paper. The purpose of the validation is to give an example of using CMS Open Data for educational applications and show the potential of its usage for research applications.

The decay modes chosen in the validation are Higgs to four muons, four electrons or two muons and two electrons. Most of the cuts are the same as the original cuts in the CMS paper. However, some of the more advanced analysis methods and systematic uncertainties of the discovery paper were skipped for simplicity. Nevertheless, it provides a qualitative insight into how the original result was obtained. The work shows that the Higgs peak is two standard deviations, compared to the original publication, which is 3.2 standard deviations in this channel alone. The corresponding example code was publicly released together with the CMS primary dataset for 2012. Many people have used it as a reference or as a starting point on the further usage of the CMS Open Data.

Since the systematic uncertainties have not yet been fully estimated in our measurement, an outlook on the systematics evaluations is provided. One of the systematic evaluations that still need to be completed in this analysis is the signal extraction method. One can study the systematic uncertainty by comparing the background subtraction and fitting function method. Another evaluation is the particle's selection cut. One can check the cut of the variables in each bin of p_T and $|y|$ to study their impact on the results.

Some possible improvements can be made in this analysis. The recent study by ALICE mentioned that the fragmentation fraction might not be universal; thus, a study in this direction can also be interesting, for example a study on different charm mesons and baryons. Within this analysis, the differential charm cross section can be used to study whether the fragmentation fraction is p_T and/or $|y|$ dependent. Since QCD parameters can be different for different theoretical predictions, the measurement of total charm cross section can help to constrain the parameters such as charm quark mass and proton parton density functions. This can be done by performing QCD analysis at higher-order accuracy using available frameworks such as `XFITTER` and comparing the results with this analysis.

Appendices

APPENDIX

A

ΔM PERFORMANCE WITH AND WITHOUT BEAMSPOT CONSTRAINT

In this appendix, the performance of Δm with and without beamspot constraint is studied. The purpose is to know how large is the difference between these variables in data and MC and which variable is better to use for the analysis. The study was conducted at 5 TeV (low pileup) MinimumBias data as it has more statistics and with prompt D^0 MC. Figure A.1 and Figure A.2 show the full kinematic phase space of Δm where the blue and red distribution represent with and without beamspot constraint respectively. The result shows that there are some differences at low p_T and central |y| regions. Figure A.3 is the zoom-in plot of phase space p_T 2-3 GeV in data, (a), in MC, (b) and $p_T > 3.5$ GeV in data, (c). The benefit of using beamspot constraint as mentioned in Section 6.4 helps in improving the analysis.

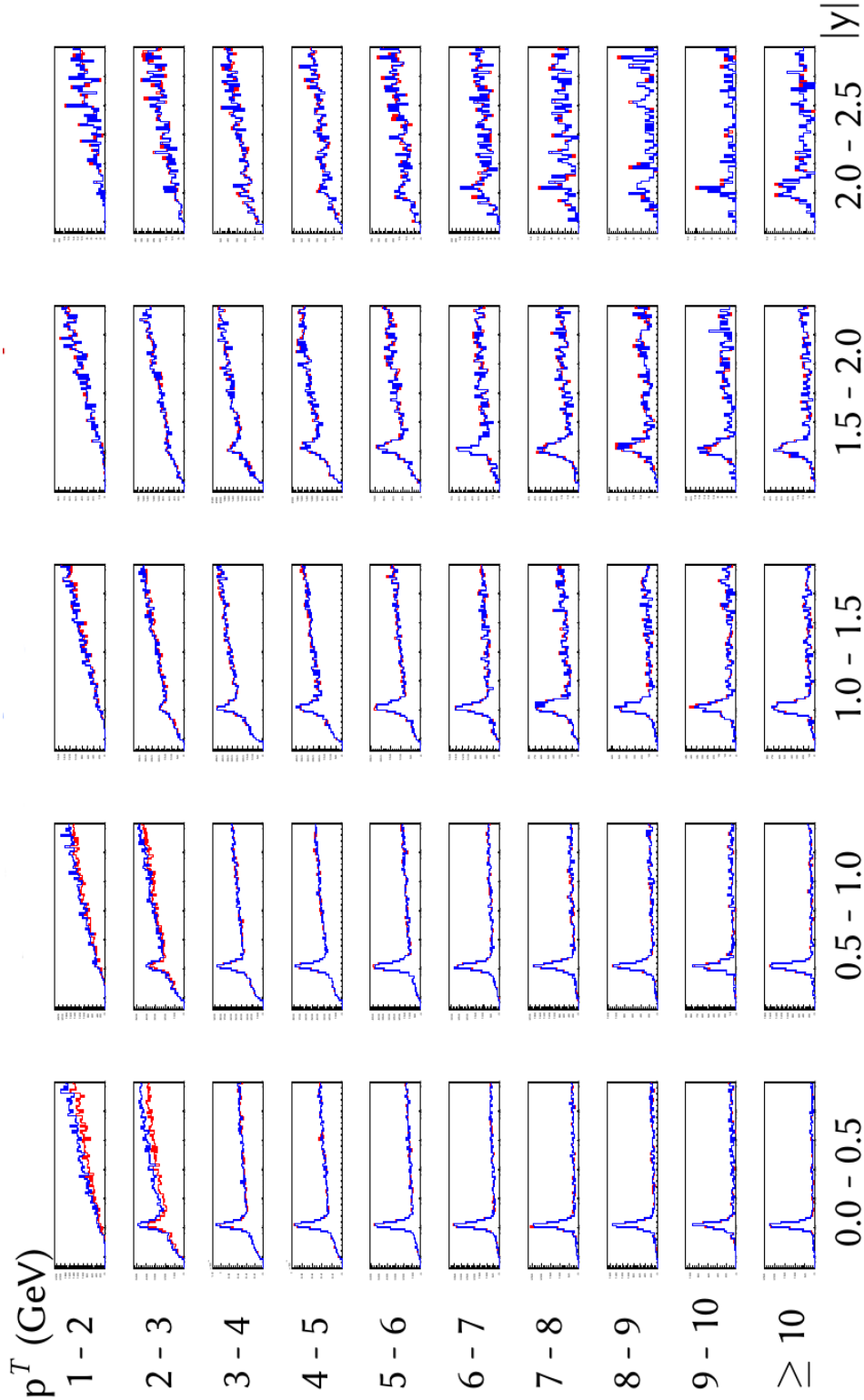


Figure A.1: Full phase space of Δm distribution using with (blue) and without (red) beamspot constraint in 5 TeV data.

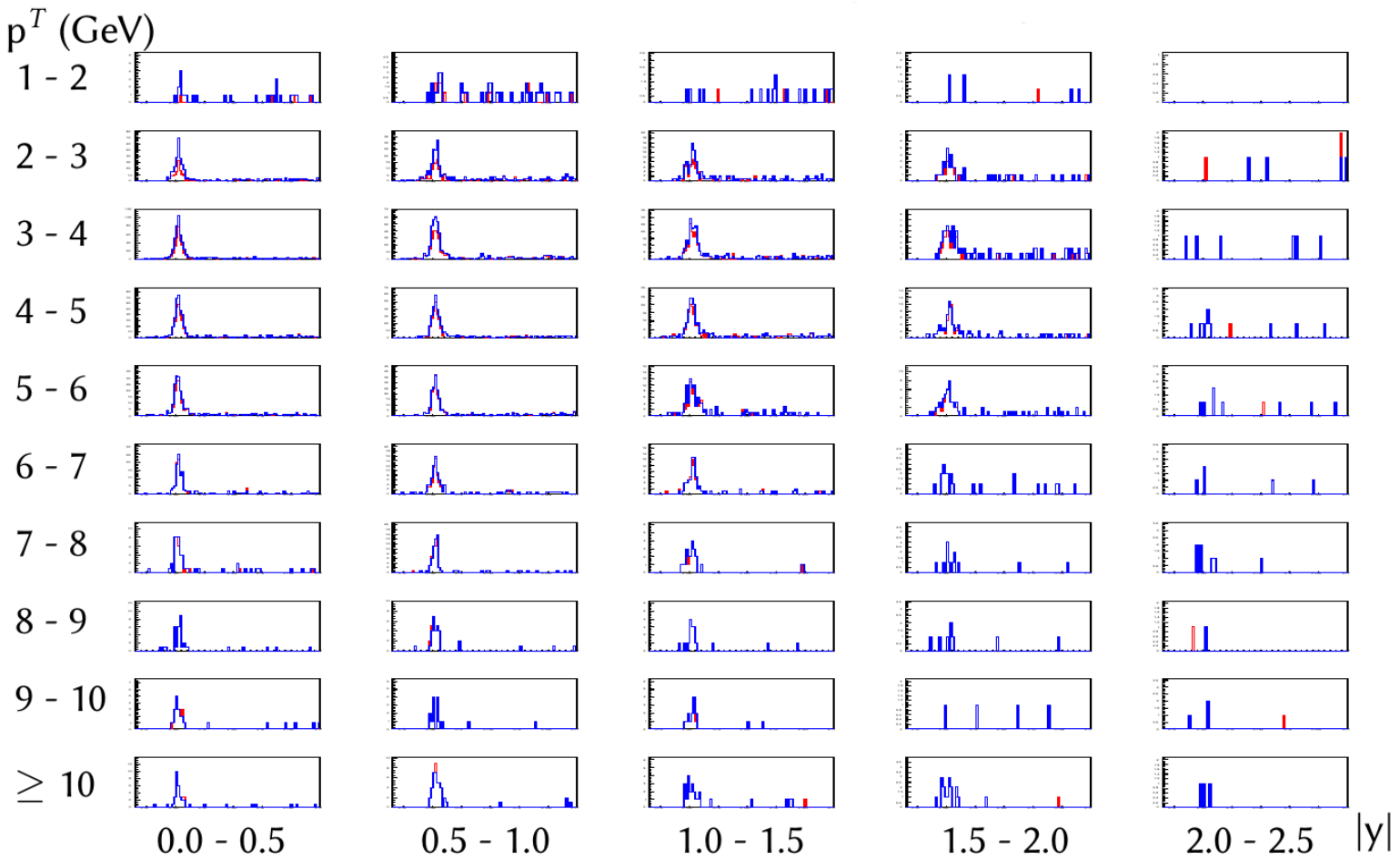
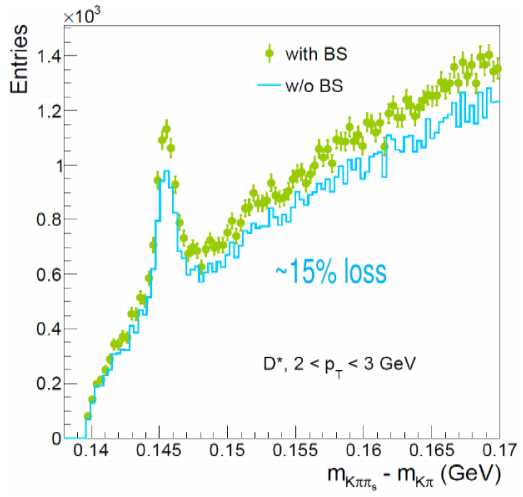
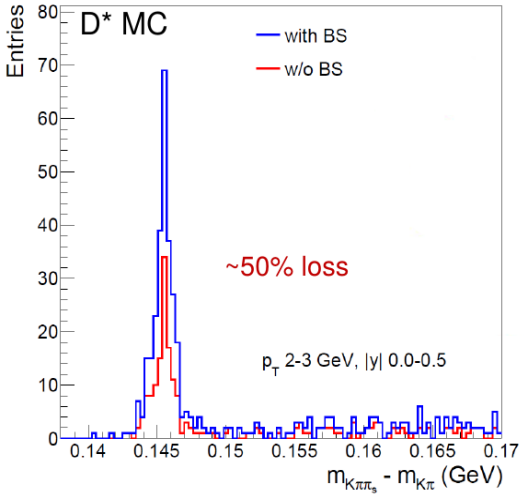


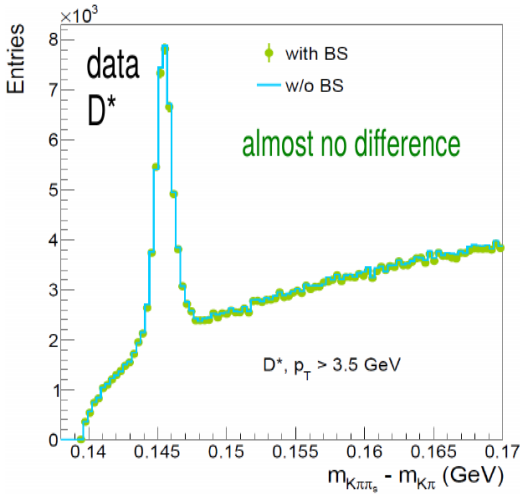
Figure A.2: Full phase space of Δm distribution using with (blue) and without (red) beamspot constraint in 5 TeV MC.



(a) Mass of Δm for p_T 2-3 GeV at 5 TeV data.



(b) Mass of Δm for p_T 2-3 GeV at 5 TeV MC.



(c) Mass of Δm for $p_T > 3.5$ GeV at 5 TeV data.

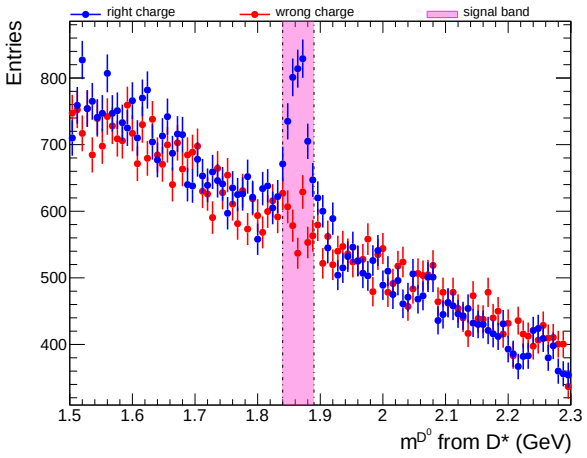
Figure A.3: The performance of D^* reconstruction with (blue) and without (red) beam spot constraint at 5 TeV data and MC for p_T 2-3 GeV in (a) and (b), for $p_T > 3.5$ GeV in (c).

APPENDIX

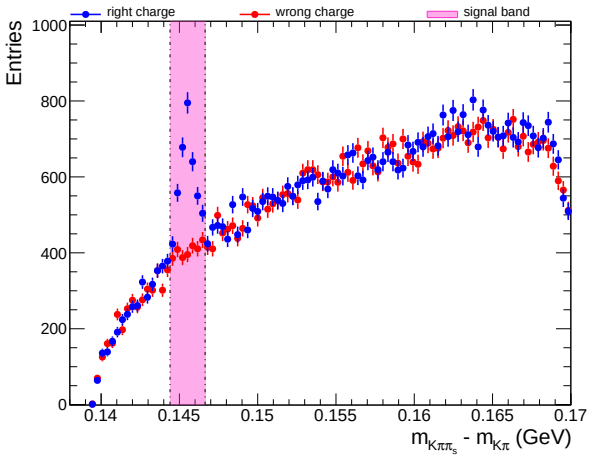
B

MASS CUT OF D^0 AND D^* AT LOW P_T

In this appendix, D^0 and D^* mass cut for $p_T < 3.5$ GeV are given. The plots show the chosen signal region for D^0 mass and Δm is fine at low p_T compared to at high p_T of D^0 mass which a bit too narrow.



(a) m^{D^0} spectrum ($p_T < 3.5$ GeV) in data indicate Δm cut.



(b) Δm spectrum ($p_T < 3.5$ GeV) in data indicate m_{D^0} cut.

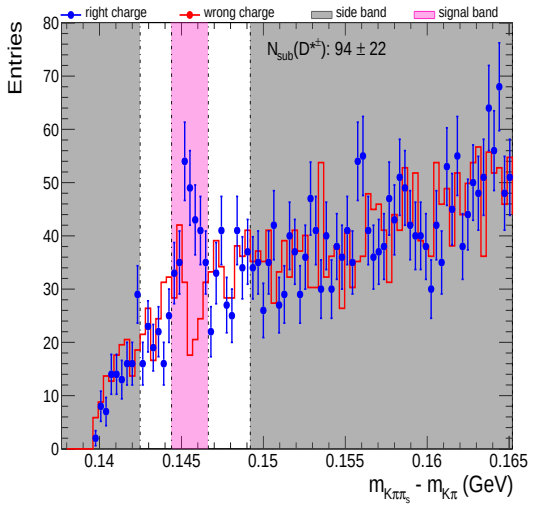
Figure B.1: The figures show the chosen signal region for Δm and D^0 mass is fine at low p_T.

APPENDIX

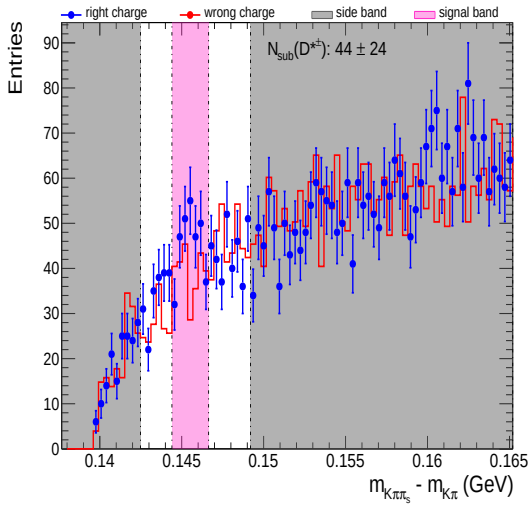
C

SIGNAL EXTRACTION FOR DIFFERENT P_T AND Y PHASE SPACE

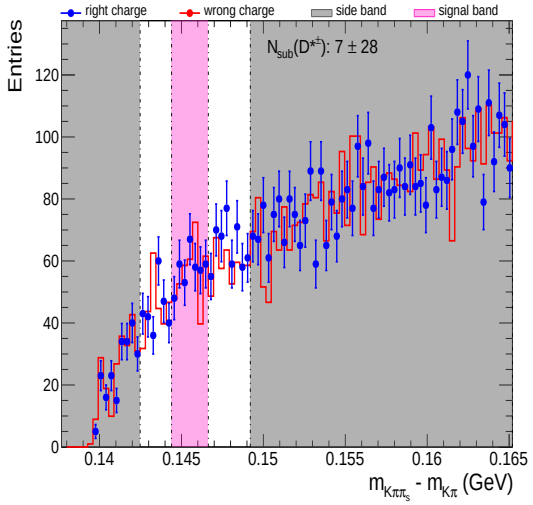
In this appendix, individual D^* signal extraction for each p_T and |y| phase space mentioned in Figure 6.12 are given. At some of the higher |y| phase space, especially at |y| 2.0 -2.5, the plots do not have a signal peak but these regions were covered by LHCb measurements.



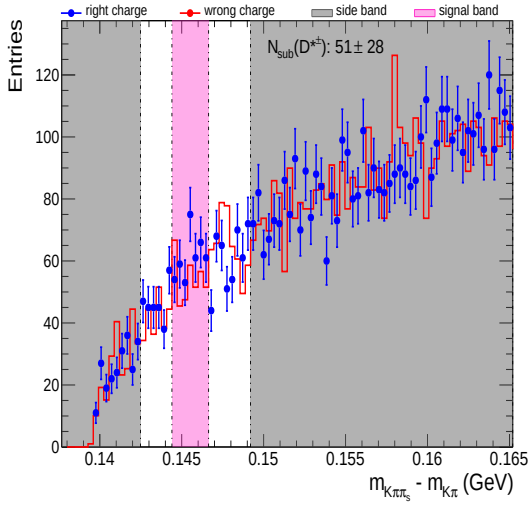
(a) p_T 1 - 2 GeV, |y| < 0.5



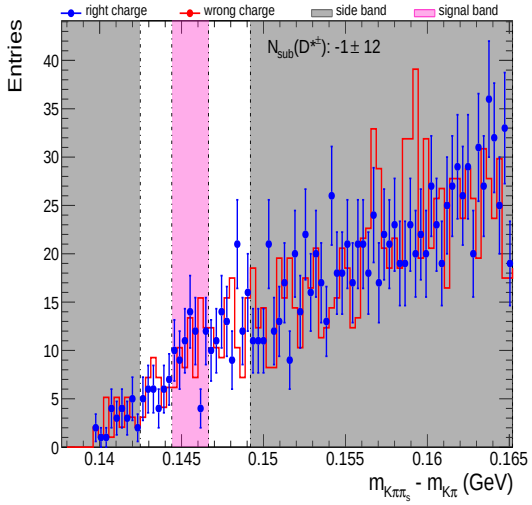
(b) p_T 1 - 2 GeV, |y| 0.5 - 1.0



(c) p_T 1 - 2 GeV, |y| 1.0 - 1.5

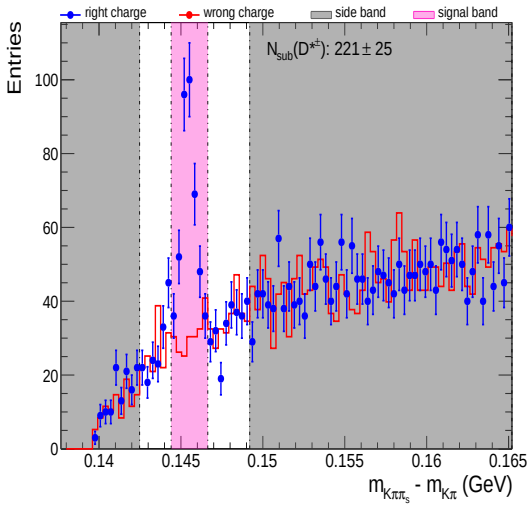


(d) p_T 1 - 2 GeV, |y| 1.5 - 2.0

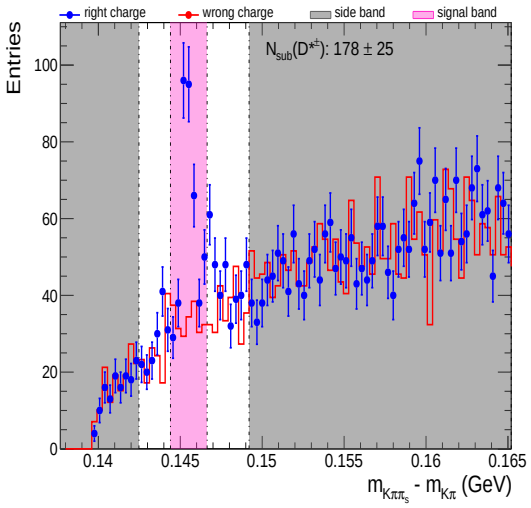


(e) p_T 1 - 2 GeV, |y| 2.0 - 2.5

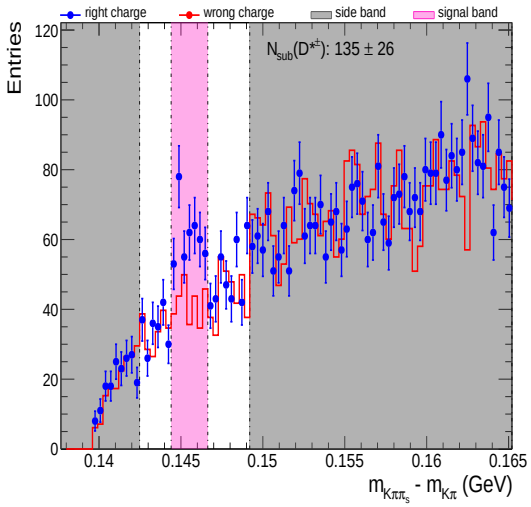
Figure C.1: No. of D* signal in p_T 1 - 2 GeV and |y| 0 - 2.5



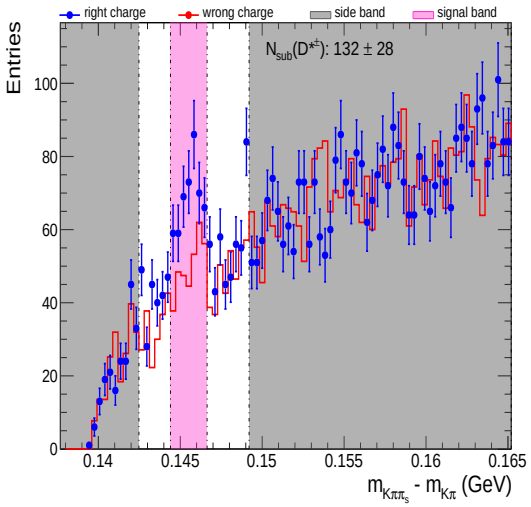
(a) p_T 2 - 3 GeV, |y| < 0.5



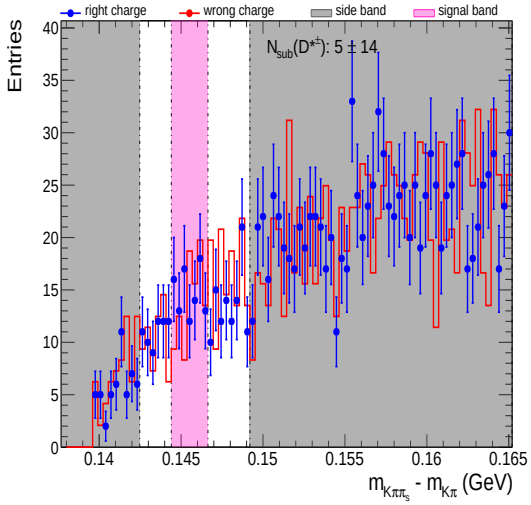
(b) p_T 2 - 3 GeV, |y| 0.5 - 1.0



(c) p_T 2 - 3 GeV, |y| 1.0 - 1.5

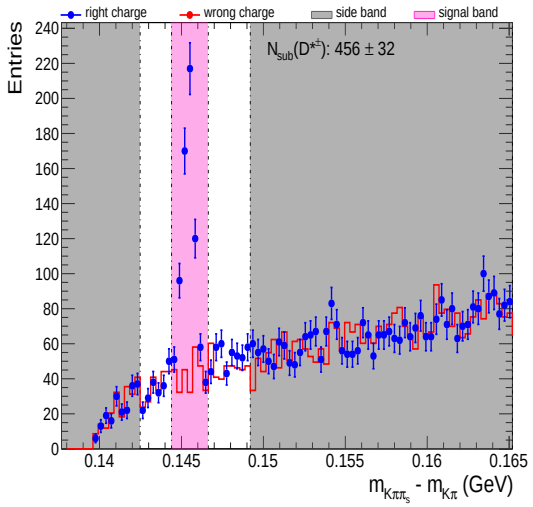


(d) p_T 2 - 3 GeV, |y| 1.5 - 2.0

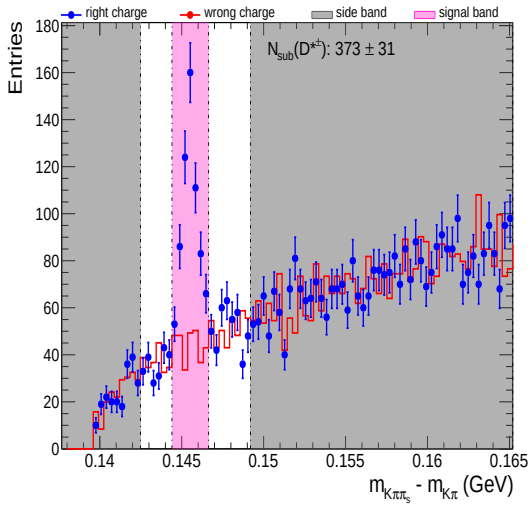


(e) p_T 2 - 3 GeV, |y| 2.0 - 2.5

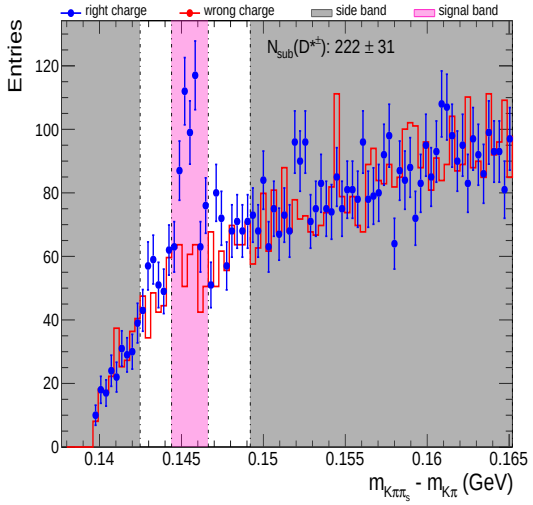
Figure C.2: No. of D* signal in p_T 2 - 3 GeV and |y| 0 - 2.5



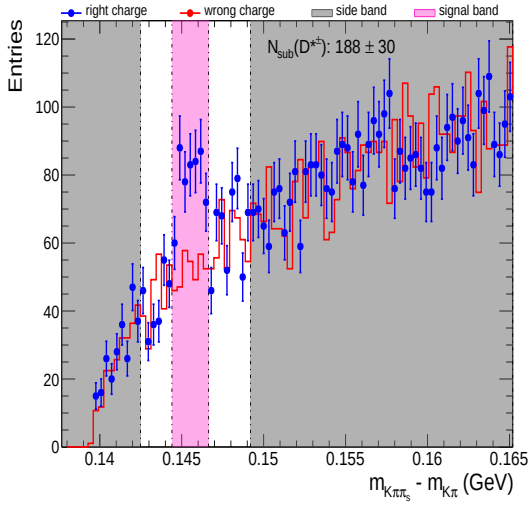
(a) p_T 3 - 4 GeV, |y| < 0.5



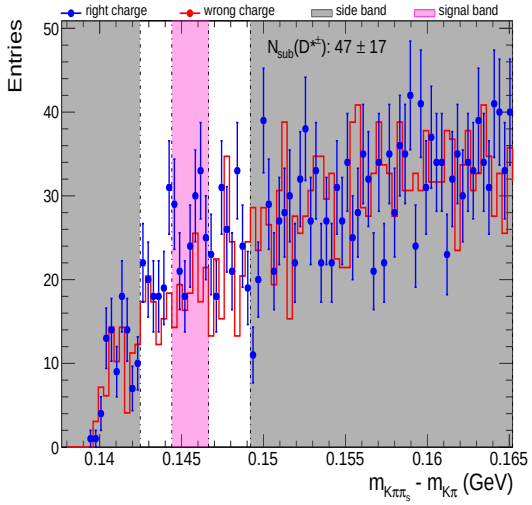
(b) p_T 3 - 4 GeV, |y| 0.5 - 1.0



(c) p_T 3 - 4 GeV, |y| 1.0 - 1.5

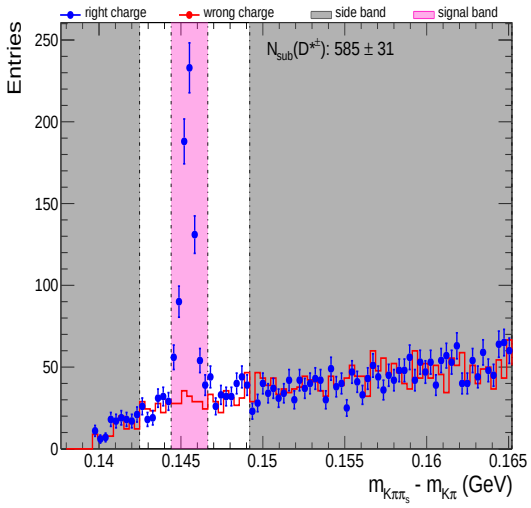


(d) p_T 3 - 4 GeV, |y| 1.5 - 2.0

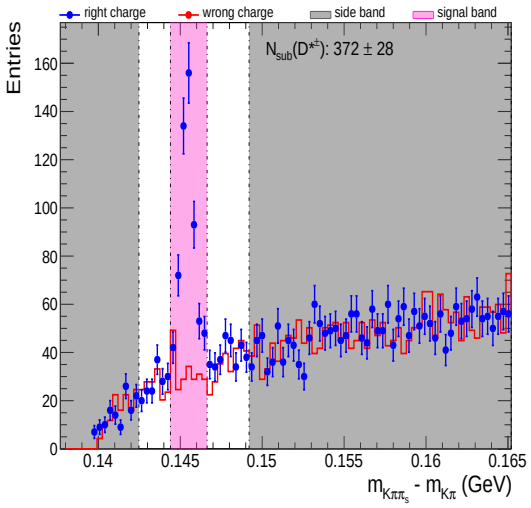


(e) p_T 3 - 4 GeV, |y| 2.0 - 2.5

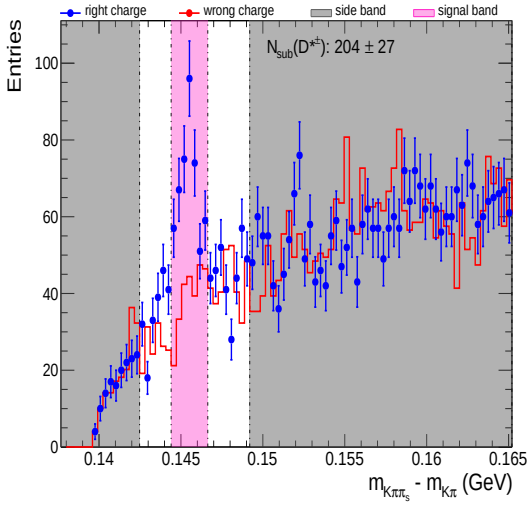
Figure C.3: No. of D* signal in p_T 3 - 4 GeV and |y| 0 - 2.5



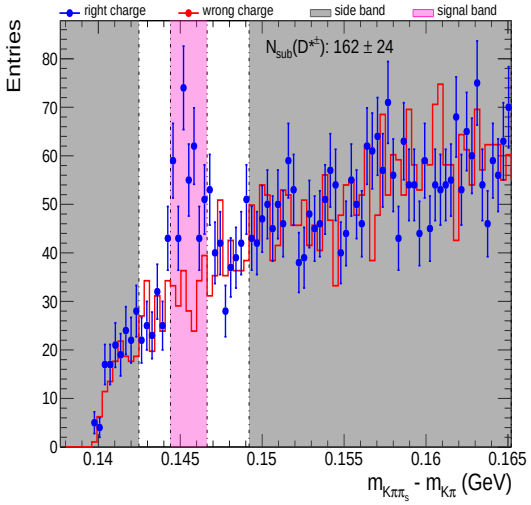
(a) p_T 4 - 5 GeV, |y| < 0.5



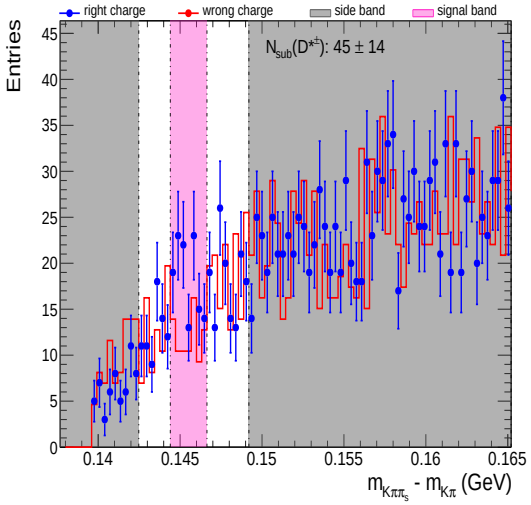
(b) p_T 4 - 5 GeV, |y| 0.5 - 1.0



(c) p_T 4 - 5 GeV, |y| 1.0 - 1.5

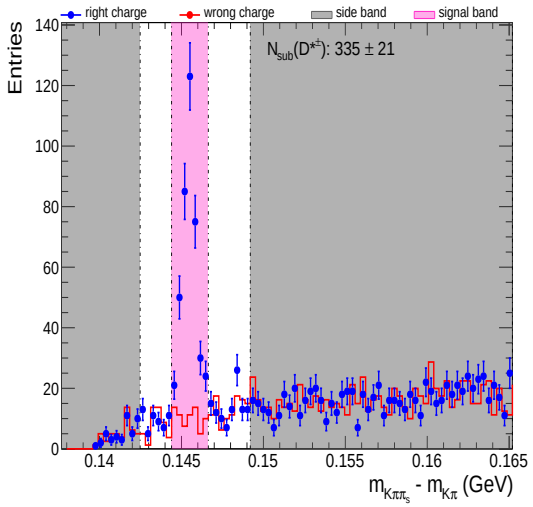


(d) p_T 4 - 5 GeV, |y| 1.5 - 2.0

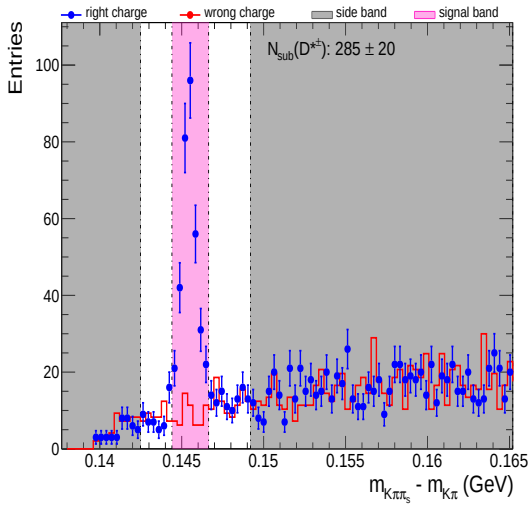


(e) p_T 4 - 5 GeV, |y| 2.0 - 2.5

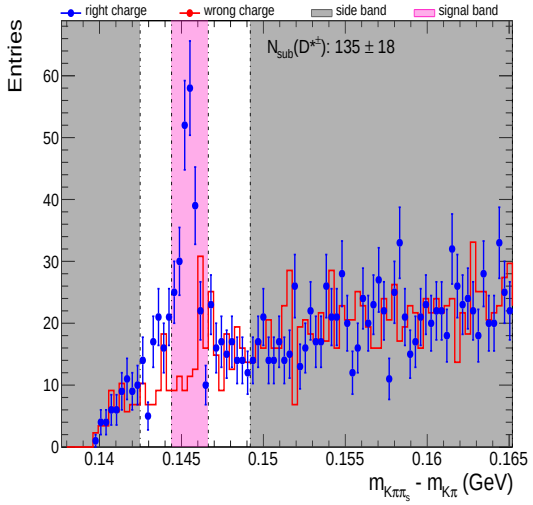
Figure C.4: No. of D* signal in p_T 4 - 5 GeV and |y| 0 - 2.5



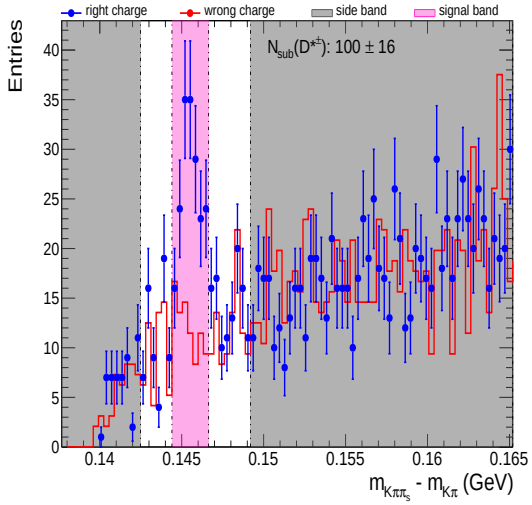
(a) p_T 5 - 6 GeV, |y| < 0.5



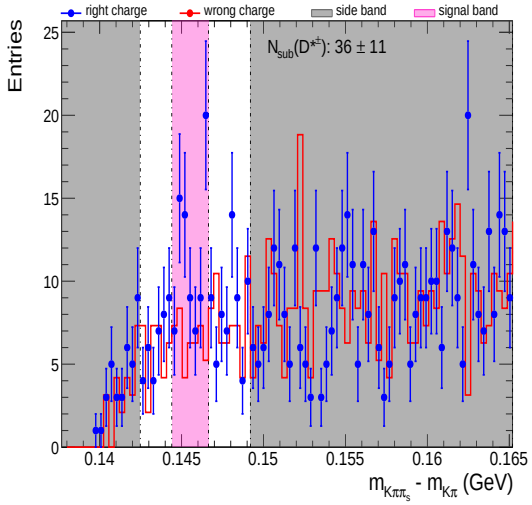
(b) p_T 5 - 6 GeV, |y| 0.5 - 1.0



(c) p_T 5 - 6 GeV, |y| 1.0 - 1.5

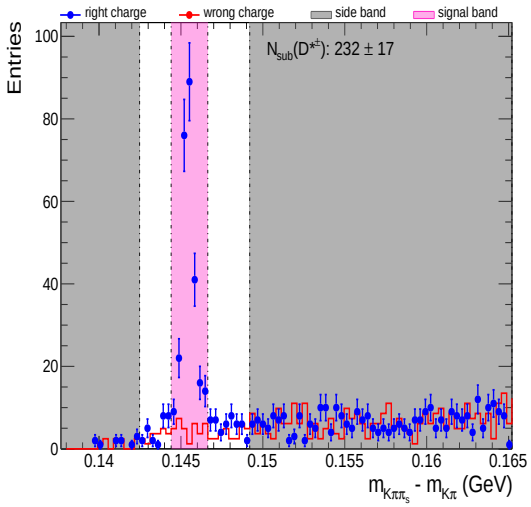


(d) p_T 5 - 6 GeV, |y| 1.5 - 2.0

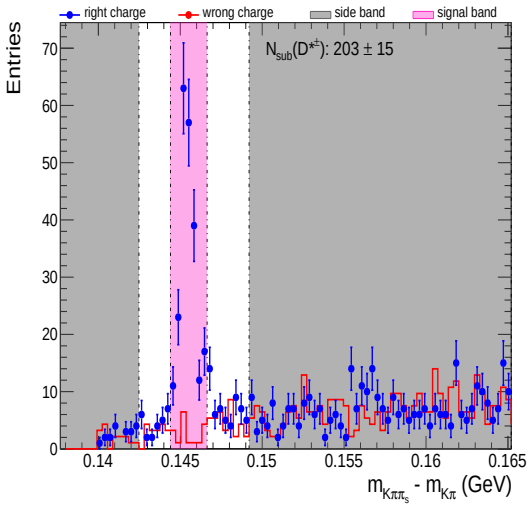


(e) p_T 5 - 6 GeV, |y| 2.0 - 2.5

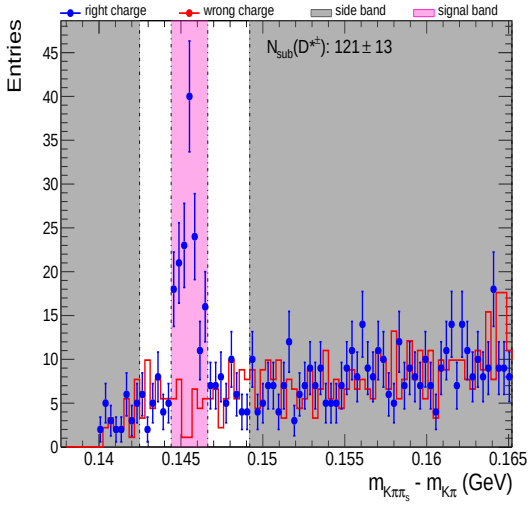
Figure C.5: No. of D* signal in p_T 5 - 6 GeV and |y| 0 - 2.5



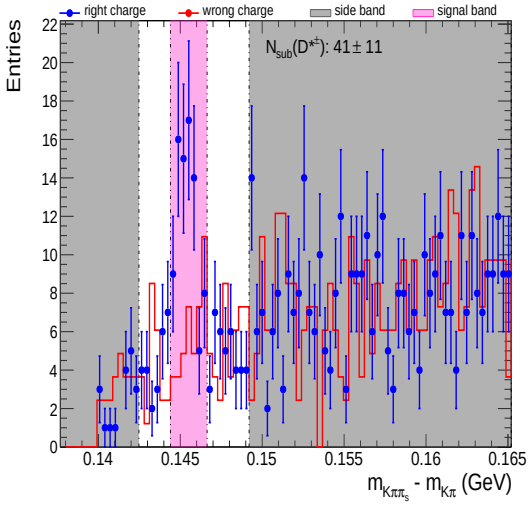
(a) p_T 6 - 7 GeV, |y| < 0.5



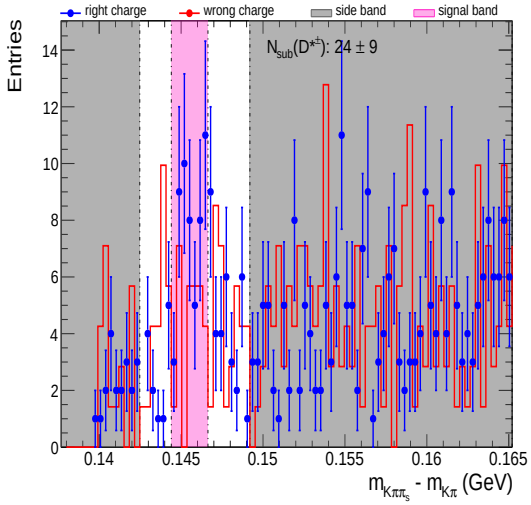
(b) p_T 6 - 7 GeV, |y| 0.5 - 1.0



(c) p_T 6 - 7 GeV, |y| 1.0 - 1.5

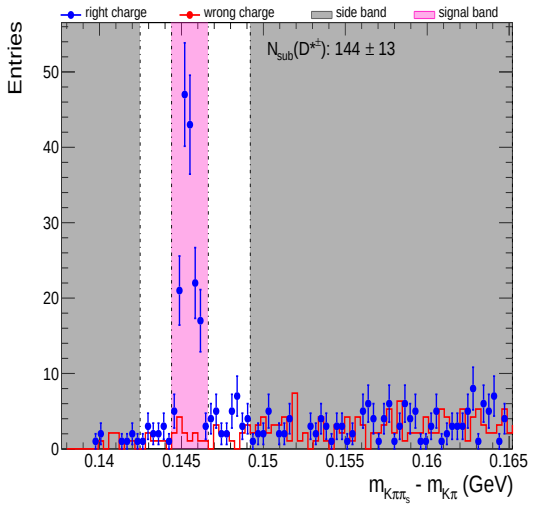


(d) p_T 6 - 7 GeV, |y| 1.5 - 2.0

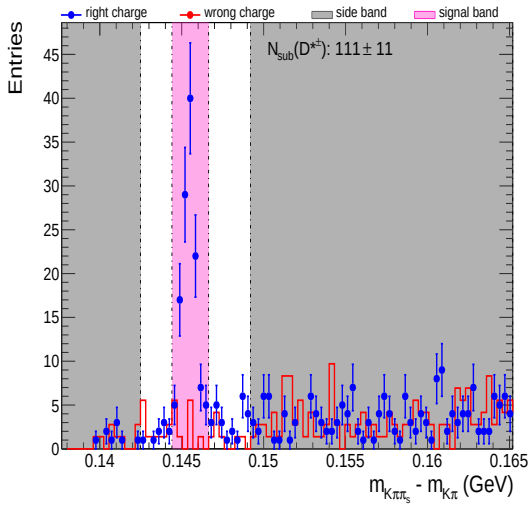


(e) p_T 6 - 7 GeV, |y| 2.0 - 2.5

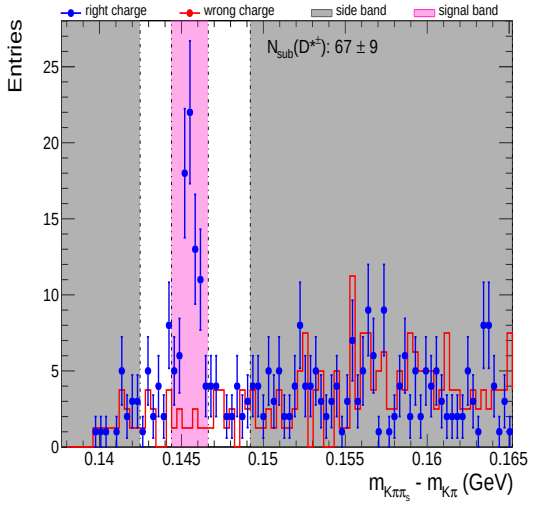
Figure C.6: No. of D* signal in p_T 6 - 7 GeV and |y| 0 - 2.5



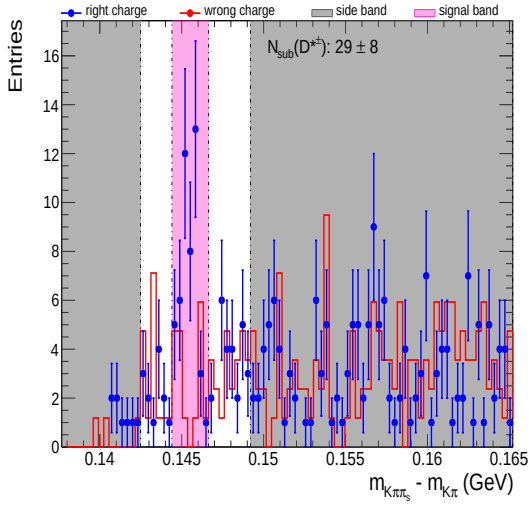
(a) p_T 7 - 8 GeV, |y| < 0.5



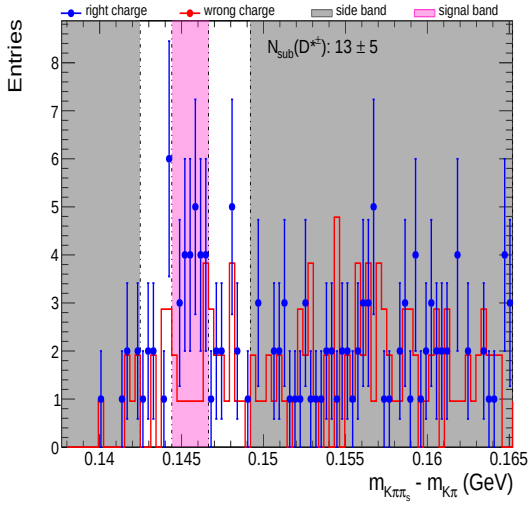
(b) p_T 7 - 8 GeV, |y| 0.5 - 1.0



(c) p_T 7 - 8 GeV, |y| 1.0 - 1.5

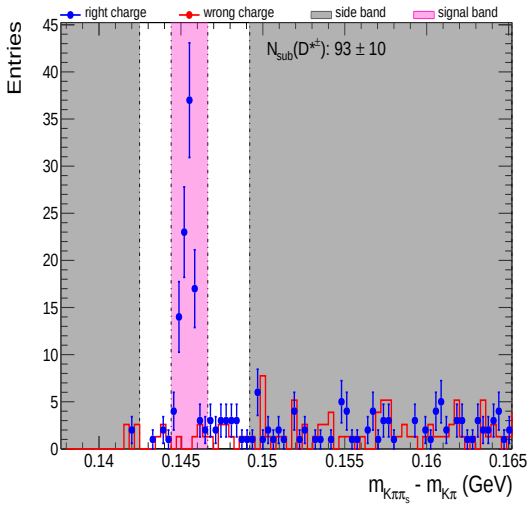


(d) p_T 7 - 8 GeV, |y| 1.5 - 2.0

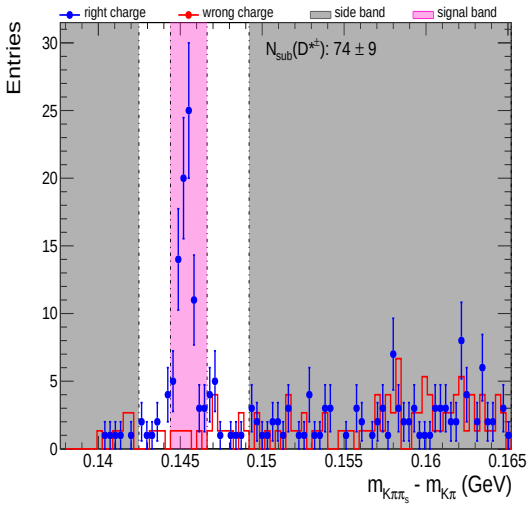


(e) p_T 7 - 8 GeV, |y| 2.0 - 2.5

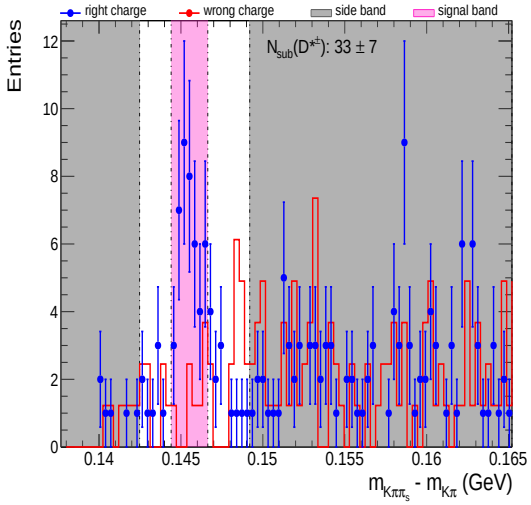
Figure C.7: No. of D* signal in p_T 7 - 8 GeV and |y| 0 - 2.5



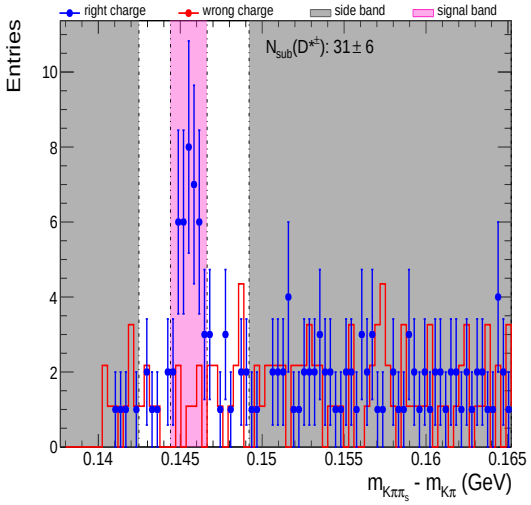
(a) p_T 8 - 9 GeV, |y| < 0.5



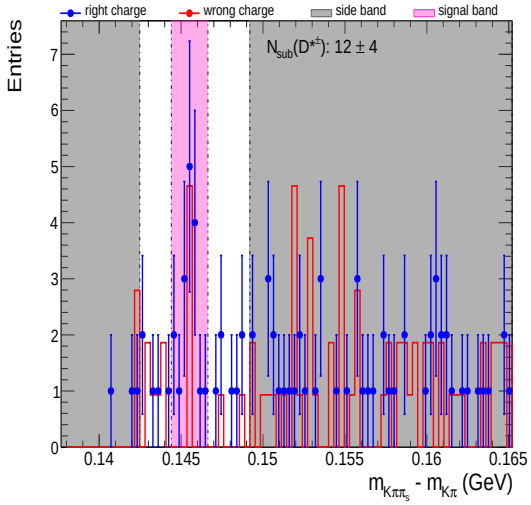
(b) p_T 8 - 9 GeV, |y| 0.5 - 1.0



(c) p_T 8 - 9 GeV, |y| 1.0 - 1.5

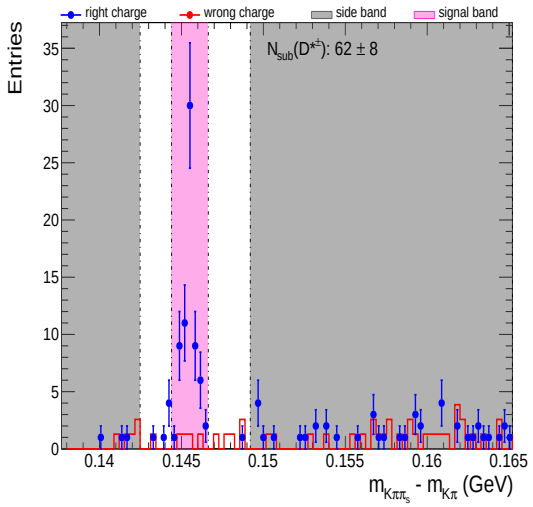


(d) p_T 8 - 9 GeV, |y| 1.5 - 2.0

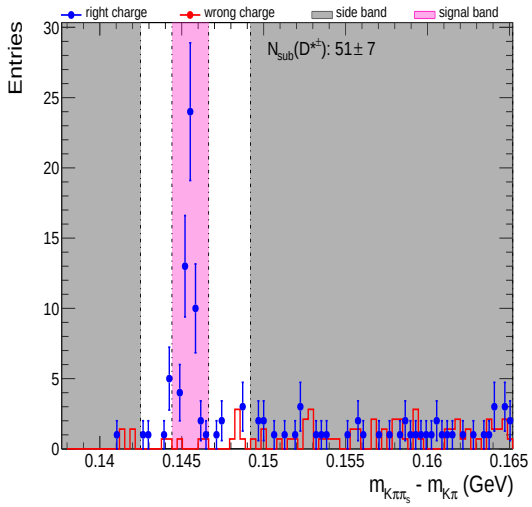


(e) p_T 8 - 9 GeV, |y| 2.0 - 2.5

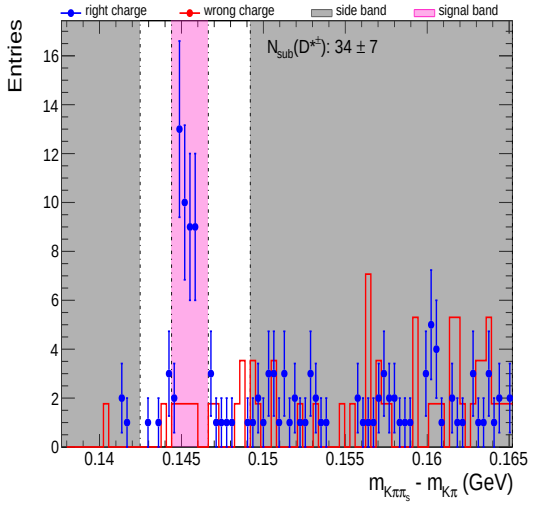
Figure C.8: No. of D* signal in p_T 8 - 9 GeV and |y| 0 - 2.5



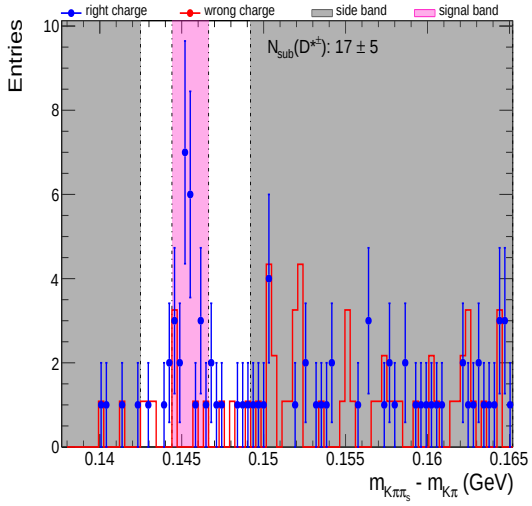
(a) p_T 9 - 10 GeV, |y| < 0.5



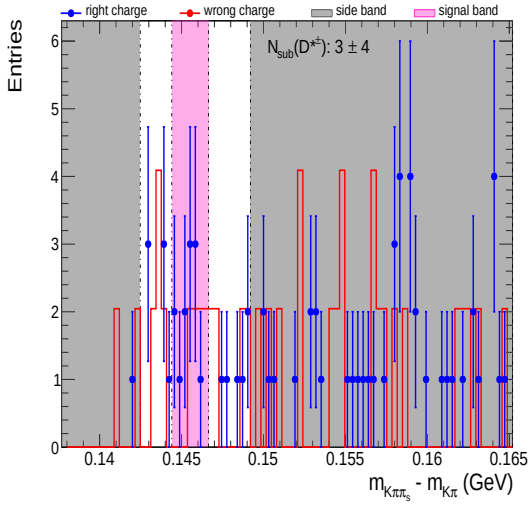
(b) p_T 9 - 10 GeV, |y| 0.5 - 1.0



(c) p_T 9 - 10 GeV, |y| 1.0 - 1.5

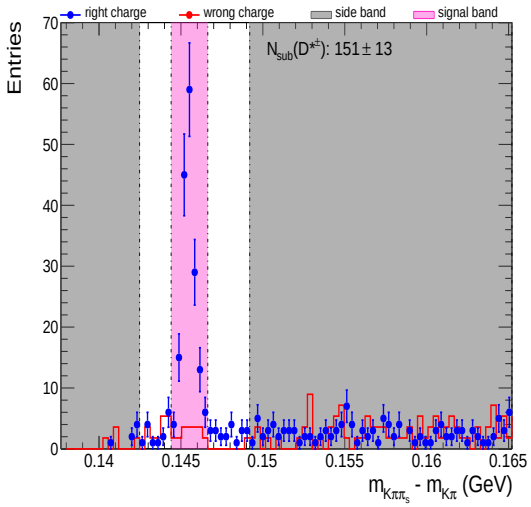


(d) p_T 9 - 10 GeV, |y| 1.5 - 2.0

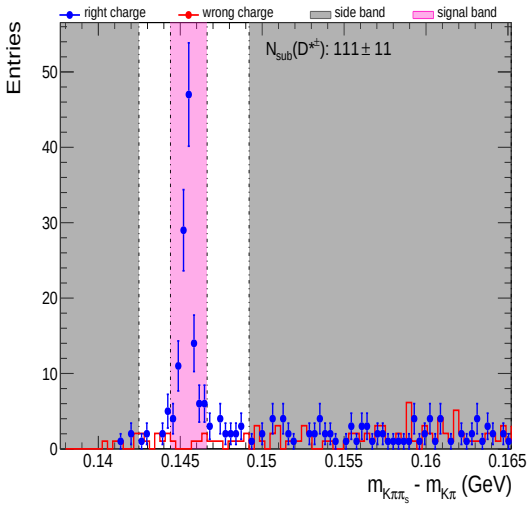


(e) p_T 9 - 10 GeV, |y| 2.0 - 2.5

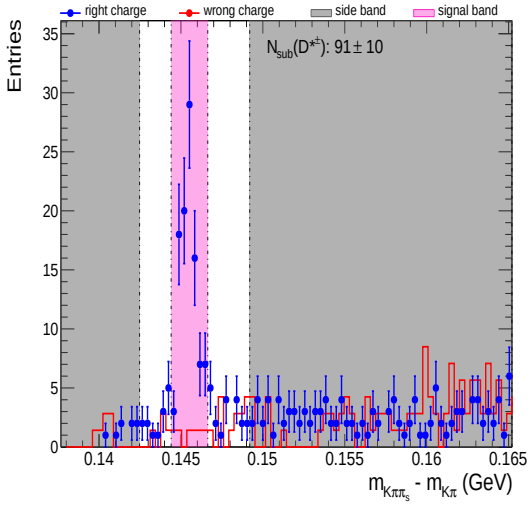
Figure C.9: No. of D* signal in p_T 9 - 10 GeV and |y| 0 - 2.5



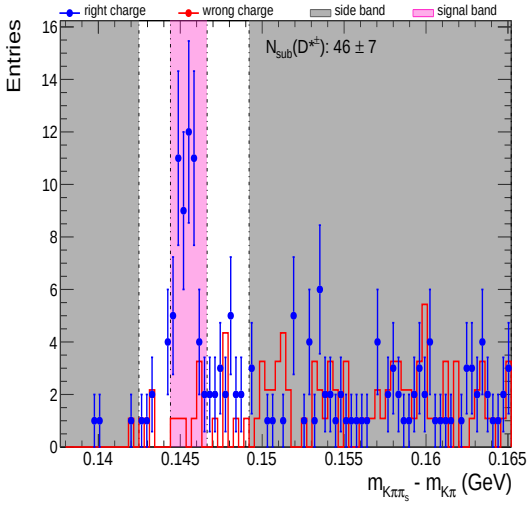
(a) $p_T > 10$ GeV, $|y| < 0.5$



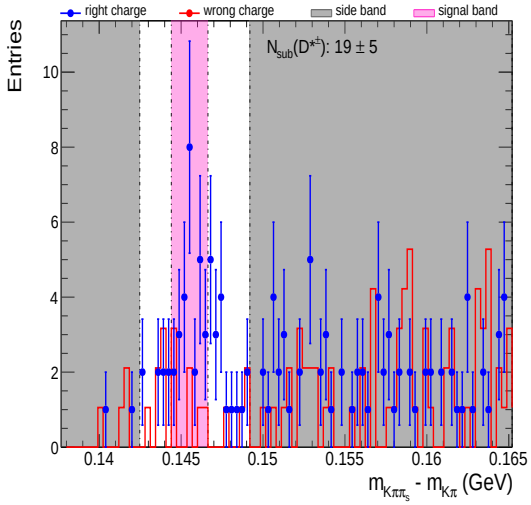
(b) $p_T > 10$ GeV, $|y|$ 0.5 - 1.0



(c) $p_T > 10$ GeV, $|y|$ 1.0 - 1.5



(d) $p_T > 10$ GeV, $|y|$ 1.5 - 2.0



(e) $p_T > 10$ GeV, $|y|$ 2.0 - 2.5

Figure C.10: No. of D* signal in $p_T > 10$ GeV and $|y|$ 0 - 2.5

APPENDIX

D

$N_{RECO\&TRUE}$ AND N_{TRUE} FOR EFFICIENCY CALCULATION

This appendix is a complimentary material to the efficiency calculation in Subsection 6.5.2. It shows the value of $N_{reco\&true}$ in Figure D.1 and N_{true} in Figure D.2 in MC D^0 sample for full kinematic phase space of p_T and $|y|$.

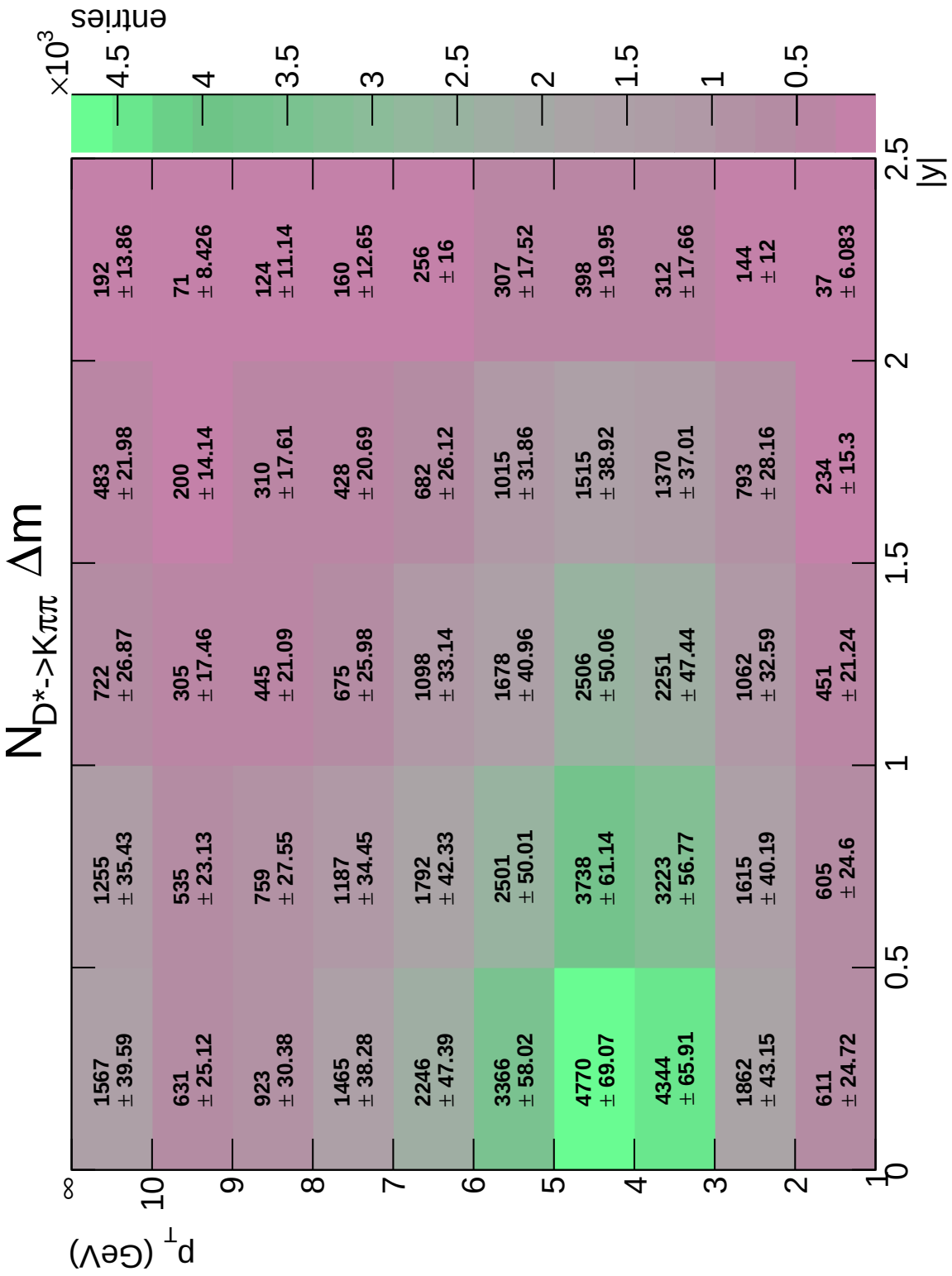


Figure D.1: Figure shows the number of reconstructed $D^{*\pm}$ that match a generated $D^{*\pm}$ inside the bin ($N_{reco\&true}$) of all kinematic phase space. See Subsection 6.5.2.

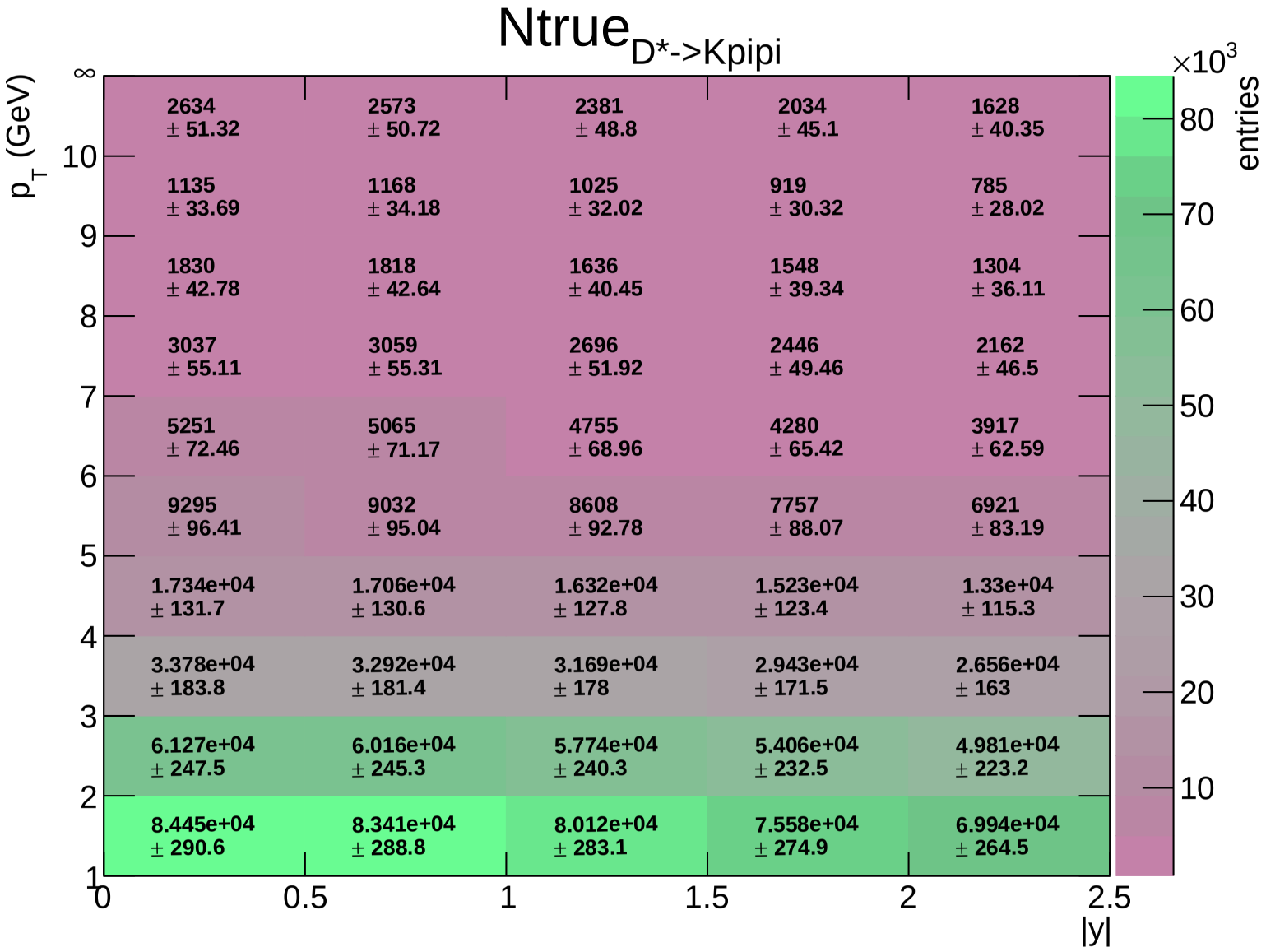


Figure D.2: Figure shows the number of generated $D^{*\pm}$ in the bin (N_{true}) of all kinematic phase. See Subsection 6.5.2.

APPENDIX

E

D^{*+} CROSS SECTION COMPARISON WITH ATLAS

In this appendix, a comparison of the D^* cross section as a function of p_T with ATLAS measurement (see Subsection 6.5.3) is conducted as a cross check during the early stage of the analysis. The kinematic range follows ATLAS phase space plus p_T below 3.5 GeV. For p_T, the ranges are 0.0-0.5, 0.5-2.0, 2.0-3.5, 3.5-5.0, 5.0-6.5, 6.5-8.0, 8-12, 12-20, 20-30, 30-40, 40-60, 60-100 (GeV). For pseudorapidity, it is $|\eta| < 2.1$. Both CMS and ATLAS in this result has charm and beauty (c+b) contributions. The MC used for this comparison is the charm MC.

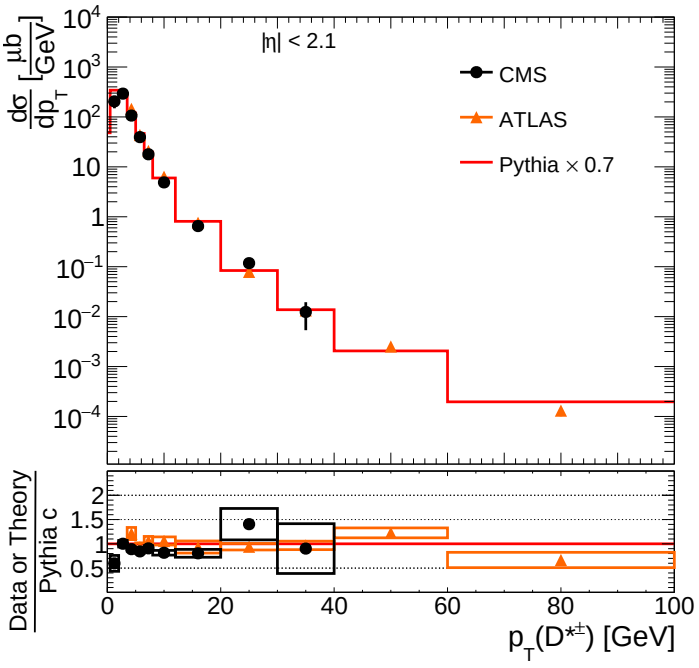


Figure E.1: D* cross section comparison with ATLAS during early stage of the analysis. The measurement between CMS and ATLAS is comparable.