UNDERLYING EVENT MEASUREMENT IN $t\bar{t}$ DILEPTON CHANNEL WITH CMS DETECTOR

FAHMI MAULIDA

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNDERLYING EVENT MEASUREMENT IN $t\bar{t}$ DILEPTON CHANNEL WITH CMS DETECTOR

FAHMI MAULIDA

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF PHYSICS
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **FAHMI MAULIDA**

Registration/Matric No.: **SGR 140050**

Name of Degree: **MASTER OF SCIENCE**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

UNDERLYING EVENT MEASUREMENT IN $t\bar{t}$ DILEPTON CHANNEL WITH CMS DETECTOR

Field of Study: **EXPERIMENTAL PHYSICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:
Designation:

UNDERLYING EVENT MEASUREMENT IN $t\bar{t}$ DILEPTON CHANNEL WITH CMS DETECTOR

ABSTRACT

The measurements of the charged particles can shed a light on the non-perturbative QCD regime at the LHC and on the universality of the behaviour of the Underlying Event (UE) in proton-proton collisions. In this thesis, the proton-proton collision data acquired by the CMS experiment at a centre-of-mass energy of 13 TeV and corresponding to an integrated luminosity of 35.9 fb^{-1} are used. The kinematics of the reconstructed charged particles are analysed after subtracting the estimated contribution from the dilepton and the two b jets as the decay product of the $t\bar{t}$ events. The data are also subtracted from the dominant backgrounds. The results are presented in both detector and particle level and compared to different MC setups to probe the underlying event tune at the energy scale of $t\bar{t}$ production. An unfolding method is implemented to correct the measured data to particle level. Good agreement is found between the data and the simulation used in CMS.

Keywords : Underlying Event, QCD, $t\bar{t}$ events, proton-proton collisions, CMS experiment, dilepton.

PENGUKURAN PERISTIWA PENDASAR PADA SALURAN $t\bar{t}$ DUA-LEPTON DENGAN PENGESAN CMS

ABSTRAK

Pengukuran zarah-zarah bercas boleh menerangkan kromodinamik quantum tanpa-usikan (*non pertubative QCD*) di LHC dan juga kepada Peristiwa Pendasar (*Underlying Event*, UE) dalam perlanggaran proton-proton. Dalam thesis ini, data perlanggaran proton-proton didapati oleh eksperimen CMS pada tenaga pusat jirim 13 TeV dengan nilai kilauan bersepadu 35.9 fb^{-1} telah digunakan. Kinematik zarah bercas yang dibina semula dianalisis selepas penolakan anggaran penghasilan dari dua-lepton dan dua *b*-jet hasil produk pereputan peristiwa $t\bar{t}$. Data juga ditolak daripada gangguan dominan. Hasil kajian ditunjukkan pada segi pengesan dan partikel dan juga dibandingkan kepada beberapa jenis Monte-Carlo yang berbeza untuk menyelami Peristiwa Pendasar pada tenaga pengeluaran $t\bar{t}$. Kaedah enyahlipat (*unfolding*) diterapkan untuk membetulkan data yang telah diukur kepada tahapan partikel. Persetujuan yang baik telah ditemui diantara data dan simulasi yang digunakan di CMS.

Kata kunci : Peristiwa pendasar, kromodinamik quantum, peristiwa $t\bar{t}$, pelanggaran proton-proton, eksperimen CMS, dua-lepton

ACKNOWLEDGEMENTS

A very first special gratitude goes to Allah because His Mercy I could get a chance to meet a very kind supervisor Prof Wan Ahmad Tajuddin. "I am grateful for having you as my advisor and continuous support of my study and related research, for your patience, motivation, and immense knowledge. Your guidance helped me in all the time of research. Thank you for the opportunity that you give to me, this chance leads me to taste so many ways and experiences to make my dream come true". I also want to thank Dr Ijaz Ahmed who taught and helped me to understand my analysis also Prof Zainol Abidin Ibrahim who has approved the fund for my stay at NCPP. A very special grateful for Pedro Silva who guided me along the way and provided the comprehensible analysis. He has been a constant inspiration, always providing original ideas and constructive comments and suggestions.

I am grateful to my husband who gave me moral and emotional support along the way and cherished my every single moment. I am also thankful to all my family members especially to my mother and my father, without their love and pray impossible for me to finish this thesis. Many thanks also to all NCPP's friends, especially Ka Mira, Ka Khasmi, and Atikah Ecah, their positive energy gave me a positive motivation.

I am also using this opportunity to express my gratitude to everyone who supported me throughout my study. I am thankful for their aspiring guidance, invaluable constructive criticism and friendly advice during the project work.

Thanks for all your encouragement!

TABLE OF CONTENTS

ABSTRACT ..... **iii**

ABSTRAK ..... **iv**

ACKNOWLEDGEMENTS ..... **v**

TABLE OF CONTENTS ..... **vi**

LIST OF FIGURES ..... **ix**

LIST OF TABLES ..... **xi**

LIST OF SYMBOLS AND ABBREVIATIONS ..... **xii**

LIST OF APPENDICES ..... **xiv**

CHAPTER 1: INTRODUCTION **1**

1.1 Project Statement 2

1.2 Objectives 2

1.3 Thesis Outline 3

CHAPTER 2: THEORETICAL BACKGROUND **4**

2.1 The Standard Model 4

2.2 Quantum Chromodynamics (QCD) 7

2.2.1 The Quantum Chromodynamics (QCD) Couplings, α_s 9

2.3 Monte Carlo Event Generators 13

2.4 The Underlying Event 15

2.4.1 The Top Pairs Event 16

2.4.2 The Underlying Event Models 18

2.4.3 Underlying Event Tunes 19

CHAPTER 3: THE CMS EXPERIMENT 21

3.1 The Large Hadron Collider (LHC) 21

3.2 Compact Muon Solenoid (CMS) 22

3.3 The Reconstruction Event in the CMS Experiment 27

3.3.1 Jet in CMS Experiment 28

3.3.2 CMS Trigger System 29

CHAPTER 4: MEASUREMENT OF UNDERLYING EVENT IN TOP PAIRS DILEPTON CHANNEL 30

4.1 Analysis Strategy 30

4.2 Monte Carlo Samples and Data 36

4.2.1 Data 36

4.2.2 Monte Carlo Simulation 37

4.2.3 Corrections Applied to the Simulation 39

4.3 Event Selection 43

4.3.1 Background Determination 46

4.4 Unfolding Method 51

4.4.1 Unfolding Test 57

4.5 Systematic Uncertainties 62

4.5.1 Experimental Uncertainties 62

4.5.2 Theoretical Uncertainties 64

CHAPTER 5: RESULT 67

CHAPTER 6: DISCUSSION 74

CHAPTER 7: CONCLUSION 75

REFERENCES **76**

APPENDICES **83**

LIST OF FIGURES

Figure 2.1: The parton distribution function 6

Figure 2.2: Distribution of the partons after getting some extra energy 7

Figure 2.3: Quark-gluon interaction 9

Figure 2.4: One correction by a virtual gg and $q\bar{q}$ pair 10

Figure 2.5: The running coupling constant 11

Figure 2.6: High energy collision processes from collision to detection 14

Figure 2.7: Top-antitop ($t\bar{t}$) productions 16

Figure 2.8: Multi Parton Interaction 19

Figure 3.1: Large Hadron Collider 22

Figure 3.2: An illustration of the CMS coordinate system 23

Figure 3.3: CMS detector in slice 24

Figure 3.4: The charged particle hits the silicon tracker 25

Figure 4.1: An ilustration of the segmentation of ϕ-space 32

Figure 4.2: Resolutions for the multiplicity 33

Figure 4.3: Resolutions for the p_{T} and p_{z} fluxes 34

Figure 4.4: Resolutions for the average p_{T} and p_{z} 35

Figure 4.5: Distributions of the number of primary vertices in *ee*, $\mu\mu$, and $e\mu$ channels 40

Figure 4.6: Median energy density computed from FastJet in *ee*, $\mu\mu$, and $e\mu$ channels 41

Figure 4.7: Tracking scale factors used in our analysis 42

Figure 4.8: Inclusive charge multiplicity distributions 48

Figure 4.9: Inclusive scalar sum of the p_{T} and p_{z} of the charged particles 49

Figure 4.10: Inclusive average p_{T} and p_{z} per charged particles 50

Figure 4.11: Dilepton invariant mass in *ee*, $\mu\mu$, and $e\mu$ channels 51

Figure 4.12: Migration matrices 54

Figure 4.13: Purity and stability curves 55

Figure 4.14: The results distributions of the comparison of the folded data and recontructed data 59

Figure 4.15: The result distributions of the pulls in the individual bins 60

Figure 4.16: The results of the regularised unfolding test 61

Figure 5.1: The results of the N_{ch} 69

Figure 5.2: The results of the p_T flux 70

Figure 5.3: The results of the p_z flux 71

Figure 5.4: The results of the average p_T 72

Figure 5.5: The results of the average p_z 73

Figure A.1: Control distributions after selection for global event variables 82

Figure A.2: Two system boost related variables to profile the underlying event activity 83

LIST OF TABLES

Table 2.1: The Forces 4

Table 4.1: Typical underlying event observables and their symbolic representation ... 31

Table 4.2: The regions defined according to the p_T of $t\bar{t}$ system 32

Table 4.3: Primary datasets used in this analysis 36

Table 4.4: MC simulation samples 38

Table 4.5: MC simulation samples used for systematics 39

Table 4.6: Trigger paths used for online selection 45

Table 4.7: Expected and observed event yields in the different categories 45

Table 4.8: Drell Yan background 48

Table 4.9: Condition numbers of the migration matrices of the different variable 57

Table 4.10: Median of the bin-by-bin systematices for different variables 66

LIST OF SYMBOLS AND ABBREVIATIONS

$E_{\mathrm{T}}^{\mathrm{miss}}$	:	Missing Energy Transverse
α_s	:	Strong Coupling Constant
$\mathcal{L}$	:	Luminosity
σ	:	Cross Section
b	:	bottom quark
gg	:	gluon-gluon
p_z	:	Momentum at longitudinal direction
p_{T}	:	Momentum Transverse
pp	:	Proton-Proton
$q\bar{q}$	:	quark anti-quark
$t\bar{t}$	:	Top-antitop
ATLAS	:	A Toroidal LHC Apparatus
BBR	:	Beam-Beam Remnant
CMS	:	Compact Muon Solenoid
CMSSW	:	CMS Software Framework
CR	:	Color Reconnection
DY	:	Drell-Yan
ECAL	:	Electromagnetic Calorimeter
FSR	:	Final Sate Radiation
GeV	:	Giga Electron Volt
HCAL	:	Hadronic Calorimeter
HLT	:	High Level Trigger
ISR	:	Initial State Radiation

LHC	:	Large Hadron Collider
LO+PS	:	Leading-Order + Parton Shower
MC	:	Monte Carlo
MPI	:	Multi Parton Interaction
NLO+PS	:	Next-to-Leading-Order + Parton Shower
PDF	:	Parton Distribution Function
PF	:	Particle Flow
QCD	:	Quantum Chromodynamics
QED	:	Quantum Electrodynamics
SM	:	Standard Model
UE	:	Underlying Event

LIST OF APPENDICES

Appendix A: Control Distribution and Event Yields 82

CHAPTER 1: INTRODUCTION

The fundamental structure of matter has been described well by Standard Model (SM) (ATLAS Collaboration, 2012). This model explains how fundamental particles build everything in the universe governed by three of four known elementary forces: strong, weak, electromagnetic, and gravitational force. At Large Hadron Collider (LHC), there exist two main detectors, A Toroidal LHC Apparatus (ATLAS) and Compact Muon Solenoid (CMS), which are also the biggest detectors in the world that were created to verify the Standard Model of particle physics and to search for new physics beyond it by studying the Proton-Proton (pp) collision.

Due to the complexity of the proton composition that have quarks and gluons (generally called partons), the collision between two protons leads to busy parton scattering at different energy scales. The scattering can be classified into two processes: Hard and soft proccess. Besides any important hard interaction processes are searched at high energies momentum scale, the bulk of the particle activities are due to the soft processes at low energies momentum between the several partons also interest to study. At high energy transfer, perturbative QCD is the appropiate theory to explain the hard scattering processes whereas at low energy momentum the perturbative approach is no longer feasible, so the non-perturbative QCD guides to describe the soft scattering proccess (Vallero, 2012).

One possible access to measure the soft activity as a function of hard proccess is Underlying Event (UE) which relies on non-perturbative QCD models to investigate its behaviour. Due to the soft event-activities domination in the pp scattering, UE measurement is mandotary to draw the whole description of everything during collision. The accuracy of the LHC measurements not only depends on the models of the hard scattering part but also on the improvement of the current UE models, even further it also depends on the

development of the new phenomenological models of soft QCD processes to get the more precise results.

1.1 Project Statement

In this thesis, the UE analysis is defined as the charged particles activity except for the particles that come from Top-antitop ($t\bar{t}$) in dilepton channel. In this analysis, the dilepton and the two b jets originate from the decay products of the $t\bar{t}$ are subtracted from the event to reconstruct the charge particles. Then, the multiplicity and the summed Momentum Transverse (p_{T}) of these charged particles used in approaching the observable to study the UE. This study was motivated due to one of the $t\bar{t}$ decay products is the bottom quark (b) which carries colour flow after decaying. The colour charges of the b are expected to connect to the remnant or other coloured final states produced along with the hard processes which need the UE to explain it. Therefore, this thesis used data from proton-proton collisions at a centre-of-mass energy of 13 TeV taken with the CMS detector at LHC. The results can be used to improve the precise measurement of the top quark mass from systematic error. The improvement of the current UE models is necessary to get the result in simulating the pp collision at the LHC more accurate, even further to develop new models of soft QCD processes to increase the understanding of the behaviour of low energy strong interactions.

1.2 Objectives

The thesis intended to achieve the objective of studying the behaviour of the Underlying Event in $t\bar{t}$ dilepton channel and comparing to the non-perturbative QCD Monte Carlo (MC) models.

1.3 Thesis Outline

The thesis is structured as follows: The overview of the proton-proton collision, the $t\bar{t}$ in the dilepton channel, and Underlying Event models are given in Chapter 2. The Chapter 3 gives a brief introduction of the LHC and CMS experiments. In the Chapter 4 provides the general ideas of characterisation the UE properties, summarises the data and simulations employed to calibrate and perform our measurement and explains the data correction from detector limit using unfolding methods. Then the results and discussions are presented sequentially in Chapter 5 and in Chapter 6 whereas the conclusion of this analysis is delivered in Chapter 7.

CHAPTER 2: THEORETICAL BACKGROUND

This chapter gives the short introduction about Standard Model (SM) of particle physics, the theoretical background of the Quantum Chromodynamics (QCD), Monte-Carlo (MC) event generators, and the underlying event in $t\bar{t}$ dilepton studies.

2.1 The Standard Model

The SM describes the interaction of the six type of leptons (e, ν_e, μ, ν_μ, τ, ν_τ) and six type quarks (u, d, c, s, t, b) particles governed by the three of four fundamental forces: electromagnetic, weak, and strong forces where the electromagnetic and weak described in a unification as electroweak force. These forces are caused by exchanging numbers of gauge bosons (Z^0, $W^\pm$, g, and γ). For the strong nuclear force, it exchanged gluons (g), the photon (γ) for the electromagnetic force, and the W and Z bosons are the barter particles for the weak nuclear force. See Table 2.1.

Table 2.1: The Forces

Force	Strength	Gauge Boson	Theory
Strong	10^{38}	Gluon (g)	QCD
Electromagnetic	10^{36}	Photon (γ)	QED
Weak	10^{25}	Vector Bosons ($W^\pm, Z$)	EWT
Gravitational	1	Graviton	GR

In Gauge Theories, SM means the unification of the theory of Glashow, Weinberg, and Salam (G.W.S) and QCD. To match experimental observations, we pick the gauge group (G) for transforming the Lagrangian of the SM to be invariant in local symmetries. As a consequence transforming the Lagrangian under this group, create the existence of the propagators such as photon for the electromagnetic interaction. The SM can be expressed

in a non-Abelian gauge symmetry

$$G = SU(3)_C \times SU(2)_L \times U(1)_Y, \tag{2.1}$$

where the $SU(3)_C$ stands for the vector particles are called gluons for strong interactions. The electroweak symmetry $SU(2)_L \times U(1)_Y$ is spontaneously broken by $<H> \neq 0$ before $U(1)_{\mathrm{EM}}$, as a consequence, the $W^{\pm}$ and Z^0 vector particles become massive while the photon γ remains massless. The photon couples to the electric charge

$$q = T^3 + Y, \tag{2.2}$$

and stands for the electromagnetic interactions, the $W^{\pm}$ and Z^0 are responsible for the weak interactions which act only on the left-handed component of fermions. The (left-handed) neutrino and the left-handed part of the electron form an $SU(2)$ doublet, whereas the right-handed components of the electron and neutrino are $SU(2)$ singlets. The "gauge" concept will be constructed so that the gauge bosons have self-interactions.

In particle physics, the strong force binds the quarks and gluons together inside a proton. To describe the momentum distribution of these partons inside a proton, there is a model called Parton Distribution Functions (PDFs), $f(x, Q^2)$, which depends on two kinematic variables: $x = p_{\mathrm{parton}}/p_{\mathrm{proton}}$ is the fraction of the total momentum of the hardon carried by the parton, Q^2 is roughly the square of the amount of transverse momentum transferred between the two partons. These two variables related to the energy of the proton collision.

In two protons collision, the x is distinguished to describe the fraction of the parton from the proton's forward momentum, x_f and the x_t is from the proton's target, the

mathematical relationship for the total energy is

$$(\text{Total Energy})^2 = s = \frac{Q^2}{x_f x_t}. \tag{2.3}$$

At lower energies, the collision between two protons is considered as the collision between two pointlike particles. The protons just deflect each other which is called elastic scattering. Otherwise, at higher energy, the protons are forced to hit each other and then the partons in one proton interact with another parton from the other proton. As we can see in Figure 2.1, at small Q and small x, the protons look like a group of many partons, each of which is relatively large.

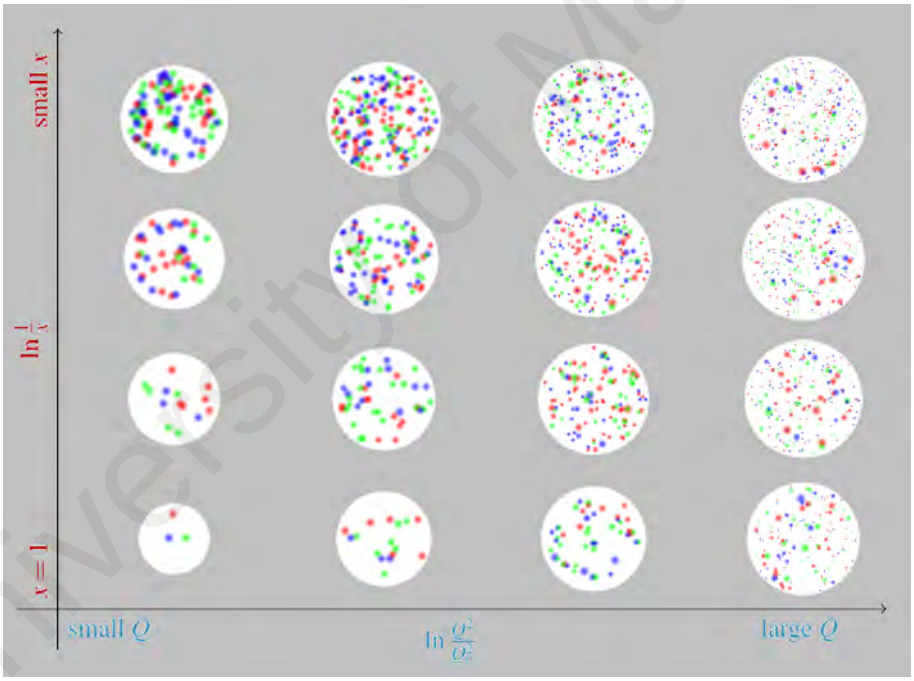


Figure 2.1: The parton distribution function

The idea of accelerating the protons at LHC is to increase the energy collision to reach the protons look like a group tiny partons before smashing them. This acceleration reaches enough energy to allow the partons with minimal values of x_f or x_t to participate in the collision and easily make it out to be detected. Other than that, at the certain energy, the interaction of the partons produce the particle with a large amount of transverse momentum.

At this energy, the interaction is not between two pointlike protons but expected between the partons inside the protons to see what the particles produce after this partons collision. More particles came out than particle went in. The new particle creations are from the additional energy coming from the acceleration of particles. Therefore, the production of particle has been one of the interesting topics to be investigated in high energy experiment. By giving the extra energy to the protons, it will create the new particles with higher multiplicity. The Figure 2.2 explains giving extra energy to the protons means the partons look smaller than at less energy. The proton contains many and tiny partons at small x and large Q.

Figure 2.2: Distribution of the partons after getting some extra energy

The composite of the protons is different after accelerating them in the accelerator. When these two bunch of high energy protons smash in the detector, in parallel, the hard and soft scattering occur during this collision. The Quantum Chromodynamics (QCD) is an underlying theory which can explain the scattering that happens in this parton level and can be calculated using perturbative techniques.

2.2 Quantum Chromodynamics (QCD)

As mentioned above, force occurs when a particle carrying some charge emits another particle transmitting a force. The gluon is an exchanged particle of strong force which can be explained by QCD or $SU(3)_C$ in gauge theory. Unlike Quantum Electrodynamics (QED) which a photon has no charge, a gluon carries the strong charge and has colour. This gluon makes a huge consequence for this strong interaction. As gluons interact with coloured particles, the gluon can interact among each other and it changes the behaviour

of the strong force. In QED, forces between two magnets (or electrons) stronger when they are close to one other and weaker when they are far apart. But in QCD, the strong force becomes more powerful when two quarks pull apart, and it is like free quarks when they are close to each other.

QCD is therefore still only partially solved due to our inability to explain strongly coupled field theories (Skands, 2019). However, all its features, across all distance scales, are believed to be encoded in a one-line equation of alluring simplicity; the Lagrangian of QCD.

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{free} + \mathcal{L}_{int} \\ &= \bar{\psi}_q^i[(i\gamma^\mu)(D_\mu)_{ij}\psi_q^j - m_q\psi_{qi}] + 1/4F_{\mu\nu}^a F^{a\mu\nu}, \end{aligned} \tag{2.4}$$

where ψ_q^i denotes a quark field with colour index i, $\psi_q = (\psi_{qR}, \psi_{qG}, \psi_{qB})^T$, γ^μ is a Dirac matrix that expresses the vector nature of the strong interaction, with μ being a Lorentz vector index, m_q allows for the possibility of non-zero quark masses (included by standard Higgs mechanism or similar), $F_{\mu\nu}^a$ is the strength of gluon field tensor for a gluon with colour index a (in the adjoint representtration, i.e., $a \in [1, ..., 8]$), and D_μ is the covariant derivative in QCD,

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu - ig_s t_{ij}^a A_\mu^a, \tag{2.5}$$

with g_s is the strong coupling which can be expressed as α_s by $g_s^2 = 4\pi\alpha_s$, A_μ^a the gluon field with (adjoint-representation) colour index a, and t_{ij}^a proportional to the hermitean and traceless Gell-Mann matrices of SU(3). These generators are just the SU(3) like of the Pauli matrices in SU(2). By convention, the constant is normally taken to be

$$t_{ij}^a = \frac{1}{2}\lambda_{ij}^a. \tag{2.6}$$

This choice in turn determines the normalization of the coupling g_s, via Equation (2.5), and fixed the values of the SU(3) Casimirs and structure constants, f^{abc}. This structure constants appear in Equation (2.4) expanding the $F_{\mu\nu}F^{\mu\nu}$,

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu. \tag{2.7}$$

The Feynman diagrams for QCD interactions are shown in Figure 2.3 that has different fact from QED which is in the strong interaction which gluons interact with each other. This gluon self-interaction is the main difference between QED and QCD. This fact leads to a lot of important physical consequences. Two consequences are the "asymptotic freedom" and the "confinement".

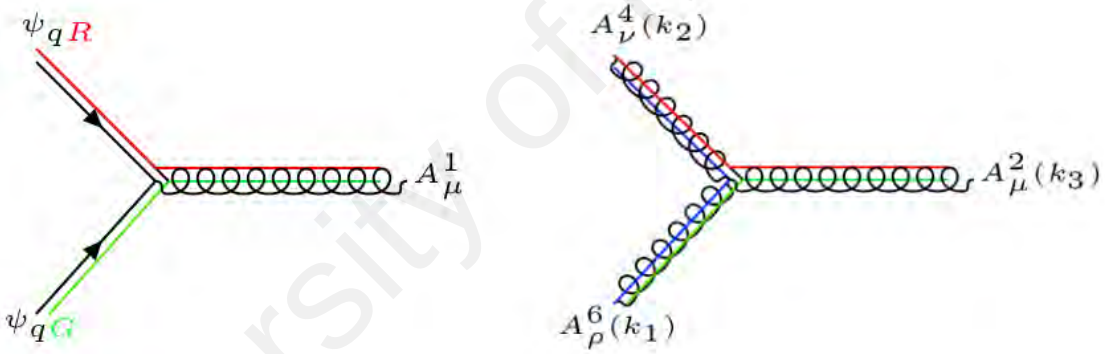


Figure 2.3: Quark-quark-gluon interaction (left) and gluons interaction (right)

2.2.1 The QCD Coupling, α_s

The analogue of the Strong Coupling Constant (α_s) in QCD, Equation 2.5, is the structure constant α in QED. But, due to the gluon self-interaction, the strength of this α_s becomes high at large distance or at low momentum transfers, this fact is opposite from what is in QED which an α is a constant. (Bethke, 2009).

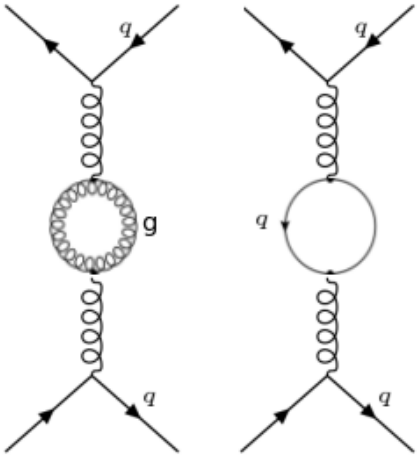


Figure 2.4: One correction by a virtual gg (left) and $q\bar{q}$ (right) pair

QCD does not predict the actual value of α_s, however this variation is due to the energy dependence of α_s. While large coupling at small energy scales leads to the confinement of quarks and gluons inside hadrons, at short-distance reactions or high-energy, the coupling becomes small. In this small coupling quarks and gluons are said to be asymptotically free, i.e. at $Q \to \infty$ the $\alpha_s \to 0$. Let's introduce a function called beta-function to describe the strength of the coupling constant due to the variation of the Q^2. This beta function can be written as a series expansion in powers of the coupling constant,

$$\frac{d\alpha(Q^2)}{d\ln(Q^2)} \equiv \beta(\alpha) = -(\beta_0\alpha^2 + \beta_1\alpha^3 + \beta_2\alpha^4 + ...), \tag{2.8}$$

where the first term corresponds to the leading log approximation. This beta function is part of the formula for the one-loop running constant which is

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \beta_0\alpha_s(\mu^2)\ln(Q^2/\mu^2)}, \tag{2.9}$$

where the μ^2 can be considered as a specific input at some arbitrary reference scale μ^2 which is also a function of renormalisation coupling.

From the Equation (2.9), the β_0 for one-loop QED and QCD coupling are

$$\beta_0(\text{QED}) = -1/3\pi \qquad \beta_0(\text{QCD}) = \frac{11N_c - 2n_f}{12\pi}, \tag{2.10}$$

where the $N_c = 3$ is the number of colours and $n_f = 6$ is the number of flavours in the standard model. In QCD the effective strong coupling constant, e.g. between two quarks exchanging a gluon, is modified by higher order loop corrections in the perturbation theory as shown in the Figure 2.4. The first term $11N_c/12\pi$ in β_0 comes from the left diagram that causes "screening". For the second term $-2n_f/12\pi$ comes from the right diagram and causes "anti-screening". The quarks loop give rise to a positive contribution while the gluon loop contributes to negative sign, which means that the α_s decreases with Q^2 as we can see in the Figure 2.5.

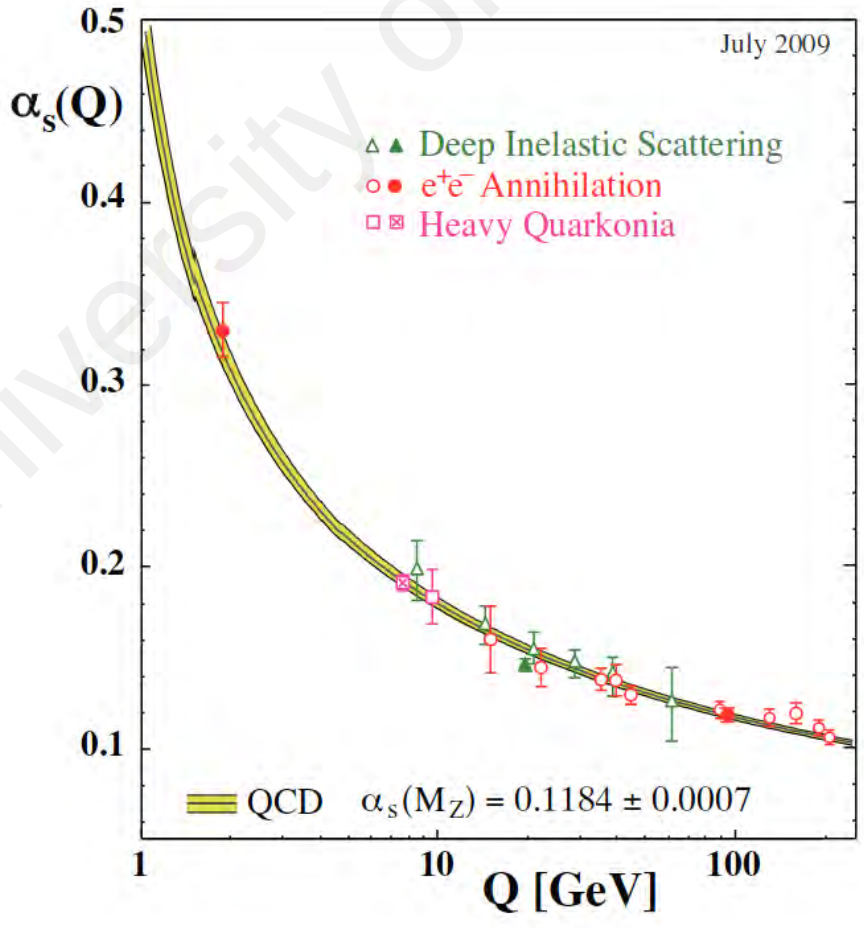


Figure 2.5: The running coupling constant. Figure taken from: arxiv:0908.1135v2

The strong coupling between the two quarks for the quark anti-quark ($q\bar{q}$) has a similar effect as in the QCD. Otherwise, the gluon-gluon (gg) loop has an opposite effect that the interaction strength between two quarks is decreasing in higher energy. In other words, the fermion loop increases and the gluon loop reduces the strong coupling. In very short distances (high Q^2), the quarks and gluons behave like free particles, which is called asymptotic freedom. As coupling strength decreases at higher energy, it gives access to describe QCD physics using perturbation theory which becomes better at higher absolute energies.

Now consider what in the opposite direction of this Figure 2.5, when the coupling α_s becomes stronger in smaller energies. Otherwise, when the quarks and gluon behave like free particles at higher energies, at a certain distance when the force becomes strong enough, the quarks and gluons bind into colourless hadrons (confinement), as isolated free quarks cannot exist in nature.

At certain low energy level, the value of Q^2 is close to Λ, where the Λ is QCD scale parameter as follows at the other expression of the Equations (2.9) changed to,

$$\alpha_s(Q^2) = \frac{1}{\beta_0 \ln(Q^2/\Lambda^2)}. \tag{2.11}$$

with the $\Lambda \approx 250$ MeV, at this condition the perturbation theory breaks down. This Λ is also known as Λ_{QCD} is the scale which non-perturbative effects take over. At this regime, calculation in QCD is very difficult at present time. One way that physicists know how to solve the problem in the strong coupling regime is to simulate the theory with *lattice* QCD (Davies, 2002).

2.3 Monte Carlo Event Generators

To describe everything that happens at LHC during the collision, physicist built up the Monte Carlo (MC) event generators which are software that generates high-energy physics simulation. These generators are expected to allow theoretical and experimental studies of complex collider experiment. Some of general-purpose of Monte Carlo generators, such as, SHEPRA (Gleisberg et al., 2004), HERWIG (Corcella et al., 2001), HERWIG++ (Bahr et al., 2008), PYTHIA 6 (Sjostrand et al., 2006) and PYTHIA 8 Sjostrand et al., 2008, are built to play an essential role in QCD modeling (in particular for aspects beyond fixed-order perturbative QCD) that implement hard scattering process, parton showers, hadronisation and underlying event, and widely used to simulate $t\bar{t}$ events at LHC. In particular, HERWIG and PYTHIA simulate the hadronisation transition according to the cluster and string models. These event generators are used with the detector simulation such as GEANT 4 (GEANT4 Collaboration, 2003) to produce a realistic estimate of the detector response to collision events or in case to plan new experiments. These event generators are built from several components, that describes the physics starting from very short distances scales, up to the typical scale of hadron formation and decay. The structure of proton-protons collision can be described by a few main steps. At the Figure 2.6 illustrates the variety of the levels of the hard scattering processes until the detector detects it.

Hard scattering process is when a parton from one proton collide with another parton from the proton's target and it investigates distance scales far below the radius of the proton to probe the collision between the partons of the proton. Hard scattering part is usually specified regarding the momentum component that is transverse to the beam, p_{T}. These processes are the key access in understanding what is happening at LHC collision and also known as the necessary primary processes to study a new physics. At the heart of

the collision, perturbative QCD handles all the calculation to explain everything here.

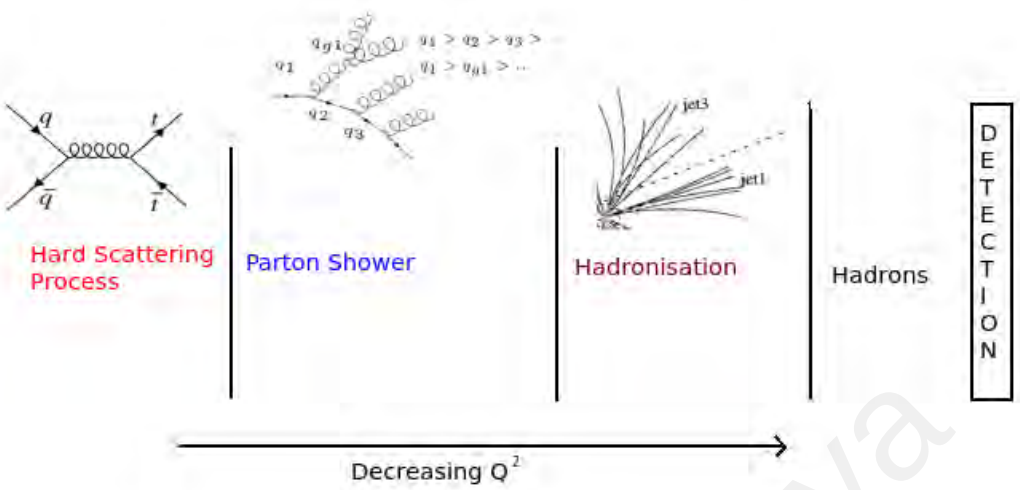


Figure 2.6: High energy collision processes in a few steps: from collision to detection

But, the particle production from the quarks and gluons interaction in this hard process level is different as what is detected by the detector. This is due to the lower-energy gluons that might be radiated a few second after hard scattering processes occured. This stage is known as Parton Shower (PS). The gluons radiate often at small angles from the original parton until the perturbative QCD breaks down to a level when the partons bind into colour-neutral hadrons or hadronizarion. Hadronisasion is a long-distance process, involving only small momentum transfers, and can not be calculated using perturbative QCD. The parton shower and hadronisasion level rely on non-perturbative QCD models to study their characteristics.

Along with this hard process, other activities from the remnant partons accompany during the collision. These activities are categorised as soft proccesses and classified also as non-perturbative QCD part, but to explain these particles the Underlying Event (UE) take over. The detailed description about the UE is in the next section.

2.4 The Underlying Event

In events containing a hard parton-parton interaction, the underlying event represents the additional activity which is not directly associated with that hard interaction. Here, we shall define the underlying event to represent the additional activity after all bremsstrahlung off the hard interaction has already been taken into account. The soft particles that are categorised as UE activities are dominant in the protons collision. It is not achievable to simulate all possible collision events, but the simulation needs to be structured with a focus on what hard process is needed. For example, in CMS experiment, the UE studies using various leading objects such as a leading (highest p_{T}) charged-particle jet (First UE Measurement) (CMS Collaboration, 2010), Drell-Yan lepton pair (CMS Collaboration, 2012), and a leading charged particle (CMS Collaboration, 2011). In our UE studies, the $t\bar{t}$ event is chosen as the leading object. This UE analysis used dilepton channel and b jets as the final decay products of the $t\bar{t}$, which means these final decay products are subtracted from the event.

Measurement of the UE using different leading objects has been studied at the Tevatron in proton-antiproton collision (CDF Collaboration, 2002, 2004, 2010) and at the LHC, the ALICE, ATLAS, and CMS experiments in proton-proton collision (ATLAS Collaboration, 2011; CMS Collaboration, 2013, 2015). Note that results from the two colliders may not be directly comparable, as the $p\bar{p}$ collisions at the Tevatron will give rise to more high-momentum $q\bar{q}$ interactions than at the LHC (where there are no valence anti-quarks).

The only tiny fraction which the events of pp collision plays an important role as an interesting process, the domination processes come from the leftover partons that are not associated with the hard scattering. If the perturbative QCD explains the hard part, the soft part needs the UE phenomenological models to describe the performance. These UE models consist of particles from the hadronisation of Beam-Beam Remnant (BBR),

Multi Parton Interaction (MPI), Color Reconnection (CR) and their associated Initial State Radiation (ISR) and Final Sate Radiation (FSR) which do not come from the hard scattering processes. The detailed explanation about these models can be found in the Section 2.4.2.

2.4.1 The Top Pairs Event

Top quark is well known as the heaviest quark in the Standard Model particle, with the mass is, $m_t = 173.29 \pm 0.10(\text{stat.}) \pm 0.92(\text{syst.})$ Giga Electron Volt (GeV) (Markus & Mulders, 2017), close to the electroweak scale. This top quark mass property is one of the interesting topic to play a special role in electroweak symmetry breaking especially in many precision measurements. Precise measurement of the top quark mass provide a constraint on the Higgs boson mass (D∅ Collaboration, 2011). The top mass measurements at LHC are becoming more and more precise, with the most recent ones achieving a precision of less than 1 GeV (Vos, 2016). Other important property of the top quark is it short lifetime ($\sim 5 \times 10^{-25}$ sec), it makes the top will decay before hadronisation. This condition gives us a rather unique access to study more about top quark such as mass, charge, spin, or coupling.

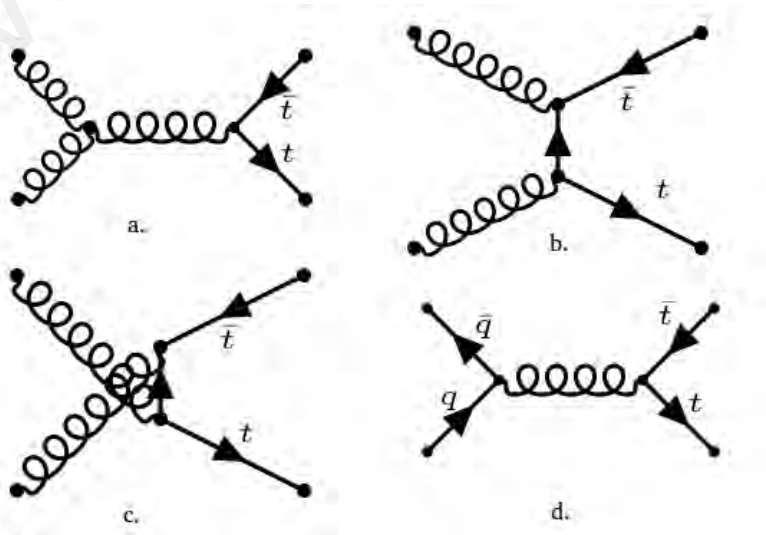


Figure 2.7: $t\bar{t}$ productions

At LHC, the dominant production mechanism is gluon fusion into a $t\bar{t}$ pair ($\sim$ 85 %), the Feynman diagrams are shown in Figure 2.7 (a, b, and c), while the remaining 15 % are from annihilation quark and antiquark contribution (Butterworth et al., 2012), see Figure 2.7 (d). The top quark decays almost 99 % to a b quark and a W boson, which gives rises to a set of final states which are classified according to the W decays.

- Dilepton channel: In this channel, the final state contains the eν or a $\mu\nu$ as the final decay product of W boson from both top and antitop. The missing energy that belongs to the two neutrino candidates besides this dilepton signature also as W decays. Even though only 5 %, this dilepton channel is an interesting process because of the benefits from low background levels (D$\varnothing$ Collaboration, 2011).
- Semileptonic channel: This channel is characterised when one of the W boson decays into a pair of quark and anti-quark while the other W to a charged lepton and a neutrino of the same flavour. The branching ratio for the semileptonic channel is about 30 % and the background contribution mainly from W production with additional jets. Due to the tau lepton lifetime that is very short, the lepton candidates for this channel are possible for electrons and muons only.
- Fully hadronic channel: In this channel, both W bosons decay into a pair of quark and anti-quark of the different flavour. As a result, this channel produces six quarks in the final state. About 44 % the final state of the $t\bar{t}$ decay is on this channel. However, this channel has a high background coming from QCD processes.

The fact that one of the top quark decays, namely the b quark, carries the color flow after decaying. This b quark is expected to fragment and hadronize which are the main source of uncertainty in the measurements of the top-quark mass (Corcella & Mescia, 2010). The fragmentation expects from this b quark and generates the B hadron which can be surrounded by collimated products of the fragmentation, called b-jet. Owing to this

fact the full reconstruction of the initial kinematics may be compromised by the ability of the jet algorithm to capture the products of the fragmentation of the *b*, the connection between the *b* and the proton remnants or the other *b* in the event. An accurate description of both the fragmentation and hadronisation of the *b* quark is underlying event.

2.4.2 The Underlying Event Models

As mentioned above, the underlying event processes are described by phenomenological models. The goal is potentially improving understanding of soft-QCD, if the underlying event can be accurately modelled, then it can be removed from the analyses of hard scattering processes where it is an inconvenient background (Wynne, 2012). The UE can not be calculated perturbatively but relies on phenomenological models that "tuned" to all relevant available data. Four of those UE models are: the ISR and FSR and are considered as part of the UE due to one of the partons emits radiation, such as a photon, before/after the interaction, as long as is not clustered as the two *b* candidates. These radiative emissions are not due to the annihilation of the particle. The BBR can be treated as UE due to the left over partons that were not classified as a hard process or ISR. Viewing hadrons as "bunches" of quarks and gluons, it is possible in a single proton-proton collision has multiple partons collisions, as depicted in Figure 2.8. This phenomenon is called MPI. MPI has access descriptions for both processes, hard and soft regimes. For hard scattering process, the MPI studies mostly focus on two simultaneous parton scatterings. It affects the soft regime in a single collision. The UE can be explained by the small transverse MPI in Monte Carlo generators such as PYTHIA.

In fact, by extending simple perturbation theory, high-energy hadronic collisions are guaranteed to have several perturbatively calculable interactions. Furthermore, such interactions - even when soft= can be highly significant causing non-trivial changes to the colour topology of the colliding system as a whole, will donate drastic consequences for

the particle multiplicity in the final state.

CR is a formed mechanism to describe the interactions that can occur between colored field, from top decay products, during the hadronisation process. At the LHC, due to the high number of colored partons are produced during collision of the two protons, this process is expected to occur at a significant rate. CR significantly contribute to the UE process since there is no first-principles models explain in this process. In this UE studies, the color reconnection is expected happened between the b jets candidate and the partons remnant or the other coloured final states produced along with the $t\bar{t}$ production during the b fragmentation.

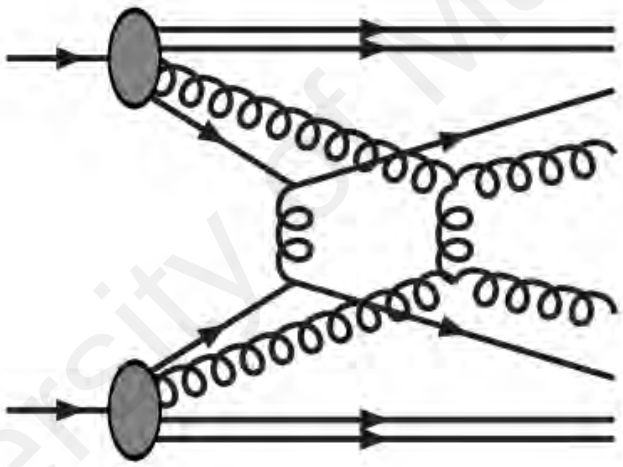


Figure 2.8: Multi Parton Interaction

2.4.3 Underlying Event Tunes

The MC generators of QCD have other parameters that can be adjusted to control modelling of the properties of the events, and a specified set of such parameters adjusted to fit certain prescribed aspects of the data referred to as a "tune" (CMS Collaboration, 2016a). This tune is a group of various parameters that can be changed from the previous experiment to be tuned to recent experimental data. Technically, the variation of the parameters are independent of each other, but the physical requirement of a sensible

description of a set data needs to correlate and anti-correlate between them. Hence the tunes are produced by a group of the parameters simultaneously.

The RIVET (Buckley et al., 2013) software is a tool for generating predictions of physics quantities produced from MC event generators. It is used for providing sets of MC predictions with a different choice of parameters related to the UE simulation. PROFESSOR framework includes the predictions, which parametrises the generator response and returns the set of tuning parameters that best fits the input measurements.

QCD MC generators such as PYTHIA and HERWIG ++ have parameters that may be adjusted to control the behaviour of their event modelling. For example, the parameters in PYTHIA 6 and PYTHIA 8, together with the selected PDF, determine the energy dependence of MPI. Recently, in HERWIG++ the same formula has been adopted to provide an energy dependence to their MPI cutoff. The other difference between PYTHIA 8 and HERWIG++ is the model of hadronisation process. The PYTHIA 8 uses string hadronisation model while HERWIG++ uses cluster model to explain the hadronisation behaviour (Gieseke, 2012). The UE tunes used in these simulations are CUETM2P4 and EE5C with PYTHIA 8 and HERWIG++, respectively, and we also compared the other tunes which are varied depending on the UE models.

CHAPTER 3: THE CMS EXPERIMENT

This chapter covers explanation about the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) experiment.

3.1 The Large Hadron Collider (LHC)

The LHC is the particle accelerator in the world. LHC is designed to collide proton beams with center-of-mass energies of 14 TeV and luminosity of 10^{34} $cm^{-2}s^{-1}$, which means that in the LHC experiment the collisions are produced about 10^{34} per second per cm^3. It has two rings superconducting hadron and collider in a 26.7 km long tunnel of eight straight sections and eight arcs between 45 m and 170 m below the surface. As the LHC is a proton-proton collider, both beams cannot share the same phase space in a single ring, but it needs two rings with counter-rotating beams. The LHC can also collide heavy ions, such as gold, with an energy of 2.8 TeV per nucleon and a peak luminosity of 10^{27} $cm^{-2}s^{-1}$.

The LHC is separated into eight arcs and eight straight section of 528 m length. The two high-luminosity experiments, A Toroidal LHC Apparatus (ATLAS) and Compact Muon Solenoid (CMS), are located opposite to each other, at points 1 and 5. Point 2 houses the A Large Ion Collider Experiment (ALICE) and point 8 houses the (Large Hadron Collider beauty) LHC-B experiment. At point 6 locates the beam dump insertion: horizontally-deflecting fast-pulsed magnets combined with vertically-deflecting double steel septum magnets extract the beams vertically. It is shown in the Figure 3.1.

23 regular arc cells of 106.9 m length build up the arcs of the LHC lattice. Each cell is subdivided into two half cells, each containing a cold mass, a short straight section and three dipole magnets. The design of an LHC has been chosen such as to optimise the maximum integrated dipole field along with the smallest possible number of magnet

interconnections and minimum beam envelopes.

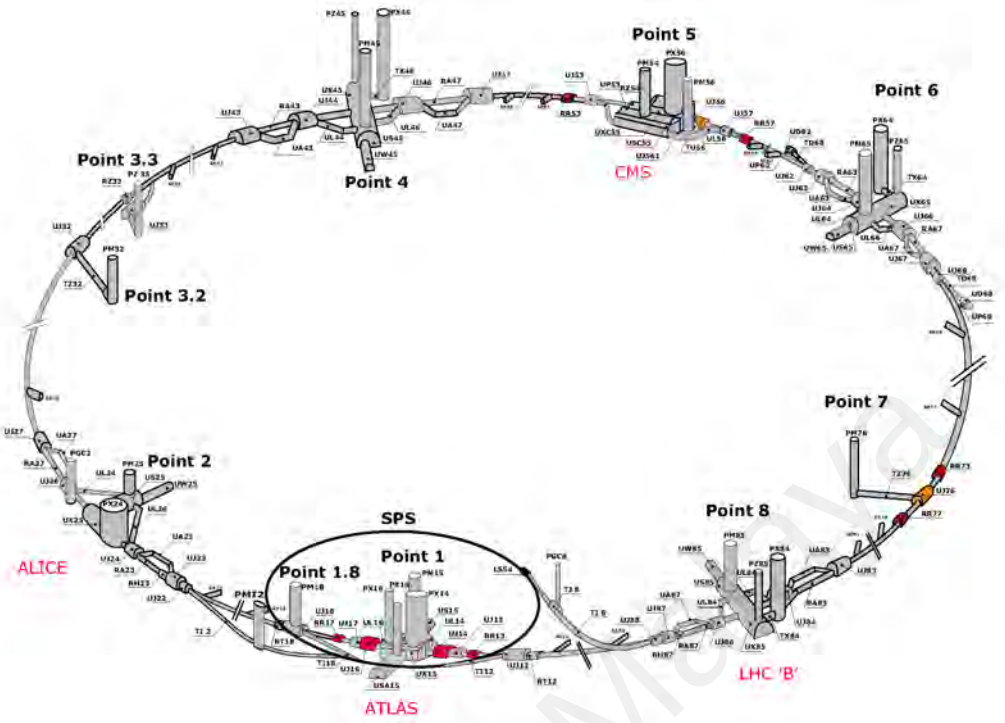


Figure 3.1: Large Hardon Collider: The figure was taken from http://te-epc-lpc.web.cern.ch

Luminosity ($\mathcal{L}$) is one of the most important variables used in accelerator physics. It is a measurement of the number of collisions that can be produced in a detector per cm^2 and per second. The larger the value of $\mathcal{L}$, the higher the number of collisions. Cross Section (σ) is considered to calculate the number of collisions. The number of events N (event) per second generated in collisions with cross section σ (event) and luminosity $\mathcal{L}$ is

$$N/s = \mathcal{L}\sigma. \tag{3.1}$$

3.2 Compact Muon Solenoid (CMS)

This part describes the CMS detector at the LHC that is operated at Point 5. The CMS was built to study proton-proton collisions at center-of-mass energies of 14 TeV and at luminosities of up to 10^{34} cm^{-2}s^{-1}. According to its name Solenoid and Muon, the main features of this detector is a superconducting solenoidal magnet enclosing an all-silicon

internal tracking system. An important aspect driving the CMS detector design and layout is the choice of the magnetic field configuration for identifying muons and reconstruct charged particles with high efficiency and good momentum resolution. Another main design choice is to resolve missing transverse energy and dijet masses to better levels. The CMS detector weighs 12500 tonnes, is 21.6 m long and 14.6 m for the diameter.

The coordinate system of the CMS detector origins at the interaction point, the y-axis pointing vertically upward and the x-axis pointing toward to the centre of the LHC. Along the beamline defined the z-axis. Cylindrically, sets ϕ as the azimuthal angle on the x-axis the direction in the xy-plane and θ as the polar angle on the z-direction commonly used the pseudorapidity $\eta = -\ln\tan(\theta/2)$. It illustrates in Figure 3.2.

On average 10^9 inelastic collisions per second will be observed by CMS detector with a total proton-proton cross section of 100 mb at 14 TeV center-of-mass energy. The CMS online event selection (trigger) and offline analysis will reduce to 100 events per second for storage. The short time between two LHC bunch spacings, 25 ns, gives a challenging task to design the readout and trigger systems. On average 20 inelastic collisions produce approximately 1000 charged particles that influence the event measurement in every 25 ns.

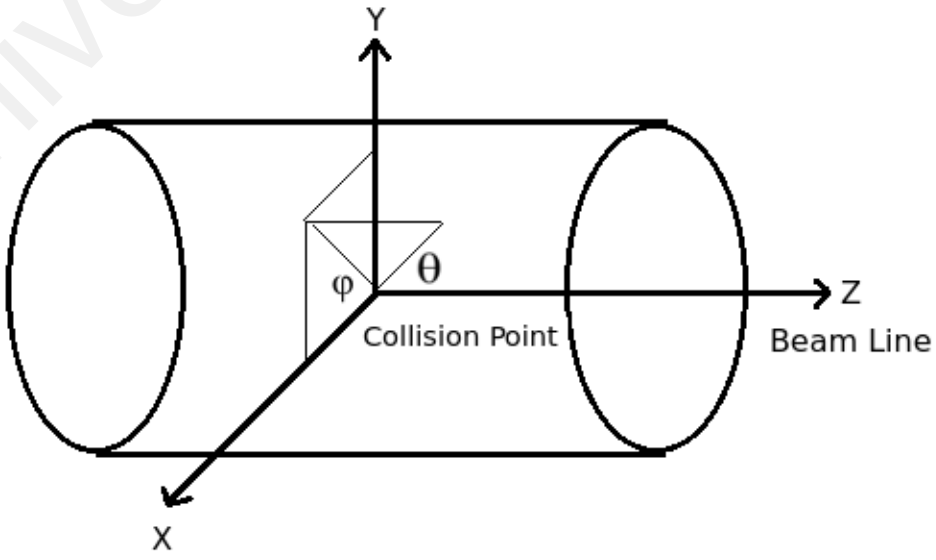


Figure 3.2: An illustration of the CMS coordinate system

Frequently, these collisions have low-p_T and low-multiplicity processes. Therefore, trigger systems work to select "interesting" events efficiently. The particles from both the scatter of interest and additional interactions in the same bunch-crossing (pile-up) will be read out. To reduce the effect of the detector performance due to the increase in the detector occupancy from pile-up needs a high detector granularity with a fast time resolution. Although, this kind of detector requires a significant number of detector channels with perfect synchronisation.

In the Figure 3.3 illustrates a perspective view of the CMS experiment. At the heart of the CMS sits a superconducting solenoid provide a strong magnetic field of 3.8 T, with 13 m long, and 5.9 m inner diameter. The CMS magnet allows electricity to flow without resistance and creating a powerful magnetic field. A high magnetic field was chosen to achieve good momentum resolution of the charged particles with high precision. For a higher charged particle's momentum, the path is less curved in the magnetic field. So, the CMS detector is designed to measure muon that has high momentum more accurate. Outside the solenoid are the yokes for the return magnetic field and the muon system. The Figure 3.3 taken from (Barney, 2016).

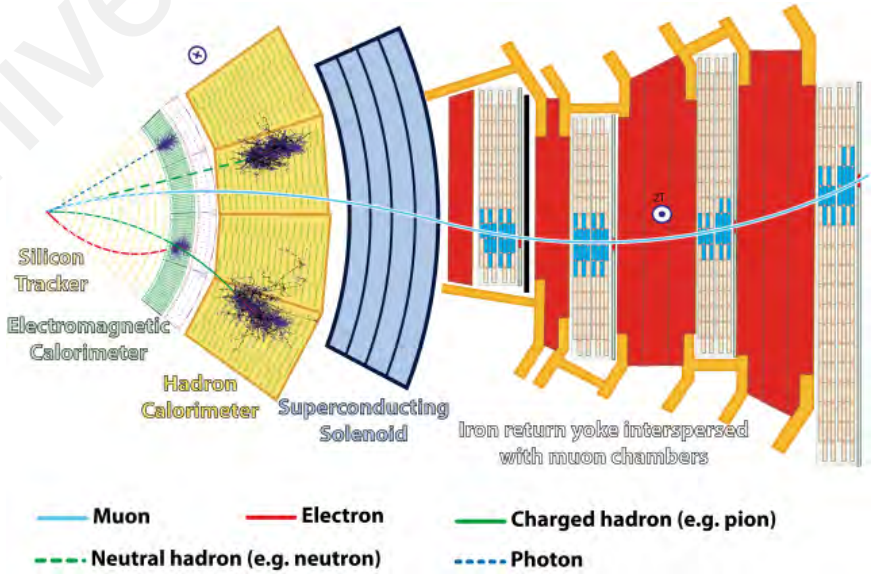


Figure 3.3: CMS detector in slice

CMS internal tracking system measures charged particles after the collision. The charged particles pass through and interact with the tracking device then reveal the paths of the particles. These tracks of the particles are visible by small record electrical signals as they move through the tracking instrument. This tracker reconstructs the signal of the high-energy muons, electrons, and hardons that hit through the device. The path of the particles hit the device can be seen as the curve pattern if the momentum of the particle is small due to Lorenz force law. See the Figure 3.4. The higher quality track has more hits within the Silicon Tracker. The particle paths need to be recorded accurately due to the busy events at the heart of the collision. By taking position measurements using a few reconstructed hits, the track of the particles is measured accurately.

The CMS tracker consists of 13 layers of silicon pixel detector in the central part and 14 layers in the endcaps. This CMS Tracker is the largest silicon detector in the world. It has 205 m^2 of silicon sensors comprising 76 million channels.

In the interest to find Higgs boson and other new physics, the electrons and the photons are the importance particles to measure at LHC. CMS must find energies of emerging particles to build up the picture of events collision. The part of the CMS detector that measures these kinds of particles is Electromagnetic Calorimeter (ECAL).

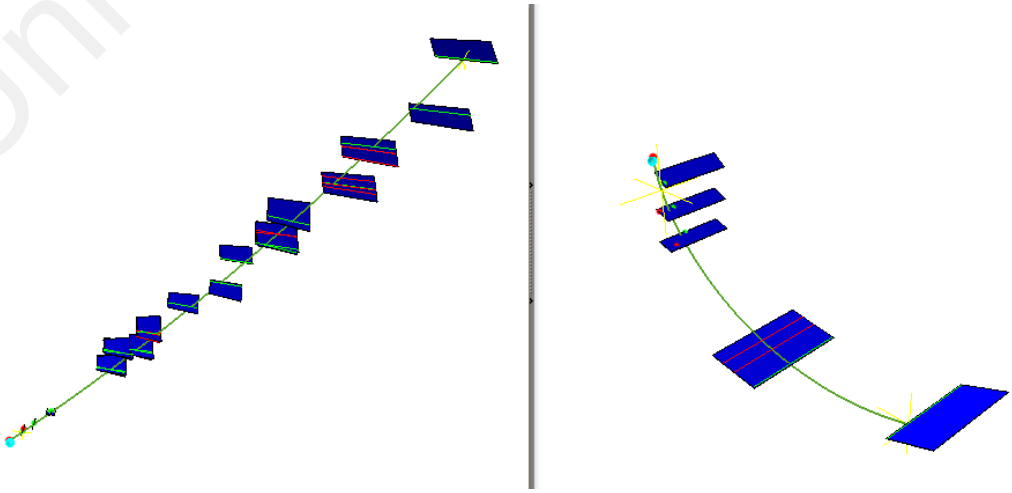


Figure 3.4: The charged particle hits the Silicon Tracker

The ECAL is built up between the tracker and the Hadronic Calorimeter (HCAL), made up of a barrel section and two endcaps. The barrel consists of 61,200 crystals cylindrically, formed into 36 supermodules. The weight of each supermodule around three tonnes and be composed of 1700 crystals. The flat ECAL endcaps close off the barrel at either end and are made up of almost 15,000 further crystals.

The next layer external ECAL is the HCAL. It is designed to measure the energy of "hadrons", all the particles that made up of quarks and gluons. Additionally, it provides the missing energy transverse (E_T^{miss}) from non-interacting uncharge particle such as neutrino indirectly. Some of the particles decay and produce new particles without leaving the track on the detector. The way to see untracked particles uses the imbalance the momentum and the energy of the shooting particle to one side of the detector, but not to the other. Measuring these particles is important to find new particles such as supersymmetric particles.

The layers of the HCAL were built in a "compact" way to make sure there are no gaps in a direct line that will make the particle might escape undetected. The HCAL can find a particle's position, energy and arrival time using alternating layers of an absorber and fluorescent scintillator materials that release a rapid light pulse when the particle passes through.

The HCAL consists of four parts: the Hadron Barrel (HB), the Hadron Endcap (HE), the Hadron Outer (HO) and the Hadron Forward (HF) sections. There are 36 barrels with the weight is 26 tonnes each.

Detecting of muons is one of the important mission that the CMS has as its name "Compact Muon Solenoid". Muons are the charged particles which 200 times heavier than electrons and positrons. Muons are expected to decay of some potential new particles, such as one of the clearest signatures is Higgs to four muons.

The Muons detector were placed at the outer layer of CMS detector due to the muons that can penetrate several metres of iron without interacting. Muons can not be blocked by any of CMS's calorimeters like the most particles.

To measure the Muons, the hists among the four muon stations are fitted to a curve. These muon stations are from the outside the magnet coil and interleaved with the iron "return yoke" plates. The particle's path is detected precisely through the combination of the muons tracking position through the multiple layers of each station and the tracker measurements. This path is considered as the particle momentum because that particle travelling with less momentum curve in a magnetic field. The momentum of the muon is not only measured by the inner tracking but also by the muon system. The different methods for these two layers are the muon system uses the bending angle leaving solenoid, and inner tracking system uses the track curvature fit. By merging the information from tracking and the muon system, it enhances the accuracy in the full momentum range. Thanks to the CMS magnet that is very powerful so we can bend the paths of very high energy muons to calculate their momentum.

This detector has 1400 muon chambers, includes 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) track the particles' positions and set with a trigger, while 610 resistive plate chambers (RPCs) form an additional trigger system, which quickly decides to keep the collected muon data or not. The many layers of detectors have the different type of each speciality, and this is naturally able to filter the background noise out.

3.3 The Reconstruction Event in the CMS Experiment

The reconstruction of the physics object at CMS experiment uses The Particle Flow (PF) algorithm (CMS Collaboration, 2017c). The reconstructions are stated from the final state of the leading objects. This algorithm combines the signals from all CMS sub-detectors to enhance the reconstruction performance, and it allows to identify the stable particles from

pp collision production, such as photons, muons, charged hadron, and neutral hadrons.

The information from the tracker reconstructs the PF event considering as good momentum resolution and precise measurement of the charged-particle direction of the vertex production. On the other hand, the tracker does not give the information of the stable neutral particles, such as photon and neutral hadron. These particles use the information from the calorimeter to find its energy. Performance and development algorithm for the PF event reconstruction is specified in each sub-detector. The reconstruction and the identification processes are used to build the jets, determining the missing transverse energy, decay products reconstruction, etc.

3.3.1 Jet in CMS Experiment

Jets are experiment signatures of partons produced in high energy protons collisions. The colour-confinement as a consequence of the gluon self-interaction leads the partons to bind together to form colour-neutral hadrons (hadronisation). This hadronisation form to a collimated spray of hadrons called a jet. The signals of charged-hadrons are determined as the combination of tracks stored in tracker layer associated ECAL and HCAL as the energy deposits. However, the neutral-hadrons is determined from the remaining energy deposits at calorimeter. The information from all these sub-detector layers is combined using the jet algorithm to reconstruct a jet. Commonly, CMS used anti-k_{T} algorithm (Cacciari et al., 2008) with the radius parameter is R = 0.5, but other jet algorithm also supported to reconstruct the jet on the event. The PF also is used prior using the jet clustering based on information from all the sub-detectors. Jet studies give access to the precise understanding of the underlying partonic process and their properties which are crucial in many physics analyses.

3.3.2 CMS Trigger System

The rate of the proton-proton and heavy ion collisions is in high intensity. About one billion proton-proton interactions happen in every second inside the detector. During the collision, impossible for the detector to detect all final interaction products because only a small fraction of these collisions contain an event of the interest and only a small fraction of those event interest can be stored to use in offline analysis. This selection is a job of the trigger system which collects events of potential physics interest from the collision. The CMS trigger utilizes two levels. The first level built up as the hardware to select the event that contains muon, lepton, Missing Energy Transverse ($E_{\mathrm{T}}^{\mathrm{miss}}$), or jets. In this level, the trigger selects the best 100,000 events each second from the billion available. This selection is the upper limit imposed by the CMS readout electronics. In the High Level Trigger (HLT), the entire event is created by assimilating and synchronising information from different part of the sub-detectors. The HLT, implemented in software, selects the average rate about 400 Hz and stores it to the computers for offline user (CMS Collaboration, 2017a).

CHAPTER 4: MEASUREMENT OF UNDERLYING EVENT IN TOP PAIRS DILEPTON CHANNEL

4.1 Analysis Strategy

The objective of this analysis is to estimate the contribution of the underlying event to the hard scattering processes. Because of that, this study was started by choosing the hard scattering processes. As mentioned in the Section 2.4, $t\bar{t}$ in dilepton channel was chosen as the hard part, which means, the event were categorised according to the final state of $t\bar{t}$ decay products and the flavour of the dilepton candidate (ee, eμ, $\mu\mu$). The kinematic reconstruction is explained further ahead in Section 4.3.

The underlying event in the pp collision is a mixing of several contributions from the models: Initial- and final-state radiation (ISR/FSR), multiple interactions (MPI), beam-beam remnants (BBR), and also colour reconnection (CR). Other than that, most underlying event observables only probe the sum of these models. So, for the second step, these models were chosen to describe the UE behaviour as a direct test of these models. It allows their relative merits to be examined and their subsequent performance improving event processes with the free parameters adjusted to match experimental results (see Section 2.4.2). If the underlying event can be accurately modelled, the understanding of soft-QCD processes will be improved. Together with the hard processes, these improved models would explain better the whole story about what happened during the collision at the collider.

Underlying event observables are expressed as the collective properties of all the selected charged-particles. As the third step, some of observables were selected in this analysis to characteris the UE activity. The set of variables are related to the multiplicity and momentum of the charged particles candidates are listed in the Table 4.1.

Table 4.1: Typical underlying event observables, and their symbolic representation

Observables	Interpretation
N_{ch}	Number of charged particles
Σp_{T} or Σp_{z}	Scalar p_{T} or Momentum at longitudinal direction (p_z) sum of charged particles
$\bar{p}_{\mathrm{T}}$ or $\bar{p}_{\mathrm{z}}$	Average p_{T} or p_z per charged particles with $\bar{p} = \Sigma p/N$

The conceptual separation of the component of "hard-scattering" and "underlying-event" is not necessarily equivalent to a clean separation in terms of "short-distance" and "long-distance" physics (Skands, 2010). So for the next step, we defined the function of different variables related to the kinetics of the $t\bar{t}$ system in order to profile the UE which is called "slicing" variables. The first "slicing" variable is the number of additional jet (=0, =1 or ≥2) with $p_{\mathrm{T}} > 30$ GeV, and $|\eta| < 2.5$. The second one is the reconstruction of the $t\bar{t}$ system direction as:

$$p_{\mathrm{T}}(t\bar{t}) = p_{\mathrm{T}}(b_1) + p_{\mathrm{T}}(b_2) + p_{\mathrm{T}}(l_1) + p_{\mathrm{T}}(l_2) + p_{\mathrm{T}}^{\mathrm{miss}}, \tag{4.1}$$

where the $p_{\mathrm{T}}(b)$ and $p_{\mathrm{T}}(l)$ are the p_{T} for the b-jets and the lepton respectively, with the index 1 and 2 represent the lepton and the b jet that come from the top and anti-top event, and the $p_{\mathrm{T}}^{\mathrm{miss}}$ is the p_{T} for the missing tranverse energy. The variables chosen are of interest for top mass measurements and thus, profiling the effect of soft QCD along these, may help to bring further insight to the uncertainties affecting the measurement of the mass.

We have chosen the kinematics observables which are expected to be reconstructed in high resolution. The Figures 4.2 - 4.4 show the expected resolution on the UE variables, as a function of the generator level quantities. The results are obtained from a $t\bar{t}$ MC sample. The entries in the 2D histograms are normalised per column to 100 % and the points represent the bias and width of the distributions as estimated from a Gaussian fit. The variables are expected to be reconstructed with a slight bias owing mostly to the

inefficiency in the track reconstruction, but also to residual contamination from the pileup, nuclear interactions, track splitting, etc.

In the p_{T}-slicing variables can define the regions in the ϕ-space on an event-by-event basis which is expected to be correlated with the direction of the $t\bar{t}$ system:

$$\phi(t\bar{t}) = \arctan\frac{p_y(t\bar{t})}{p_x(t\bar{t})}. \tag{4.2}$$

These regions that used the event topology in the position to observe the underlying event domination (see the Figure 4.1). The following region is defined from this ϕ-slicing are listed in the Table 4.2:

Table 4.2: The regions defined according to the p_{T} of $t\bar{t}$ system

Regions	Categorized
$\lvert\Delta\phi\rvert < 45^o$	toward region
$90^o < \lvert\Delta\phi\rvert < 270^o$	transverse region
$\lvert\Delta\phi\rvert > 270^o$	away region

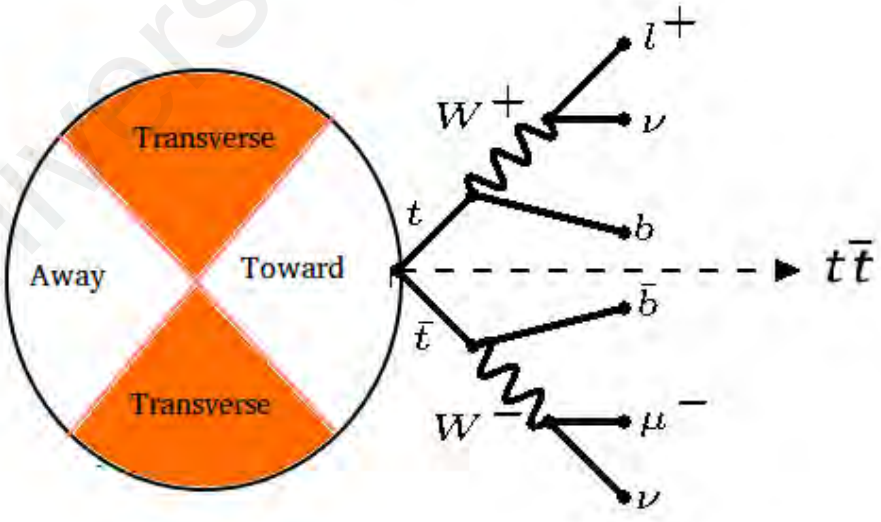


Figure 4.1: An ilustration of the segmentation of ϕ-space

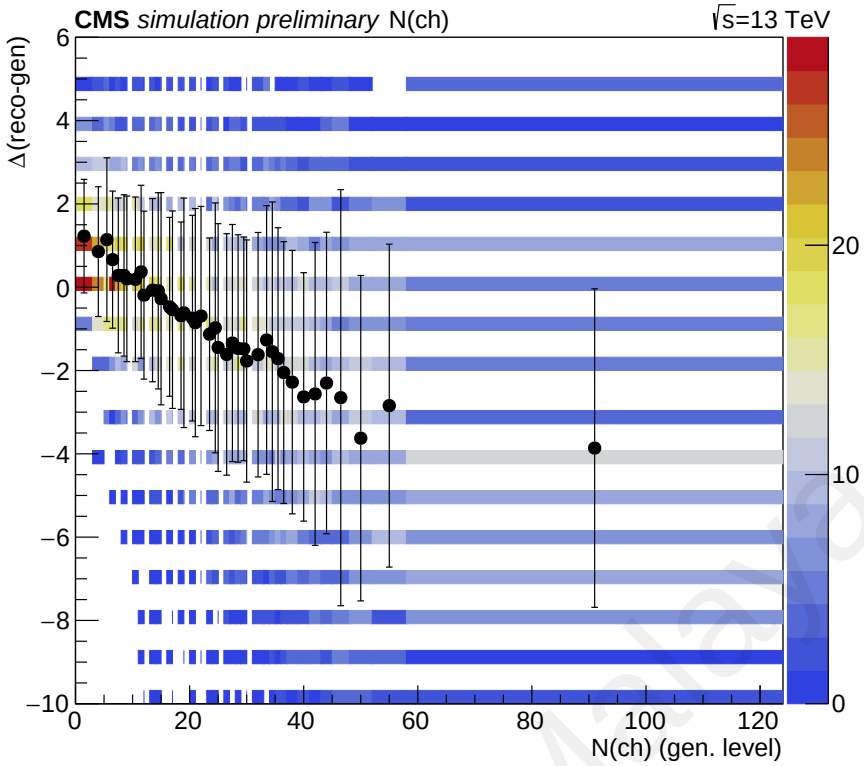
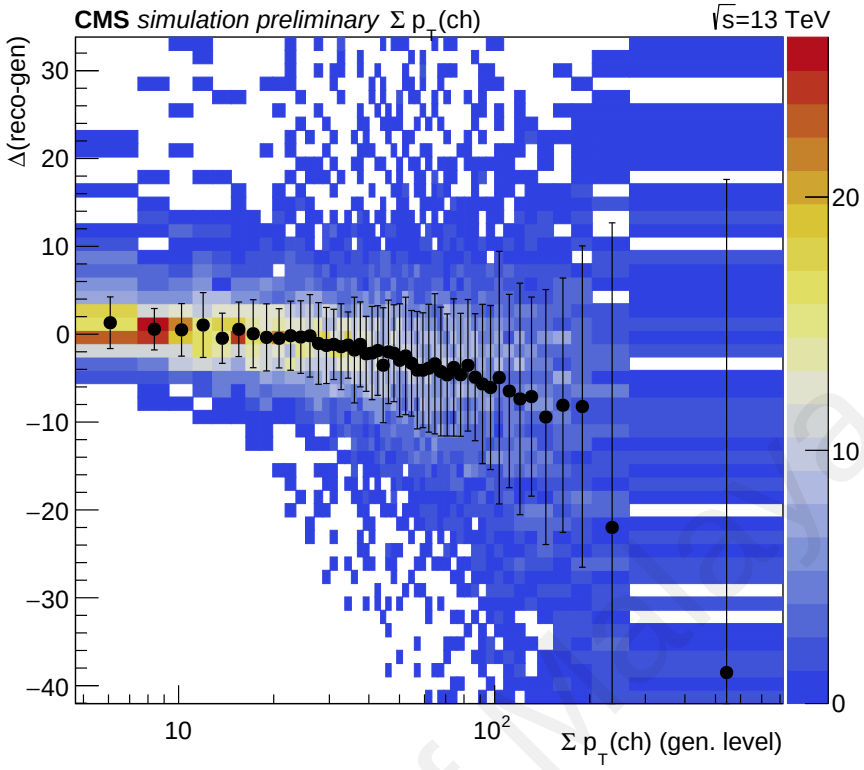
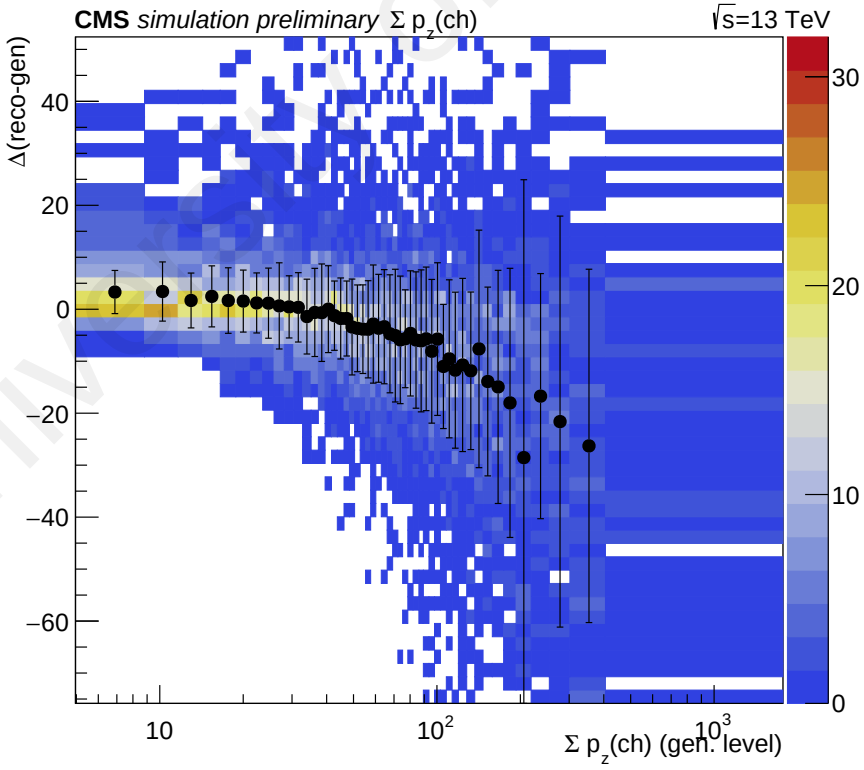


Figure 4.2: Resolutions for the multiplicity as function of the generator level quantities

The basic idea about this segmentation was from the assumption that the transverse region was dominated by underlying event while the toward and the away regions are less contaminated. These reasons are due to the toward and away regions contain the particles that come from the hard processes. This technique was developed and used in the previous studies of the underlying event at the Tevatron (Kar & CDF Collaboration, 2009). This segmentation allows us to investigate the distribution of the models such as from MPI (transverse region), ISR/FSR (away region) contributions and to test the different production models of the $t\bar{t}$ system. We also investigated the observables that reflect the underlying activity in these regions to study the particle production from the number of charged particles, the transverse momentum of the particles, and the spreaded momentum among the particles.

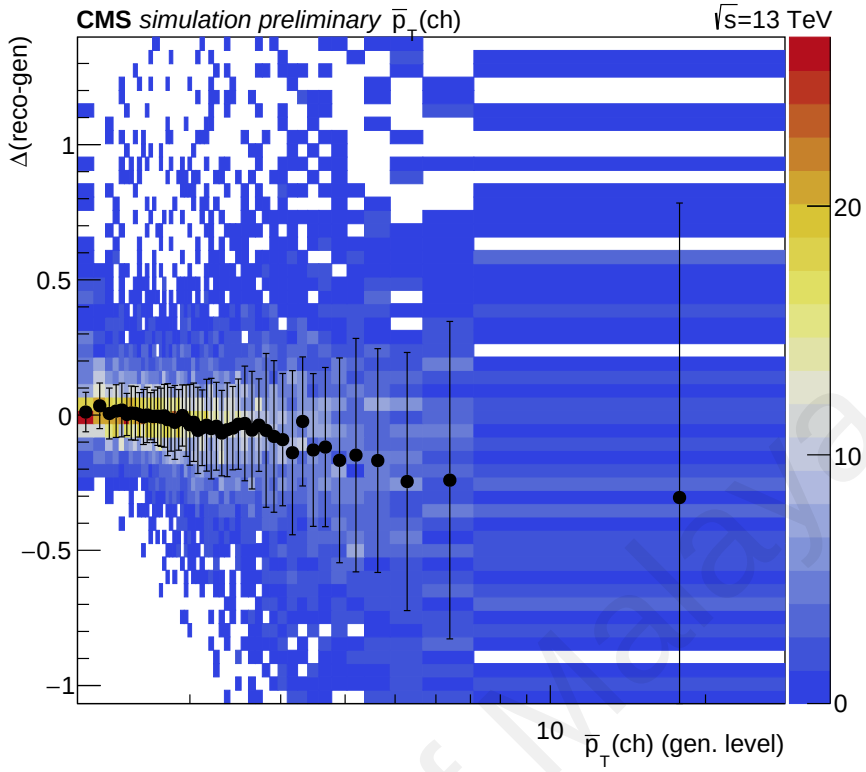


(a)

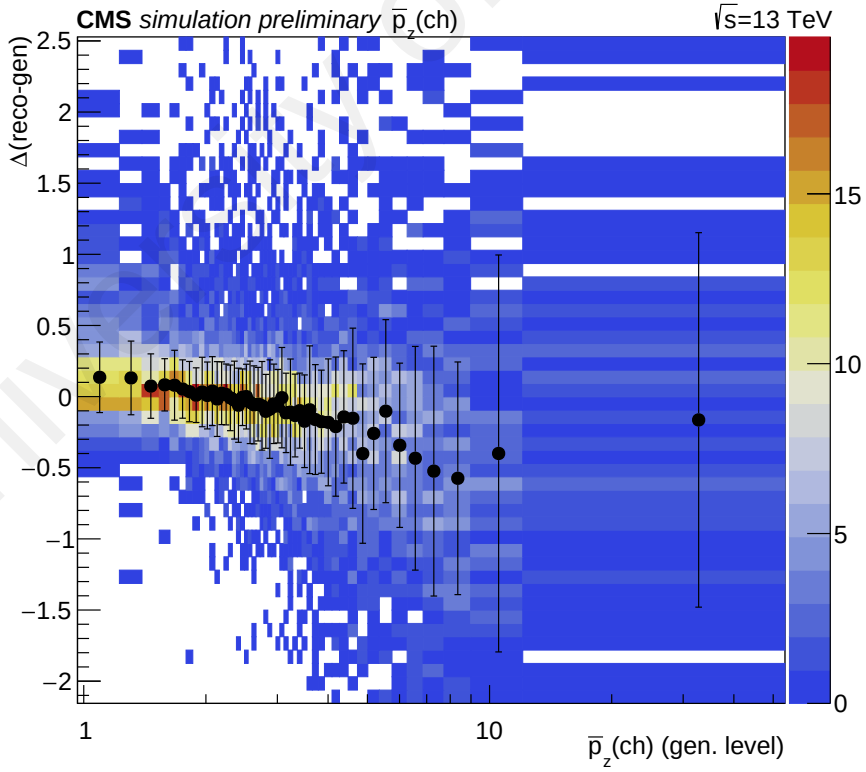


(b)

Figure 4.3: Resolutions for the p_T (a) and p_z (b) fluxes, as function of the generator level quantities



(a)



(b)

Figure 4.4: Resolutions for the average p_T (a) and p_z (b), as function of the generator level quantities

4.2 Monte Carlo Samples and Data

This analysis has been done using CMS Software Framework (CMSSW), series 8_0_26 and MINIOAD. Our code is publicised in Silva and Seidel (2017).

4.2.1 Data

For all the channel, the data used for this study consists of 2016B-H based on data taking periods for a total certified luminosity of 35.9 fb^{-1} collected with the CMS detector in 2016 at a centre-of-mass energy of 13 TeV. All the data are listed in the Table 4.3.

Table 4.3: Primary datasets used in this analysis. PD is an abbreviation for SingleMuon or SingleElectron

Primary dataset	Integrated Luminosity
/PD/Run2016B-23Sep2016-v3/MINIAOD	$35.9 fb^{-1}$
/PD/Run2016C-23Sep2016-v1/MINIAOD	
/PD/Run2016D-23Sep2016-v1/MINIAOD	
/PD/Run2016E-23Sep2016-v1/MINIAOD	
/PD/Run2016F-23Sep2016-v1/MINIAOD	
/PD/Run2016G-23Sep2016-v1/MINIAOD	
/PD/Run2016H-PromptReco-v2/MINIAOD	
/PD/Run2016H-PromptReco-v3/MINIAOD	

The cross sections used as reference are taken from LHCtopWG Collaboration (2015a) and GENPOG Collaboration (2015b) except for $t\bar{t}$ for which the generator cross section is quoted according to GENPOG Collaboration (2015a). At NNLO the expected $t\bar{t}$ cross section is 832^{+20}_{-29} (scale) $\pm$ 35 (PDF $+\alpha_s$) (LHCtopWG Collaboration, 2015b). In this version of the analysis, we used the NNLO reference to normalise all $t\bar{t}$ samples. Our measurement is independent of the total cross section as we will subtract the background contributions from data and compare normalised distributions. These subtracted data were compared to the events that were simulated in a few generators simulation. The simulated events are described in the following subsection.

4.2.2 Monte Carlo Simulation

The signal process was created using POWHEG v2 (Nason, 2004; Frixione et al., 2007; Alioli et al., 2010) with the top quark mass 172.5 GeV was used. Parton showering was simulated using PYTHIA 8.219 (Sjöstrand et al., 2015). The signal and the single top background simulated using the CUETP8M2 underlying event tune (CMS Collaboration, 2016b), whereas all the other processes used CUETP8M1 event tune. The $t\bar{t}$+W and the $t\bar{t}$+Z backgrounds were generated using MG5_aMC@NLOv2.2.2 (Alwall et al., 2014) while WZ, W+jets. The $ZZ(\rightarrow 2l2q)$ background processes were simulated using MG5_aMC@NLO with combining of several Next-to-Leading-Order + Parton Shower (NLO+PS) samples which differ by final-state multi-plicities ($FxFx$ merged) (Frederix & Frixione, 2012), and Drell-Yan background combines with several Leading-Order + Parton Shower (LO+PS) samples (MLM matching) (Alwall et al., 2008). WW, $ZZ(\rightarrow 2l2\nu)$, and tW backgrounds were generated using POWHEG v2 . The single top t-channel background was simulated using POWHEG v2+ madspin (Madspin Collaboration, 2017). The generated events were processed through the CMS detector simulation based on GEANT 4 and event reconstruction. The simulation samples are listed in the Table 4.4. In addition, other samples were used for systematic uncertainties or alternative signal simulation for comparison with the measurement are listed in the Table 4.5.

PDFs and QCD scale uncertainties used event weights by a factor of two or one half that represent the changes of factorization and renormalization scales. POWHEG v2 + PYTHIA 8 samples with varied parameters were used to estimate the parton-shower and UE uncertainties. This analysis used the samples where the parton-shower ISR or FSR scales were varied independently by factors of 2 or 0.5 with respect to their central values that correspond to $\alpha_s = 0.11108$. Additionally, the other tunes have been selected based on different colour reconnection models where the central values of the MPI and MPI-based

colour reconnection model were varied by their uncertainties in the CP0 tune. The other simulation where HERWIG++ with the EE5C tune was used instead of PYTHIA 8 or where MG5_aMC@NLO with $FxFx$ merging was employed instead of POWHEG have been used. The tunes were validated using RIVET framework to ensure that the selections and definitions of the objects at generator level are fully synchronised with those used in RIVET routines.

Table 4.4: Simulation samples are from the RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6 production. We quote the cross section used to normalise the sample in the analysis

Process	Dataset	$\sigma[pb]$
Signal		
$t\bar{t}$	TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	832
Background		
$t\bar{t}$ +W	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.20
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.41
$t\bar{t}$ +Z	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.53
	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.25
WZ	WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	5.26
WW	WWToLNuQQ_13TeV-powheg	50.0
	WWTo2L2Nu_13TeV-powheg	12.2
ZZ	ZZTo2L2Nu_13TeV_powheg_pythia8	0.564
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
W+jets	WToLNu_0J_13TeV-amcatnloFXFX-pythia8	51923
	WToLNu_1J_13TeV-amcatnloFXFX-pythia8	7342
	WToLNu_2J_13TeV-amcatnloFXFX-pythia8	2260
tW	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.9
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.9
t-ch	ST_t-channel_top_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powhegV2-madspin	44.3
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powhegV2-madspin	26.4
Drell-Yan	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	18610
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6025

Table 4.5: Simulation samples used for systematics, from the RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIVv6 production

Signal variation	Dataset
Parton shower scale	TT_TuneCUETP8M2T4_13TeV-powheg-isrup-pythia8
	TT_TuneCUETP8M2T4_13TeV-powheg-isrdown-pythia8
	TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8
	TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8
Underlying event	TT_TuneCUETP8M2T4up_13TeV-powheg-pythia8
	TT_TuneCUETP8M2T4down_13TeV-powheg-pythia8
ME-PS matching scale (hdamp)	TT_hdampUP_TuneCUETP8M2T4_13TeV-powheg-pythia8
	TT_hdampDOWN_TuneCUETP8M2T4_13TeV-powheg-pythia8
Colour reconnection	TT_TuneCUETP8M2T4_erdON_13TeV-powheg-pythia8
	TT_TuneCUETP8M2T4_QCDbasedCRTune_erdON_13TeV-powheg-pythia8
Top mass	TT_TuneCUETP8M2T4_mtop1665_13TeV-powheg-pythia8
	TT_TuneCUETP8M2T4_mtop1785_13TeV-powheg-pythia8
HERWIGpp	TT_TuneEE5C_13TeV-powheg-herwigpp

4.2.3 Corrections Applied to the Simulation

Different sets of corrections were applied to the simulated events are:

- **Pileup re-weighting**: All generated samples were passed through a full simulation of the CMS detector based on GEANT 4 and include additional minimum bias interactions superimposed in order to include the effect of in-time and out-of-time pileup in the events. At generator level, we reweighted the number of extra interactions to match that estimated in data. The generated pileup distribution was based on the configuration SimGeneral.MixingModule.mix_2016_25ns_Moriond17MC_PoissonOOTPU_cfi. A minimum bias cross section of 69 mb was used to estimate the pileup distribution following the recommendations from PVT Collaboration (2015). A 5 % uncertainty assigned to the minimum bias cross section assumed. Figures 4.5 and 4.6 show the distribution of data compared to expectations in the number of primary vertices reconstructed in the event and median energy density from FastJet (Cacciari et al., 2012), correspondingly. The top panels show the sum of the expected contributions from each process (histogram stack) compared to the

data (points). The bottom panel shows the data to expectation ratio. The dashed band represents the uncertainty associated with the statistics of the simulated samples and the integrated luminosity.

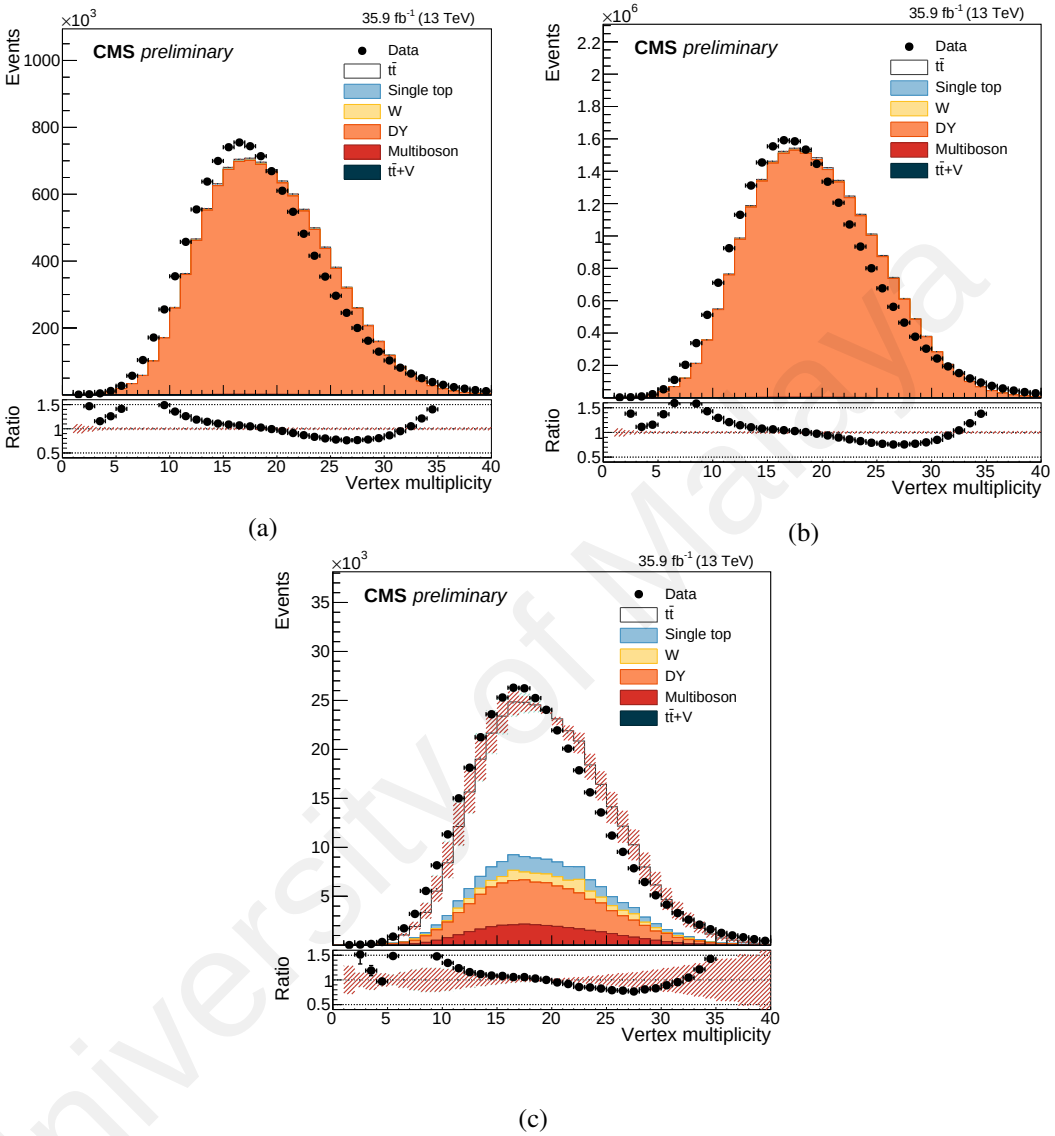


Figure 4.5: Distributions of the number of primary vertices in the *ee* (a) $\mu\mu$ (b) and $e\mu$ (c) channels

- **Trigger efficiency**: We corrected for the difference in performance in data and simulation of trigger paths used. A p_{T}, η-dependent scale factor was applied to the data. The values used for this correction were the ones proposed by the TOP Collaboration (2017).

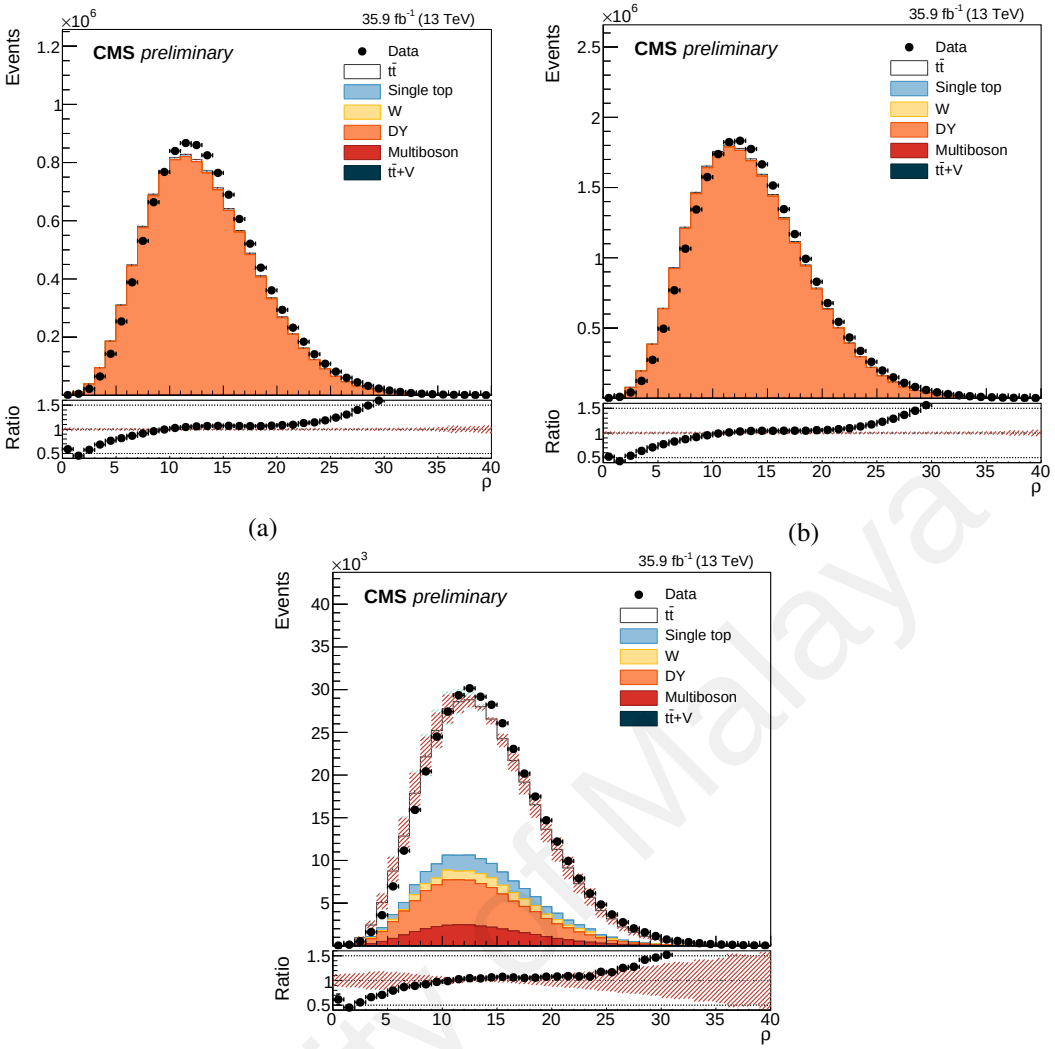


Figure 4.6: Median energy density computed from FastJet in the ee (a) $\mu\mu$ (b) and $e\mu$ (c) channels

- **Lepton id+isolation efficiency**: We corrected for the difference of the performance for the lepton id and isolation between data and simulation, by applying a p_{T}, η-dependent scale factor. We used of tight muons (Muon POG Collaboration, 2015a) and electrons (Egamma POG Collaboration, 2015a). The official values recommended by the Muon POG Collaboration (2015b); Egamma POG Collaboration (2015b) were therefore used in this correction. For muons, we included the tracking efficiency scale factor corrections provided by the muon POG.

- **Tracking efficiency**: The simulated tracking efficiency of the reconstructed charged particles was corrected with scale factors which depend on the reconstructed vertex multiplicity. The scale factors were run-dependent (BCDEF and GH are separated), see Figure 4.7: for each event, a random era is assigned based on the relative contribution to the total integrated luminosity. The scale factors were provided by the Muon POG Collaboration (2015a) and were expected to be consistent with the ones derived by the TRK POG using $D^0 \to K\pi\pi$ and $D^0 \to K\pi$ decays (Mariani, V. and CMS Collaboration, 2017). The correction is independent of the p_{T} of the charged PF candidates and, it was applied by removing the reconstructed tracks randomly in an event with a random probability sampled from a uniform PDFs. If the probability exceeds the scale factor, the track was removed from the analysis, if not it was kept. We assign a 100 % uncertainty on the correction applied.

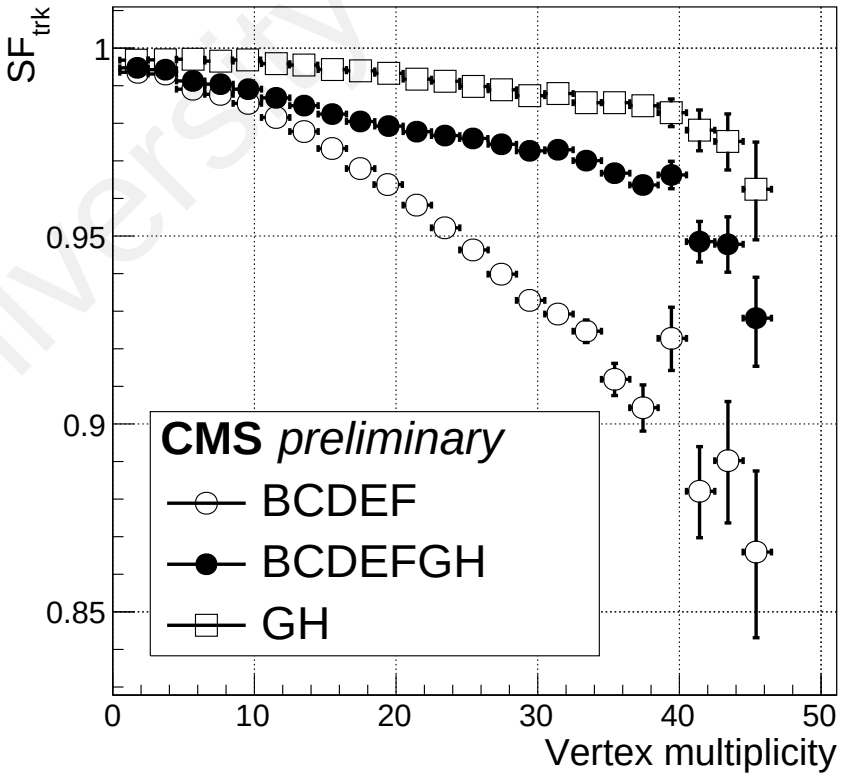


Figure 4.7: Tracking scale factors used in our analysis

- **Jet energy scale and resolution** : Jet energy scale corrections were applied using the so called Summer16_23Sep2016V4_Data,MC corrections (JetMET POG Collaboration, 2017b). In simulation the nominal jet energy resolution is smeared using a p_{T}, η-dependent parameterization (JetMET POG Collaboration, 2017a). In both cases alternative scenarios generated by shifting the corrections according to their uncertainties are considered and shall be discussed in detail.
- b **tagging efficiency** : The difference in performance of the b tagging algorithm used in the analysis is accounted for by applying a p_{T}-dependent scale factor. We used the medium working point of the Combined Secondary Vertex (v2) algorithm. The scale factors were used to correct a-posteriori the b tagging decision in the simulation (BTV POG Collaboration, 2015).
- **Generator level weights** : The simulated processes were scaled according to its effective integrated luminosity which can be computed from the sum of the per-event weights. The general formula applied to determine the number of events expected for a given process ($\hat{N}$), is:

$$\hat{N} = \mathcal{L} \cdot \sigma \cdot \frac{\sum_{i=1}^{N_{sel}} w_i}{\sum_{i=1}^{N_{gen}} w_i} \quad , \tag{4.3}$$

where $\mathcal{L}$ is the integrated luminosity, σ is a reference theory prediction for the inclusive cross section and w_i are the per-event generator level weights.

4.3 Event Selection

This thesis designed the underlying event analysis using high-purity dilepton channel. It means the kinematics of the reconstructed charged particles was analysed after subtracting the estimated contribution from the leptons and b-jets which are the $t\bar{t}$ decay products. First, the analysis used the particle flow (PF), see Chapter 3.3, plays a role to recontruct

the final state objects as important as the particle identification to determine the particle energy and direction. The selection was based on a decay topology of $t\bar{t}$ where the W boson in the event decay into charged leptons and neutrinos. We assumed that top quarks decay to a W boson and a b quark. The W bosons from the top and the antitop quarks both decay to $l^+\nu$ or $l^-\nu$. So, the final state of $t\bar{t}$ reconstruction consists of two charged leptons, missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) from the undetected neutrinos, and two jets from the fragmentation of the b quarks. Promptly charged leptons were dressed with nearby photons using anti-K_{T} algorithms with the R = 0.1 cone. For each event, jets were clustered using the particle-flow with the anti-K_{T} algorithm (Cacciari et al., 2008).

To maximise the efficiency, single lepton and dilepton triggers are used when collecting the data online. The triggers used are listed in Table 4.6. In the offline selection, for one lepton requires at least $p_T > 25$ GeV and another with $p_T > 20$ GeV both having $|\eta| < 2.5$ while for two leptons need have opposite sign with the invariant mass $M_{ll} > 12$ GeV. For more than two reconstructed leptons, the dilepton candidate built from the highest p_{T} leptons in the event. For e^+e^- and $\mu^+\mu^-$ events, the Z pole mass region is vetoed with $|M_{ll} - 91| > 15$ GeV.

Two additional b-jets required $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.5$ in the event with two of them at least were to be b-tagged By the Combined Secondary Vertex algorithm (CSVv2). Events are then categorised according to the flavour of the dilepton candidate (ee, eμ, $\mu\mu$).

The offline selection of the event in the simulation was mimicked at particle level using the PseudoTopProducer tool (Goh & Seidel, 2017; CMS Collaboration, 2017b). The preferred tool is based on the RIVET framework. This procedure was therefore expected to ensure that the selections and definitions of the objects at generator level are fully synchronized with those used in RIVET routines. Charged leptons and jets were reconstructed using the anti-K_{T} algorithm with R = 0.1 and R = 0.4 cone respectively. Jets

were clustered after removing the leptons from the hard-process as well as neutrinos and included B-hadrons in the clustering could identify the flavour of the jets at the particle level. The neutrinos candidate in the event were estimated from the $E_{\mathrm{T}}^{\mathrm{miss}}$.

Table 4.6: Trigger paths used for online selection in the analysis

Final state	Path	Run range
eμ	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	up to 280385
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	278273-280385
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	>280385
$\mu\mu$	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v	all
ee	HLT_DoubleEle24_22_eta2p1_WPLoose_Gsf_v HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	all
ee or eμ	HLT_Ele32_eta2p1_WPTight_Gsf_v	all
$\mu\mu$ or eμ	HLT_IsoMu24_v HLT_IsoTkMu24_v	all all

Table 4.7 summarizes the event yields observed in data and expected in the different categories used in the analysis. We observed a good agreement for all categories within 1-2 % which is well within the expected uncertainty from the integrated luminosity.

Table 4.7: Expected and observed event yields in the different categories used in the analysis. The uncertainties quoted are of statistical nature

Process	Final state		
	ee	$\mu\mu$	$e\mu$
$t\bar{t}$	$[14.23 \pm 0.05] \times 10^3$	$[27.80 \pm 0.08] \times 10^3$	$[49.85 \pm 0.10] \times 10^3$
Single top	437 ± 9	816 ± 13	1496 ± 17
W	-9 ± 6	0 ± 0	-10 ± 6
DY	$[0.45 \pm 0.15] \times 10^3$	$[1.04 \pm 0.25] \times 10^3$	45 ± 10
Multiboson	14.5 ± 1.0	28.0 ± 1.3	16.7 ± 1.7
$t\bar{t}$ +V	45.8 ± 1.6	81.3 ± 2.3	125.0 ± 2.9
Total	$[15.17 \pm 0.16] \times 10^3$	$[29.77 \pm 0.26] \times 10^3$	$[51.52 \pm 0.10] \times 10^3$
Data	15269	29769	52645

Inclusive distributions of the charged particle multiplicity, the total momentum flux in the transverse plane and longitudinal direction, and the average p_{T} and p_z per particle are shown in Figures 4.8, 4.9, and 4.10 correspondingly. The selected particles are expected to have $p_{\mathrm{T}}>$ 0.9 GeV and $|\eta| < 2.5$. Other requirements were made regarding the association to the primary vertex of the event. In addition, in these distributions we have not included the contributions from particles matched to the two lepton and the two b-jet candidates. In data we observed less charged particles than expected and consequently the total fluxes in the transverse plane and along z are also softer in data, with respect to the MC predictions. In the computation of the average momentum, the trends tend to cancel out and good agreement with the CUETP8M2T4 tune for the average p_{T} and p_z of the charged particles have been obtained. At very low p_z we observed some deficit in data with respect to the simulated predictions.

4.3.1 Background Determination

All backgrounds are determined from simulation, after the corrections mentioned in Section 4.2.3. The only exception is the Drell-Yan (DY) background which is expected to be the dominant one in the dilepton channel.

The DY background was estimated making use of the Z pole region. The ratio of same-flavour dileptons (ll) reconstructed outside or inside a mass window [76,106] GeV ($R_{ll}(\mathrm{out/in})$) was estimated from simulation and used to extrapolate the DY contamination in the signal region, after multiplying by the observed number of events in the mass window. Contamination from non-DY backgrounds can still be presented in the Z mass window and were subtracted using a similar region constructed in the (eμ) channel. The number of events outside the Z mass window can then be computed as

$$N_{ll}(\mathrm{out}) = R_{ll}(\mathrm{out/in}) \cdot [N_{ll}(\mathrm{in}) - \frac{1}{2} N_{e\mu}(\mathrm{in}) k_{ll'}], \tag{4.4}$$

where the scaling factor $R_{ll}(\text{out/in})$ was estimated from a DY MC sample as

$$R_{ll}(\text{out/in}) = \frac{N_{\text{DY MC}}(\text{out})}{N_{\text{DY MC}}(\text{in})} \,. \tag{4.5}$$

Furthermore the $k_{ll'}$ correction factor used in Equation 4.4 was applied to take into account the differences between electron and muon reconstruction. It was calculated using the events in the Z peak region passing the standard dilepton selection, and can be expressed as:

$$k_{ll'} = \sqrt{\frac{N_{ll}(\text{in})}{N^{l'l'}(\text{in})}} \tag{4.6}$$

with l = e and $l' = \mu$ or vice-versa.

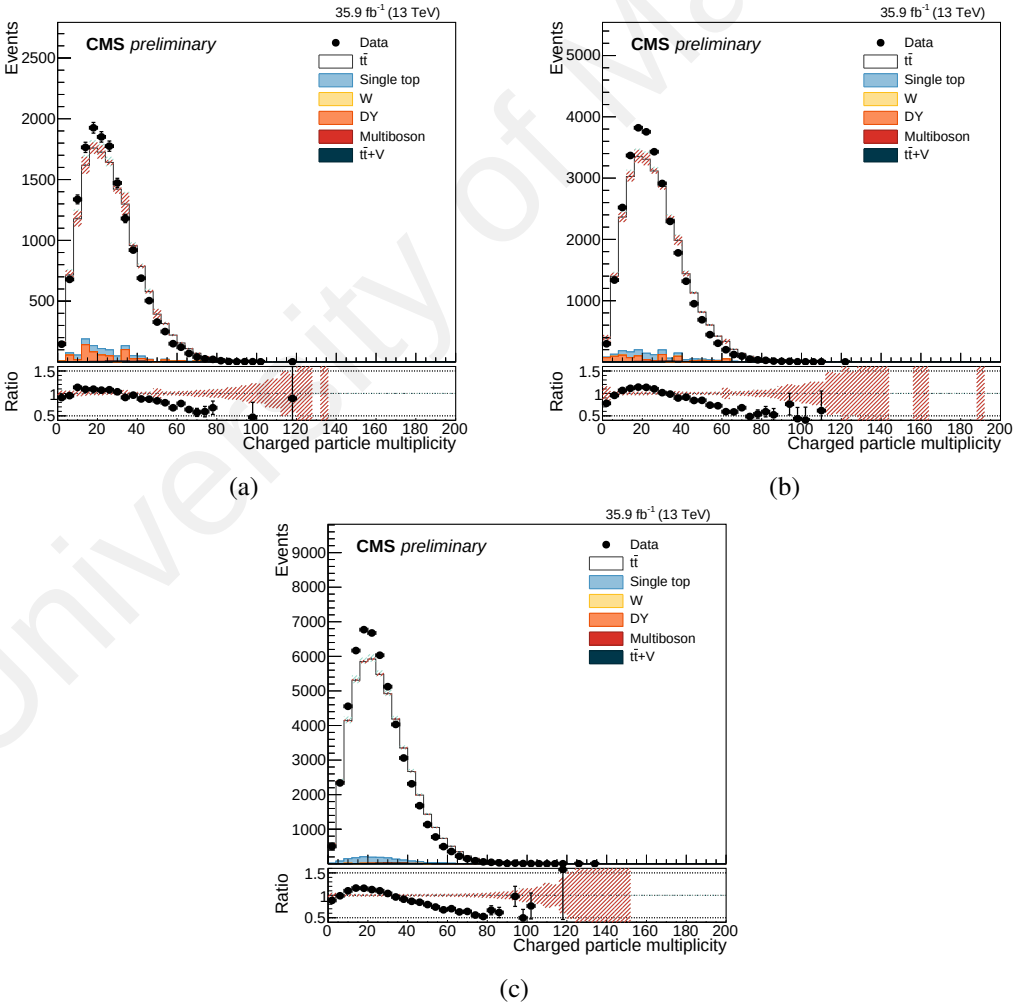
The ratio of the yields predicted by Equation 4.4 to the expectations defines the scale factor to be applied to normalize the DY MC in the ee and $\mu\mu$ channels. A combined Scale Factor (SF) was applied to the eμ channel, defined as:

$$SF_{e\mu} = \sqrt{SF_{ee} \cdot SF_{\mu\mu}}. \tag{4.7}$$

The computations described above are carried out after performing the full event selection. Table 4.8 summarizes the results. Overall we obtain factors which scale the simulated expectations by $O(30\%)$. If the exercise is repeated in the 0-tag (1-tag) regions, the scale factors are 1.02 (1.14) for the ee and 1.00 (1.11) for the $\mu\mu$ channel, indicating that there is a strong dependency of this correction factor on jet multiplicity and the heavy flavour content of the event. An overall 30 % uncertainty (i.e. on the magnitude of the scale factor itself) is assigned to the DY normalisation. The DY estimation is summarised in the Table 4.8

Table 4.8: DY **estimation for the categories of events used in our analysis. Uncertainties are of statistical nature. The last two rows report the data/MC scale factors for control regions in events with no b-tagged jet and one b-tagged jet**

	ee	$\mu\mu$	$e\mu$
$R_{ll}(out/in)$	$[70.14 \pm 0.13] \times 10^{-3}$	$[90.90 \pm 0.09] \times 10^{-3}$	
$k_{ll'}$	0.711 ± 0.004	1.407 ± 0.009	
$N_{ll}(in)$	10297	20370	
N_{out}	352 ± 8	902 ± 16	
SF (Data/MC) $\geq 2b$	1.327 ± 0.030	1.340 ± 0.024	1.333 ± 0.019
SF (Data/MC) $=0b$	1.024	0.996	1.010
SF (Data/MC) $=1b$	1.138	1.106	1.122



(a) (b)

(c)

Figure 4.8: Inclusive charge multiplicity distributions. The distributions are shown for the ee (a) $\mu\mu$ (b) and $e\mu$ (c) channels

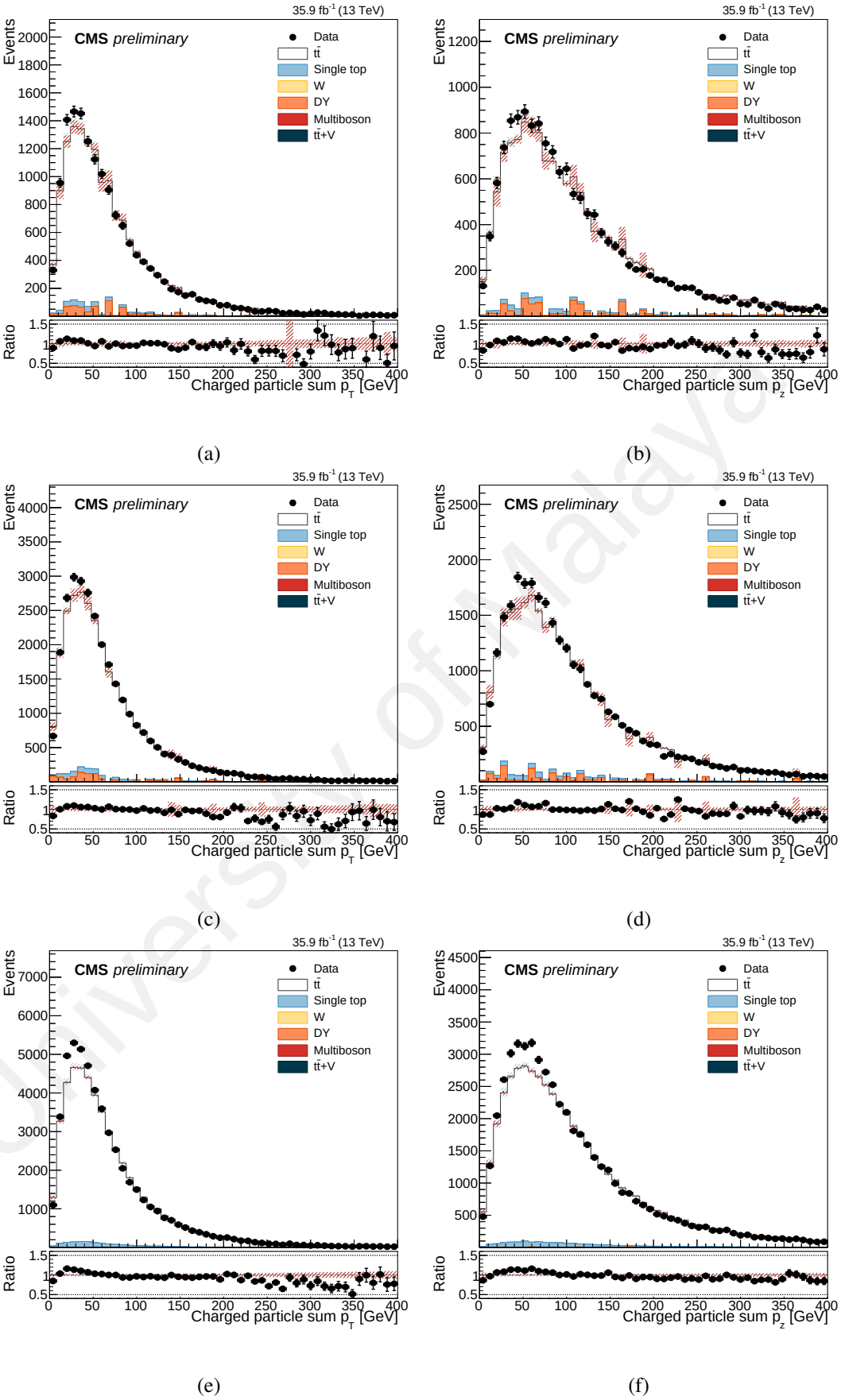


Figure 4.9: Inclusive scalar sum (flux) of the p_T (left) p_z (right) of the charged particles. The distributions are shown for the ee (a) $\mu\mu$ (b) and $e\mu$ (c) channels

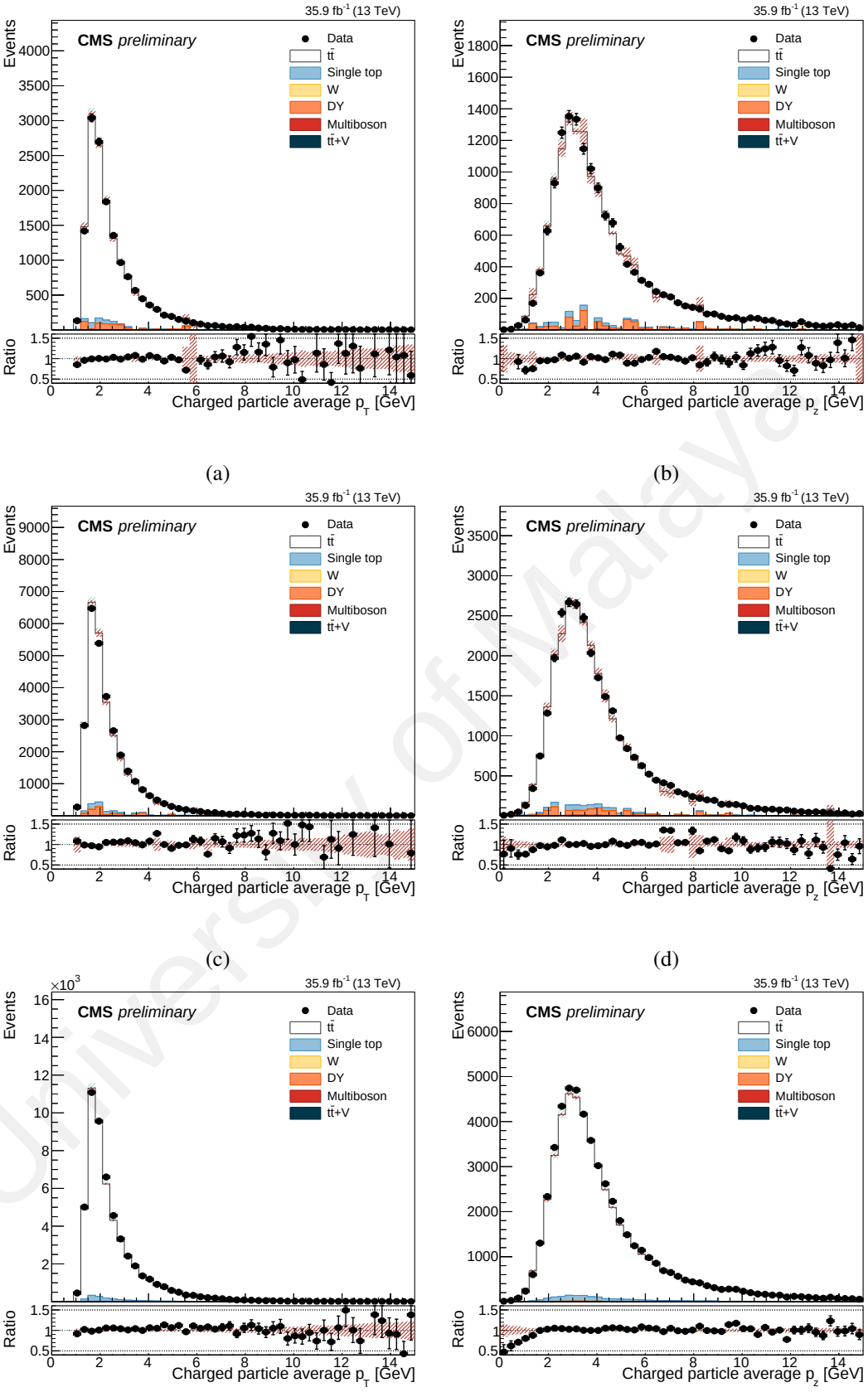


Figure 4.10: Inclusive average p_T (left) and p_z (right) per charged particle. The distributions are shown for the ee (a) $\mu\mu$ (b) and $e\mu$ (c) channels

Figure 4.11 shows the distributions of the invariant mass of the dilepton pairs after applying the scale factors derived from data. The fair agreement is observed between data and expectations.

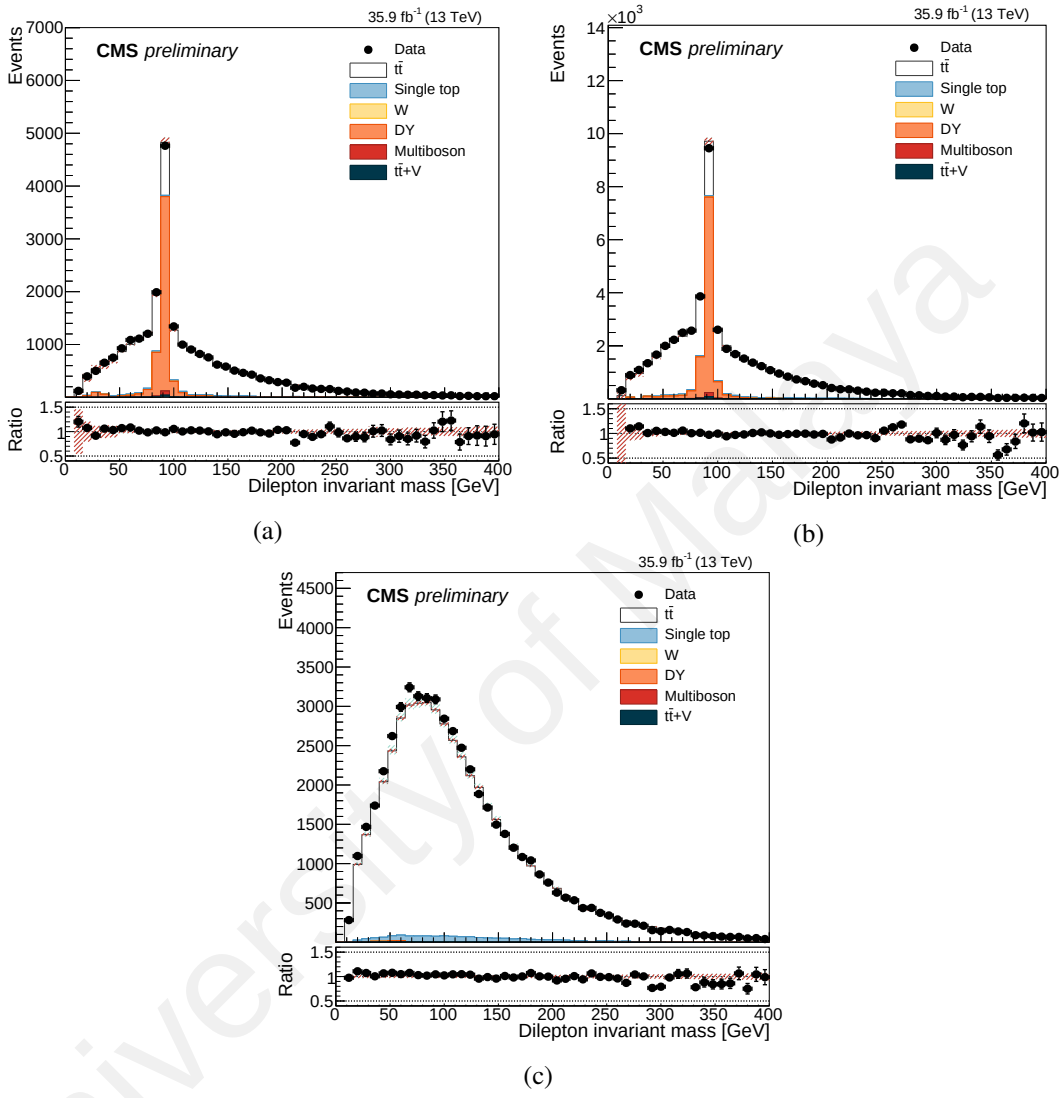


Figure 4.11: Dilepton invariant mass in the ee (a) $\mu\mu$ (b) and $e\mu$ (c) channels

4.4 Unfolding Method

The complex configuration of the CMS detector causes finite acceptance of the particle detection. The particles that interact with the detector potentially losing energy, changing direction or creating the additional particles. This condition may differ measurement substantially from the actual happens during a collision with what is accepted in the detector.

A finite resolution of the detector leads to bin-to-bin migrations, which means that the number of events that are counted in the reconstructed bin will be different from that are counted in the generated bin. Some of the measured data that were generated in the certain bin could migrate to neighbouring bins. In order to correct the effects of the detector acceptance and the migration bin, the detector and its response were modelled by simulation and applied to the measured data using the unfolding technique. By implementing the unfolding procedure, we correct the detector effect from the measured data to get the "true" value of the observables that can be compared to the theoretical prediction. This unfolding problem may be written as

$$y = K\lambda + b, \tag{4.8}$$

where the λ is the particle level expectation (true distribution), the K is a matrix of probabilities describing the migrations from any certain bin to any other neighbour bins on detector level. y is the average expected event count at detector level whereas b is the background.

In order to get the true distribution (λ), we can easily solve it by inverting the Equation 4.8,

$$\lambda = K^{-1}(y - b), \tag{4.9}$$

but this Equation 4.9 causes statistical problem with showing large fluctuations in the distribution (Britzger, 2013). To solve this statistical problem, we applied the regularised unfolding method that is implemented in TUnfoldDensity package (Schmitt, 2016, 2012) which used to implement the Tikhonov regularisation (Tikhonov, 1963) to inverse the migration scheme (Brobel, 2002). We got the true generated distribution by using a least

square minimization approach which have two terms

$$\chi^2 = \chi^2_A + \chi^2_L, \tag{4.10}$$

for

- the χ^2_A term is $(y - K\lambda)^T V_{yy}^{-1}(y - K\lambda)$, where V_{yy}^{-1} is the covariance matrix of y.
- the regularisation χ^2_L term is $\tau^2|L|\lambda - \lambda_0||^2$, where the $|L|\lambda - \lambda_0||^2$ term is a penalty function with the λ_0 being estimated from the simulated samples and the τ is the regularisation parameter.

This regularised unfolding needs some important recipes to get the expected true distributions. The first two keys are the migration matrix as a representative for the matrix K and the binning choice. The matrix K should be capable of describing all the detector effects accurately that is produced from Monte Carlo and detector simulation. A migration matrix is defined through the included observables for the description of the migrations, by mapping their binning on generator and detector level, etc. Unfolding applies this scheme to be used in order to recover the true values of observables from the detector effect contribution. The migration matrix could include as many observables as possible with very fine binning on generated and reconstructed level. The migration matrix for each variable showed in the Figure 4.12 while the purity and the stability value in every bin showed in Figure 4.13.

The binning choice also plays an important role as it must be compatible with the expected resolution from a simulation. Also one must ensure that the bin yields enough counts in each bin. In this analysis, for each bin, we computed the expected resolution by comparing the reconstructed value with the one expected at generator level. The bins merged such that each one corresponds to $\approx 2\sigma$, i.e. two times the estimated resolution.

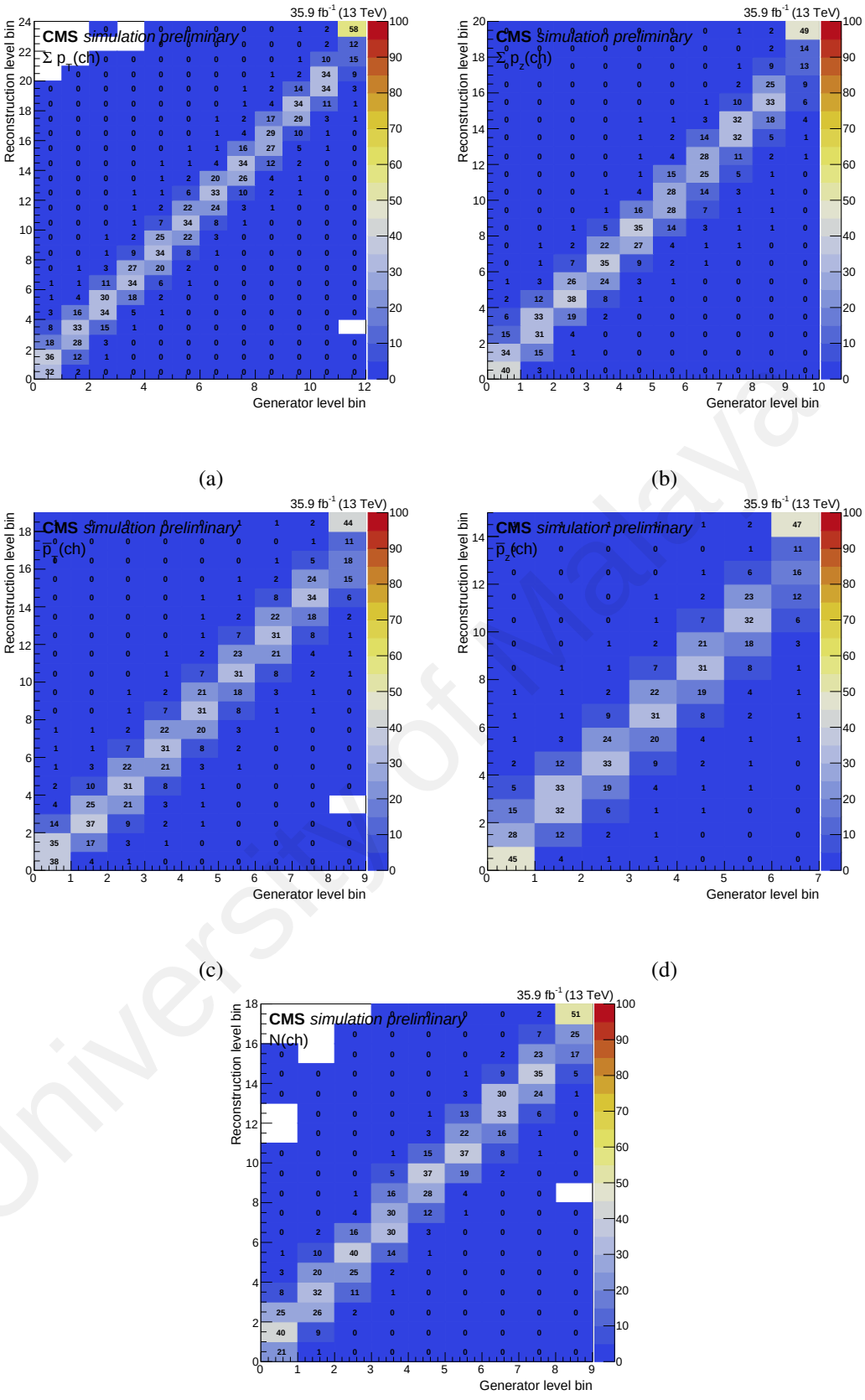


Figure 4.12: Migration matrices in the variables correspond to p_T flux (a), p_z flux (b), average p_T (c), average p_z (d), and charged multiplicity (e)

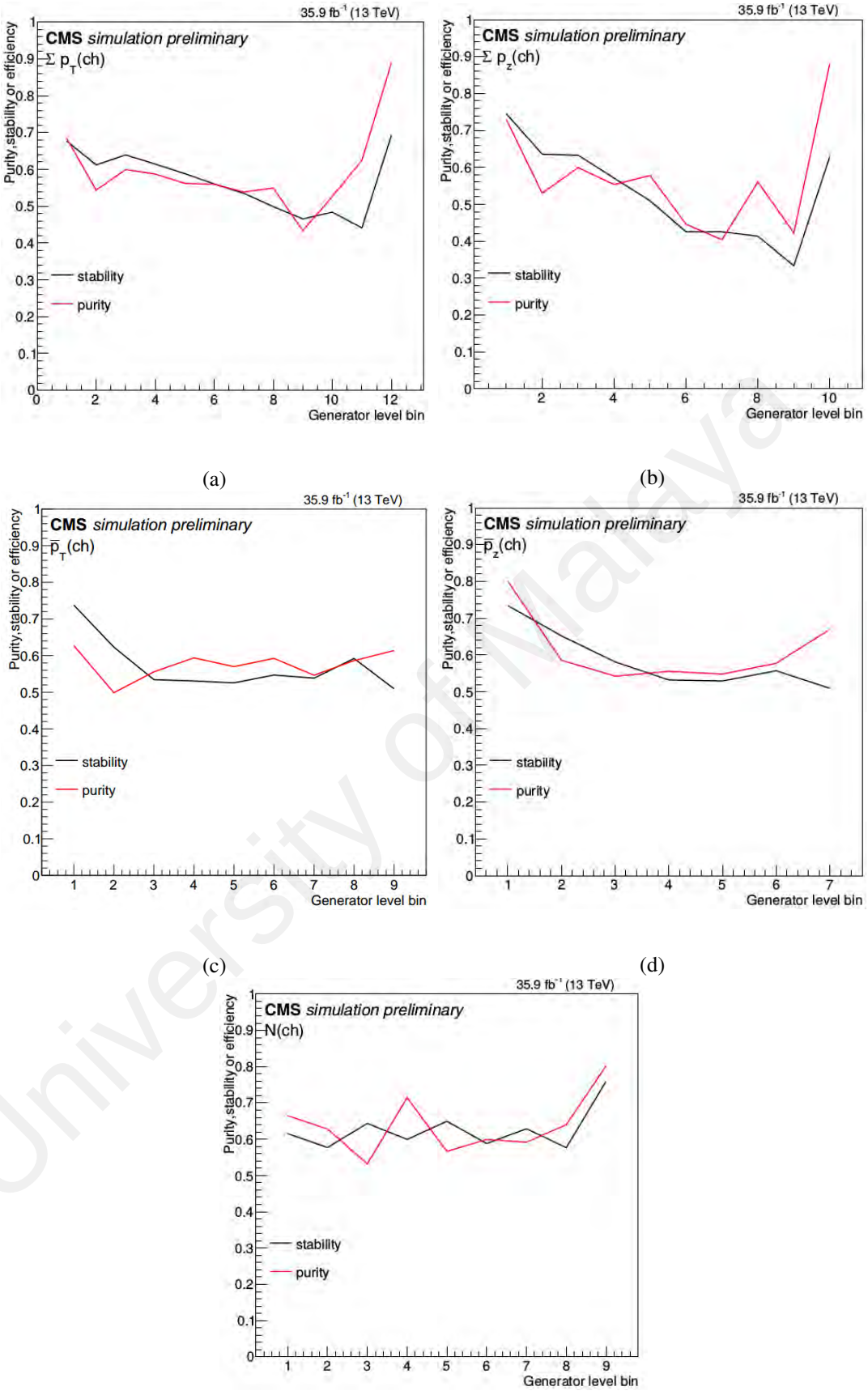


Figure 4.13: Purity and stability curves in the variables correspond to p_T flux (a), p_z flux (b), average p_T (c), average p_z (d), and charged multiplicity (e)

What it means, in every bin is considered two times finer grain reconstruction level bins. This procedure ensures a uniform distribution of the statistics.

The other keys that are not less important than the first two are the regularisation strength and the regularisation matrix. This regularisation term is used to damp statistical fluctuations in the reconstructed level with the $\tau > 1$ regulated how strong the penalty term conditions to the minimisation of the equation 4.10. The regularisation matrix L can also introduce smoothness conditions between bins of the type of an n-th discrete derivative. In our setup, we chose the function L to be the second derivative (TUnfold::rRegModeCurvature option). However, the regularisation parameter τ was undefined and needs to be optimised depending on the distribution being studied.

Several methods are implemented in order to optimise the regulation parameter τ, which is a priori an unknown parameter. In our analysis, the τ was tested by performing a singular value decomposition (SVD) and by computing the condition number (cond(K))= $\frac{\sigma_{max}}{min(0,\sigma_{min})}$ (Wikipedia, 2016) of each migration matrix. The condition number gives a bound on how inaccurate the solution of the inversion will be after numerical approximation. The smaller the number condition, the more accurate results that we get. Because the larger value of this number potentiates the enhancement of statistical fluctuation. Table 4.9 summarises the values obtained for the condition numbers of the migration matrices. The (cond(K)) values are in general > 40 and typically not exceeding three orders of magnitude. These values indicate that a mild regularisation of the inversion needs to be made.

One of the methods that was proposed in the reference Schmitt (2012, 2016) to choose the regularisation strength is to minimise global correlation coefficients that present in Tunfold package. The global correlation the i-th component y that given the covariance

matrix V_{yy} is defined as

$$\rho_i = \sqrt{1 - \frac{1}{(V_{yy}^{-1})_{ii}(V_{yy})_{ii}}}. \tag{4.11}$$

The regularisation strength τ^2 was chosen when the average global correlation $\Sigma_i \rho_i / m$, where m is the dimension of y. We performed a scan for this regularisation parameter to find the optimal value in the range $10^{-4} < \tau < 10^{-1}$. The results in the Table 4.9 also shows that the τ in general found in $< 10^{-3}$. The consistent values are showed also by minimising the average square of the correlation coefficient. The variation of the average correlations are from $40 - 70$ % depending on the distribution.

Table 4.9: Condition numbers of the migration matrices of the different observables. The second and third column show the maximum and minimum singular values correspondingly, after an SVD decomposition of the matrix. The fourth column shows the condition number obtained

Observable	$\sigma_{\mathbf{max}}$	$\sigma_{\mathbf{min}}$	**cond(K)**	$\boldsymbol{\rho}$	$\boldsymbol{\tau}$	$\log_{\mathbf{10}}\boldsymbol{\tau}$
N_{ch}	0.000973	$2.627 \cdot 10^{-5}$	37.0	0.471	0.0010	10^{-3}
$\sum p_T$	0.000828	$9.198 \cdot 10^{-6}$	90.0	0.432	0.0011	$\sim 10^{-3}$
$\sum p_z$	0.000901	$6.429 \cdot 10^{-6}$	140.1	0.472	0.0010	10^{-3}
$< p_T >$	0.001097	$6.070 \cdot 10^{-6}$	180.9	0.467	0.0010	10^{-3}
$< p_z >$	0.001073	$1.850 \cdot 10^{-6}$	579.7	0.478	0.0009	$\sim 10^{-3}$

4.4.1 Unfolding Test

As mentioned in the previous section, the regularisation parameter is an undefined parameter that needs to be optimised. The optimised τ values performed using minimising the average global correlation that was summarised in Table 4.9. These τ values should be checked to validate the unfolding procedure that we applied is correct or not. The test started with folding back the unfolded data and comparing to the reconstructed data. This test can explain the modification of the original distribution because of the unfolding procedure. The results show that the folded and original data distributions agree with each other within 1-5%. This agreement is better than the total systematic uncertainty which

will be explained in the next section. The results in the Figure 4.14 show that some bins differ 5 – 10 % and these are typically populated with fewer statistics.

The second test consists of evaluating the statistical coverage of the unfolding procedure. For this purpose, we constructed ensembles from subsets of generated $t\bar{t}$ MC events that treated as pseudo-data and applied the unfolding method to each ensemble. The unfolded distribution was compared to the truth generator-level distribution in the ensemble, and the differences, as well as the pull (difference/uncertainty), were evaluated for each bin. In general, we found that the results were compatible with no bias and the pull was close to 1 in most bins indicating that the statistical coverage of the uncertainty is expected to be correct (see the Figure 4.15). The global pull distribution was also consistent with the bin-by-bin observation. The distribution of the pulls in the individual bins and the inclusive distribution of the pulls is shown in the Figures 4.15 and 4.16 correspondingly. The points in the centre plot correspond to the mean and width obtained from a Gaussian fit to the pull distribution in each bin.

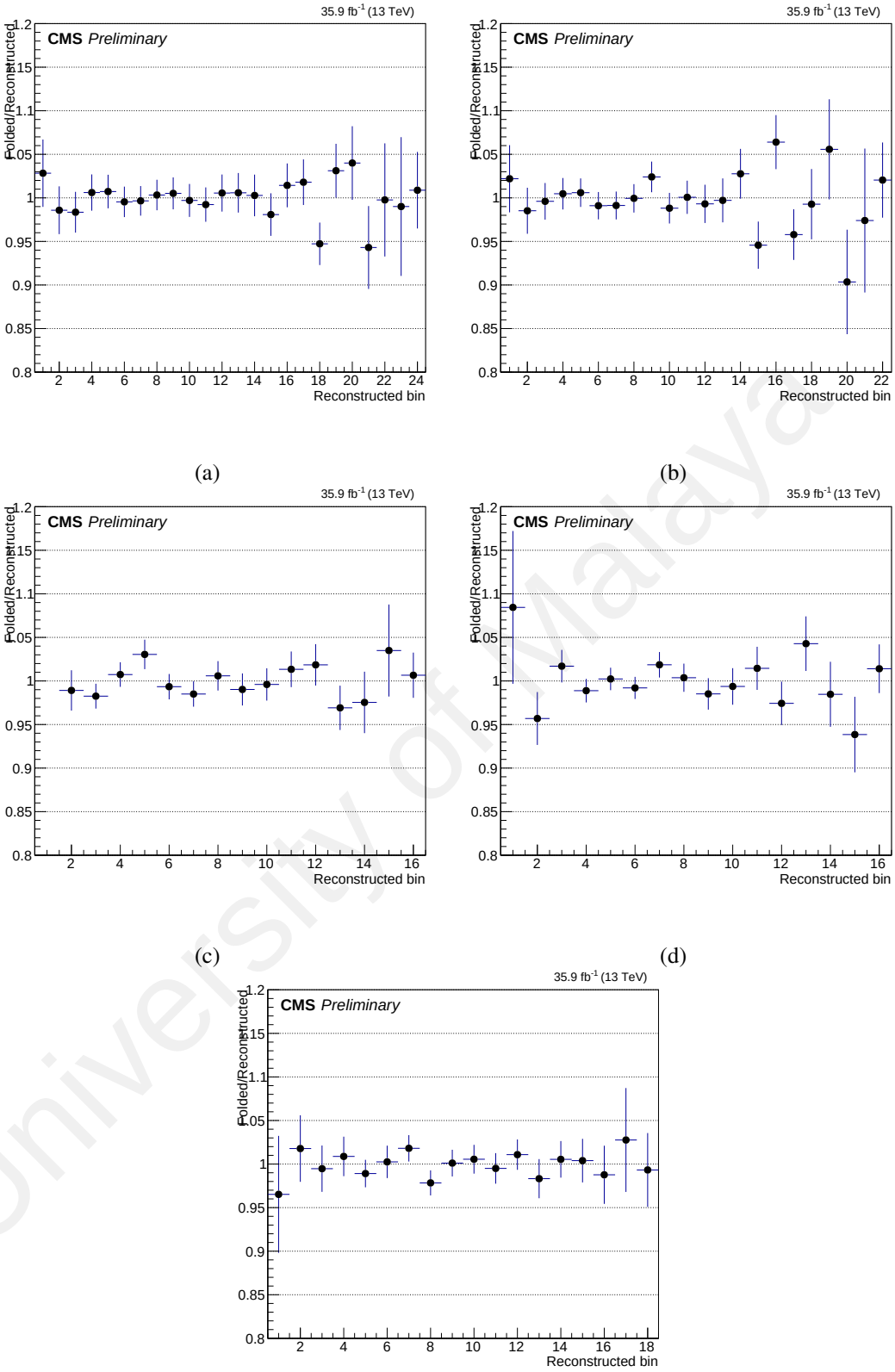


Figure 4.14: The result distributions of the comparison of the folded data to the reconstructed data for the variables correspond to p_T flux (a), p_z flux (b), average p_T (c), average p_z (d), and charged multiplicity (e)

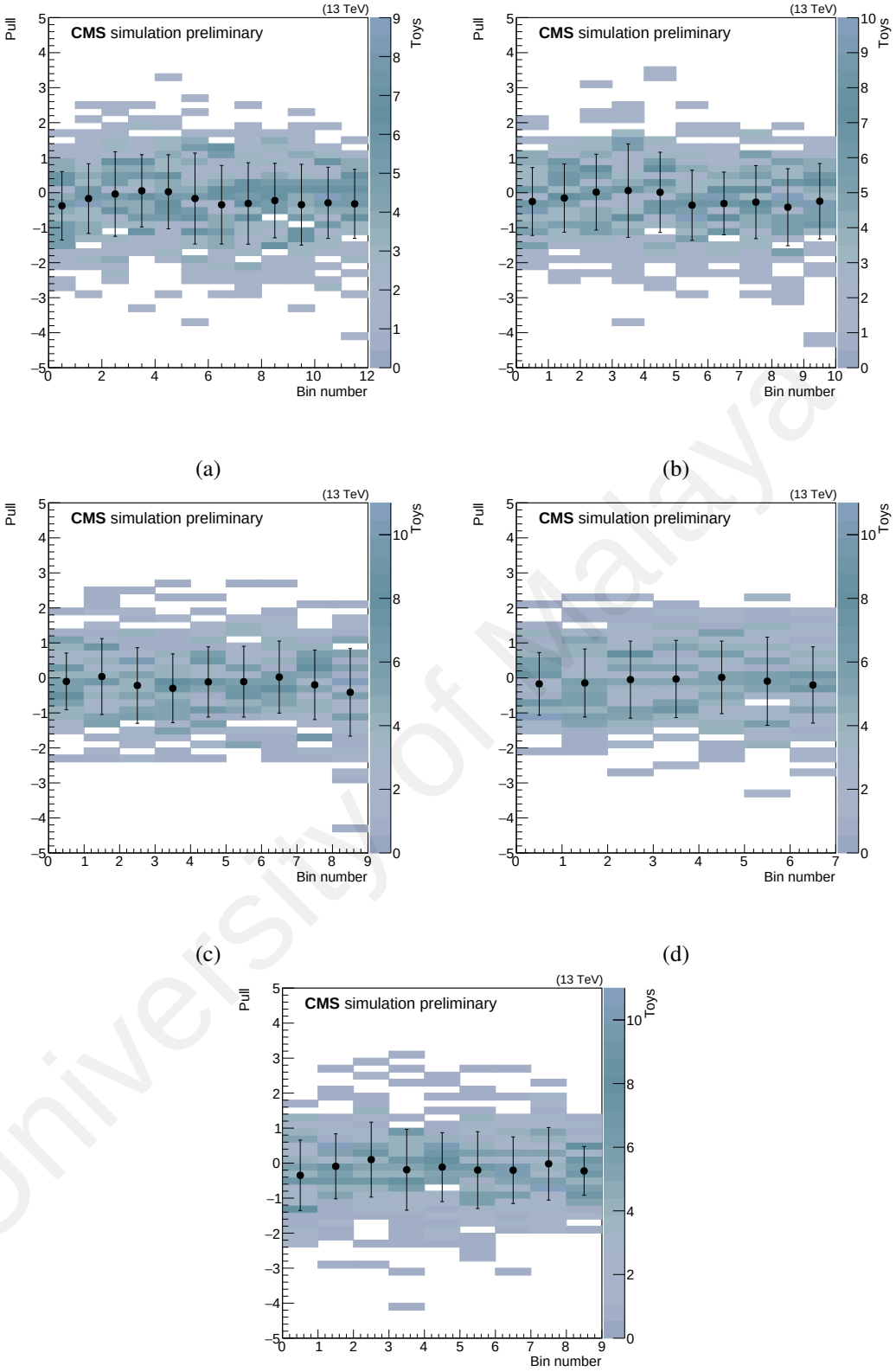


Figure 4.15: The result distributions of the pulls in the individual bins for the variables correspond to p_T flux (a), p_z flux (b), average p_T (c), average p_z (d), and charged multiplicity (e)

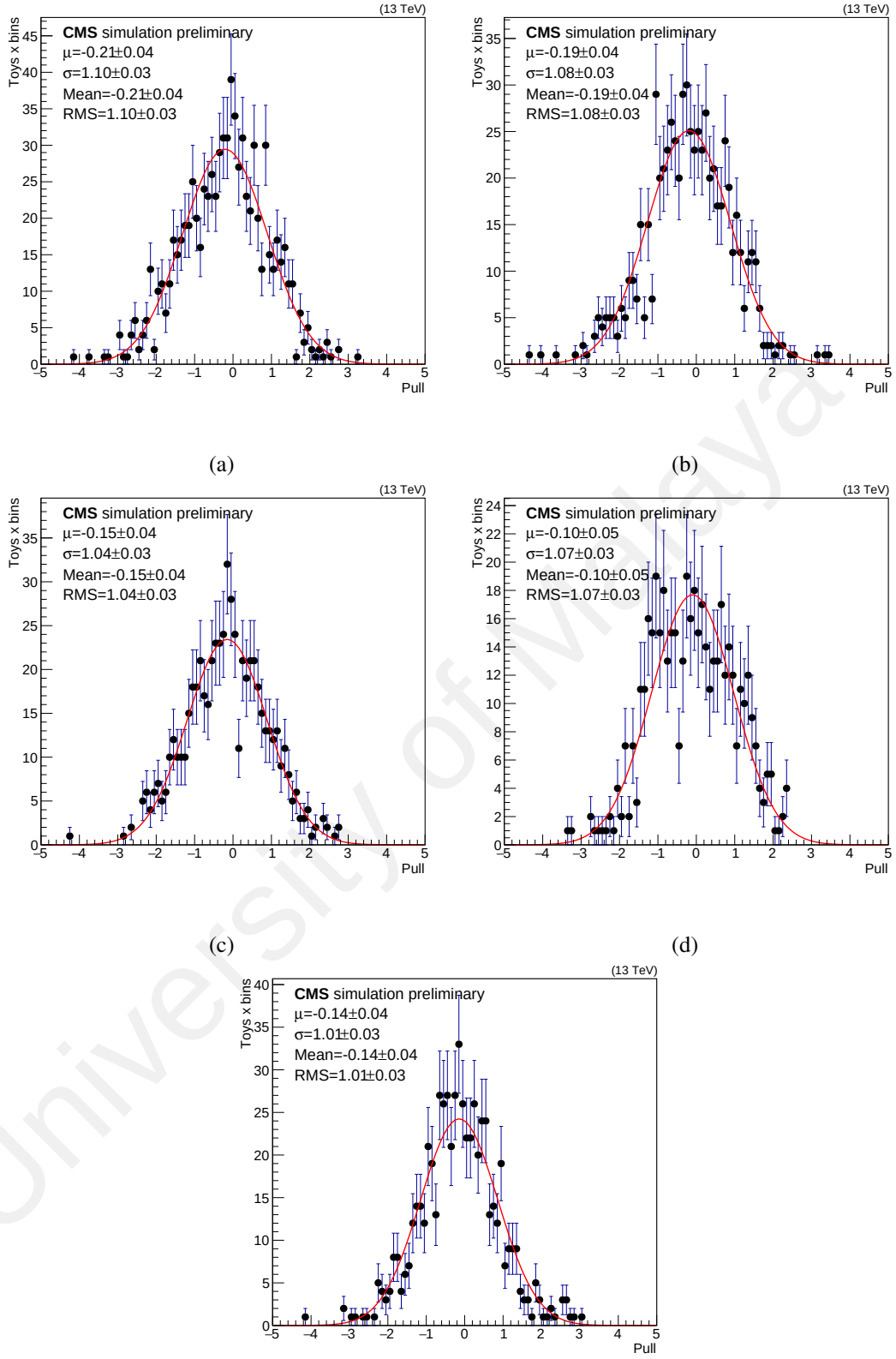


Figure 4.16: The result distributions of the inclusive distribution of the pulls for the variables correspond to p_T flux (a), p_z flux (b), average p_T (c), average p_z (d), and charged multiplicity (e)

4.5 Systematic Uncertainties

4.5.1 Experimental Uncertainties

The following experimental uncertainties considered in this analysis are:

- **Pileup**: Although pileup wss included in the simulation, there is an intrinsic uncertainty in modeling it appropriately. To estimate the effect of miss modelling the pileup, we varied the average pileup scenario, through the choice of the minimum bias cross section parameter, by 5 % with respect to its initial estimate.
- **Trigger and selection efficiency**: The uncertainty on the trigger efficiency and the lepton identification and isolation efficiency scale factors were propagated by re-weighting the simulation after shifting the nominal values up or down. The uncertainty on the muon tracker efficiency was included in this category and added in quadrature, although its effect was expected to be negligible. The impact on the rate is fully absorbed by normalising the distributions in the end, and only the impact on the shape (by weighting more/less some events) is relevant in this analysis.
- **Lepton energy scale**: Given the muon scale has been corrected using the Kalman fit method we considered the corresponding uncertainties. For electrons we have used the uncertainty after applying the electron energy corrections and smearing. The main effect of this systematic is related to the migration of events and to the uncertainty in some of the "slicing" variables, most notably $p_{\mathrm{T}}(l^+,l^-)$.
- **Jet energy scale**: A p_{T}, η-dependent parameterisation of the jet energy scale is used to vary the calibration of the jets in the simulation. We considered the full jet energy scale uncertainty sources in our analysis. When modifying the jet energy scale, the $E_{\mathrm{T}}^{\mathrm{miss}}$ estimate was updated. The main effect of this systematic is related to the exclusion/inclusion of events with jets near the offline thresholds.

- **Jet energy resolution**: Each jet is further smeared up or down depending on its p_{T} and η with respect to the central value measured in data. The main effect of this systematic is related to the exclusion/inclusion of events with jets near the offline thresholds.
- **b-tagging and misidentification efficiencies**: The nominal efficiency expected in the simulation was corrected by the p_{T}-dependent scale factors provided. Depending on the flavour of each jet, the b-tagging decision was updated according to the scale factor measured. The scale factor was also varied according to its uncertainty. The main effect of this systematic was the demotion/promotion of candidate b-jets and thus a migration of events used for the analysis.
- **Drell-Yan normalisation**: The simulated Drell-Yan background events were rescaled by the expected yield calculated from the Z pole mass region, from 76 to 106 GeV. Other backgrounds were taken from the simulation. In addition, the $t\bar{t}$ events that pass the reconstruction selection but are observed to fail the selection at generator level (fakes) are also subtracted. Given that in this analysis we were looking for the variables which are not necessarily expected to be reproduced in shape by MC, although we subtracted the expected background contribution, we should be careful in the claim that we have subtracted an unbiased estimate of the background and that no bias was introduced in the measurement. Rate effects were induced at the level of 8-10 % owing to the high purity of the selected sample. Shape effects were, also, expected to occur, in particular in the regions where the default UE tune used disagrees more with the data. As such we assigned in this analysis the total background uncertainty as for the difference between subtracting or not the background+fake $t\bar{t}$ contributions from the data, before unfolding. These were observed to be up to 5 % in some variables.

- **Tracking efficiency**: We assigned a variation of the tracking efficiency scale factor to evaluate the difference in the tracking efficiency in data and simulation. With this prescription we expected to cover also differences in performance in the different data taking eras. The relative change induced in the charge multiplicity distribution was expected to be of the same order of that observed in the data. Despite the change being single-sided, we shall mirror this uncertainty for the final result.

4.5.2 Theoretical Uncertainties

The following theoretical uncertainties are considered in this analysis:

- **QCD scale choices**: We considered anti-correlated variations of the factorisation and renormalisation scales (μ_R/μ_F) in the $t\bar{t}$ and W+jets, by factors of 0.5 and 2. These differences were saved in the simulated events as alternative sets of weights which were used in the evaluation of this systematic. The envelope of 6 variations were considered as a systematic, excluding the most extreme cases of $\mu_{R,F}$=(2,0.5) and $\mu_{R,F}$=(0.5,2) which could potentially result in large logarithms in resummation. (Cacciari et al., 2004; Catani et al., 2003).
- **Top** p_{T}: The uncertainty due to the modelling of the top quark p_{T} distribution in simulation is evaluated by re-weighting at generator level.
- **Top mass**: CMS had analysed the most precise measurement of the top quark mass yields a total uncertainty of ±0.49 GeV (CMS Collaboration, 2016c). In the fits we considered variations of the top mass due to this uncertainty as a nuisance parameter, in a 6 σ band at $m_t = 172.5 \pm 3.0$ GeV. Template distributions were generated from Monte Carlo simulated datasets where the generator-level top mass is fixed at either edge of the band.

- **ME-PS matching scheme**: The default simulation was based on POWHEG. The so-called hdamp parameter was varied from it's central value of $1.58m_{top}$ using samples with hdamp=$2.24m_{top}$ and hdamp=$0.99m_{top}$.
- **Parton shower scale**: Alternative POWHEG+PYTHIA 8 samples where the parton shower scale choice was varied by a factor of 0.5 and 2 used in the analysis. This variation affects the fragmentation and hadronisation of the jets initiated by the matrix element calculation as well as the emission of extra jets by the hardroniser. Shape and rate effects are separated in two nuisances. The problems were assumed to be log-normally distributed.
- **Colour reconnection model**: We varied the colour reconnection model with respect to the default using alternatives including the resonant decay products in possible reconnections to the UE. The default simulation (MPI-based colour reconnection) has this effect excluded. We examined two alternative models for CR: the so-called gluon move and the QCD-inspired models. The envelope of the differences was considered as a systematic uncertainty.
- **UE variations**: The default parameters in the CUETP8M2T4 were varied according to their uncertainty and the effect on the unfolding was taken as an estimate of the systematic uncertainty.

For all the uncertainties we considered a maximum-based prescription: i.e. built the envelope of the contributions to each source, in each bin. In each bin, we took the $\max(|\delta^+|, |\delta^-|)$ as the final uncertainty estimation as some of the sources are single-sided. Theory uncertainties are a limiting factor for some variables and certain regions of phase space. In particular, for the average momentum, the uncertainty in the smeared phase space is typically small, but is strongly affected when one attempts to correct for detector level events. The corrections are observed to be strongly dependent on the hadroniser

model. This is expected due to the significantly different predictions from HERWIG++ with respect to PYTHIA 8 for both the underlying event and the $t\bar{t}$ kinematics. In Table 4.10 listed the percentage of the uncertainties in every different variables.

Table 4.10: Median of the bin-by-bin systematics for different variables in the unfolded phase space. The values are expressed in percentage

Uncertainty source	**Variable**				
	Multiplicity (N_{ch})	**p_{T} flux**	**p_z flux**	**<p_{T}>**	**<p_z>**
JER	0.1	0.2	0.0	0.1	0.1
JES	0.1	0.0	0.2	0.1	0.0
ME-PS	0.7	1.6	0.6	0.7	1.0
ISR	0.3	1.2	1.3	1.1	0.3
Trk. eff.	0.3	0.3	0.5	0.3	0.5
UE variations	0.7	0.8	0.8	0.5	0.8
FSR	1.3	0.9	1.3	1.2	0.7
$p_{\mathrm{T}}(t)$	0.7	0.2	0.5	0.2	0.4
Trigger/Sel.	0.5	0.3	0.6	0.2	0.1
μ_R/μ_F	0.8	0.6	0.6	0.5	0.2
Background	1.9	1.4	1.3	0.9	0.8
b-tag	0.3	0.2	0.3	0.3	0.2
Pileup	0.4	0.5	0.5	0.2	0.6
LES	0.2	0.2	0.4	0.2	0.2
m_t	0.7	0.9	0.3	0.7	0.9
total	2.93	3	2.8	2.3	2.1

CHAPTER 5: RESULTS

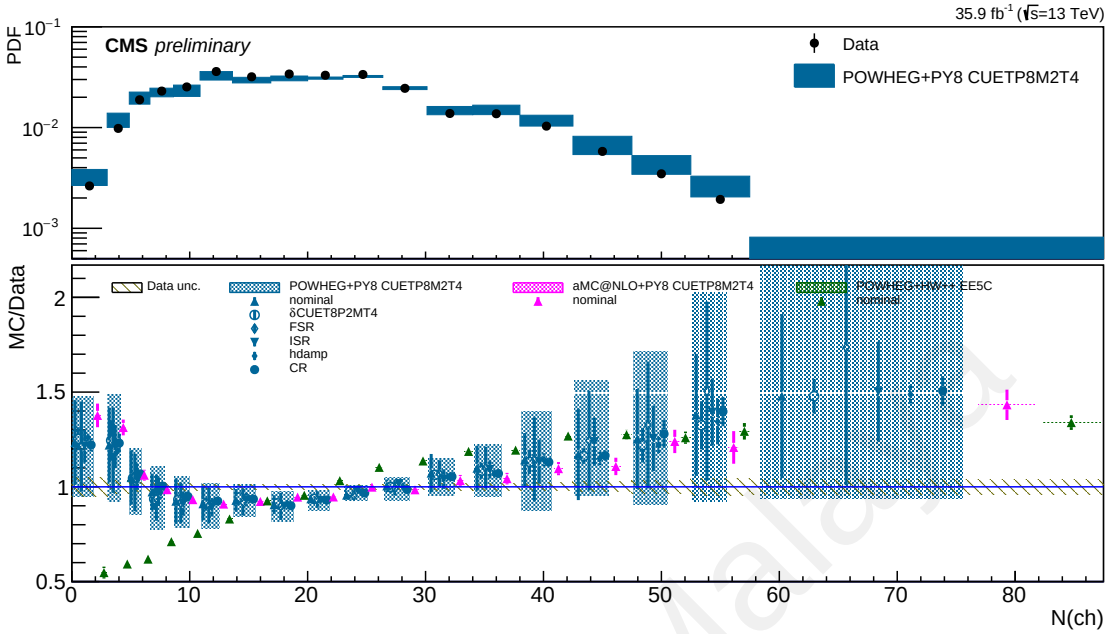
The distribution of the underlying event observables (N_{ch}, Σp_{T}, Σp_z, $<p_{\mathrm{T}}>$, and $<p_z>$) are shown in this section. The results are presented in the reconstructed level and particle level inclusively. The results in particle level presented in merged bins as a consequence of the binning choices in unfolding procedures. All figures at the upper-panels compare the $t\bar{t}$ simulation to the data which are subtracted from the background. The checking effect of the UE variations tunes also presented in the sub panels that showed the ratio of the data to the different varieties of the $t\bar{t}$ simulation with the total uncertainty is shown as the blue band. To be highlighted, the distributions we have not included the contributions from the two leptons and two b-jets which are matched to the $t\bar{t}$ decay products.

The number of charged particles (N_{ch}) in the Figure 5.1 obeserved a good agreement between the PYTHIA 8 tune and the ones reconstructed in data. Even the prediction shows slightly higher number than the data in lower charged multiplicity, but it is reasonable agreement within the error bar. The bottom-panels compared the data not only to the POWHEG+PY8 with CUETP8M2T4 tune but also to the aMC@NLO+PY8 with also applied the CUETP8M2T4 tune and POWHEG+HW++ sample with the EE5C tune. The phenomenological models that described the underlying event behaviour such as ISR, FSR, and also colour reconnection also presented and compared with the data. From all the variations of this tune, FSR scale variation can change more significantly this prediction and eventually bring the tune to be in better agreement with the data. The higher uncertainties found in, the higher multiplicity.

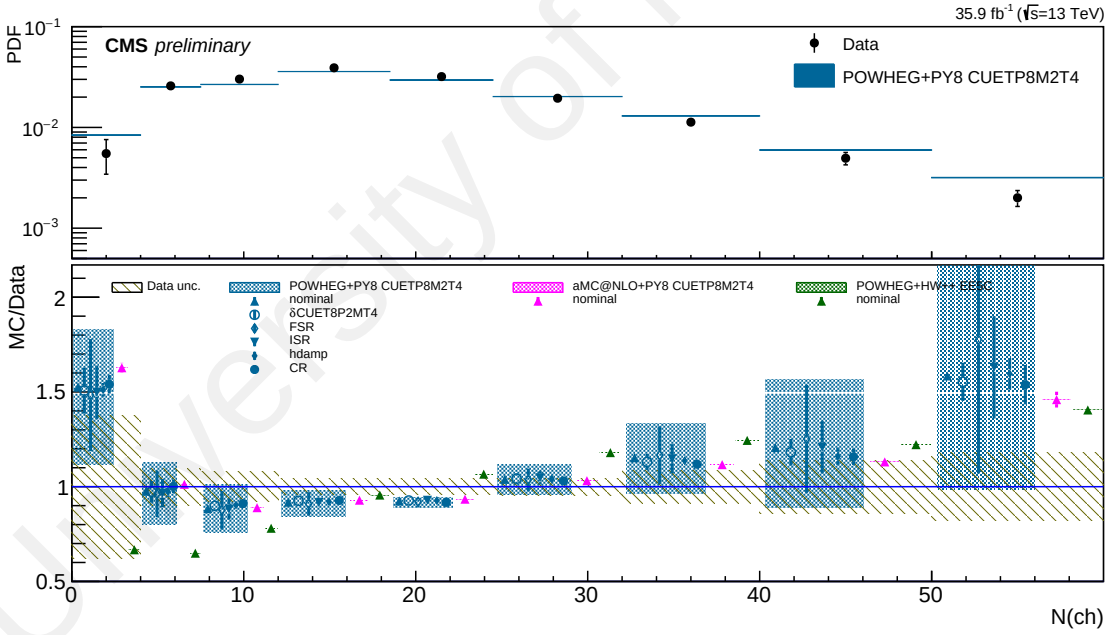
The total fluxes (Σp) in the Figures 5.2-5.3 are observed to disagree in the low momentum range both in the transverse plane (p_{T}) and longitudinal direction (p_z). This disagreement apparently correlated with the un-fitted low charge multiplicity. However,

the differences concerning data tend to cancel out when measuring the average momentum of the particles (<p>), and better agreement is found between data and the CUETP8M2T4 tune (See the Figures 5.4-5.5) This is proven by the smallest uncertainty values of <p> obtained shown in the Table 4.10.

We also notice that the interface of PYTHIA 8 with either POWHEG or aMC@NLO with PYTHIA 8 does not seem to interfere significantly with the underlying event. The central predictions of HERWIG++ tend to be significantly off concerning the data in most observables. In overall distributions, the cummulative relative uncertainties are shown in Table 4.10 less than 3 %. The conclusion hold for both the smeared and corrected phase spaces.

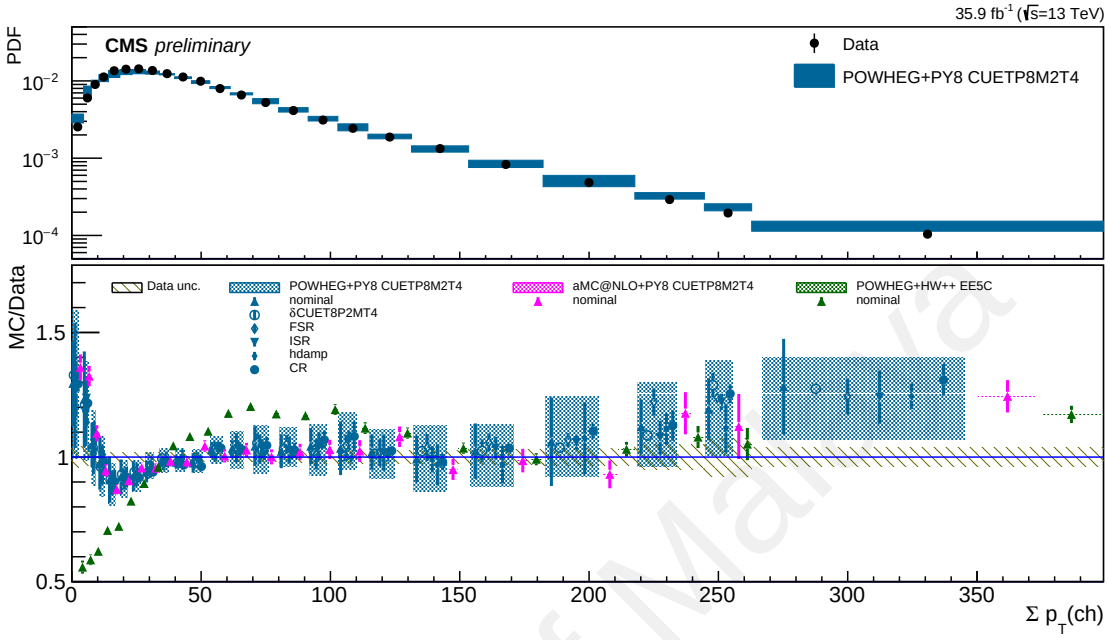


(a)

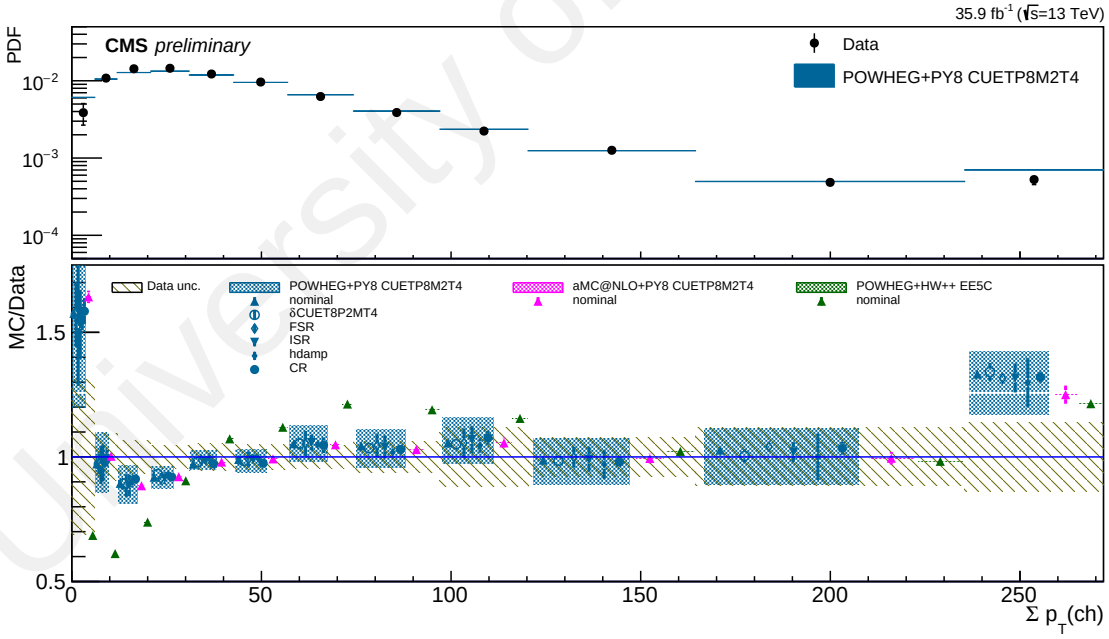


(b)

Figure 5.1: Average charge multiplicity observed at reconstruction (a) and particle (b) level, compared to different $t\bar{t}$ simulations. The top panel shows the data and a reference MC sample. The bottom panel displays the intrinsic uncertainty of the data compared to the MC to data ratio for the nominal simulation and variations of it where available. The filled bands encapsulating different variations of the same MC setup correspond to the total uncertainty envelope assigned to that prediction

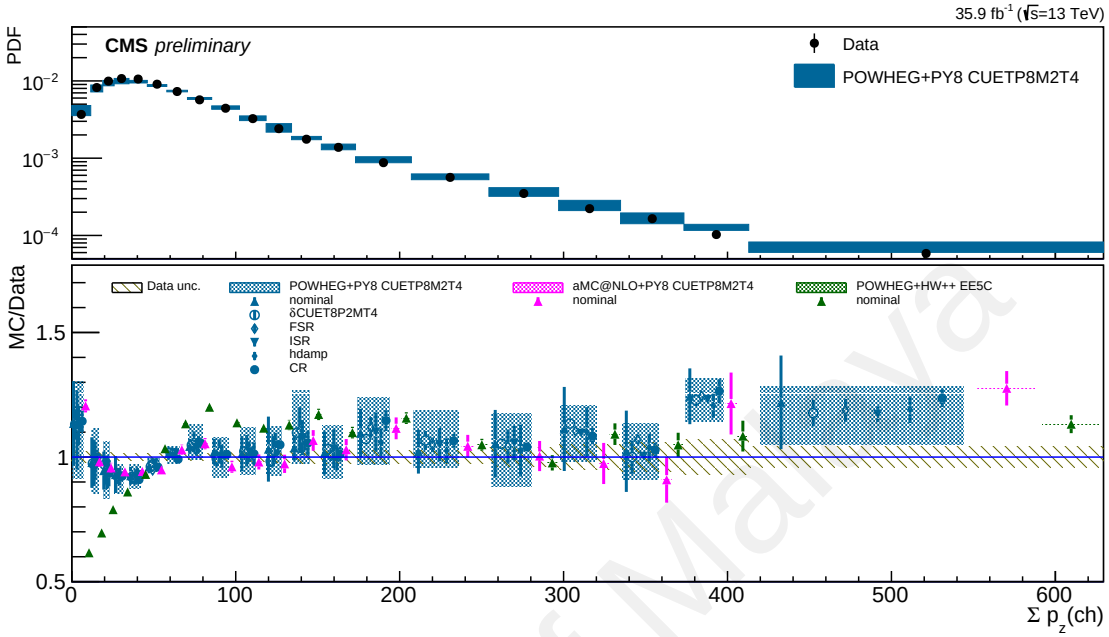


(a)

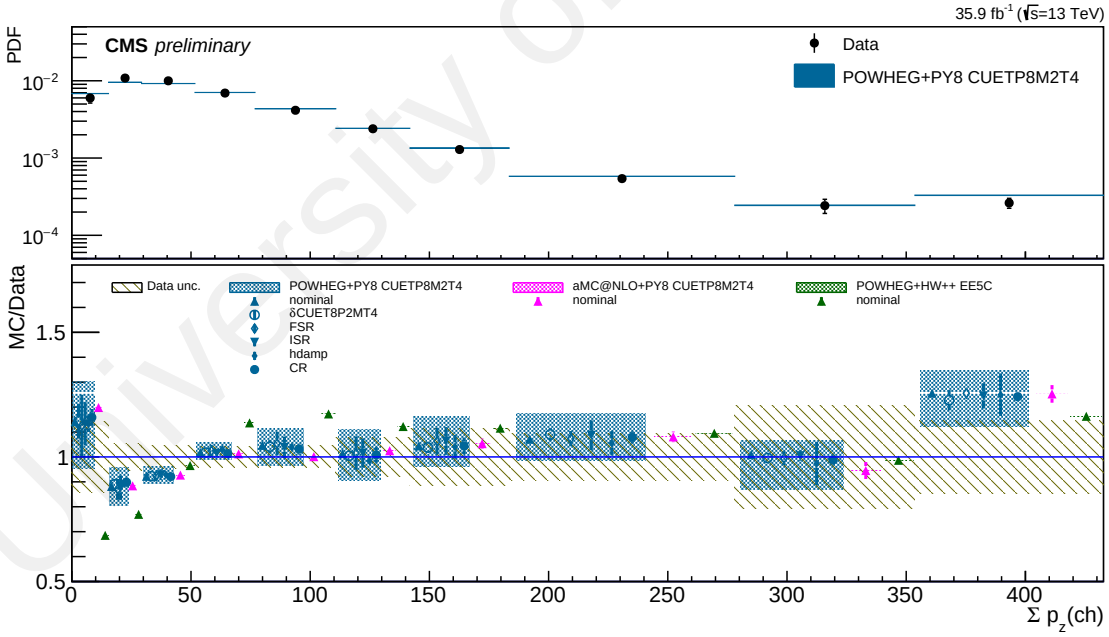


(b)

Figure 5.2: p_T flux observed at reconstruction (a) and particle (b) level, compared to different $t\bar{t}$ simulations

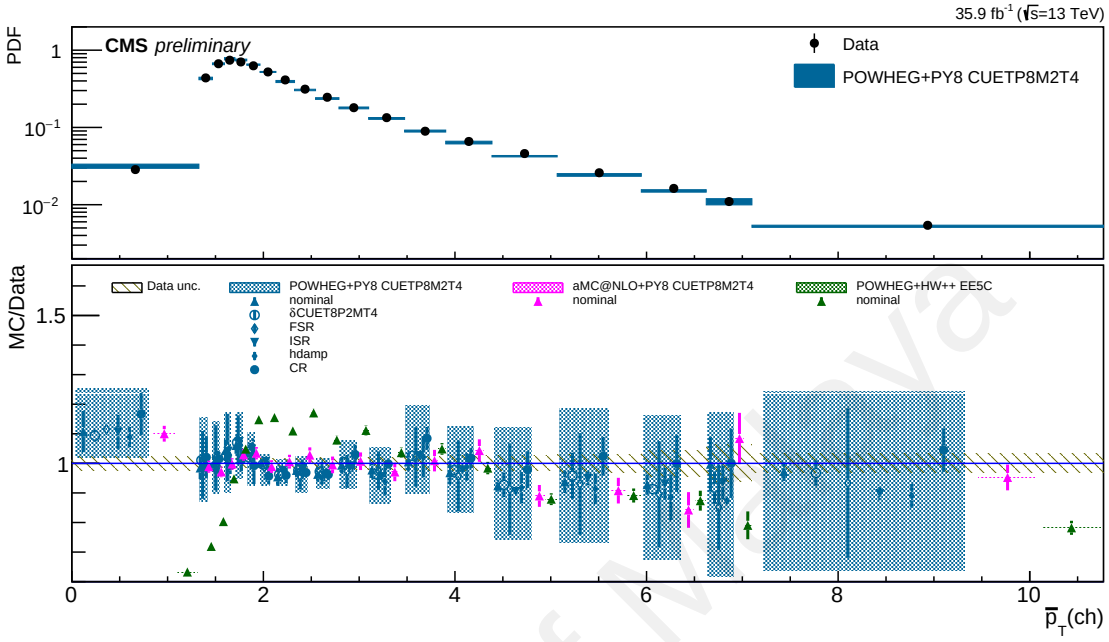


(a)

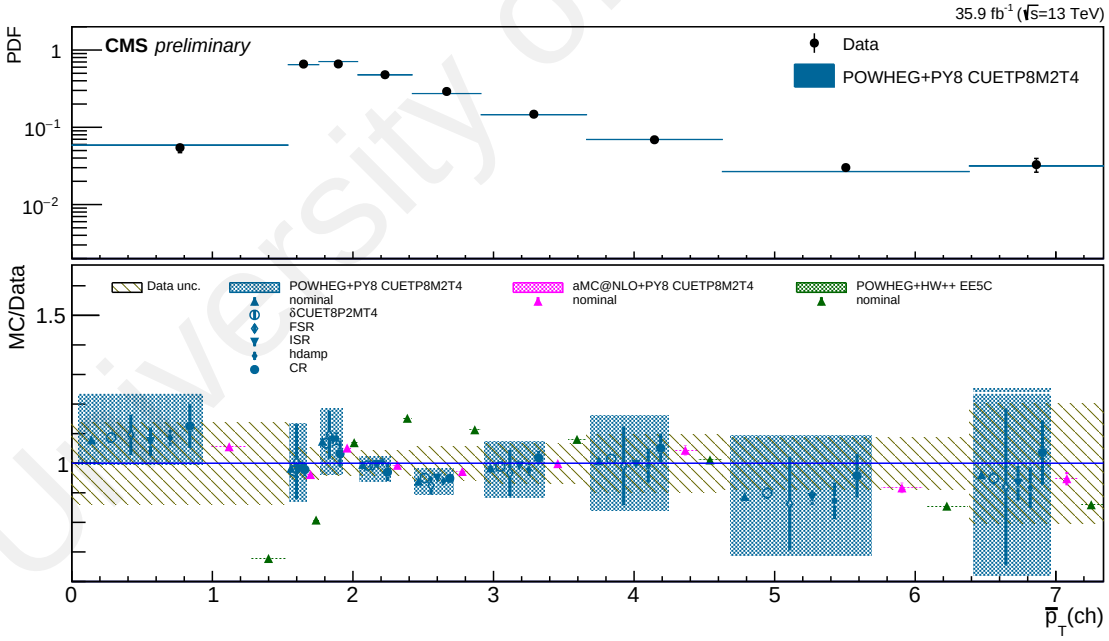


(b)

Figure 5.3: p_z flux observed at reconstruction (a) and particle (b) level, compared to different $t\bar{t}$ simulations

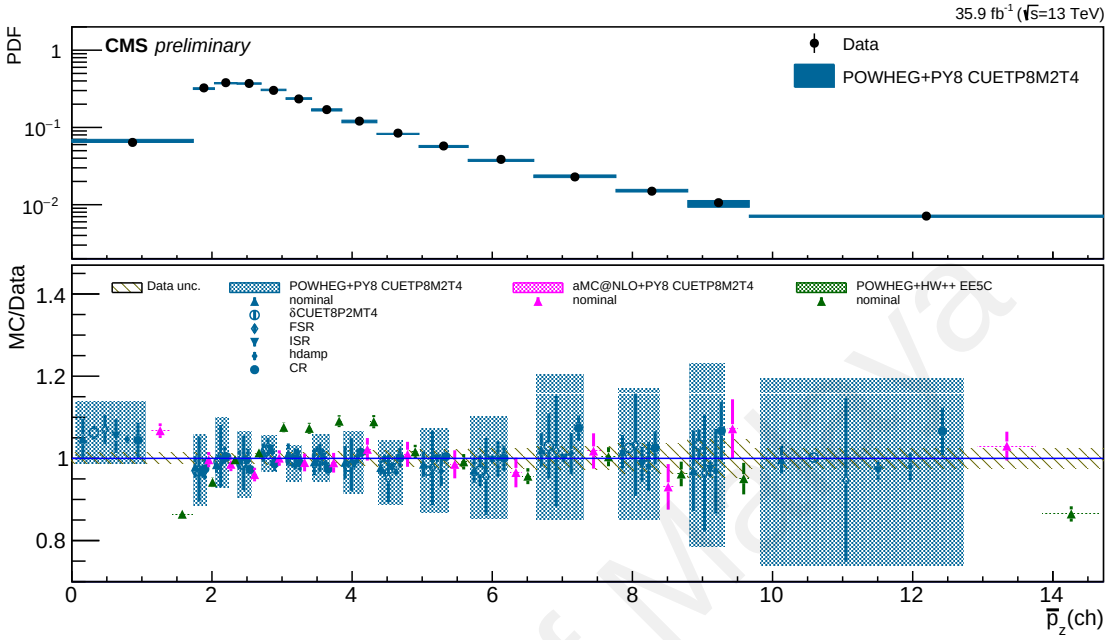


(a)

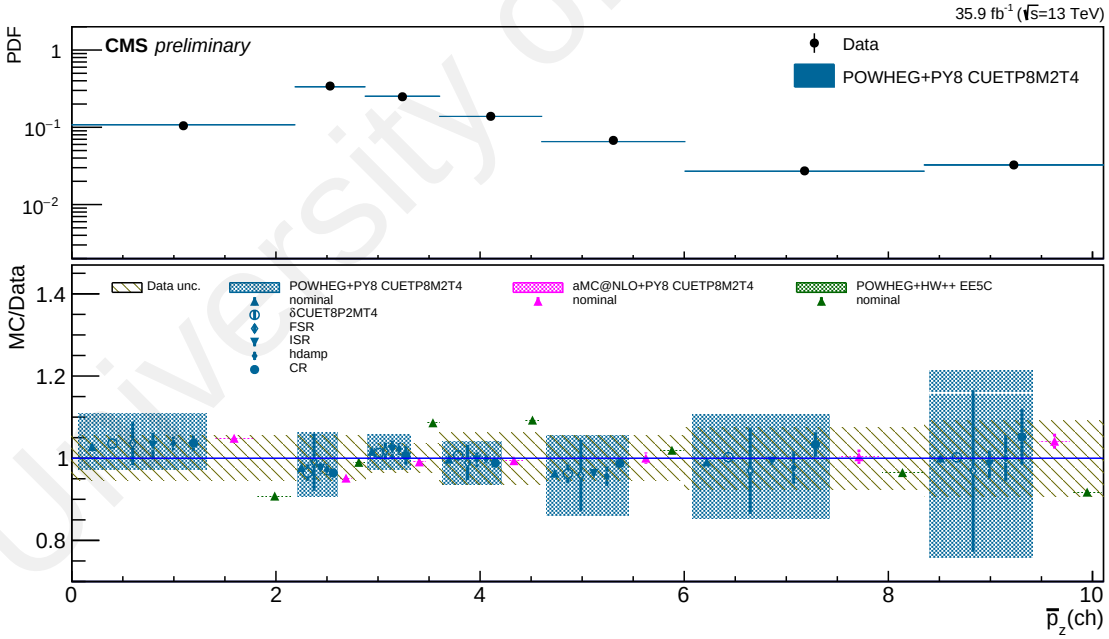


(b)

Figure 5.4: Average p_T observed at reconstruction (a) and particle (b) level, compared to different $t\bar{t}$ simulations



(a)



(b)

Figure 5.5: Average p_z observed at reconstruction (a) and particle (b) level, compared to different $t\bar{t}$ simulations

CHAPTER 6: DISCUSSIONS

The variations of the interfacing the PYTHIA 8 with POWHEG or MG5_aMC@NLO does not give the significantly impact to the UE. It shows that the NLO matrix-element generators, which are applied in MC@NLO and POWHEG, does not significantly effect in characterising the UE. But, disagreement found when interfacing the POWHEG with HERWIG++. The different hadronisation models in the HERWIG++ and PYTHIA 8 simulation, see Section 2.4.3, seems as a reasonable explanation to describe the disagreement. The string model which used in PYTHIA 8 is assumed to be a good description of the hadronisation model in this UE channel. This assumption linked to the hadronisation and fragmentation of the b-jets as the $t\bar{t}$ decay product.

As the first measurement of the UE in $t\bar{t}$ dilepton channel, until these steps, the results show the tunes which describe the UE behaviour better and provide the variations tune that can be used for the next analyses. It is a good achievement to start the UE studies in top quark sector since most of the UE measurements in the previous studies used high p_{T} charged particle or charged jets as the leading objects. The list of the previous UE studies can be seen in the Chapter 2.4.

Further UE studies in the $t\bar{t}$ channel can be measured by presenting the tested tunes and the various variables in this analysis with respect to the p_{T} of the $t\bar{t}$ system. Even further, the results can be shown in the slicing regions to understand the characteristic of the UE modelling distribution in the events. The limit study was faced when coming to the unfolding in multi-dimensional.

We hope, the results can be used to improve the assessment of systematics in the top quark related analysis as well as contribute to test the universality of the UE hypothesis at a high energy scales.

CHAPTER 7: CONCLUSION

The underlying event activity in $t\bar{t}$ dilepton channel has been measured using proton-proton collision data recorded by CMS detector at a centre-of-mass energy of 13 TeV and corresponding to an integrated luminosity of 35.9 fb^{-1}. The hard part of the collision subtracted using the matcing $t\bar{t}$ decay candidates and the left charged particles are used to characterise the UE. Observables N_{ch}, total p_{T} and p_z, and also average p_{T} and p_z have been presented in both detector and particle level. The majority of the distributions showed a good agreement between the data and the POWHEG+PYTHIA 8 CUETM2P4 setup and disfavor the POWHEG+HERWIGpp EE5C setup. The QCD scale variations for FSR tune seem to give impact contribution to the UE modelling with PYTHIA 8, while for ISR, the QCD scale variations do not significantly contribute.

REFERENCES

Alioli, S., Nason, P., Oleari, C., & Re, E. (2010). A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, *06*, 043.

Alwall, J., Frederix, R., Frixione, S., Hirschi, V., Maltoni, F., Mattelaer, O., … Zaro, M. (2014). The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, *07*, 079.

Alwall, J., Hoeche, S., Krauss, F., Lavesson, N., Lonnblad, L., Maltoni, F., … Worek, M. (2008). Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *The European Physical Journal*, *C53*, 473-500.

ATLAS Collaboration. (2011). Measurement of underlying event characteristics using charged particles in *pp* collision at $\sqrt{s}$ = 900 GeV and 7 TeV with the atlas detector. *Physical Review D*, *83*, 112001.

ATLAS Collaboration. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letter B*, *716*, 1-29.

Bahr, M., Gieseke, S., Gigg, M. A., Grellscheid, D., Hamilton, K., Latunde-Dada, O., … Webber, B. R. (2008). Herwig++ Physics and Manual. *The European Physical Journal*, *C58*, 639-707.

Barney, D. (2016). *CMS Detector Slice* (CMS Collection). Retrieved from https://cds.cern.ch/record/2120661?ln=en

Bethke, S. (2009). The 2009 world average of α_s. *The European Physical Journal*, *C64*, 689-703.

Britzger, D. (2013). *Regularized unfolding of jet cross sections in deep-inelastic ep scattering at hera and determination of the strong coupling constant* (Doctoral Thesis). Retrieved from https://www-h1.desy.de/psfiles/theses/h1th-831.pdf

Brobel, V. (2002). *An unfolding method for high energy physics experiments* (Technical

Report: Desy Report No. DESY 02-078). Retrieved from https://arxiv.org/abs/hep-ex/0208022v1

BTV POG Collaboration. (2015). *Btagcalibration documentation* (Twiki Page). Retrieved from https://twiki.cern.ch/twiki/bin/view/CMS/BTagCalibration

Buckley, A., Butterworth, J., Grellscheid, D., Hoeth, H., Lonnblad, L., Monk, J., ... Siegert, F. (2013). *RIVET user manual* (Technical Report: Manual No. MC-net/10/03). Retrieved from https://arxiv.org/abs/1003.0694

Butterworth, J., Dissertori, G., & Salam, G. (2012). Hard processes in proton-proton collisions at the large hadron collider. *Annual Review of Nuclear and Particle Science*, *62*, 387-405.

Cacciari, M., Frixione, S., Mangano, M. L., Nason, P., & Ridolfi, G. (2004). The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence. *Journal of High Energy Physics*, *04*, 068.

Cacciari, M., Salam, G. P., & Soyez, G. (2008). The anti-K_{T} jet clustering algorithm. *Journal of High Energy Physics*, *04*, 063.

Cacciari, M., Salam, G. P., & Soyez, G. (2012). FastJet user manual (for version 3.0.2). *The European Physical Journal*, *C72*, 1896.

Catani, S., de Florian, D., Grazzini, M., & Nason, P. (2003). Soft gluon resummation for Higgs boson production at hadron colliders. *Journal of High Energy Physics*, *07*, 028.

CDF Collaboration. (2002). Charged jet evolution and the underlying event in proton-antiproton collision at 1.8 TeV. *Physical Review D*, *65*, 092002.

CDF Collaboration. (2004). Underlying event in hard interactions at the fermilab. *Physical Review D*, *70*, 072002.

CDF Collaboration. (2010). Studying the underlying event in drell-yan and high transverse momentum jet production at the tevatron. *Physical Review D*, *82*, 034001.

CMS Collaboration. (2010). Measurement of the Underlying Event Activity in Proton-

proton Collisions at 0.9 TeV. *The European Physical Journal*, *C70*, 555-572.

CMS Collaboration. (2011). Measurement of the underlying event activity at the lhc at sqrt(s)= 7 TeV and comparison with sqrt(s) = 0.9 TeV. *Journal of High Energy Physics*, *09*, 109.

CMS Collaboration. (2012). Measurement of the underlying event in the Drell–Yan process in proton–proton collisions at $\sqrt{s}$ =7 TeV. *The European Physical Journal*, *C72*, 2080.

CMS Collaboration. (2013). *Study of the underlying event, b-quark fragmentation and hadronization properties in $t\bar{t}$ events* (Techical Report: CMS Physics Analysis Summary No. CMS-PAS-TOP-13-007). Retrieved from http://cds.cern.ch/record/1600599

CMS Collaboration. (2015). Measurement of the underlying event activity using charged-particles jets in proton-proton collision at $\sqrt{s}$ = 2.76, TeV. *Journal of High Energy Physics*, *09*, 137.

CMS Collaboration. (2016a). Event generator tunes obtained from underlying event and multiparton scattering measurements. *The European Physical Journal*, *C76*, 155.

CMS Collaboration. (2016b). *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of* $\mathrm{t\bar{t}}$ *at* $\sqrt{s}$ = 8 *and 13 TeV* (Technical Report: CMS Physics Analysis Summary No. CMS-PAS-TOP-16-021). Retrieved from http://cds.cern.ch/record/2235192

CMS Collaboration. (2016c). Measurement of the top quark mass using proton-proton data at sqrt(s) = 7 and 8 TeV. *Physical Review D*, *93*, 2004.

CMS Collaboration. (2017a). The CMS Trigger System. *Journal of Instrumentation*, *12*, P01020.

CMS Collaboration. (2017b). *Object definitions for top quark analyses at the particle level* (Technical Report: CMS Note No. CMS-NOTE-2017-004). Retrieved from https://cds.cern.ch/record/2267573

CMS Collaboration. (2017c). *Particle-flow reconstruction and global event description with the cms detector.* Retrieved from http://iopscience.iop.org/article/10.1088/

Corcella, G., Knowles, I. G., Marchesini, G., Odagiri, K., Richardson, P., Seymour, M. H., & Webber, B. R. (2001). HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes). *Journal of High Energy Physics*, *01*, 010.

Corcella, G., & Mescia, F. (2010). A phenomenological study of bottom quark fragmentation in top quark decay. *The European Physical Journal*, *C65*, 171-180.

Davies, C. (2002). *Lecture note: Lattice QCD.* Retrieved from https://arxiv.org/abs/hep-ph/0205181

D∅ Collaboration. (2011). Precise measurement of the top quark mass in the dilepton channel at D0. *Physics Review Letter*, *107*, 082004.

Egamma POG Collaboration. (2015a). *Cut Based Electron ID for Run 2* (Twiki page). Retrieved from https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2

Egamma POG Collaboration. (2015b). *Electron scale factors for run 2* (Twiki Page). Retrieved from https://twiki.cern.ch/twiki/bin/view/CMS/ElectronScaleFactorsRun2

Frederix, R., & Frixione, S. (2012). Merging meets matching in MC@NLO. *Journal of High Energy Physics*, *12*, 061.

Frixione, S., Nason, P., & Oleari, C. (2007). Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *Journal of High Energy Physics*, *11*, 070.

GEANT4 Collaboration. (2003). GEANT4: A simulation toolkit. *Nuclear Instruments and Methods in Physics Research*, *A506*, 250-303.

GENPOG Collaboration. (2015a). *McM Monte-Carlo Request Management* (site). Retrieved from https://cms-pdmv.cern.ch/mcm/

GENPOG Collaboration. (2015b). *Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing* (Twiki page). Retrieved from https://twiki

.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns

Gieseke, S., Rohr, C., & Siodmok, A. (2012). Colour reconnections in Herwig++. *The European Physical Journal*, *C72*, 2225.

Gleisberg, T., Hoeche, S., Krauss, F., Schaelicke, A., Schumann, S., & Winter, J. (2004). SHERPA 1.alpha, a proof-of-concept version. *Journal of High Energy Physics*, *02*, 056.

Goh, J., & Seidel, M. (2017). *Pseudotopproducer* (Github repository). Retrieved from https://github.com/intrepid42/cmssw/tree/pseudotoprivet_80x/TopQuarkAnalysis/TopEventProducers

JetMET POG Collaboration. (2017a). *Jet energy resolution measurement* (Twiki Page). Retrieved from https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution

JetMET POG Collaboration. (2017b). *Recommended jet energy corrections and uncertainties for data and mc* (Twiki Page). Retrieved from https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC

Kar, D., & CDF Collaboration. (2009). *Measurement of the underlying event at tevatron.* Retrieved from https://arxiv.org/abs/0905.2323

LHCtopWG Collaboration. (2015a). *NLO single-top channel cross sections : ATLAS-CMS recommended predictions for single-top cross sections using the Hathor v2.1 program* (Twiki page). Retrieved from https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec

LHCtopWG Collaboration. (2015b). *NNLO+NNLL top-quark-pair cross sections : ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program* (Twiki page). Retrieved from https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO

Madspin Collaboration. (2017). *MadSpin* (site). Retrieved from https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/MadSpin

Mariani, V. and CMS Collaboration. (2017). *Measurement of the relative tracking efficiency for charged pions using the ratio of branching fractions for D0 decays to $K\pi\pi$ and $K\pi$* (Technical Report: CMS Analysis Note No. CMS-AN-2015/048).

Retrieved from http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2015_248_v4.pdf

Markus, C., & Mulders, M. (2017). Top-quark physics at the large hadron collider. *Journal of Physics G: Nuclear and Particle Physics*, *44*, 063001.

Muon POG Collaboration. (2015a). *Reference muon id, isolation and trigger efficiencies for Run-II* (Twiki page). Retrieved from https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2

Muon POG Collaboration. (2015b). *Reference muon id, isolation and trigger efficiencies for Run-II* (Twiki Page). Retrieved from https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2

Nason, P. (2004). A New method for combining NLO QCD with shower Monte Carlo algorithms. *Journal of High Energy Physics*, *11*, 040.

PVT Collaboration. (2015). *Utilities for accessing pileup information for data* (Twiki page). Retrieved from https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData

Schmitt, S. (2012). TUnfold: an algorithm for correcting migration effects in high energy physics. *Journal of Instrumentation*, *7*, T10003.

Schmitt, S. (2016). *Tunfold software package website* (site). Retrieved from http://www.desy.de/~sschmitt/tunfold.html

Silva, P. and Seidel, M. (2017). *Topljets2015* (Github repository). Retrieved from https://github.com/pfs/TopLJets2015

Sjostrand, T., Mrenna, S., & Skands, P. Z. (2006). PYTHIA 6.4 Physics and Manual. *Journal of High Energy Physics*, *05*, 026.

Sjostrand, T., Mrenna, S., & Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1. *Computer Physics Communications*, *178*, 852-867.

Sjöstrand, T., Ask, S., Christiansen, J. R., Corke, R., Desai, N., Ilten, P., ... Skands, P. Z. (2015). An Introduction to PYTHIA 8.2. *Computer Physics Communications*, *191*,

159-177.

Skands, P. (2012). *Lecture note: QCD for collider physics.* Retrieved from https://arxiv.org/abs/1104.2863

Skands, P. Z. (2010). Tuning Monte Carlo Generators: The Perugia Tunes. *Physical Review D*, *82*, 074018.

Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics - Doklady*, *4*, 1035.

TOP Collaboration. (2017). *Top trigger (run2)* (Twiki page). Retrieved from https://twiki.cern.ch/twiki/bin/view/CMS/TopTrigger

Vallero, S. (2012). *Study of the underlying event in pp collisions with the alice detector at the lhc* (Doctoral Thesis). Retrieved from http://archiv.ub.uni-heidelberg.de/volltextserver/13114/

Vos, M. (2016). *Top-quark mass measurements at the LHC: alternative methods* (Tech. Rep.). Retrieved from https://arxiv.org/abs/1602.00428

Wikipedia. (2016). *Condition number* (site). Retrieved from https://en.wikipedia.org/wiki/Condition_number

Wynne, B. (2012). *Measurement of the underlying event in pp collisions using the atlas detector and development of a software suite for bayesian unfolding* (Doctoral Thesis). Retrieved from http://cds.cern.ch/record/1626590/files/CERN-THESIS-2013-200.pdf

APPENDIX A: CONTROL DISTRIBUTION AND EVENT YIELDS

The Figure A.1 below shows control distributions after selection for global event variables: the *b*-tagging multiplicity (prior to the requirement of two *b*-tags), the additional jet multiplicity (the two *b*-tagged jets are excluded) and the $E_{\mathrm{T}}^{\mathrm{miss}}$. We observe overall fair agreement in all the distributions.

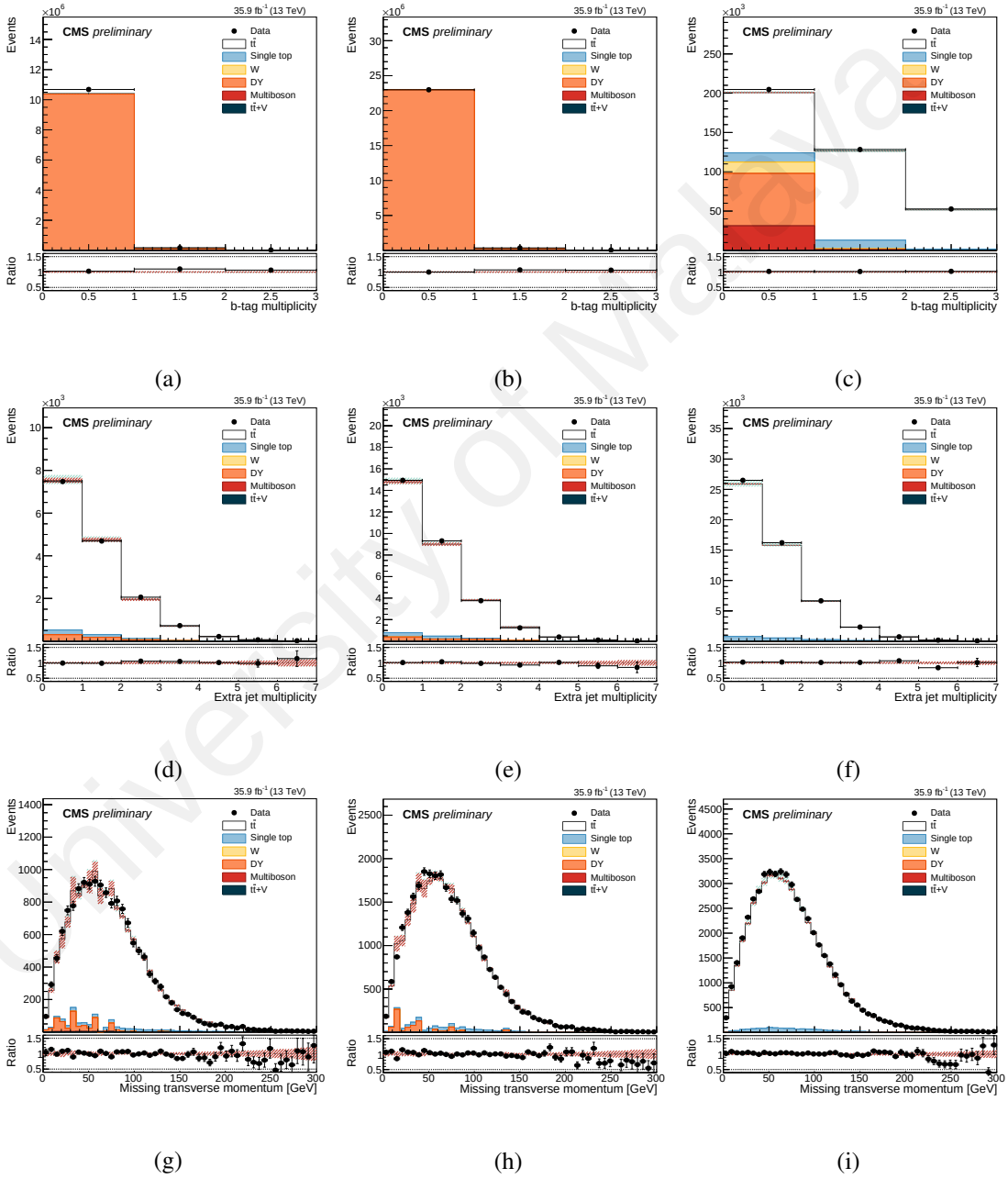


Figure A.1: Top: *b*-tagging multiplicity; center: jet multiplicity; bottom: $E_{\mathrm{T}}^{\mathrm{miss}}$. The distributions are shown for the ee (left) $\mu\mu$ (center) and $e\mu$ (right) channels

Two system boost related variables, which will be used later to profile the underlying event activity are shown in Figure A.2. The p_{T} of the $t\bar{t}$ system, computed from the vectorial sum of the b-jets p_{T}, the two leptons p_{T} and the $E_{\mathrm{T}}^{\mathrm{miss}}$, and the dilepton p_{T} are shown. The first is expected to be directly sensitive to the recoil in the event and thus can be used to enhance or suppress the contributions from ISR to the event. The latter also preserves some of this sensitivity and it is expected to be reconstructed with high resolution. In both cases there is an excellent agreement between data and simulation.

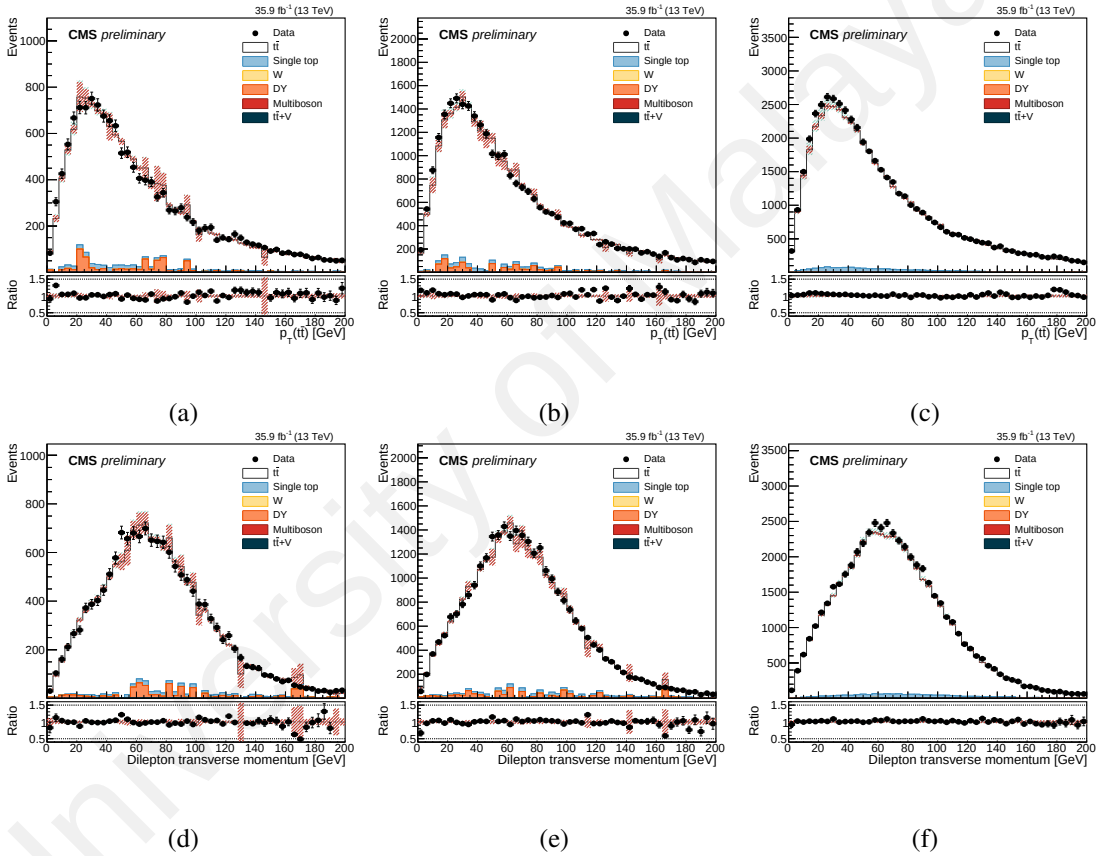


Figure A.2: Top: p_{T} $(t\bar{t})$; bottom: p_{T} (l,l). The distributions are shown for the ee (left) $\mu\mu$ (center) and $e\mu$ (right) channels