Study of Jets Production Association with a Z boson in pp Collision at 7 and 8 TeV with the CMS Detector

by

Kittikul Kovitanggoon, B.S., M.S.

A Dissertation

In

PHYSICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

Sung-Won Lee
Chair of Committee

Nural Akchurin

Teruki Kamon

Roger Lichti

Richard Wigmans

Mark A. Sheridan
Dean of the Graduate School

May, 2014

Acknowledgments

I would like to express my gratitude to my advisor, Sung-Won Lee, for his precious guidance and encouragement. He taught me about physics as a physicist and about life as a big brother. I am honored to be his student and his young brother. I am grateful to Nural Akchurin for his supervision and teaching in both physics and life.

I really appreciate all of my TTU professors (in particular, Richard Wigmans, Roger Lichti, Igor Volobouev) and all physics department faculty and staff. To accomplish these researches in a big collaboration, I would like to thank my colleagues at the CMS collaboration, Teruki Kamon, Stephen Linn, Shin Shu Yu, Tomislav Seva, Sudhir Malik, and others. It was a joy to collaborate with them and to benefit from their expertise.

I also would like to thank old and new TTU HEP group members for their helps, Efe Yazgan, Jordan Damgov, Christopher Cowden, Cosmin Dragoiu, Phillip Dudero, Chiyoung Jeong, Youn Roh and others. I appreciate all the support from CMS secretaries especially Yasemin Yazgan. I am thankful to Alan Sill and HPCC crews for providing us with better computational excellent.

I want to say thanks all my friends, Bugra Bilin, Van Rusuriye, Jacob Ajimo, James Faulkner, Terence Libeiro, Rick Mengyan, Luke Whittlesey and others for all their support for these years.

Finally, it would not be possible without my family. Thank to my mother Kamoljit Kovitanggoon, my father Kuekkong Kovitanggoon, and my sister Nattavadee Kovitanggoon for supporting me in this long journey. Thank you, my grandparents (in particular, Darunwan Kovitanggoon) for making these study abroad possible. Thanks, all my friends

in Thailand, US, and Europe.

Contents

Abstract

This study presents the measurement of the rapidity distributions in events containing a Z boson and a jet in proton-proton collisions at a center-of-mass energy of 7 TeV. The data correspond to an integrated luminosity of approximately 5 fb^{-1}, recorded by the CMS detector at the Large Hadron Collider (LHC). The measured angular distributions are compared with the predictions from next-to-leading order perturbative QCD calculations and two generator programs that combine tree-level matrix element calculations with parton showers. We also present a measurement of jet production rates in association with a Z boson using data recorded at a center-of-mass energy of 8 TeV and with an integrated luminosity of 19.8 fb^{-1}. This measurement provides a stringent test of perturbative QCD calculations, and the result is compared with predictions from theoretical calculations.

List of Tables

List of Figures

List of Abbreviations

ALICE: A Large Ion Collider Experiment

ATLAS: A Toroidal LHC ApparatuS

CDF: Collider Detector at Fermilab

CERN: European Organization for Nuclear Research

CM: Center-of-Momentum

CMS: Compact Muon Solenoid

CMSSW: CMS SoftWare

CSC: Cathode Strip Chamber

CTEQ: Coordinated Theoretical Experimental project on QCD

DAQ: Data AcQuisition system

DT: Drift Tube

EM: ElectroMagnetic

ECAL: Electromagnetic CALorimeter

EWK: Electromagnetic and WeaK interaction

ES: Endcap preShower

PF: Particle Flow

HB: Hadronic Barrel

HCAL: Hadronic CALorimeter

HE: Hadronic Endcap

HF: Hadronic Forward

HLT: High Level Trigger

HO: Hadronic Outer

JEC: Jet Energy Correction

JES: Jet Energy Scale

JER: Jet Energy Resolution

L1: Level-1 Trigger

LO: Leading Order

LHC: Largh Hadron Collider

LHCb: Large Hadron Collider Beauty experiment

MC: Monte Carlo

MCFM: Monte Carlo for FeMtobarn processes

MET: Missing E_T

NLO: Next-to-Leading Order

NNLL: Next-to-Next-to-Leading-Log

NNLO: Next-to-Next-to-Leading Order

PAT: Physics Analysis Tool

PDF: Parton Distribution Function

PF: Particle Flow

POG: Particle Object Group

PU: PileUp

pQCD: pertubative Quantum ChromoDynamics

QCD: Quantum ChromoDynamics

QED: Quantum ElectroDynamics

QFT: Quantum Field Theory

RECO: offline RECOnstruction

RPC: Resistive Plate Chamber

SHERPA: Simulation of High-Energy Reactions of PArticles

SM: Standard Model

SUSY: SUperSYmmetry

TEC: The two End Cap detector system

TIB: The Inner Barrel detector system

TID: The Inner Discs detector system

TOB: The Outter Barrel detector system

Chapter 1

Introduction

The Standard Model (SM) is the current theory for quarks and leptons and their electromagnetic, weak, and strong interactions. Although this theory has had great success in describing the physical universe, many questions still remain unanswered, such as the verification of mass generation by spontaneous symmetry breaking. One other unsolved mystery is dark matter, which is believed to be the most dominant constituent of matter in the universe. In order to answer these questions, it is important to use the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC) to measure well known physics processes with precision in order to recognize new physics. For instance, one well known physics process important for verifying the SM is the production of a Z vector boson associated with jets, where the boson decays into two leptons.

A crucial test of the SM is measuring a cross section, or total probability of producing a particular physics process. This type of measurement on Z+jets production is crucial to the theoretical physics community and helps improve perturbative Quantum ChromoDynamic (QCD) theory. New physics searches including the Higgs boson or Supersymmetry (SUSY) benefit from studying background processes that include such a decay of Z bosons into leptons, where high momentum leptons can be misidentified as a signature of new particles. In addition, the high predicted cross section of Z+jets events at the LHC offers a possible calibration benchmark for the CMS detector.

This analysis looks at the characterization of Z + jets production within the CMS detector at $\sqrt{s}$ = 7 TeV in 2011 and at $\sqrt{s}$ = 8 TeV in 2012. The outline of this dissertation

is as follows. In Chapter 2, a brief overview of the SM of elementary particles and their interactions is given. This is followed by a short introduction to the LHC and the CMS detector in Chapter 3. Chapter 4 shows the compared shape results of Z + jet angular analysis at $\sqrt{s}$ = 7 TeV, which is included in a published journal article [1]. Chapter 5 presents the results of the ongoing measurement of the Z+jets cross section at $\sqrt{s}$ = 8 TeV. Chapter 6 is the conclusions of this dissertation.

Chapter 2

Theory

This chapter introduces electroweak and strong interactions. The SM [2] is the most comprehensive and widely-accepted theory available. It is a Quantum Field Theory (QFT), verified by data collected to date, which describes weak, strong, and electromagnetic interactions in terms of a local gauge symmetry group:

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \tag{2.1}$$

$SU(3)_c$ is a non-Abelian gauge symmetry group which describes strong interactions. It contains a structure involving eight independent matrices, which are the generators of the group, and reflects that strong interactions are carried by eight massless particles, the gluons. The gluons have a strong charge known as "color", and, like the quark, their strong interactions are well described by Quantum Chromodynamic (QCD) theory [3]. $SU(2)_L \otimes U(1)_Y$ is the weak isospin symmetry group, which was introduced by Glashow-Weinberg-Salam [4, 5, 6] and describes the unified electromagnetic and weak (EWK) interaction. The EWK force is carried by three massive vector bosons W^+ ,W^- and Z^0, plus the massless photon γ.

Although the SM fails to incorporate gravity with other forces, to predict dark matter and dark energy, or to explain the excess of matter over anti-matter, it has successfully predicted most particle physics measurements during the past 30 years.

2.1 Elementary Particles and Interactions

The overall view of SM particles is shown in Fig. 2.1.

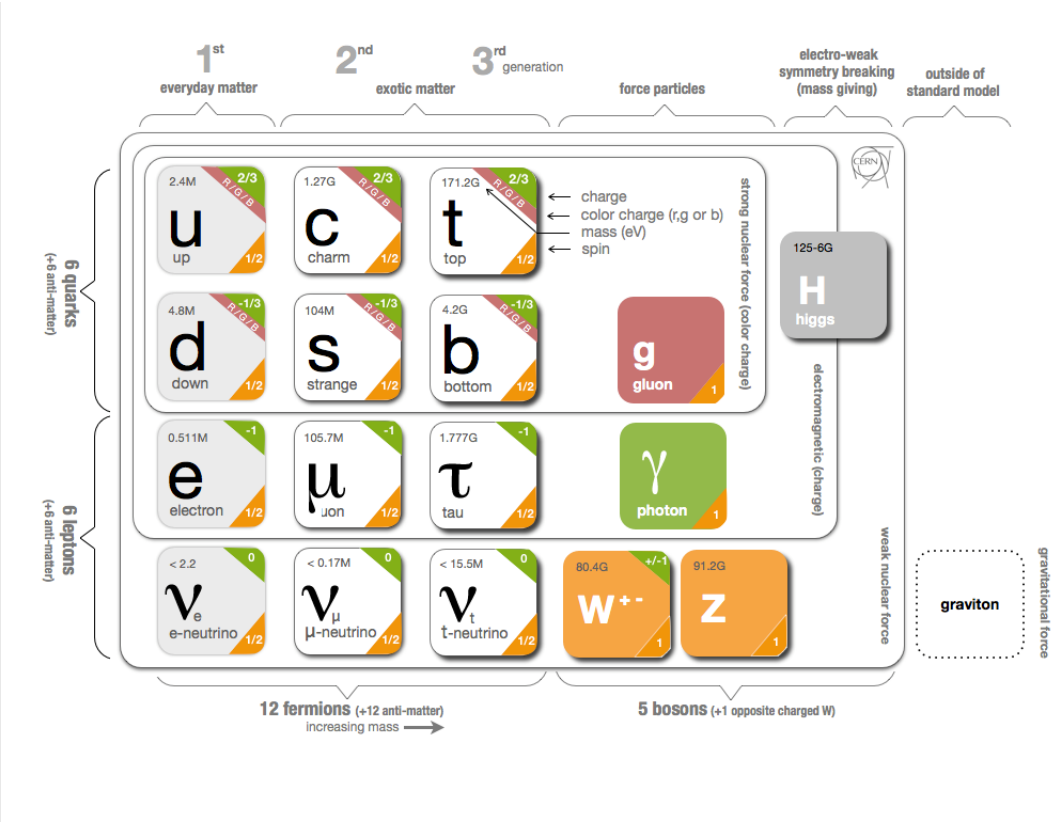


Figure 2.1: Overview on the elementary particles of the SM including their electric charges and measured masses. The mass of the Higgs boson refers to the one of the recently discovered Higgs-like boson [7].

2.1.1 Fermions

Ordinary matter consists of fermions, or particles with non-integer spin, that are classified into three generations ordered by mass. These generations are divided into two subcategories: fermions with integer electrical charge, 0 or -1, called leptons; fermions with non-integer electrical charge, +2/3 or -1/3, called quarks. This categorization allows for 12 fermion species, each with its own anti-particle. A particle shares the exact same characteristics as its anti-particle with the exception of opposite charge. Although particles and

their anti-particles have opposite charges, electrically neutral particles are their own anti-particles. However, whether or not neutrinos and their corresponding anti-neutrinos are identical particles has not yet been resolved. According to the current understanding of the universe, the visible composition is essentially made of such matter, with the reason for excess of matter over anti-matter remaining unknown.

Leptons are observed in free states while quarks exist only in bound states called hadrons which are divided into three categories: mesons composed of a quark and an anti-quark, ,baryons composed of three quarks, and anti-baryons composed of three anti-quarks.

2.2 Particle Interactions

The gauge bosons ($W^{\pm}$, Z^0, photon, gluon), with integer spin, act as mediators of force interactions. This is determined by their charges g, which are directly related to the coupling constants α of the interaction via $\alpha \propto g$. Note that in addition, the gauge bosons can partly also couple to each other depending on the respective group structures. The magnitudes of the coupling constants play a decisive role, both for understanding the fundamental principles of the interactions as well as for the calculation of physical observables like cross sections. Such calculations rely on perturbation theories, i.e. the expansion of the solution in orders of the coupling constants, which can only converge if $\alpha \ll 1$. The expansion terms can be illustrated using Feynman diagrams.

2.2.1 Electromagnetic Interaction

The electromagnetic (EM) interaction is mediated by the exchange of virtual photons γ, which couple to particles with electric charge. Photons are electrically neutral and therefore do not interact with each other except at the high enough energy in which $\gamma\gamma$ scattering can be happened. Furthermore, they are massless. The coupling constant α_{EM} increases from $\alpha_{EM} \approx 1/137$ at low energy scales to $1/129$ at the Z boson mass scale of 91 GeV/c^2, but remain $\ll 1$ up to very high energy scales. The underlying theory is called Quantum

ElectroDynamics (QED) [8] based on the $U(1)_{EM}$ symmetry group. In order to leave the Lagrangian invariant under the action of the transformation Λ:

$$\psi(x) \rightarrow e^{i\Lambda x}\psi(x), \tag{2.2}$$

a field $A_\alpha(x)$ is introduced which is modified by the gauge transformation $A_\alpha(x) \rightarrow A_\alpha(x) + \frac{1}{e}\partial_\alpha\Lambda$. The derivative must change to the covariant derivative $\partial_\alpha \mathscr{D}_\alpha = \partial_\alpha - ieA_\alpha$, in order to let the term $\bar{\psi}(i\gamma^\alpha \mathscr{D}_\alpha - m)\psi$ be invariant under the transformation 2.2, because the derivatives do not transform linearly under a transformation depending on a specific point of the spacetime x^μ.

The term corresponding to the propagation of the gauge field is also introduced,

$$-\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}, \tag{2.3}$$

with

$$F^{\alpha\beta} = \frac{i}{e}[\mathscr{D}_\alpha, \mathscr{D}_\beta] = \partial_\alpha A_\beta - \partial_\beta A_\alpha, \tag{2.4}$$

where $F^{\alpha\beta}$ is the gauge field tensor.

Thus, the resulting Lagrangian is invariant under local gauge transformation:

$$\mathscr{L}_{QED} = \bar{\psi}(i\gamma^\alpha \mathscr{D}_\alpha - m)\psi - \frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}, \tag{2.5}$$

where the $F^{\alpha\beta}$ contains A_α and A_β as representing the gauge boson, the photon, and the kinematic term it is propagating.

Although QED is a very successful and precise theory on its own, it turns out that, for a fundamental description, the electromagnetic interaction needs to be united with the weak interaction.

2.2.2 Weak Interaction

The weak interaction is mediated by the exchange of gauge bosons, $W^{\pm}$ and Z^0. They couple to other particles with weak charge g_w. The first observations of weak interaction were from the β decay, where an electron couples with a neutrino through charged gauge boson, W. The helicity is chosen as the associated conserved parameter because the weak force only acts with the left helicity particles. This is the case for all fermions as well as the W and Z bosons due to the non-Albelian nature of the underlying $SU(2)$ symmetry. An unitary non-Abelian group of off-diagonal generators, known as Pauli matrices τ_α, go as:

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.6}$$

where τ_1 and τ_2 correspond to the contribution of $W^{\pm}$ while τ_3 indicates the existence of another neutral boson.

The fermion's field can be decomposed into its left- and right-handed components as

$$\psi = \psi_L + \psi_R, \tag{2.7}$$

where right-handed components are represented as singlets, and left-handed components are doublets:

$$L_L = \frac{1}{2}(1-\gamma^5)\begin{pmatrix} \nu_l \\ l \end{pmatrix} = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, L_R = \frac{1}{2}(1+\gamma^5)(l). \tag{2.8}$$

Equation 2.8 represents general leptons. $L_{L,R}$ are the left- and right-handed components of the lepton spinor L. $(1-\pm\gamma^5)$ is the projectors on the chirality states by $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, where γ^μ with μ = 0,1,2,3 are the four Dirac matrices. Parity transformations of the weak interaction restrict only definite chirality states that can be involved in the interaction [9].

Quarks are described in an equivalent representation as:

$$Q_L = \frac{1}{2}(1-\gamma^5)\begin{pmatrix} u \\ d \end{pmatrix} = \begin{pmatrix} u \\ d \end{pmatrix}_L \tag{2.9}$$

where

$$u_R = \frac{1}{2}(1+\gamma^5)(u), d_R = \frac{1}{2}(1+\gamma^5)(d) \tag{2.10}$$

Moreover, the W bosons also couple to the photon due to their electric charges. Consequently, to fully understand the properties of the weak interactions and its gauge bosons the unification with the electromagnetic interaction in form of $SU(2)_L \otimes U(1)_Y$ symmetry group, called electroweak theory, is needed. Y stands for a new quantum number, the hypercharge, whereas L denotes the special role of left-handed particles, to which the W bosons couple exclusively and violate parity. The unification states that the weak charge and electric charges are related to each other by a weak mixing angle. The bare weak charge is of the same order as the electric charge. However, the W and Z bosons are observed to be very massive (80 GeV/c^2 and 91 GeV/c^2 respectively), which leads to a suppression of the coupling strength at low energies. The other unique property of the weak interaction is the coupling of the W bosons to different fermion types, or so called flavour, at the same interaction vertex, namely either to a neutrino and a charged lepton or to an up-type and a down-type quark. This situation normally occurs within the same generation, except for quarks, and has been observed in experiments. Mixing favour probability is explained by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix [10, 11].

2.2.3 Electroweak Interaction

The electroweak interaction is based on the $SU(2)_L \otimes U(1)_Y$ Lagrangian:

$$\mathscr{L}_{SU(2)\otimes U(1)} = \mathscr{L}_{gauge} + \mathscr{L}_{\phi} + \mathscr{L}_{f} + \mathscr{L}_{Yukawa}. \tag{2.11}$$

The gauge Lagrangian is represented by

$$\mathscr{L}_{gauge} = \frac{1}{4} W^i_{\mu\nu} W^{\mu\nu i} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{2.12}$$

where $W^i_{\mu\nu}$ (i = 1,2,3) and $B_{\mu\nu}$ are field strength tensors for the weak isospin ($SU(2)_L$) and the weak hypercharge ($U(1)_Y$) fields, respectively. Field strength tensors are defined as:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\varepsilon_{ijk} W^j_\mu W^k_\nu, \tag{2.13}$$

where g is the $SU(2)_L \otimes U(1)_Y$ gauge coupling constant and ε_{ijk} is the totally antisymmetric tensor. The presence of three and four-point self-interactions is shown in the structure of gauge fields. B is the $U(1)$ field associated with the weak hypercharge $Y = QT_3$, where Q and T_3 are the electric charge operator and the third component of the weak $SU(2)$, respectively.

The scalar Lagrangian is

$$\mathscr{L}_\phi = (D^\mu \phi)^\dagger D^\mu \phi - V(\phi), \tag{2.14}$$

where $\phi = \binom{\phi^+}{\phi^0}$ is the complex Higgs scalar field, which is a doublet under $SU(2)_L$ with $U(1)$ charge $y_\phi = +1/2$. The gauge covariant derivative is defined as:

$$D_\mu \phi = (\partial_\mu + ig\frac{\tau^i}{2} W^i_\mu + \frac{ig'}{2} B_\mu)\phi \tag{2.15}$$

where τ_t are the Pauli matrices. The square of the covariant derivative leads to three and four-point interactions between the gauge and scalar fields.

$V(\phi)$ is the Higgs potential. The combination of $SU(2)_L \otimes U(1)_Y$ is invariant and renor-

malized according to: V to the form of

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.16}$$

There is spontaneous symmetry breaking due to the term μ^2. λ describes a quartic self-interaction between the scalar fields, with vacuum stability requiring that $\lambda > 0$.

$\mathscr{L}_f$ and $\mathscr{L}_{Yukawa}$ describe the free lepton Lagrangians and the Yukawa couplings between the single Higgs doublets, ϕ, and the various flavors of quarks and leptons, respectively. These couplings are needed to make the quark and lepton mass term generations in the SM possible.

2.2.4 The Strong Interaction

The quark model proposed by Gell-Mann [12] was not accepted because there was neither experimental evidence of quark existence nor the predicted particles formed by three u quarks, thus violating Pauli's exclusion principle (i.e. Δ^{++}). Later the idea of "color" quantum number (red, green, and blue) was introduced by Han and Nambu in 1965. This idea tells us that quarks can exist in three different color states and only colourless quark bound states can exist. It is then possible to explain non-observable free quarks and the existence of Δ^{++} that does not violate Pauli's exclusion principle, because the three quarks differ by their color charge.

Strong interaction is mediated by massless gluons. This is why the underlying theory is called QCD with its symmetry group $SU(3)_C$ [13, 14, 15]. The QCD coupling constant ranges over several orders of magnitude when moving from hard, i.e. large momentum transfer processes, to soft processes. Its value grows as the momentum transfer decreases. This effect is known as asymptotic freedom, and it justifies the use of perturbation theory (perturbative QCD or pQCD) when describing hard processes. At small energies (large distances), where the value of the coupling constant becomes large, the theory behaves in

a non-perturbative way; in such a regime the isolated quark or gluon cross sections vanish and are replaced by bound state dynamics. This effect is known as "confinement" and it justifies the non-observation of free quarks and gluons.

The strong interaction Lagrangian is written as:

$$\mathscr{L}^{QCD} = \mathscr{L}_{inv} + \mathscr{L}_{gaugefix} + \mathscr{L}_{ghost} \tag{2.17}$$

$\mathscr{L}_{inv}$ is invariant under local $SU(3)_C$ transformations and can be expressed as:

$$\mathscr{L}_{inv} = \sum_f \bar{\psi}_f (i\gamma_\mu D^\mu - m_f)\psi_f - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{2.18}$$

where f runs over the six quark fields, and D^μ is the covariant derivative as:

$$D_\mu = \partial_\mu + igA_{\mu a} T_a \tag{2.19}$$

and the field tensor

$$F_{\mu\nu} = \partial_\mu A_{\nu a} - \partial_\nu A_{\mu a} - gC_{abc} A_{\mu b} A_{\nu c}, \tag{2.20}$$

where $A_{\mu a}$ are the fields of the eight coloured gluons, T_a are the eight generators of $SU(3)$, C_{abc} are the structure constants that define the commutation rules of the $SU(3)$ generators, and $\alpha_s = g_s/4\pi$ is the strong coupling constant.

$\mathscr{L}_{gaugefix}$ and $\mathscr{L}_{ghost}$ are needed for technical reasons connected to how the quantization of the QCD Lagrangian is performed [16].

2.2.5 Parton Distribution Function

Since quarks do not exist in an isolated state outside hadrons, any interactions between them must take into account the structure of those hadrons, in this case protons. The quark structure of protons is given by a set of functions called "Parton Distribution Functions", or PDFs. At a pp collision, the cross section for a $2 \to N$ process is conveniently expressed

as

$$d\sigma_{pp\to N} = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1,\mu_F^2) f_2(x_2,\mu_F^2) d\hat{\sigma}_{pp\to N}(\mu_F^2) \tag{2.21}$$

The expressions of each term are

- $\hat{\sigma}$ is the parton level cross section,
- x_1 and x_2 are the momentum fraction of the proton momentum carried by the two colliding partons,
- $f_{1,2}$ are the PDFs that describe the probability of partons carrying momentum fraction $x_{1,2}$,
- μ_F is the factorization scale that is the scale at which the separation between the hard perturbative interaction and the long distance, non-pertubative, evolution of the produced partons take place.

PDFs for quarks and gluons at a scale μ^2 = 10 GeV^2 were calculated by the MRST [17] collaboration in 2004 are shown in Fig. 2.2. A PDF evolution with differing scales is governed by the DGLAP equation [18], as long as $\alpha_s(Q)$ remains in the perturbative validity region. DGLAP equation allows global fits of a variety of data taken from different experiments, at different scales. Two collaborations, CTEQ [19] and MRST are the main providers of global PDF fits using the DGLAP equation, which proves useful when comparing a variety of data taken from different experiments or scales.

2.3 Angular Correlations of a Jet Associated with the Z Boson

In the SM, high energy pp collisions can give rise to final states in which a Z boson is produced in association with a parton. At the LHC, these Z+jet events are predominantly produced by quark exchange processes (i.e. $q\bar{q} \to Z^0 g$ and $qg \to Z^0 q$, shown in Fig. 2.3). The angular distribution of the Z in the parton-parton center-of-momentum (CM) frame,

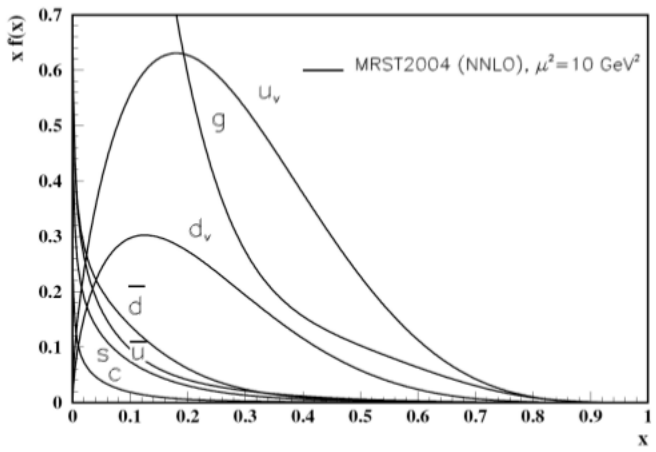


Figure 2.2: Distribution of $xf(x)$ as a function of the momentum fraction x at μ^2 =10 GeV^2 for different partons[20].

$\mathrm{d}\hat{\sigma}/\mathrm{d}\cos\theta^*$, is sensitive to the spin of the exchanged particle ($\mathrm{d}\hat{\sigma}$ is the partonic cross section and θ^* is the CM polar scattering angle). For example, a spin-$\frac{1}{2}$ propagator produces an angular distribution of the form $(1-|\cos\theta^*|)^{-1}$. At large momentum transfers, these events may be used to test the predictions of pQCD and to search for signatures of new physics.

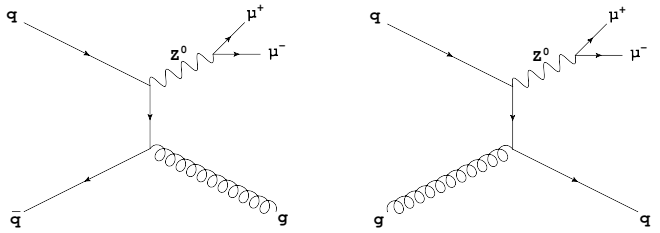


Figure 2.3: Two of the leading order (LO) t-channel Feynman diagrams for Z+jet production at the LHC.

A schematic representation of Z+jet production is shown in Fig. 2.4. The initial state partons carry a fraction of the incident proton's momentum, x_i. They interact to form a boosted state of mass M moving with momentum fraction $x = x_1 - x_2$ and rapidity $y = \frac{1}{2}\ln\left[(E+p_z)/(E-p_z)\right]$, where E is the energy and p_z is the component of the momentum

along the beam axis. This state then decays into a two body final state with measured y and transverse momenta (p_T). Distributions measured in the lab, however, are related to the cross section of an interaction.

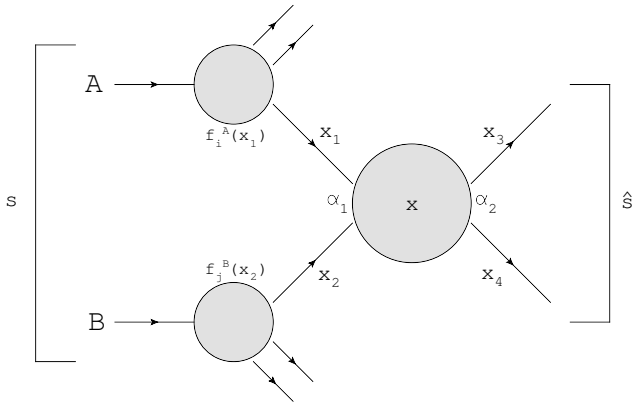


Figure 2.4: Schematic representation of the partons in a proton-proton collision, along with the kinematic variables that describe the scattering. The s represents the proton-proton CM energy. All other variables are explained in the text. For a Z+jet event, x_1 and x_2 represent the incoming partons, x_3 represents the Z boson, and x_4 represents the outgoing jet.

The invariant cross section for inclusive hadronic reactions can be decomposed as a partonic cross section multiplied by PDF [21],

$$\frac{E_3 E_4 \mathrm{d}^6\sigma}{\mathrm{d}^3 p_3 \mathrm{d}^3 p_4} \rightarrow \frac{\mathrm{d}^3\sigma}{\mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}\hat{t}} \propto \sum_{i,j} f_i(x_1) f_j(x_2) \frac{d\hat{\sigma}}{d\hat{t}}(\hat{s},\hat{t},\hat{u}), \tag{2.22}$$

where σ is the inclusive hadronic cross section; p is the particle momentum; f_i is the probability that a parton of type i $\left(i = \mathrm{u}, \bar{\mathrm{u}}, \mathrm{d}, \bar{\mathrm{d}}, \mathrm{g}, \ldots\right)$ carries a given fraction of the incident proton's momentum; and $\hat{s}$, $\hat{t}$ and $\hat{u}$ are the partonic Mandelstam variables. The partonic cross section takes the form of $\hat{\sigma} \propto (\alpha_s \alpha) \left|\mathscr{M}\right|^2 / \hat{s}$, where α_s and α are the strong and electromagnetic coupling constants, respectively. The relevant Z+jet scattering amplitudes, $\left|\mathscr{M}\right|^2$, are listed in Table 2.1. Note that the partonic cross section is solely a function of

Table 2.1: Partonic scattering amplitudes for Z+jet production at the LHC. The hat notation has been removed for simplicity.

Process	$\propto \lvert\mathscr{M}\rvert^2$
$q+\bar{q} \to Z^0+g$	$\frac{8}{9}\left(t^2+u^2+2sm_3^2\right)/tu$
$q+g \to Z^0+q$	$-\frac{1}{3}\left(s^2+u^2+2tm_3^2\right)/su$

the Mandelstam variables. In the CM frame, they simplify to,

$$\begin{aligned} s &= (P_1+P_2)^2 = (2p_1)^2 \\ t &= (P_1-P_3)^2 = m_3^2 - \frac{s}{2}(1-\cos\theta) \\ u &= (P_1-P_4)^2 = -\frac{s}{2}(1+\cos\theta), \end{aligned} \tag{2.23}$$

where P_i and p_i are the four- and three-momentum vectors, respectively (using the numbering scheme in Fig. 2.4, with the partonic notation removed for simplicity). The second equality in Eqs. 2.23 assumes that partons/jets are massless ($m_1 = m_2 = m_4 = 0$).

Therefore, one can argue that the invariant cross section is solely a function of the PDFs, $\hat{s}$ and $\cos\theta^*$. This can be written explicitly by parametrizing the phase space in Eq. 2.22 as

$$\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}\hat{t} \propto \mathrm{d}(p^*)^2\mathrm{d}y_B\mathrm{d}\cos\theta^*,$$

where y_B is the rapidity of the CM or boosted system (see Fig. 2.4) and p^* is the Z or jet momentum in the CM frame (where they are back-to-back). One can then solve for the angular distribution of the Z, such that

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta^*} \propto \sum_{i,j}\int f_i(x_1)f_j(x_2)\frac{\mathrm{d}\hat{\sigma}_{ij}}{\mathrm{d}\cos\theta^*}\mathrm{d}y_B\mathrm{d}p^* \tag{2.24}$$

$$x_{i,j} = \frac{4\left(p^*\right)^2}{s}e^{\pm y_B}, \tag{2.25}$$

Note that the variables y_B and p^* will be integrated over, which will play a major role in the accessible phase space. While QCD predicts the form of the Z's angular distribution in the CM frame, the final state particle kinematics are measured in the lab frame. We therefore apply a Lorentz boost on the Z and jet four-momentum vectors in order to transform them to the CM frame.

We can now reconstruct the boosted system (by summing the four-momentum vectors of the Z and jet in the lab) and use it to perform a Lorentz boost. This transforms the four-momentum vector of the Z+jet from the lab frame to the CM frame.

However, It is informative to study the relationship between lab and CM frame variables explicitly. Since rapidities are additive under Lorentz transformations, one can write

$$y = y_B + y^*, \tag{2.26}$$

where y^* and y are the rapidities of the Z+jet in the CM and lab frame, respectively, and y_B is the rapidity of the boosted system. One can also write the energy, E^*, and the longitudinal momentum, p_z^*, as a function of rapidity,

$$\begin{aligned} E* &= m_{\mathrm{T}} \cosh y^* \\ p_z^* &= m_{\mathrm{T}} \sinh y^* \\ m_{\mathrm{T}} &\equiv \sqrt{M^2 + p_T^2}, \end{aligned} \tag{2.27}$$

where m_{T} is the transverse mass and M is the mass of the particle ($M = 0$ for jets). Equations 2.27 holds true in all reference frames.

One can then derive the angular distribution by calculating $p_z^*/E^* = \tanh y^*$ and noting that p_z^*/E^* can be rewritten as $p^* \cos\theta^*/E^* = \beta^* \cos\theta^*$ (since $p^*/E^* = \beta^*$). Therefore, $p_z^*/E^* = \beta^* \cos\theta^*$. Finally, equating both ratios gives

$$\beta^* \cos\theta^* = \tanh y^* \tag{2.28}$$

For massless particles, β^* is equal to one. The measurement of differential cross section in $\cos\theta^*$ probes the spin of the particle exchanged between incoming partons: the spin-$\frac{1}{2}$ quark propagator in the Z +jet production introduces a specific behavior of differential cross section where

$$\frac{d\sigma}{d\cos\theta^*} \propto \frac{1}{1-|\cos\theta^*|} \tag{2.29}$$

If the observed $\cos\theta^*$ distribution deviates from predictions, it may indicate a presence of a new particle or that an improvement of pQCD is required. There are fundamental variables in the lab frame which are related to $\cos\theta^*$: the absolute rapidity difference and the absolute rapidity sum of the Z boson and the jet, namely $|0.5(\mathrm{y}_Z - \mathrm{y}_{jet})|$ and $|0.5(\mathrm{y}_Z + \mathrm{y}_{jet})|$; the former is closely related to y^* and $\cos\theta^*$ while the latter is closely related to y_B. For a 2→2 process with massless outgoing particles (12→34), one could derive that $y = 0.5(y_3 - y_4)$ and $y_B = 0.5(y_3 + y_4)$, where y_3 and y_4 are the rapidities in the lab frame of the outgoing particles 3 and 4. The $|0.5(\mathrm{y}_Z + \mathrm{y}_{jet})|$ is expected to be sensitive to the choice of PDFs, while $|0.5(\mathrm{y}_Z - \mathrm{y}_{jet})|$ is expected to be insensitive to PDFs.

The angular distribution of photon+jet, W+jet, and dijet events have been previously measured at the Tevatron CDF [22] (see Fig. 2.5) and D0 [23] detectors, as well as the LHC CMS [24] and ATLAS [25] detectors. This analysis presents, for the first time in Z+jet events, the angular distributions. The results will be good supplements of previously studies. Furthermore, this analysis includes regions of phase space that were not available to previous studies; we are able to probe larger values of $\cos\theta^*$ as a result of an increased reach in values of s and y. The data used in the analysis correspond to $\mathscr{L}_{int} \approx 5.1$ fb^{-1}, which was taken in 2011 using the CMS detector.

2.4 Production of Jets Associated with the Z Boson

Since the discovery of the Z boson at CERN in 1983, the study of jets associated with Z boson has been of increasing importance. The production of a Z in association with

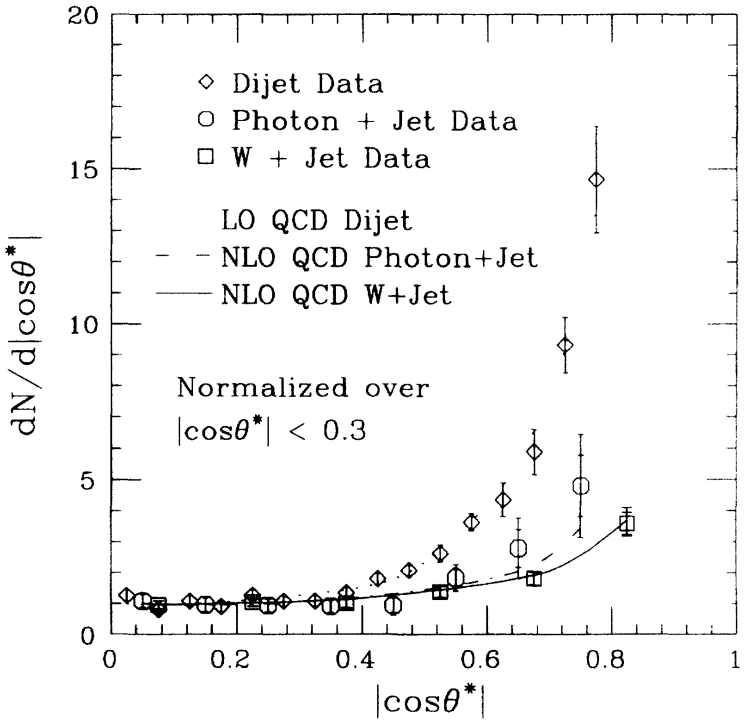


Figure 2.5: dN/d|cosθ^*| for CDF W+jet data (squares), compared to previously published measurements of $|\cos\theta^*|$ for dijet and photon+jet data. Next-to-leading order (NLO) QCD predictions are compared with the W+jet (solid curve) and the photon+jet (dashed curve) data. A LO QCD prediction (dotted curve) is compared to the dijet data. The data and theoretical predictions are all normalized to have an average value of 1 in the region $|\cos\theta^*| < 0.3$ [22].

jets provides a good opportunity to test predictions of pQCD at large momentum transfers. Moreover it is also an important physics signal at hadron colliders, such as Fermilab Tevatron and the LHC. Studying this process helps to constrain a major background to many SM signals, e.g. top pair and single top production. In addition to that, many searches for new particles such as the Higgs boson or SUSY particles involve final states containing one or more jets with a pair of opposite charged leptons as their signatures. Z+jets were studied at 1.96 TeV at Tevatron [26, 27, 28] and at 7 TeV with 36 pb^{-1} by ATLAS [29] and CMS [30]. Good agreement with the CMS and ATLAS data is shown in the LO and NLO prediction up to 4 jets in the final state. The precision varies from 10% up to 30% [31, 32] due to uncertainties on the PDFs and on the perturbative nature of the calculations.

The production of a Z boson in association with hadronic jets can be understood with: first, LO quark-antiquark annihilation, known as the weak Drell-Yan process (see next section), in hard scattering processes; second, multijet perturbation where the Z boson is balanced by gluon radiations and quark-gluon processes of the colliding partons prior to their annihilation.

2.4.1 Drell-Yan Process

A lepton-antilepton pair with large invariant mass, $M^2 = (p^{l+} + p^{l-})^2 \gg 1$ GeV^2, created though the quark-antiquark pair annihilation can be described by Drell-Yan [33] process:

$$pp \to q\bar{q} \to l^+l^- + X, \tag{2.30}$$

where X is a generic hadronic final state consistent with energy and momentum conservation, shown in Fig. 2.6.

The inclusive cross section $\sigma_{pp\to l^+l^-+X}$ can be calculated by finding all possible sub-process cross sections, $\hat{\sigma}_{q\bar{q}\to l^+l^-}$, from all quark and anti-quark combinations which are available in protons. Each sub-process cross section is then weighted with PDFs, $f_q(x)$ and $f_{\bar{q}}(x)$, of each parton carrying momentum fractions (x_1 and x_2) and summing over all

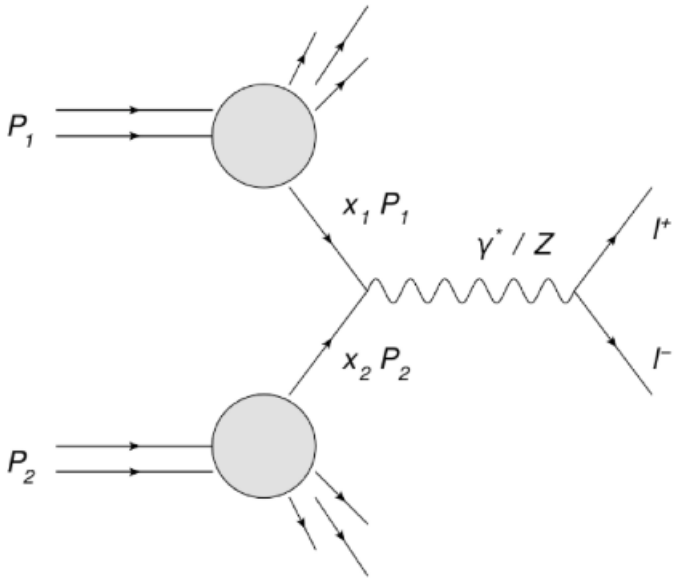


Figure 2.6: Drell-Yan process at leading order of $pp \to q\bar{q} \to l^+l^- + X$.

combinations. Thus, the Drell-Yan cross section can be written by:

$$\sigma^{DY} = \sum_q \int dx_1 dx_2 f_q(x_1) f_{\bar{q}}(x_2) \hat{\sigma}_{q\bar{q}\to l^+l^-}. \tag{2.31}$$

$f_q(x)$ and $f_{\bar{q}}(x)$ are obtained from various deep inelastic scattering experiments, including those from LHC. This inclusive cross section is valid only in the limit where quarks are asymptotically free at 0^{th} order of α_s. The lowest-order total cross section for quark-antiquark annihilation into a lepton pair via a off-mass-shell photon γ^* is given by:

$$\hat{\sigma}_{q(p_1)\bar{q}(p_2)\to l^+l^-} = \frac{4\pi\alpha^2}{2\hat{s}} \frac{1}{N_C} Q_q^2, \tag{2.32}$$

where $\hat{s} = (p_1 + p_2)^2 = x_1 x_2 s$, p_1 and p_2 are the parton four-momentum, N_C is the number of colours, Q_q^2 is a quark fraction charge and α is the electromagnetic coupling constant.

Naturally, from an incoming quark and anti-quark with a spectrum of energies $\sqrt{\hat{s}}$, it is more appropriate to consider the differential lepton pair mass distribution. For producing a lepton pair mass M, the differential cross section from incoming quark and anti-quark with

a spectrum of energies is

$$\frac{d\hat{\sigma}}{dM^2} = \frac{4\pi\alpha^2}{3M^2N_C}Q_q^2\delta(\hat{s}-M^2). \tag{2.33}$$

Substituting the Eq. 2.33 into Eq. 2.31 gives a parton model differential cross section for the Drell-Yan process at the LO:

$$\begin{aligned}\frac{d\sigma^{DY}}{dM^2} &= \int_0^1 dx_1 dx_2 \sum_q f_q(x_1) f_{\bar{q}}(x_2) + (q \leftrightarrow \bar{q}) \times \frac{d\hat{\sigma}}{dM^2}(q\bar{q} \rightarrow l^+l^-) \\ &= \frac{4\pi\alpha^2}{3M^2N_C}\int_0^1 dx_1 dx_2 \delta(x_1x_2s - M^2) \times [\sum_q Q_q^2 f_q(x_1) f_{\bar{q}}(x_2) + (q \leftrightarrow \bar{q})] \end{aligned} \tag{2.34}$$

The perturbative QCD corrections can be used to improve the LO result of Eq. 2.34 by taking into account higher order α_s. Three different types of higher α_s contributions (Fig. 2.7) are

- virtual gluon corrections to the LO contribution,
- real gluon corrections,
- quark-gluon scattering process together with the corresponding qg contribution.

It is worth nothing that virtual corrections do not modify the final state, while real corrections appear in the form of additional jets in the final state. As the result of these corrections, the PDFs acquire a logarithmic mass dependence and thus Eq. 2.31 can be expressed as

$$\sigma_{DY} = K\sum_q \int dx_1 dx_2 f_q(x_1, M^2) f_{\bar{q}}(x_2, M^2) \hat{\sigma}_{q\bar{q}\rightarrow l^+l^-} \tag{2.35}$$

The mass-dependent PDFs are obtained from various experiments of inclusive inelastic scattering, and K represents a constant factor. Currently, calculations at the next-to-next-to-leading order (NNLO) are available at [34].

The calculations so far are described by taking into account an intermediate state of γ^*. In order to adapt these calculations for the case of the Z boson production and its leptonic

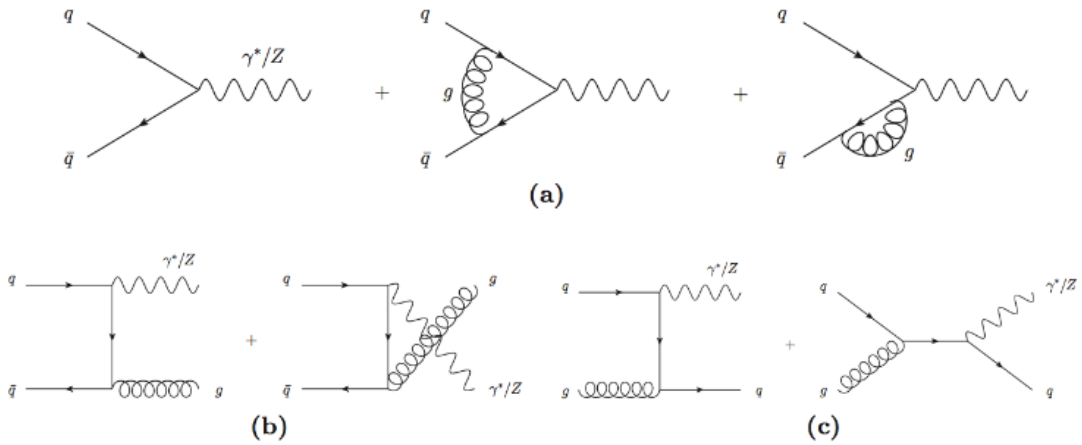


Figure 2.7: The LO and NLO diagrams for the Drell-Yan process. The processes are corresponding to a) virtual gluon corrections to the LO contribution b) real gluon corrections and c) quark-gluon scattering process together with the corresponding qg contribution.

decay, it is sufficient to substitute the $\hat{\sigma}_{q\bar{q}\to\gamma^*\to l^+l^-}$ with the cross section for the process $\hat{\sigma}_{q\bar{q}\to Z\to l^+l^-}$:

$$\hat{\sigma}_{q\bar{q}\to Z\to l^+l^-} = \hat{\sigma}_{q\bar{q}\to Z} \cdot BR(Z\to l^+l^-), \tag{2.36}$$

where $\hat{\sigma}_{q\bar{q}\to Z}$ is the production cross section of the Z boson and $BR(Z\to l^+l^-)$ is the branching ratio of the leptonic decay mode. Given that the decay width of the Z boson is small (Γ = 2.5 GeV) compared to its mass ($M_Z \approx 91$ GeV), it is sufficient to consider the production of effectively stable particles. Therefore, the on-shell Z production cross section can be approximated as:

$$\hat{\sigma}_{q\bar{q}\to Z} = \frac{\pi}{3}\sqrt{2}G_F M_Z^2 (V_q^2 + A_q^2)\delta(\hat{s} - M_Z^2), \tag{2.37}$$

where V_q^2 and A_q^2 are associated with the vector and the axial coupling constants of the neutral current interaction.

2.4.2 Multijet Production

The finite transverse momentum of the observed Z bosons can be attributed to multiple gluon radiations. Without any pertubative and non-perturbative corrections, the LO process $q\bar{q} \to Z/\gamma^*$ would result in the production of a vector boson with minimal transverse momentum. In fixed-order pQCD, the partonic cross section is calculated by expansion in powers of the strong coupling constant α_s.

The total Z production cross section can also be described by the sum of multijet components with increasing order in α_s:

$$\sigma_Z = \sigma_{Z+0\,jet} + \sigma_{Z+1\,jet} + \sigma_{Z+2\,jets} + \ldots, \tag{2.38}$$

and the jet multiplicity cross section can be perturbatively expanded by:

$$\begin{aligned} \sigma_{Z+0\,jet} &= a_0 + \alpha_s a_1 + \alpha_s^2 a_2 + \ldots \\ \sigma_{Z+1\,jet} &= \alpha_s b_1 + \alpha_s^2 b_2 + \ldots \\ \sigma_{Z+2\,jets} &= \alpha_s^2 c_2 + \ldots \\ &\ldots \end{aligned} \tag{2.39}$$

The coefficients a_i, b_j, c_k, ... in these expansions are general functions of the jet definition parameters, i.e. the cone size used to cluster the partons into jets, the transverse momentum, rapidity and separation cuts imposed on the jets or the clusters. LO contributions (a_0, b_1, c_2) can be described by the matrix elements for the parton process. It is worth noting that the sum of the parameters at each order in perturbative theory,

$$\begin{aligned} a_0 &= \delta_0 \\ a_1 + b_1 &= \delta_1 \\ a_2 + b_2 + c_2 &= \delta_2 \end{aligned} \tag{2.40}$$

is independent from the jet parameters and represents the perturbative expansion in power

of α_s of the total cross section.

The greatest contribution to the Z + jets cross section is given by the first coefficients of the exclusive multijet cross sections, $a_i, b_j, c_k, \ldots$. They can be obtained from the Feynman diagrams of the partonic processes $xy \to Z + j_1 \ldots j_n$, where x, y, j_i are quarks and gluons. The explicit calculations of these coefficients were carried out by Berends and Giele for up to 4 jets [35]. They also investigated the ratio between the cross section with n jets and the n-1 jets:

$$f_n(Z) = \frac{\sigma_{Z+n_{jets}}}{\sigma_{Z+(n-1)_{jets}}} \tag{2.41}$$

finding that this expression should be constant. It is possible to parametrized $f_n(Z)$ as

$$f_n(Z) = \alpha + \beta n_{jets}, \tag{2.42}$$

This scaling has been tested in various experiments, for example at the Tevatron [36] and LHC.

Chapter 3

The CMS Experiment at LHC

The LHC, the most powerful hadron collider, is located at the European Laboratory for Particle Physics (CERN). Slated to provide pp collisions with up to $\sqrt{s}$ = 14 TeV, the LHC is a unique tool for particle physics research since it is not only being used to probe the SM with high precision, but also for new physics searches (SUSY, Higgs production, etc.).

The four detectors at the LHC are the Compact Muon Solenoid (CMS)[38], A Large Torodial LHC Apparatus (ATLAS) [39], Large Hadron Collider b-quark experiment (LHC-b) [40], and A Large Ion Collider Experiment (ALICE) [41]. CMS and ATLAS are multi-purpose experiments, probing new physics at the TeV scale. LHC-b is a specialized experiment dedicated to the study of bottom quark physics. It will attempt to shed light on CP violation. ALICE is dedicated to heavy ion collisions and will primarily study quark-gluon plasma (QGP).

3.1 The Large Hadron Collider

The LHC is a circular pp collider, which is housed in a tunnel on the Swiss-French border. LHC was chosen to replace the Large Electron Positron collider (LEP) due to the lower amount of synchrotron radiation emitted by circulating protons. The LHC accelerator was installed in a tunnel with a circumference of 27 km about 50 to 175 m underground. A schematic explanation of the LHC accelerator and services is publicized in Fig. 3.1

The LHC has two counter-circulating proton beams which have been accelerated to a maximum energy of 8 TeV through the following stages: up to 50 MeV with a linear

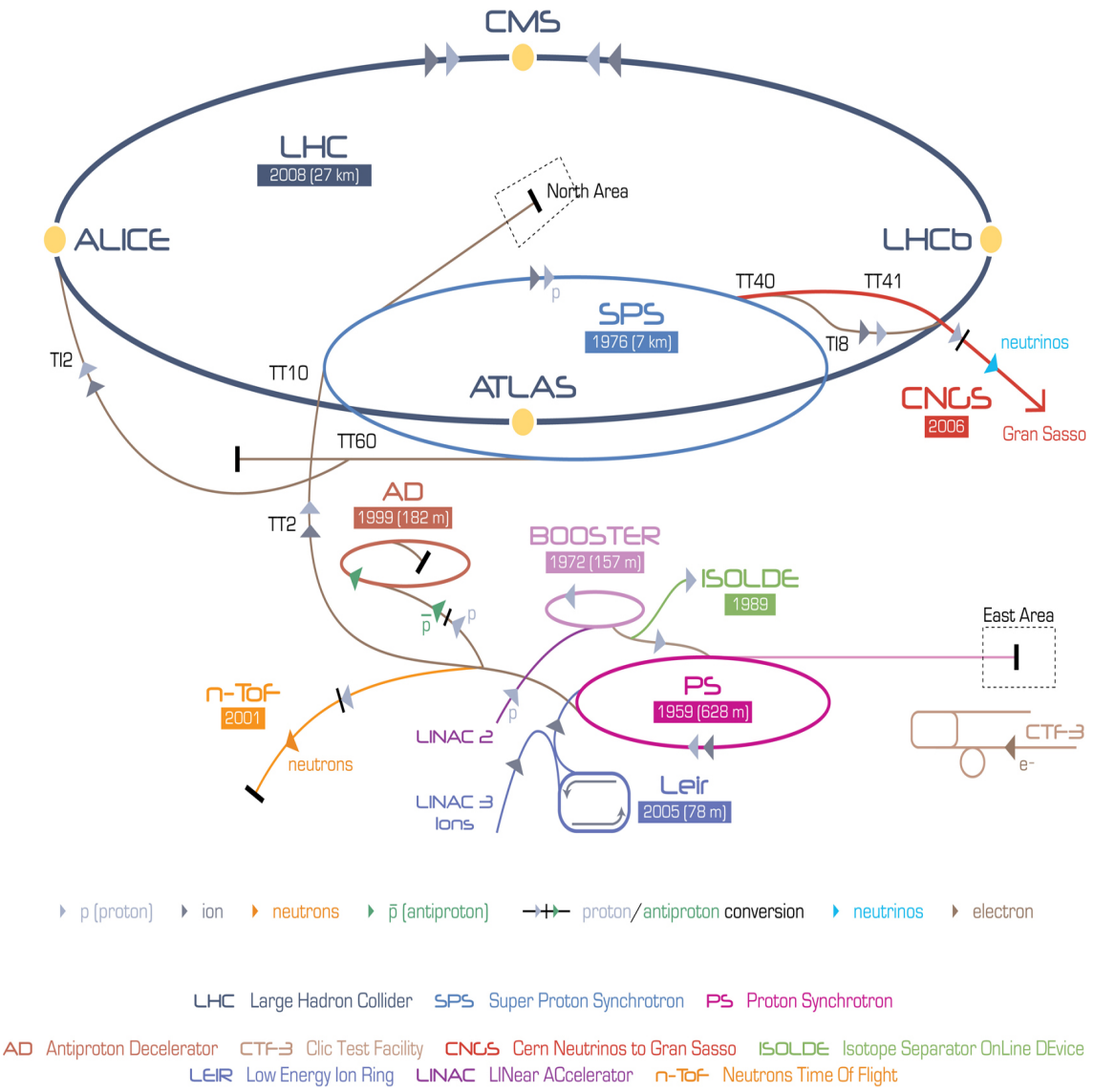


Figure 3.1: The LHC accelerator site at CERN.

accelerator (LINAC); up to 1.4 GeV by a booster; up to 25 GeV by a proton synchrotron (PS); up to 450 GeV by a super proton synchrotron (SPS); finally, up to 8 TeV after injection into the LHC ring. The two beams will then collide at interaction points built within each of the four experiments. Given the innovation of the LHC, with its high dynamic energy range, research can be performed on mass scales ranging from a few GeV, as in the case of b-meson physics, up to a few TeV to discover new vector bosons or quark compositeness.

A huge effort has been made to raise the proton momentum as much as possible in order to extend the capability of the LHC to discovering rare new physics processes. In particular, a very sophisticated magnet system is required to keep such high momentum protons in the machine orbit. The formula that connects the bending radius with the particle's momentum and the magnetic field is (upon simplification by multiplying the joule to GeV conversion factor)

$$B = \frac{p}{0.3\rho} \tag{3.1}$$

where B is the magnetic field in Tesla, p the momentum in GeV, and ρ is the orbit radius in meters. For a 27 km circumference and 7 TeV CM, the needed magnetic field for protons is about 5.4 T. In fact, since the LHC is made of curved and rectilinear sections, the bending magnetic superconductor dipoles can generate a field of 8.3 T. Since the energy of the beam is limited by the bending power of the magnetic system and the machine circumference, yet another technique for improving the search for new physics is to raise the luminosity, or rate of events occurring within the LHC. The event rate n for a process with cross section σ is

$$n = L\sigma \tag{3.2}$$

The luminosity is related to the beam properties with the approximation

$$L = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{3.3}$$

where n_1 and n_2 are the number of particles in beam 1 and 2 respectively, f is the collision frequency, σ_x and σ_y are transverse dimensions of the beams. At a frequency of about 40 MHz, the proton bunches at the LHC collide which corresponding to a spatial separation between bunches of about 7.5 m. The frequency cannot be raised further, because of the limiting requirement of avoiding further collisions on the side of each interaction region. The transverse dimensions of the beam can be squeezed down to 15 μm.

The LHC is designed to run at $L = 10^{34}$ cm^{-2} s^{-1} with 1.15×10^{11} protons per bunch and 2808 bunches per beam, separated by 25 ns (40 MHz bunch crossing rate). In 2011, a maximum of 4×10^{33} cm^{-2} s^{-1} was reached with 1.5×10^{11} protons per bunch (larger than design) and 1380 bunches per beam with a separation of 50 ns. In 2012, 23.3 fb^{-1}(21.8 fb^{-1}) were delivered to (recorded by) CMS with a maximum instantaneous luminosity of 7.7×10^{33} cm^{-2}s^{-1} [42].

In the experimental frame, the center-of-mass of the two hardly interacting partons is not motionless but it is on average boosted along the direction defined by the colliding beams. For this reason, boost invariant observables are very important to characterize the event. One of such observables is the transverse momentum p_T, defined as the projection of the momentum vector on a plane perpendicular to the beam axis.

Another useful observable is the rapidity y

$$y = \frac{1}{2} ln \frac{E + p_z}{E - p_z} = tanh^{-1}(\frac{p_z}{E}) \tag{3.4}$$

where E is the particle's energy, p_z is the particular momentum projection along the beam direction. Under a boost along z with speed β, y undergoes following the transformation: $y \rightarrow y$ - tanh^{-1} β , therefore the differences of rapidity are invariant, so the shape of the rapidity distribution dN/dy is invariant.

In the ultra-relativistic approximation the rapidity y is the same as the pseudo-rapidity

η defined as

$$\eta = -ln(tan\frac{\theta}{2}) \tag{3.5}$$

3.2 The CMS Detector

The CMS [38] is a large, high field superconducting magnet detector. The CMS main design priorities were a redundant muon tracking system, a good electromagnetic calorimeter and a high quality inner tracking system. The CMS structure consists of many cylindrical detecting layers, coaxial with the beam direction (barrel region), closed at both ends with disks (endcap region), and large pseudorapidity calorimeter close to beam line (forward region). Schematic view of the CMS detector, which is 28.7 m long, 15 m in diameter, and 14,000 tons of the total weight are shown in Fig.3.2

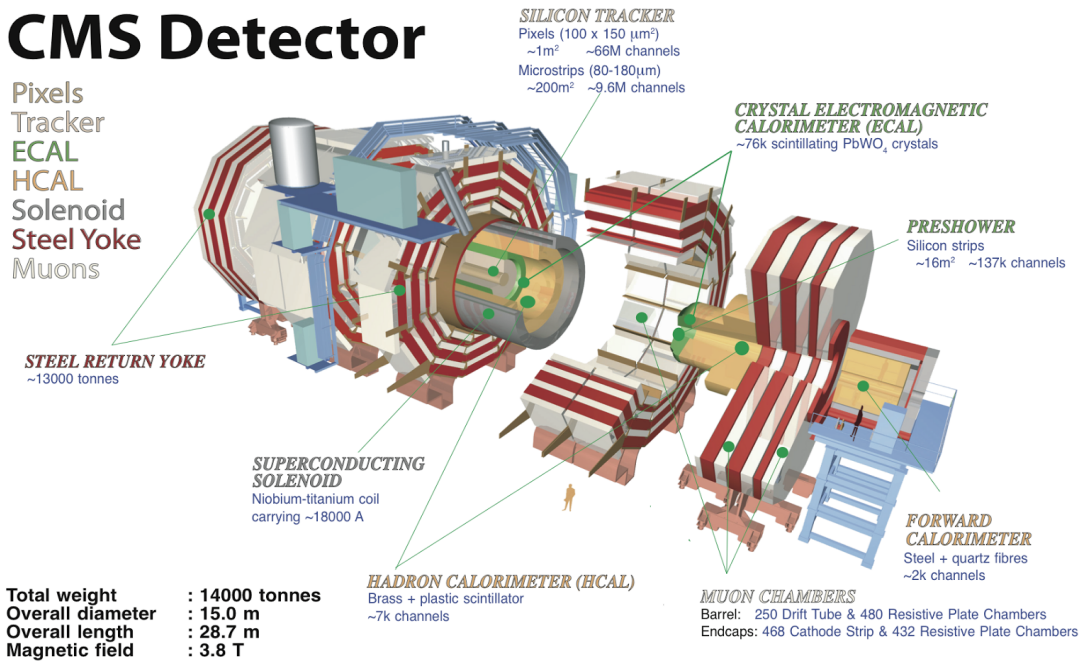


Figure 3.2: Illustration with details of the CMS detector.

A right-handed tern is used in the CMS coordinate frame. The direction of the right-handed tern consists of x axis pointing towards the LHC centre, y axis directed upward

along the vertical and z axis along the beam direction. The cylindrical symmetry of the CMS design and the invariant explanation of pp physics recommend the use of a pseudo-angular reference frame, given by the triple (r, ϕ, η) , where r is the distance from the z axis, ϕ is the azimuthal angle that measured starting from the x axis positive direction, and η is defined in Eq 3.5 where θ is the polar angle.

3.2.1 Magnet

CMS magnet is able to produce a 4 T magnetic field over a huge volume. The magnet is made of five modules, 2.5 m length, 6 m diameter, and 50 tons weight each. A Niobium-Titanium compound is used to make spires in each module. They are kept at 4 K with a liquid helium cooling system while in operation. A 20 kA current flows in the spires during process. 12000 tons of iron is built around the magnet to contain the field. The yoke consists of a barrel region which is made of five rings and two endcap regions on each side equipped with three disks. The strength of the field allows for a precise measurement of charged particle's momenta. In addition, the field in the return yoke, where muon chambers are located, permits an independent measurement of the muon's momentum. More details about magnet system used in CMS can be found in [43].

3.2.2 Tracker

The Silicon Tracker [44], the CMS innermost detector, consists of a Silicon Pixel detector and a surrounding Silicon Microstrip detector. It is designed to measure a precise momentum of charged particles and to allow an exact determination of the secondary vertice's position. Due to the complexity of LHC events, two fundamental requirements are needed to ease a complex pattern recognition problem, low detector occupancy by a highly granular detector and large hit redundancy by a large number of detecting layers. The tracker at the CMS detector is the largest exists silicon tracker of all physics experiments with a sensitive area of about 200 m^2. It has a length of 5.8 m, a radius of 1.2 m and covers

the range up to $|\eta| = 2.5$. An overview is shown in Fig. 3.3. The tracker is divided into two different parts.

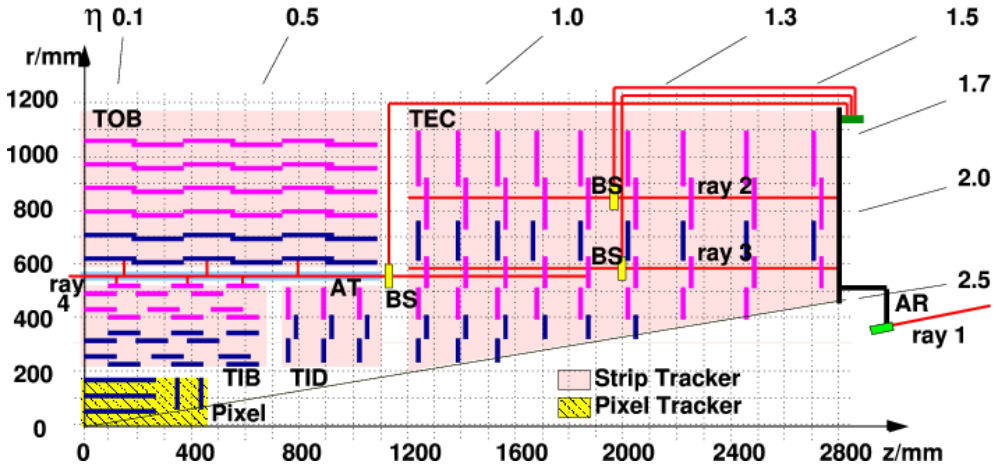


Figure 3.3: A quarter of the CMS silicon tracker in an rz view.

Pixel Tracker

The pixel detectors are chosen with a cell size of $100 \times 150\ \mu m^2$. In total, there are 1,440 modules with 66 million pixels. They are arranged in three cylindrical barrel layers with radial distances of 4.4, 7.3 and 10.2 cm to the beam line and two endcap discs at each side of the barrel (Fig. 3.4). If the particle has $|\eta| < 2.2$, it will be incident upon three combined pixel layers, otherwise with $2.2 < |\eta| < 2.5$ the particle passes through two combined pixel layers.

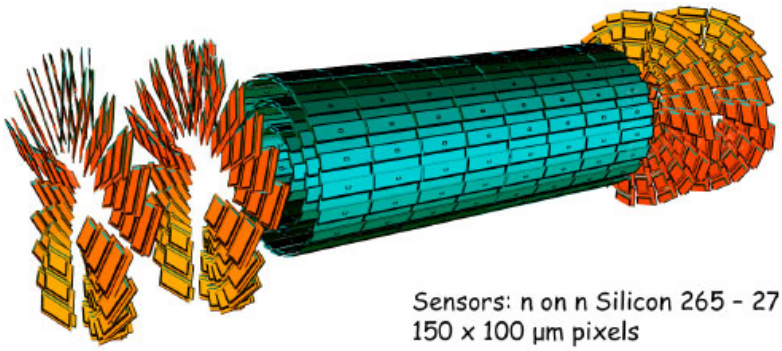


Figure 3.4: CMS pixel tracker system.

Silicon Strip Tracker

The silicon strip detectors are employed at larger radii of 20 to 116 cm and consist of 15,148 strip modules with about 10 million readout channels and are further subdivided into different parts. The Tracker Inner Barrel and Discs (TIB/TID) are composed of four barrel layers with strips parallel to the beam line and three endcap discs at each side with radial strips. The strips have a pitch of 80 to 141 μm of 10 cm and a sensor thickness of 320 μm resulting in an occupancy of 2 to 3 % and $r\phi$ resolution of 16 to 27 μm in the barrel. The Tracker Outer Barrel (TOB) has six layers and the Tracker Endcaps (TEC) has nine discs at each side use strips with 97 to 184 μm pitch, 25 cm length and 320 to 500 μm thickness. This leads to an occupancy at the percent level and $r\phi$ resolution of 25 to 41 μm in the barrel. Some of the inner layers of all strip tracker subdivisions have an additional module mounted back-to-back on the first one under a stereo angle of 100 mrad, thereby providing a position measurement in the direction along the strips (z in the barrel, r in the endcaps). The resulting z resolution is 230 μm in TIB and 530 μm in TOB [45].

3.2.3 Electromagnetic Calorimeter (ECAL)

The electromagnetic calorimeter (ECAL) [46] is designed to precisely measure energies and positions of photons and electrons. It is possible to promptly observe Higgs decay into two photons. ECAL also measures a part of energies of hadronic jets because hadrons usually begin showering in the ECAL and deposit some amount of their energy in ECAL.

ECAL is made of lead tungstate ($PbWO_4$) crystals (Fig. 3.5) which are a radiation resistant scintillating material. The radiation robustness was a core design requirement because the condition of absorbed dose per hour in high luminosity will range from 0.18 Gy/h at $\eta = 0$, to 15 Gy/h at $\eta = 3$. Lead tungstate has a short radiation length $X_0 = 8.9$ mm caused by its high density about 8.28 g/cm^2. Such a short radiation length allows a very compact design that made it possible to fit the calorimeter inside the magnetic coil. Another advantage of lead tungstate is the small Molière radius approximately 2.2 cm and

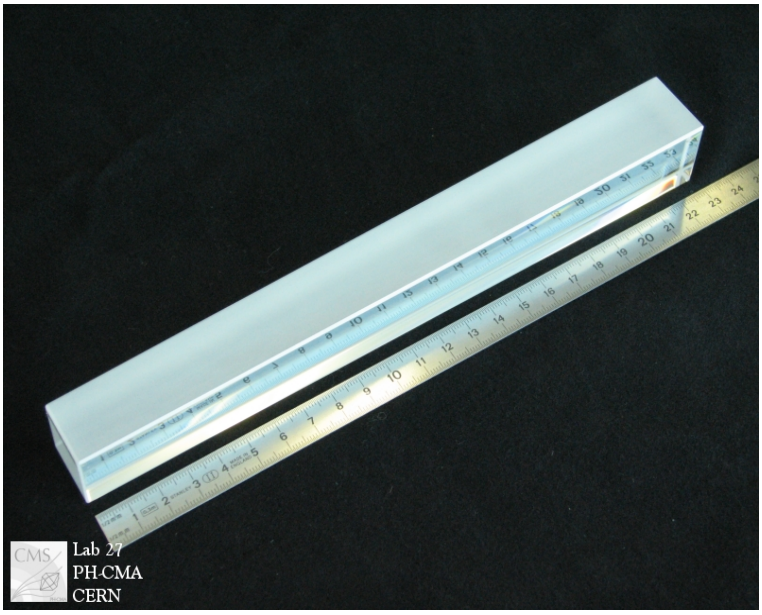


Figure 3.5: CMS electromagnetic calorimeter: lead-tungstate ($PbWO_4$) crystal.

fast scintillation decay time $\tau = 10$ ns that permits the collection of about 80% of the light production in the 25 ns interval between two bunch crossing. The main disadvantages of $PbWO_4$ are low light yield at just about 100 photons/MeV and the strong dependency of the response on the operating temperature (18° C). ECAL is divided into two parts (Fig. 3.6).A barrel region (EB) covering $|\eta| < 1.48$, and two endcap regions (EE) on both sides of EB covering $1.48 < |\eta| < 3.0$. Lead tungstate crystals are arranged in tapered shapes over the barrel area with the front face measuring 2.2 cm $\times$ 2.2 cm and 23 cm length. They are placed at a radius of 1.24 m and the radiation lengths about 26 X_0. The barrel granularity is $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$. Slightly larger crystals of 3 cm $\times$ 3 cm are used over the endcap regions of the front face. They are 22 cm long of radiation lengths about 24.7 X_0 with granularity of $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$.

The reduced depth in radiation length and the larger granularity in the endcaps with respect to the barrel are partially compensated with a preshower detector positioned in front of the endcaps. Each of preshower is made of two lead radiators and two silicon microstrip detector planes. The π^0 rejection power in the forward region is improved by the preshower.

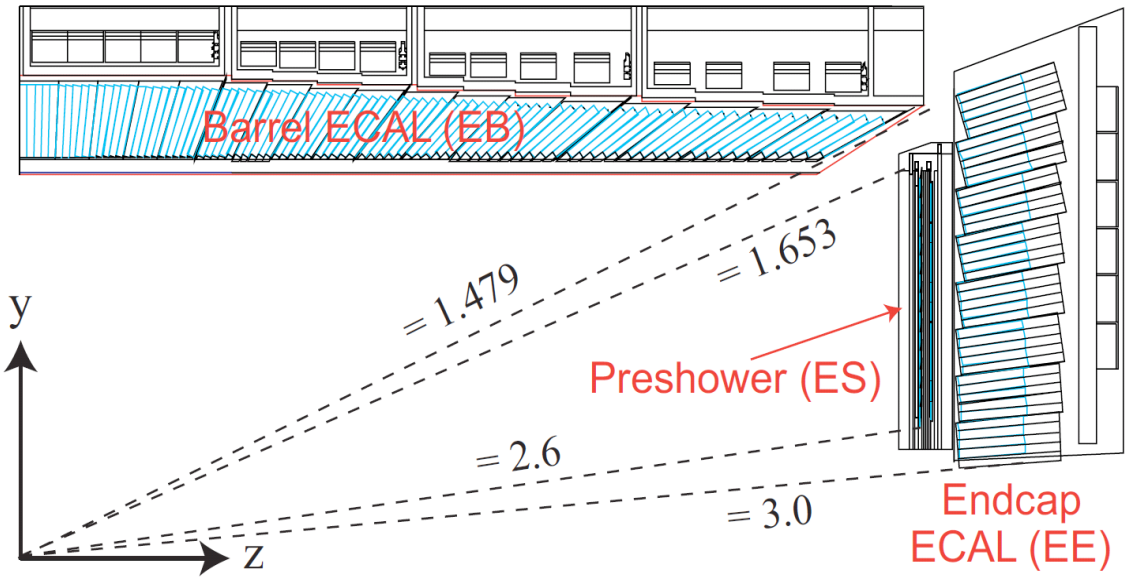


Figure 3.6: Geometric view of one quarter of the ECAL. Layout of the CMS electromagnetic calorimeter presenting the arrangement of crystal modules, supermodules, endcaps and the preshower in front.

Avalanche photodiodes (APD) are used in the system to read barrel region crystals. ECAL APDs are able to operate in the high magnetic field and can handle the low light yield of the crystals. Each crystal is equipped with two APDs that generate a total of about 4000 photoelectrons per GeV of deposited energy. Because of the higher radiation level in the endcaps, APDs would suffer from high leakage current, thus the forward crystals are readout with vacuum photodiodes (VPD) [47].

In the energy range of 25 GeV $< E <$ 500 GeV, the energy resolution σ_E of ECAL can be calculated as

$$(\frac{\sigma_E}{E})^2 = (\frac{a}{\sqrt{E}})^2 + (\frac{b}{E})^2 + c^2 \tag{3.6}$$

where:

- a is 2.7% GeV$^{1/2}$ in the barrel and 5.7% GeV$^{1/2}$ in the endcaps. It is a stochastic term and it is determined by the photoelectrons statistic.
- b is 155 MeV in the barrel and 200 MeV in thee endcaps. It is determined by electronic noise and pileup.

- c is 0.55% both in the barrel and in the endcaps. It is related to the longitudinal shower evolution containment, the light collection uniformity in the crystals and the inter-calibration precision between crystals.

3.2.4 Hadron Calorimeter (HCAL)

CMS Hadronic calorimeter (HCAL) [48] is used together with ECAL to determine energies and directions of jets as well as transverse energy E_T and imbalance missing traverse energy (E_T^{miss} or MET). It can provide good segmentation, moderate energy resolution and angular coverage up to $|\eta| < 5$. HCAL (Fig. 3.7) is divided into four parts:barrel (HB), endcap (HE), outer (HO), and forward (HF).

The HB and HE are placed between the ECAL and the magnet. They are mainly composed of 5 cm non-magnetic brass absorber interleaved with 3.7 mm plastic scintillator tiles although the first and last layers are made of steel. The HB covers $|\eta| < 1.3$, and the HE extends to $1.3 < |\eta| < 3.0$. Wavelength-shifting (WLS) fibers are embedded in the scintillator plates. Lights collected from the scintillators are readout by the Hybrid Photo Diodes (HPD) detectors. HB does not have enough depth (only 5.8 λ_I) to contain a full hadronic shower. The tail catcher outer detector (HO) had been added outside the magnetic coil in order to increase the calorimeter depth. HO is made from two scintillator layers with same granularity as HB. Thus, the HCAL in the barrel has the total depth about 11.8 λ_I. HE is located inside the magnetic coil with the same architecture and granularity of HB. The signal is read through wavelength-shift fibers and hybrid photodiodes. It has a sufficient depth of about 10 λ_I.

The forward calorimeter (HF) is the last subdetector of HCAL, located at $3.0 < |\eta| < 5.2$ outside the magnetic coil. Since the high particle flux is in this region, close to the beam line, HF requires radiation-hard materials. Steel as the absorber and Cherenkov-light-emitting quartz fibers are chosen as the active medium because of their radiation hardness. To compensate the missing ECAL in this forward region, quartz fibres of different lengths,

Short (1.43 m) and long (1.65 m) fibers are used. This arrangement makes it possible to distinguish showers generated by hadrons and electrons, which deposit a large fraction of their energy in the first 22 cm, from those generated by hadrons, which produce signals in both segments. HF calorimeter is designed to measure high energy jets with a good precision (20 % to 30 % at 1 TeV) [49].

The calorimeter is segmented and arranged in granularity $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and 0.17×0.17 at $|\eta| \geq 1.6$, whose axes point to the interaction point.

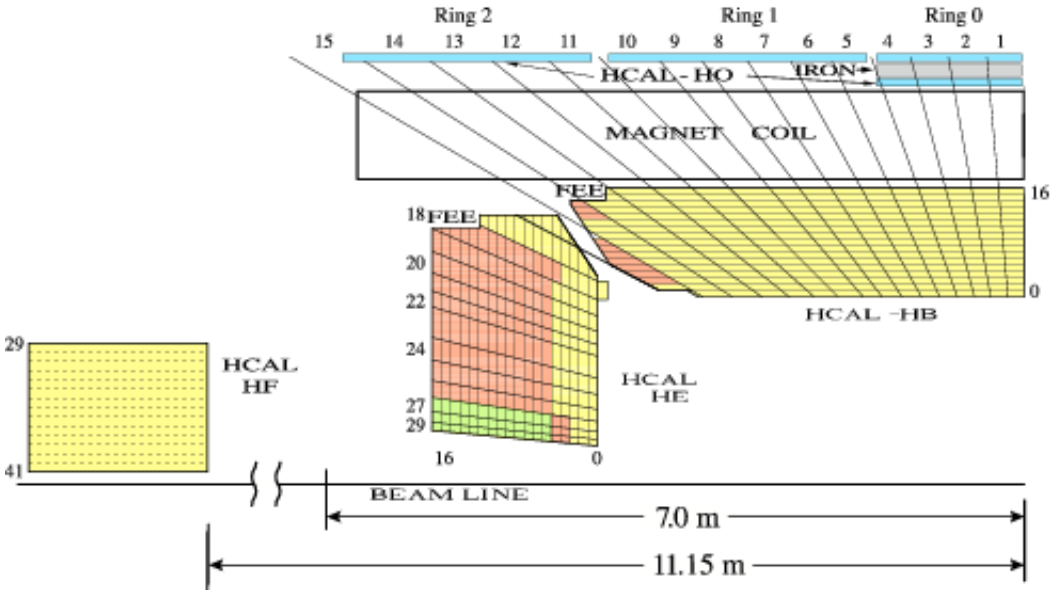


Figure 3.7: Geometric view of one quarter of the HCAL.

3.2.5 Muon System

Muons are typically considered as minimum-ionising particles (except at high energies) and traverse through whole detectors without significant energy loss in contrast to other particle types. Identifying muons correctly and reconstructing their momenta precisely is one of the design goals of the CMS. This signature often provides a good signal-to-background discrimination and therefore is used for triggering and for analysis.

The dedicated muon detectors [50] placed outside the calorimeters identify and measure high p_T muons in combination with the tracker. The system consists of three detectors

interleaved with iron return yoke plates that only muons and neutrinos can traverse. Due to the needed large area coverage at radii from 4 to 7 m, gaseous detectors are chosen. Different technologies are used as the conditions vary with pseudorapidity (Fig. 3.8):

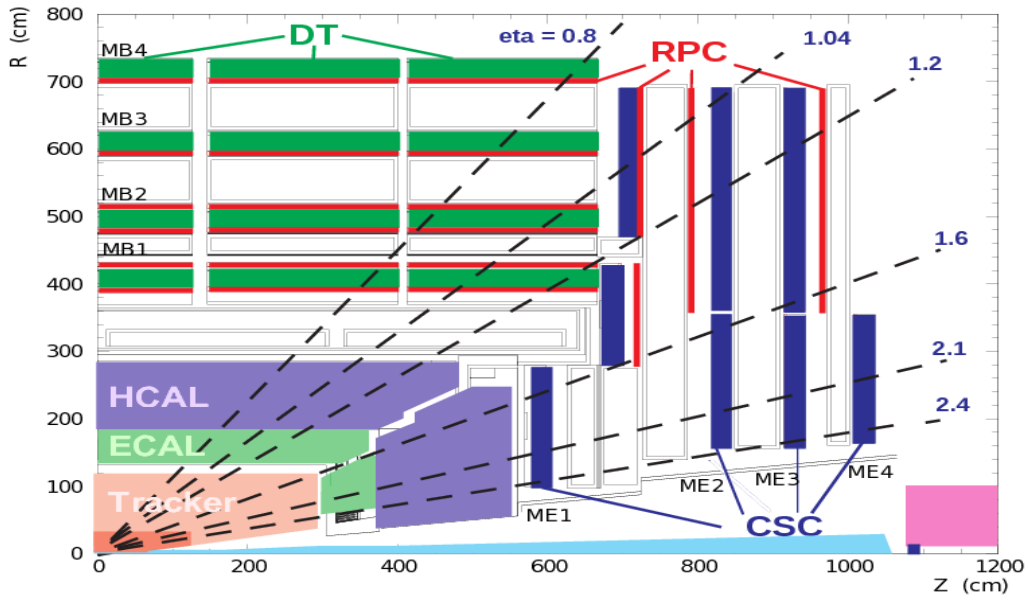


Figure 3.8: Layout of one quadrant of CMS. The four DT stations in the barrel (MB1-MB4, green), the four CSC stations in the endcap (ME1-ME4, blue), and the RPC stations (red) are shown.

- Drift Tubes (DT) are located in the barrel region ($|\eta| < 1.2$), where the occupancy is relatively low (< 10 Hz/m^2).
- Cathode Strip Chambers (CSC) are in the endcaps ($0.9 < |\eta| < 2.4$), where the occupancy is higher (> 100 Hz/m^2)
- Resistive Plate Chambers (RPC) is both in the barrel and endcaps.

The Drift Tube (DT) system is made of chambers consisting of twelve layers of drift tubes. In each layer, three independent substructures are packed, called super-layers. In each super-layer, there are four chambers of anode wires, two parallel and two perpendicular to the beam axis. Each super-layer can provide two $r\phi$ coordinate measurement and two z coordinate measurements of the track hit positions. Each chamber as shown in Fig.

3.9 is made of two parallel aluminum plates with "I" shaped spacer cathodes, isolated from the aluminum plates with polycarbonate plastic. Chambers are filled with a gas mixture of Ar (85%) and CO_2 (15%). The position resolution is about 100 μm in both $r\phi$ and rz.

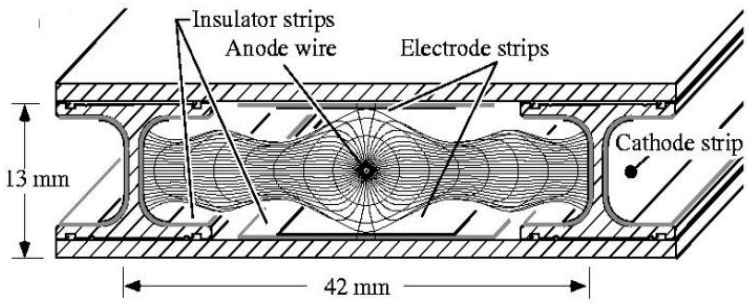


Figure 3.9: Schematic representation of a drift tube chamber.

The four stations of Cathode Strip Chambers (CSC), which can cope better with the higher particle flux and the stronger but less homogeneous magnetic field in the endcap region, are multi-wire proportional chambers with segmented cathodes as shown in Fig. 3.10. Each chamber can offer both hit position coordinates. Chambers are filled with a gas mixture of Ar (40%), CO_2 (50%), CF_4 (10%). The spatial chamber resolution is about 80-85 μm.

Both of the DTs in the barrel and the CSCs in the endcap up to $|\eta| = 1.6$ are complemented by Resistive Plate Chambers (RPC) which are made of parallel bakelite planes. The gap between the plates is filled with a mixture of $C_2H_2F_4$ (94.5%) and i-C_4H_{10}. They operate in avalanche mode with high resistivity. They provide good timing information, but a less precise position determination than the other systems. Thus, they are mainly used to improve the bunch-crossing assignment and for trigger purposes. The RPC is shown in Fig. 3.11.

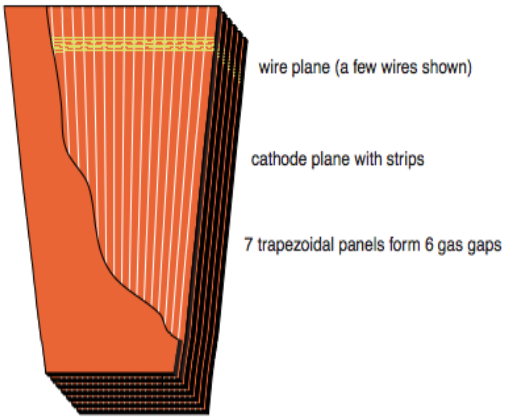


Figure 3.10: Schematic view of a CSC chamber.

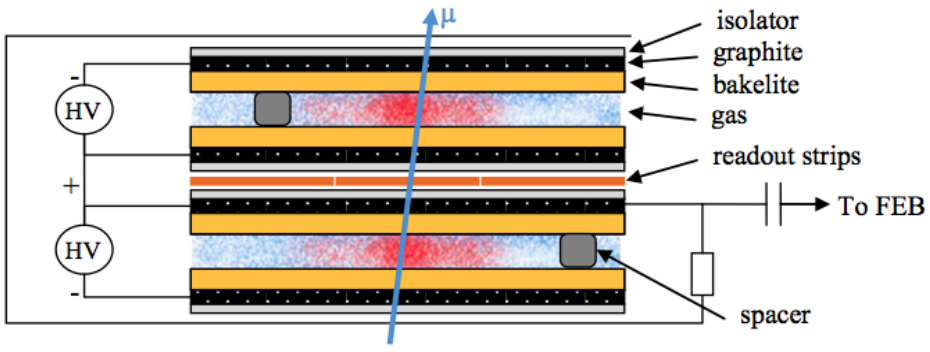


Figure 3.11: Schematic view of the RPC double-gap structure. The read-out strips in the Barrel chambers run along the beam direction.

3.3 Trigger System and Data Acquisition

The LHC crosses proton bunches 40 million times per second at a design luminosity of $L = 10^{34}$ cm^{-2} s^{-1}. Approximately 20 interactions will take place at each crossing. The data size of one event is about 1 MB. The amount of produced data exceeds what can be stored and analysed. However, only a small fraction of the interaction is worth recoding to disk. A rate suppression by more than five orders of magnitude and interesting event selections are needed. CMS has two-level trigger system, consisting of a level-1 Trigger (L1) [51] and a High Level Trigger (HLT) [52].

3.3.1 Level-1 Trigger

L1 operates on dedicated processors, and accesses coarse level granularity data from calorimetry and muon system. A L1 decision has to be taken for each bunch crossing within 32 μs. The task of L1 is used to decrease the flow of data from 40 MHz to 100 kHz. It is responsible for identifying electrons, muons, photons, jets and missing transverse energy. The output rate and speed of the L1 are limited by the readout electronics and the Data Acquisition system (DAQ) performances [52]. L1 is consisted of three subsystems: L1 Calorimeter Trigger, L1 Muon Trigger, and L1 Global Trigger. Fig. 3.12 shows the organization of CMS L1.

L1 Calorimeter Trigger

A calorimeter towers, clusters of signals collected from ECAL and HCAL, are the input for L1 Calorimeter Trigger. Towers are calculated by calorimeter high level readout circuit named Trigger Primitive Generators. The Regional Calorimeter Trigger finds out electron, photon, tau and jet candidates along with their transverse energy and drives them to the Global Calorimeter Trigger. The candidates are sorted by Global Calorimeter Trigger according to their transverse energy. The first four highest transverse energy candidates are

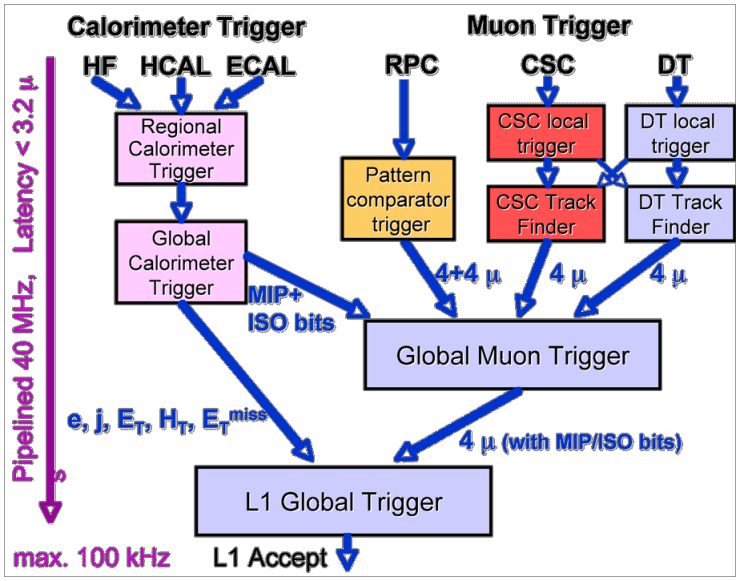


Figure 3.12: Schematic summary of Level-1 Trigger.

sent to the L1 Global trigger.

L1 Muon Trigger

RPC trigger electronic unit constructs Track Segments, gives the p_T estimation and sends these segments to the Global Muon Trigger. In addition, CSC logic unit also provides information to solve hit position ambiguities in case two or more muon tracks cross the same CSC chamber. The Local Charged Tracker (LCT), that is track segments made out of the cathode strips, is built by the CSC trigger. A p_T value and a quality flag are assigned to the LCTs. The best three from nine CSC chambers are passed to the CSC Track Finder that uses the full CSC information to build tracks, assign them to a p_T value and a quality flag, and then sends them to the Global Muon Trigger.

DTs are outfitted with Track Identifier electronic units that are able to align group hits in the four chambers of the super-layer. The track segments are sent to the DT Track Correlator that merges segments from two super-layers, measuring the ϕ coordinate. The best two segments are sent to the DT Track Finder which builds tracks and sends them to the Global Muon Trigger. The Global Muon trigger sorts the RPC, CSC and DT muon

tracks and join them. The final set of muons is sorted according to quality, and the best four quality muons are passed to the L1 Global Trigger.

L1 Global Trigger

The L1 Global Trigger has two missions. First, it is used to collect object created from the Calorimeter and Muon Trigger. Second, it is used to make a decision whether to retain the event or not. If the event is accepted, the decision is sent to the Timing Trigger and Control System which command the readout of the remaining subsystems. For taking the decision, The L1 Global Trigger sorts the rank of objects produced by calorimetry and muon system and then checks if at least one of the thresholds in the L1 requirement is approved.

3.3.2 High Level Trigger

HLT is designed for reducing the L1 output rate to the 100 events/s goal which is certainly going to be written to the mass storage. HLT code performs reconstruction using data from all subdetectors. The information read from subdetectors are assembled by a builder unit and then assigned to a switching network that dispatches events to the processor farm. The network of the CMS switching has a bandwidth of 1 Tbit/s. The time available for HLT to make a decision is approximately 300 ms. The real time nature of this selection imposes numerous constraints on the resources that an algorithm can use. The HLT algorithms reliability is extremely important, because events not selected by the HLT are lost.

In order to efficiently process events the HLT code will be able to reject not interesting events as rapidly as possible; computationally expensive algorithms must be run only on good candidates for interesting events. For achieving this requirement, the HLT code is organized in a virtually layered structure:

- Level 2: uses only muon and calorimetry information;

- Level 2.5: uses also the pixel information;
- Level 3: makes use of the full information from all the tracking detectors.

The number of events will be reduced in each step for processing in the next step. The most computationally expensive tasks are executed in the Level 3 such that time consuming algorithms, for example track reconstruction are only executed in the interest region. Moreover, since the ultimate precision is not necessary at the HLT, track reconstruction is performed on a limited hits set, and stopped once the requirement solution is achieved.

The instantaneous luminosity and along with it the event rate increase significantly as the operation goes on. To keep the total trigger rate constant, different techniques are possible. Ideally, one would find improved trigger algorithms with a higher background suppression at a constant efficiency. Alternatively, the trigger thresholds can be raised or additional identification criteria can be introduced, which mostly leads to a reduced signal acceptance or efficiency. Also, completely new trigger paths can be created by combining events, such as, single-lepton and jet trigger paths. If triggers with certain thresholds but too have rates that are too high want to be kept for dedicated purposes, they can be recorded with a prescale such that only every n-th events fulfilling the trigger requirements are stored. Fig.3.13 shows the structure of trigger and DAQ.

3.4 Luminosity Measurement

Instantaneous luminosity can be obtained by measuring the rate of a process with a precisely known cross section or other parameters that are correlated with the luminosity. HF calorimeter is used for online luminosity monitoring at the CMS [53]. Either the "zero-counting" method is applied, which measures the average fraction of empty HF towers, or the average transverse energy per tower is measured. For online luminosity determination, different methods are used, which are based on rate measurements in the HF or pixel. The luminosity used for this analysis was obtained with the "pixel-cluster-counting" method,

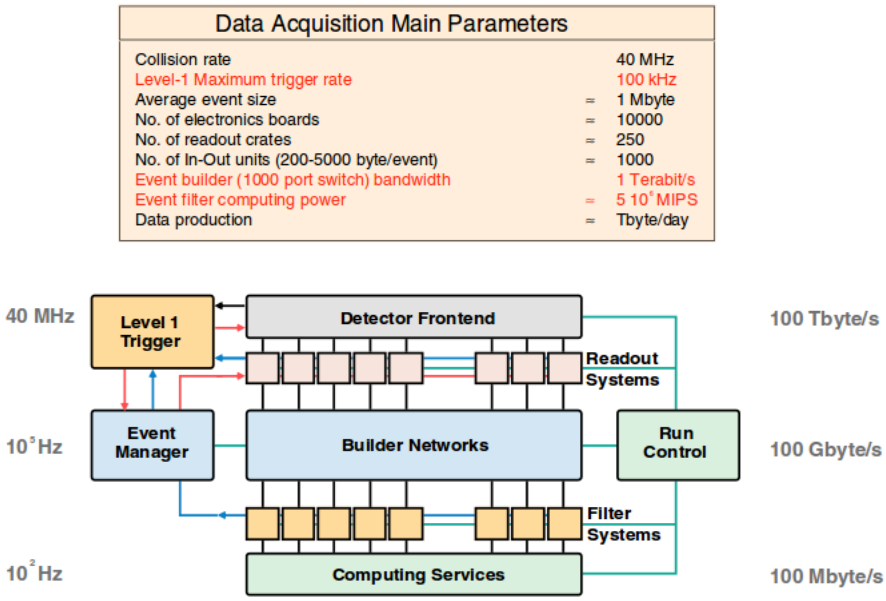


Figure 3.13: Trigger and Data Acquisition baseline structure.

which measures the number of clusters in the pixel system per bunch crossing [54].

3.5 CMS Simulation and Reconstruction Software

The prediction of the experiment outcome is an important task in developing, validating or falsifying theoretical models. Between the production of particles in a hard-scattering collision described at limited perturbative orders and the actual measurement of signals in a detector, the particle objects evolve through parton showering, hadronisation, decay and interaction with the detector materials. The complete chain of evolution of a collision event is simulated numerically with the help of Monte-Carlo (MC) event generators, utilising random numbers. The resulting output has the same format as recorded events in real data, so that the whole data analysis can be performed equally on simulated events. In fact, the simulation is often used to design and optimise the analysis strategy such as the event selections. After a thorough validation that the simulation gives reliable results, e.g. by data-to-simulation comparisons of control distributions, the simulation can be also used to correct the data to one of the intermediate levels. For example, it is used to estimate

reconstruction and selection efficiencies. Also, the background contaminations of selected samples can be estimated from the simulation.

3.5.1 Framework Implementation

The CMS simulation and reconstruction software, CMSSW [55], is a C++ framework that can be configured via Python [56] scripts. CMS Events Data Model (EDM) is based on the event concept. An event is a C++ class that contains the physics event information, both raw level data and reconstructed quantities. Reconstruction algorithms can use data from the event and put reconstructed quantities in the event. The event can be read from or written to ROOT [57] files.

The designed Python configuration script is feed into the executable "cmsRun" by CMSSW. The configuration files include the modules, such as the algorithms, which the user wants to run and they specify the order n which they require to be run. The executable reads in the configuration file and, using a plug-in manager, finds out in which libraries in the modules must be run and loads them. The CMSSW can be implemented with six types of modules and dynamically loaded via the plug-in mechanism.

There are six kinds of modules in CMSSW and promptly loaded via the plug-in mechanism, which are Source, EDProducer, EDFilter, EDAnalyzer,EDLooper, and OutputModule. More information about CMSSW framework can be found at [55].

3.5.2 Event Generation and Reconstruction

Event generation in CMSSW can be completed with lots of event generator programs. Those programs can be run from within the framework using delicate interface libraries. The event generator configurations perform feeding cmsRun with the suitable configuration file including the flags to be set in the event generator. The event generator is working to fill the HepMC [58] record with all the information about the presently generated event. The HepMC record is then captured by the CMSSW framework and stored in the event.

The smearing of the vertex position is the first step in instrumental effect simulation. The event primary vertex, which is positioned by the event generator at the origin of CMS coordinate system, is smeared according to the expected pp impact point of position distribution per bunch crossing. The next step is the simulation of the particles interaction with the detector. The interactions description is achieved using GEANT4 [59]. Once energy deposits and multiple scattering effects in the CMS subdetectors are simulated, and the signal simulation produced by the subdetectors follows. This step is called "digitization".

The chain previous described is often referred to as "full simulation" chain. The GEANT4 detector simulation effect is using the most time consuming step of the full simulation. The time required to full-simulate an event with GEANT4 can be amount of several minutes. Due to a long amount of time needed, the "fast simulation" of the detector effects has been set up. In the fast simulation, the GEANT4 and digitization steps are skipped and detector level quantities, such as the hit positions in the tracker and the energy deposits in the calorimeters, are described using parametrized functions that aim at reproducing the full simulation result.

Starting from the simulated signals in each subdetector (or from the low level reconstructed quantities produced by the fast simulation), the reconstruction of the event follows. With this approach, exactly the same algorithms that will be used on real data can be run on simulated samples.

3.6 Physics Analysis Toolkit

Physics Analysis Toolkit (PAT) [60] is a high-level analysis layer enabling the development of common analysis efforts across and within Physics Analysis Groups. It aims at fulfilling the needs of most CMS analyses, providing both ease-of-use for the beginner and flexibility for the advanced user. PAT is a layer built on and within CMSSW framework, with the aim of simplifying analysis operations by providing easier access to high-level information, as well as tools to perform common analysis tasks. CMS offline reconstruction

format is designed to provide maximum flexibility while keeping the data schema stable for long term storage of the data, to make maintenance easier by decoupling different reconstruction algorithms as much as possible, and to optimize for software performance. Consequently, the format is not optimal for analysis: basic analysis tasks can require some technical expertise (e.g. non trivial book-keeping), and not all the information is easily accessible through simple tools like ROOT macros. One of the main aims of the PAT is to address such limitations avoiding any need for a different and incompatible analysis framework or data model.

PAT defines analysis objects, provides tools to fill those objects with any high-level information required for the specific physics study, and provides tools to perform some preselection and cleaning of those objects before the really analysis specific algorithms are applied. PAT defines analysis-level objects for each basic physics object: leptons, photons, jets, missing transverse energy. In addition, objects are provided for some higher level construction like objects collected in hemispheres, and composite particle hypothesis from decay chains. Finally, a generic particle object is provided to cover most other use cases: exclusive reconstruction of hadrons in heavy flavour physics, detector objects like tracks or calorimeter clusters used as particles in inclusive analyses, or exotic physics signals that require custom event reconstruction. One important aspect of PAT is to keep the analysis objects code-wise compatible with the ones from the offline reconstruction (RECO). Template decorators are used to provide the common PAT extensions to the RECO classes in a uniform way without code duplication and avoiding issues with multiple inheritance (Fig 3.14).

PAT provides a common language among CMS physics analysis users, by means of a common high-level analysis layer. The overall goal of the PAT is to reduce the time the CMS users spend learning about computer science, and maximize their physics productivity.

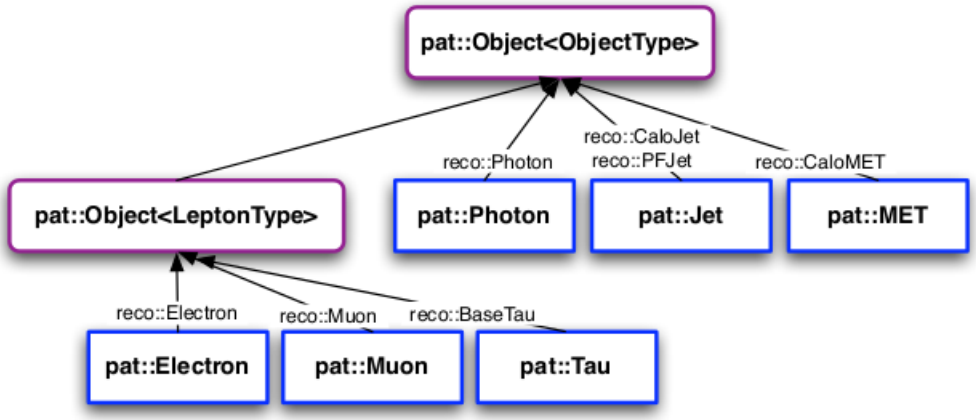


Figure 3.14: PAT objects are build by extending those from offline reconstruction. Arrows indicate the inheritance from the decorator templates, small labels are the RECO classes to which the template is applied.

Chapter 4

Angular Distributions of Z Bosons in Z+Jet Events

4.1 Introduction

Rapidity distributions of Z + jet events can be used as an important check of QCD and event generators used to simulate SM processes. Z boson decays into $\mu^+\mu^-$ is background-free and very efficient triggering. The electroweak vertex makes the perturbative calculation of dynamical quantities even more robust from a theoretical point of view. Furthermore, the detailed comparisons with data can be done because the pQCD calculations at NNLO of Z boson produced in association with four or fewer jets are available [61].

The rapidity of a particle is defined as $y = \frac{1}{2}\ln\left[(E+p_z)/(E-p_z)\right]$, where E is the energy and p_z is the component of the momentum along the beam axis. In the laboratory frame, y_Z and y_{jet} are highly correlated because Z + jet production usually involves a relatively high-momentum valence quark interacting with a low-momentum gluon or antiquark, which results in events where the Z and jet are usually on the same end of the detector. The quantity $y_{sum} = |0.5(y_Z+y_{jet})|$ is the boost from the laboratory frame to the center-of-mass frame of the Z and jet. The invariant rapidity difference can be written as $y_{diff} = |0.5(y_Z-y_{jet})|$. The rapidities y_{sum} and y_{diff} are effectively rotations in phase space of the y_Z and y_{jet} system that yield two approximately uncorrelated quantities. The distribution in y_{sum}, depends mainly on the PDF, while the distribution in y_{diff} reflects the LO partonic differential cross section.

The Z boson and a jet angular distributions in lab frame are measured in pp collisions at

$\sqrt{s} = 7$ TeV at the CERN LHC. The "shape" comparisons between data sample, recorded with the CMS detector corresponding to an integrated luminosity of approximately 5.1 fb^{-1}, to MADGRAPH, SHERPA, and Monte Carlo for femtobarn process (MCFM) simulations are analysed. Events in which there is a Z and exact one jet, with a jet transverse momentum threshold of 30 GeV and absolute jet rapidity less than 2.4, are selected for this analysis. The detail selections will be discussed in Section 4.4. Only the Z muon decay channel is studied.

4.2 Event Reconstruction

Events are reconstructed using the CMS particle-flow (PF) algorithm [1]. PF reconstructs individual particles (leptons, photons, charged and neutral hadrons) by linking tracks, ECAL clusters, and HCAL clusters. The momentum or energy of each particle is formed by using information from all sub-detectors. Broadly speaking, electrons are reconstructed from tracks and calibrated ECAL energy clusters; muons are reconstructed using tracks; individual charged hadrons are reconstructed from tracks and calibrated ECAL and HCAL clusters; and individual photons and neutral hadrons are reconstructed from calibrated ECAL and HCAL energy clusters, respectively. A detailed description of the PF algorithm may be found in [62, 63]. We highlight muon and jet reconstruction below, as they pertain to this analysis.

4.2.1 Muon Reconstruction

Muons are first reconstructed independently in the silicon tracker (*tracker track*) and in the muon spectrometer (*standalone-muon track*). Based on these, two reconstruction approaches are used:

- *Global Muon* reconstruction (outside-in): Starting from a standalone muon in the muon system, a matching tracker track is found and a *global-muon track* is fitted

[1] Using CMSSW 4_4_X

combining hits from the tracker track and standalone-muon track.

- *Tracker Muon* reconstruction (inside-out): Tracker tracks (above a certain momentum threshold) are considered as possible muon candidates and are extrapolated to the muon system, taking into account the expected energy loss and the uncertainty due to multiple scattering. If at least one muon segment (i.e. a short track stub made of DT or CSC hits) matches the extrapolated track in position, the corresponding tracker track qualifies as a *tracker-muon track*.

A detailed explanation of muon reconstruction in CMS can be found elsewhere [64].

4.2.2 Jet Reconstruction

Jets are reconstructed offline from the PF particles using the anti-k_T algorithm [65] with a size parameter $R = \sqrt{\eta^2 + \phi^2} = 0.5$. In order to ensure the best quality of the tracking information, only jets within $\eta < 2.4$, i.e. within the silicon tracker acceptance, are selected. A minimum threshold on the jet transverse momentum of $p_T > 30$ GeV/c is required, in order to reduce the contamination from the underlying event to the final state studied and to ensure a good reconstruction.

Jet Energy Correction

In order to correct the non-liner response of the detector, jet energy correction (JEC) [66] is applied on jet p_T. The JEC used are the following;

- L1FastJet uses the FastJet simulation to remove the energy coming from pileup events.

- L2Relative makes the jet response flat in all η regions.This correction can be done by MC truth or by data-driven dijet balance.

- L3Absolute makes jet response flat in p_T space. This correction is derived from MC truth or data-driven of Z/γ*+jet balance.

- L2L3Residual (only used in the data) is used as the in addition to the above sequences as the fine-tune correction.

4.3 Simulation and Data Samples

Collision data are compared to pQCD theory via MC simulations. Samples of events with a Z^0 boson, $W^{\pm}$ boson, or $t\bar{t}$ pair (all accompanied by jets) are generated by MADGRAPH [67] and interfaced with PYTHIA [68] to simulate parton showering and hadronization. MADGRAPH generates tree-level events with up to four partons in the final state on the basis of a matrix-element calculation. The PYTHIA parameters for the underlying event have been set according to the Z2 tune, which is identical to the Z1 tune described in [69], except that Z2 uses the CTEQ6L PDF. Various muon-enriched multi-jet events ("QCD backgrounds") are also simulated, although strictly with PYTHIA.

The full list of data sets used is given in Table 4.1. Whenever available, the (next-to-)next-to-leading-order ((N)NLO) cross section is used to normalize the simulated sample to the integrated luminosity of the collision data [31, 32, 70].

Table 4.1: Data sets used in this analysis, along with kinematic selections and cross sections.

Process	Generator	Kinematic Selections	Cross Section×Br (pb)
Z ($\rightarrow ll$) + jets	MADGRAPH	$m_{ll} > 50$ GeV	3048 (NNLO)
Z ($\rightarrow ll$) + jets	SHERPA	$m_{ll} > 50$ GeV	3048 (NNLO)
W ($\rightarrow l\nu$) + jets	MADGRAPH	–	31314(NNLO)
$t\bar{t}$ + jets	MADGRAPH	–	157.5(NLO)
μ-enriched QCD	PYTHIA	$p_T^{boost} > 20$ GeV, $p_T^{\mu} > 15$ GeV	3.5×10^6 (LO)

Generated events are processed through a full detector simulation based on GEANT4 [59, 71], followed by a detailed trigger emulation and the CMS event reconstruction. Several minimum bias events are superimposed to the hard interactions to simulate event pileup according to the distribution of multiple pp collisions observed during the 2011 data taking

period.

The CMS data of pp collisions have been collected during the year 2011 and corresponded to an integrated luminosity of approximately 5.1 fb^{-1}. The data sets and run lists are summarized in Table 4.2.

The JSON file is "Cert_160404-180252_7TeV_ReRecoNov08_Collisions11_JSON.txt".

Table 4.2: 2011 Data sets, run ranges and recorded luminosity.

Dataset Name	Run range	Recorded Luminosity (fb^{-1})
/DoubleMu/Run2011A-08Nov2011-v1	160404-175770	2.33
/DoubleMu/Run2011B-19Nov2011-v1	178078-180252	2.76

4.4 Event Identification

Signal events in collision data are identified through a series of trigger and identification selections. The L1 of the CMS trigger system, composed of custom hardware processors, is designed to select events of interest (in less than 1 μs) using information from the calorimeters and muon detectors [72]. The High Level Trigger (HLT) processor farm further decreases the event rate from up to 100 kHz to 100 Hz before data storage. For this study, events were selected if they passed any of the muon triggers listed in Table 4.3. The rates of various triggers were prescaled since the instantaneous luminosity increased throughout the 2011 run. Therefore, events were accepted only if they passed an unprescaled trigger.

Table 4.3: Trigger paths used in selecting events, including HLT and L1 p_T thresholds.

HLT Path	L1 Seed	HLT/L1 p_T Threshold (GeV)
HLT_DoubleMu6	L1_DoubleMu3	6 / 3
HLT_Mu13_Mu8	L1_DoubleMu3p5	13, 8 / 3.5
HLT_Mu17_Mu8	L1_DoubleMu3p5	17, 8 / 3.5

The selections used in this analysis are listed in Table 4.4 and detailed below. Since the

goal of this analysis is to study the normalized Z + 1jet angular distribution, the triggers and selections used must not introduce angular biases (e.g. η dependence).

Table 4.4: Event and particle selections used in this analysis, along with the motivation.

Category	Selection	Motivation
Vertex	$NDOF_{PV} > 4$ $\|z_{PV}\| < 15$ cm $\rho_{PV} < 2$ cm	Non-collision and beam-related background rejection
Muon Acceptance	$p_T^{\mu} > 20$ GeV $\|\eta^{\mu}\| < 2.1$	Cosmic ray muon rejection Muon trigger acceptance
Muon Quality	Global Muon & Tracker Muon Opposite Charges Di-muons Number of Pixel Hits > 0 Number of Silicon Hits > 10 Number of Muon Hits > 0 Number of Muon Stations > 1 Normalized $\chi^2 < 10$ $\|d_{xy}\| < 0.2$ cm Muon Isolation R03 sum p_T	Decay-in-flight, punch-through, noise and cosmic ray rejection
Jet Acceptance	AK5 PF Jets $p_T^{jet} > 30$ GeV $\left\|\eta^{jet}\right\| < 2.4$	Rejection of pure PU jets and better jet resolution Tracker and muon spectrometer acceptance
Z Selection	$I_{rel}^{\mu} < 15\,\%$ $76 < M_{\mu\mu} < 106$ GeV $p_T^{\mu\mu} > 40$ GeV	QCD background rejection Z signal selection Also require for lab frame analysis

To reject non-collision and beam-related backgrounds, all events are required to have a primary vertex (PV) consistent with the measured transverse position of the beam (referred to as the beam spot). Specifically, the fit for the PV must include at least four associated tracks (i.e. five degrees of freedom), the z-coordinate of the PV must lie within the luminous collision region, and the radial distance of the PV must be less than 2 cm from the beam spot. These selections are greater than 99% efficient [73].

Our signal muon selections are those used in the measurement of the W and Z cross sections [74], with modifications motivated by the vector boson + jet ratio measurements [75].

We define a relative isolation variable $I_{rel} = \sum \left(p_T^{track} + E_T^{ECAL} + E_T^{HCAL}\right)/p_T^{\mu}$, which

consists of the p_T for tracks and E_T for ECAL/HCAL towers within a cone of $R < 0.4$ (centered around the muon axis). The muon and its energy deposits are excluded from this sum by ignoring the energy within a smaller "veto cone" ($R < 0.01$, 0.7 and 0.1 for the tracker, ECAL and HCAL, respectively). A muon is isolated if $I_{rel} < 15\%$.

According to the reasons described above, the The Z+jet event selection requires the presence of an energetic ($p_T > 20$ GeV), isolated muon in the region $|\eta| < 2.1$. The dimuon invariant mass is between 76 and 106 GeV. In lab frame analysis, we require dimuon p_T grater than 40 GeV. Accompanying jets must have a p_T of at least 30 GeV within the muon system and tracker acceptance ($|\eta| < 2.4$).

4.4.1 Kinematic Distributions of Candidate Events

Figures 4.1 and 4.2 show kinematic distributions for Z and jets candidates respectively. After selections, all distributions agree with simulations within statistical and systematic uncertainty. Note that the jet mass in Fig. 4.2 is non-zero; this is attributed to the finite angular spread of the jet in the calorimeters. The Z mass distribution shown in Fig. 4.1 was created before applying the Z mass selection; the discrepancy in collision data and simulation (for $M < 50$ GeV) comes from a generator-level invariant mass selection (see Table 4.1)

Figure 4.3 shows the number of jets accompanying a Z (note that adding a jet drops the number of events by $\sim \alpha_s$; the cross-section is proportional to the number of strong-interaction vertices in the Feynman diagram). Also note $\approx 20\%$ of jet events have two or more jets.

4.5 Muon Identification Efficiency Corrections

Muon Particle Object Group (Muon POG) have studied the efficiencies of muon identification sections by using "tag-and-probe" technique which is discussed in Appendix A. These efficiencies are our scale factors. They are calculated in bins of p_T and η. The

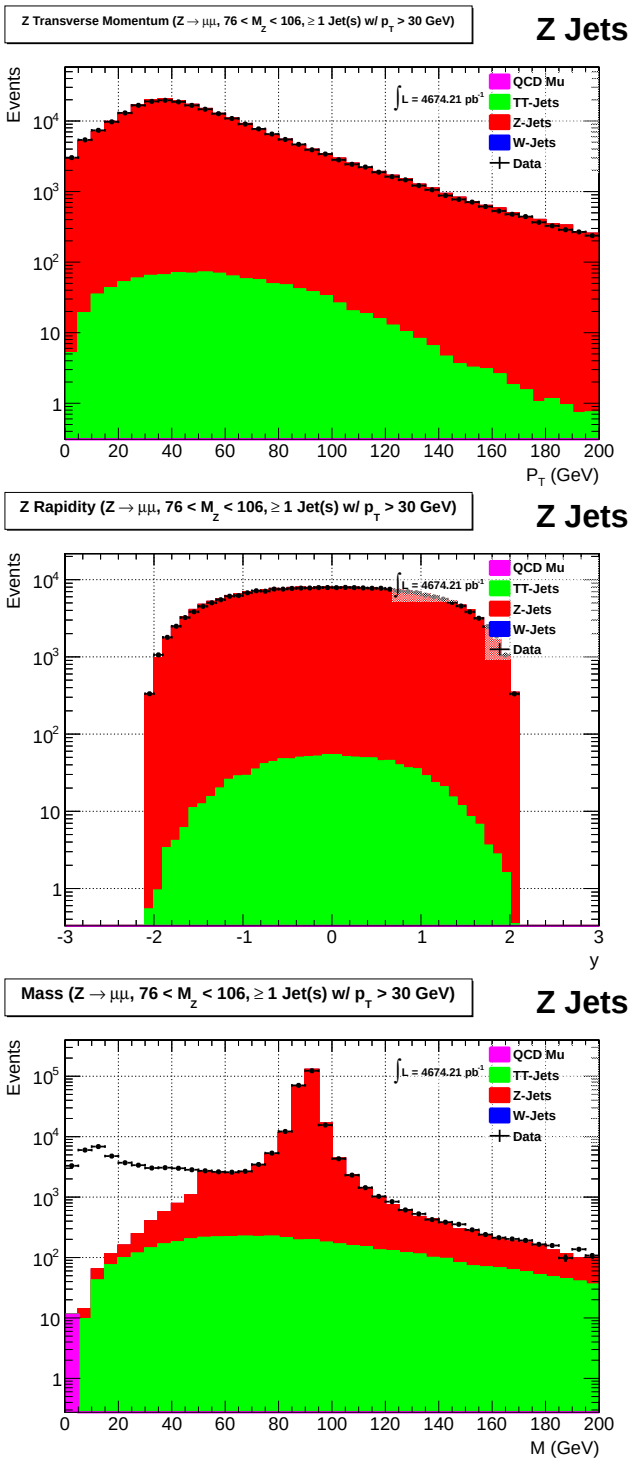


Figure 4.1: Kinematic distributions for Z candidates in collision data; transverse momentum (top), rapidity (middle), and mass (bottom). Also shown are the corresponding distributions for signal (red) and various background (other) simulations.

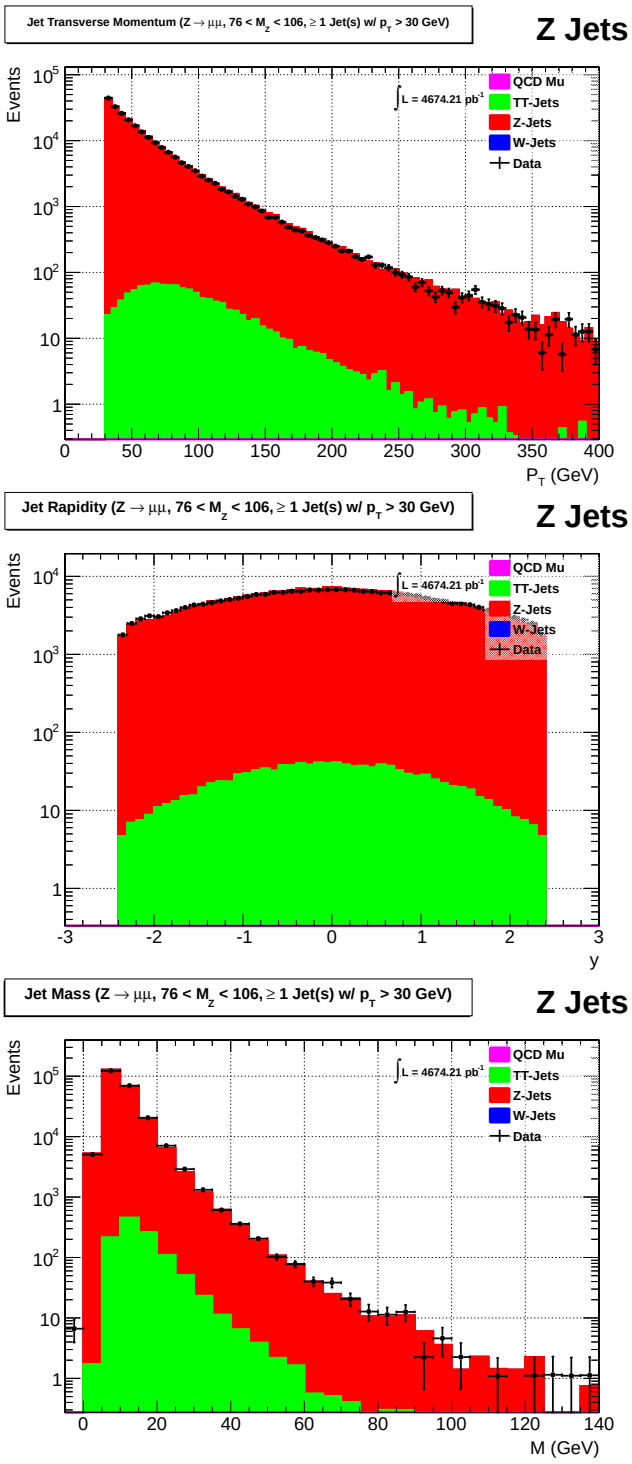


Figure 4.2: Kinematic distributions for candidate jets in collision data; transverse momentum (top), rapidity (middle), and mass (bottom). Also shown are the corresponding distributions for signal (red) and various background (other) simulations.

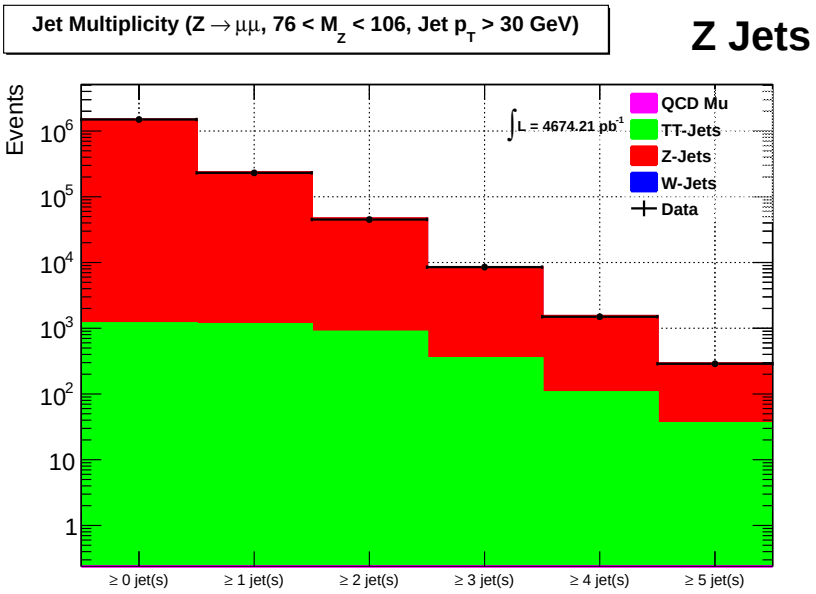


Figure 4.3: The number of jets accompanying Z events in collision data. Also shown are the corresponding distributions for signal (red) and various background (other) simulations.

scale factors are provided centrally by Muon POG [76]. We weigh the distributions of $|y_Z|$, $|y_{jet}|$, $y_{diff} = |0.5(y_Z - y_{jet})|$, and $y_{sum} = |0.5(y_Z + y_{jet})|$ that pass the identification selection with these factors. The identification efficiency correction (Figure 4.4) is the reciprocal of the ratio between weighted distribution with identification selection and distribution without identification selection.

4.6 Unfolding

The detail about unfolding procedures is discussed in Appendix B. The response matrices of $|y_Z|$, $|y_{jet}|$, $y_{diff} = |0.5(y_Z - y_{jet})|$, and $y_{sum} = |0.5(y_Z + y_{jet})|$ are shown in Figure 4.5. The correction factors (Figure 4.6) on CMS data using these matrices are consistent with unity within the statistical uncertainty except $|y_{jet}|$ which be unfolded before theory comparison.

In order to confirm the correctness of unfolding procedures, the closure test is preformed. We use the response matrices of MADGRAPH to unfold the independent MC

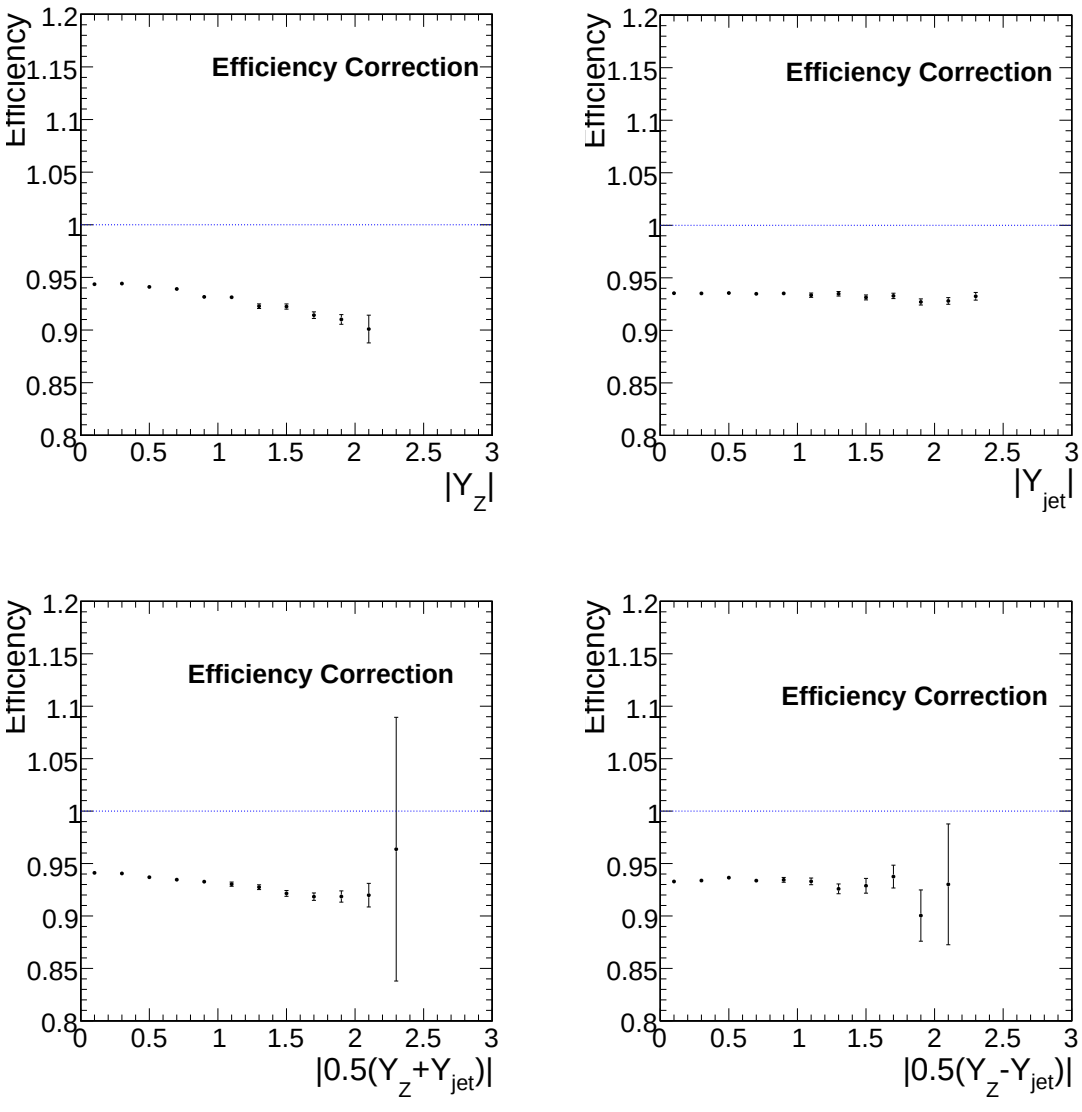


Figure 4.4: The ratio between weighted distribution with identification selection and distribution without identification selection of $|y_Z|$, $|y_{jet}|$, y_{diff} =$|0.5(y_Z$-$y_{jet})|$, and y_{sum} = $|0.5(y_Z$+$y_{jet})|$.

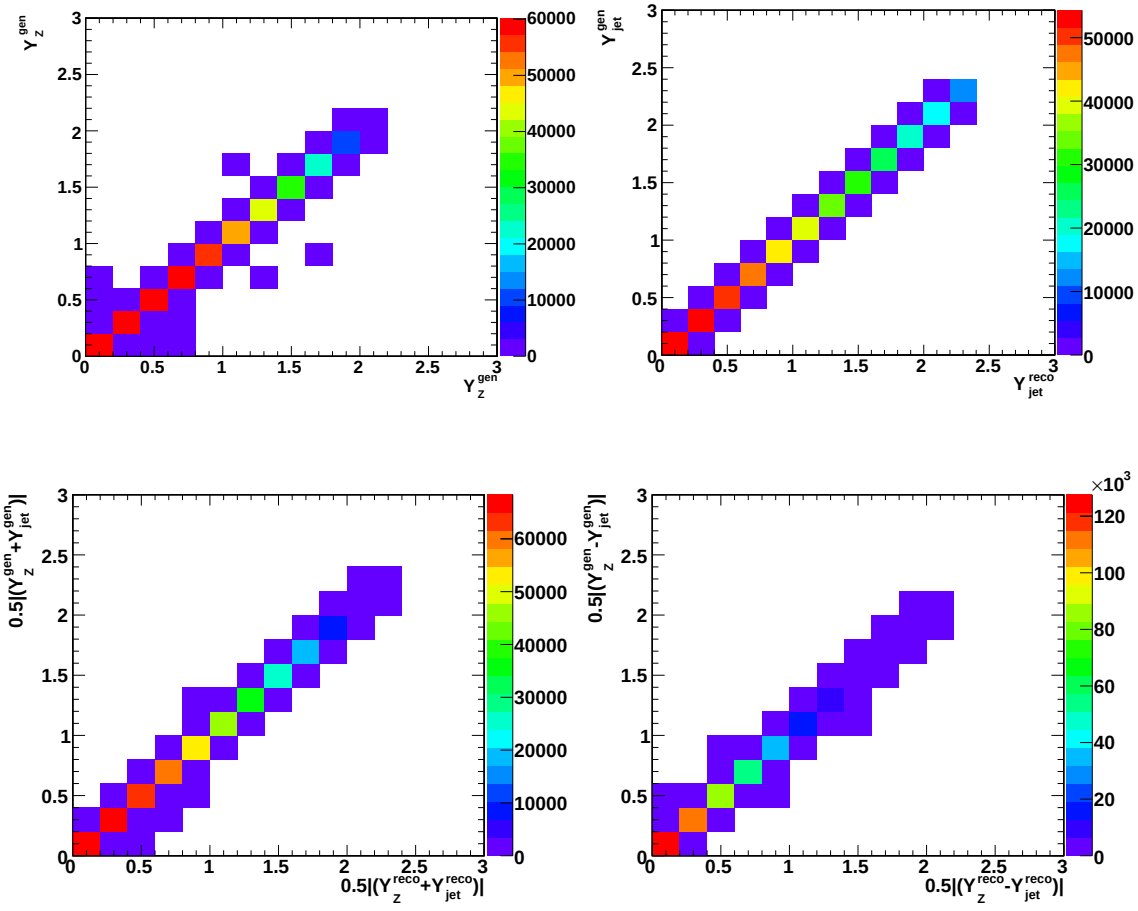


Figure 4.5: The response matrices from MADGRAPH show mostly diagonal relation between generator-level and reconstruction-level rapidities.

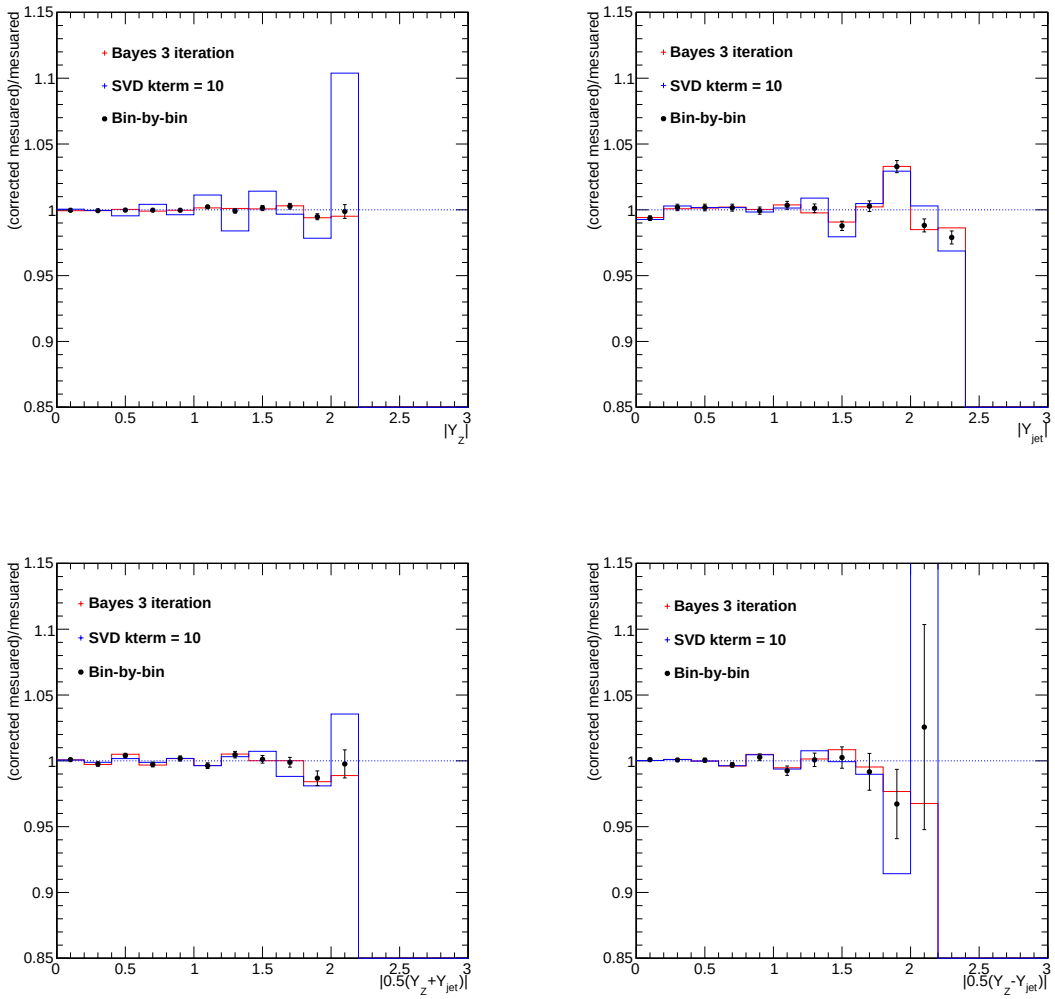


Figure 4.6: The unfolding corrections on CMS data using MADGRAPH respone matrices are close to unity except $|y_{jet}|$

prediction of Z+jets, SHERPA [77]. The closure test (Figure 4.7) shows consistency between MADGRAPH and SHERPA after unfolded.

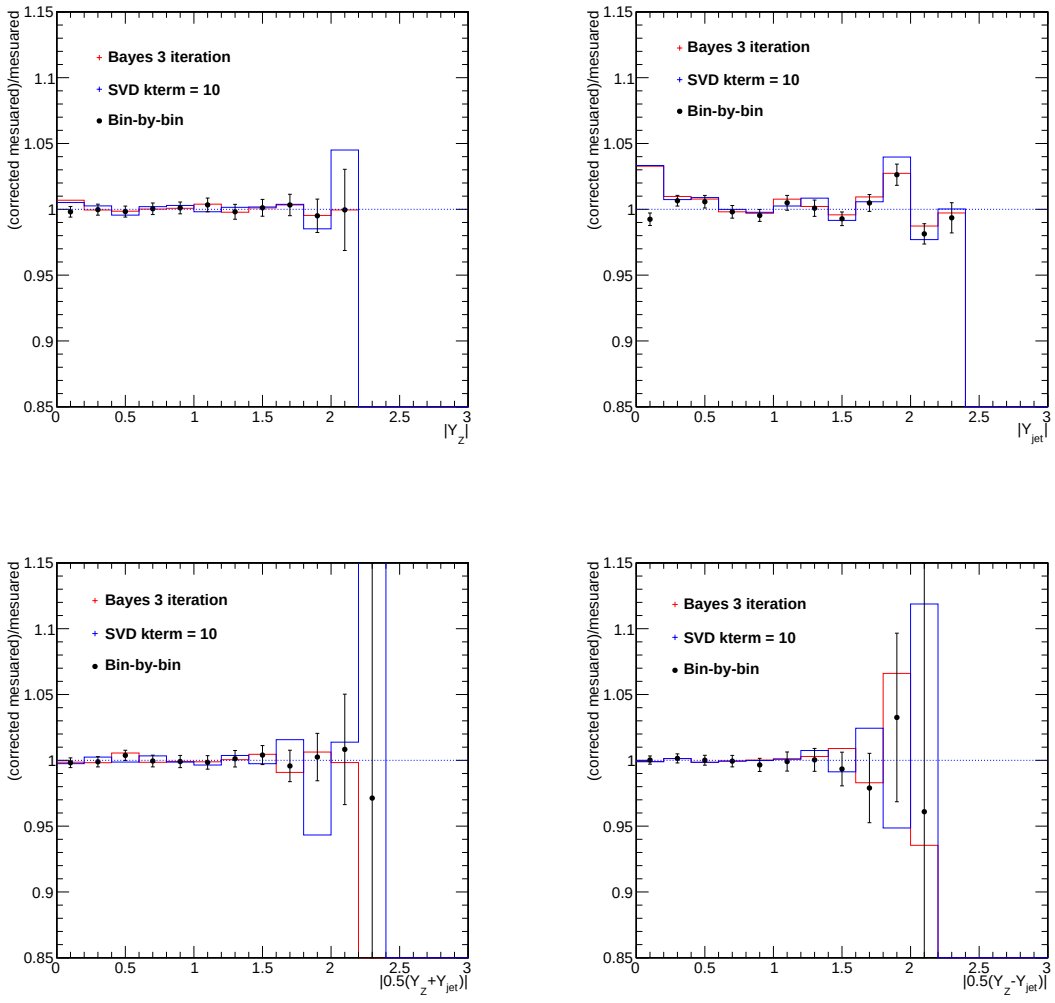


Figure 4.7: With SHERPA, the unfolding correction is close to unity except $|y_{jet}|$ like we see on MADGRAPH.

4.7 Uncertainty

For a normalized angular distribution analysis, many of the typical uncertainty studies are not required. Of the usual sources of uncertainty, we consider only those that affect the shape of the angular distribution (or that have an angular bias). We evaluate how varying

the PDF parameters in our simulation affects the shape of the angular distributions. We also study the effects of uncertainty in the jet energy scale (JES) and take the finite jet momentum and angular resolution of the detector into account. Compared to the JES and resolution, the uncertainty in the muon measurement is negligible (less than 1%). Relative uncertainty values as the final results in [1] are summarized in Table 4.5.

Theoretical Uncertainty

Crucial to NLO and above cross section predictions, PDFs are obtained by global fits to measurements from deep-inelastic scattering, Drell-Yan, and jet data. The PDF uncertainties reflect three characteristics: the choice of dataset, the type of uncertainty estimator used, and the form and size of parton parametrization. Details can be found in [78], with a brief summary below.

One method of determining PDFs is based on a Hessian approach [79], which minimizes a suitable log-likelihood χ^2 function and accounts for correlated uncertainties by means of a covariance matrix. The best fit is the point in parameter space at which χ^2 is minimum. The PDF uncertainties are found by diagonalizing the Hessian matrix (second derivatives of the χ^2 at the minimum) and then determining the range of each orthonormal Hessian eigenvector that corresponds to a prescribed increase (e.g. a 68% confidence level) of the χ^2 function with respect to the minimum.

Another source of theoretical uncertainty is the QCD factorization (μ_F) and renormalization (μ_R) scales. The estimation is done by changing these scales up and down by a factor of two. The theoretical uncertainties are studied at [80]

4.7.1 Jet Energy Scale Uncertainty

As mentioned in Section 4.2.2, the measured jet energy is different from the true particle-level jet energy. The difference is primarily caused by the non-uniform and non-linear response of the calorimeters. Therefore, corrections are made to the energy scale

of the reconstructed jets. Uncertainty studies on the JES were performed in [66] and the combined JES uncertainty[2] as a function of p_T for various η values are shown in Fig. 4.8.

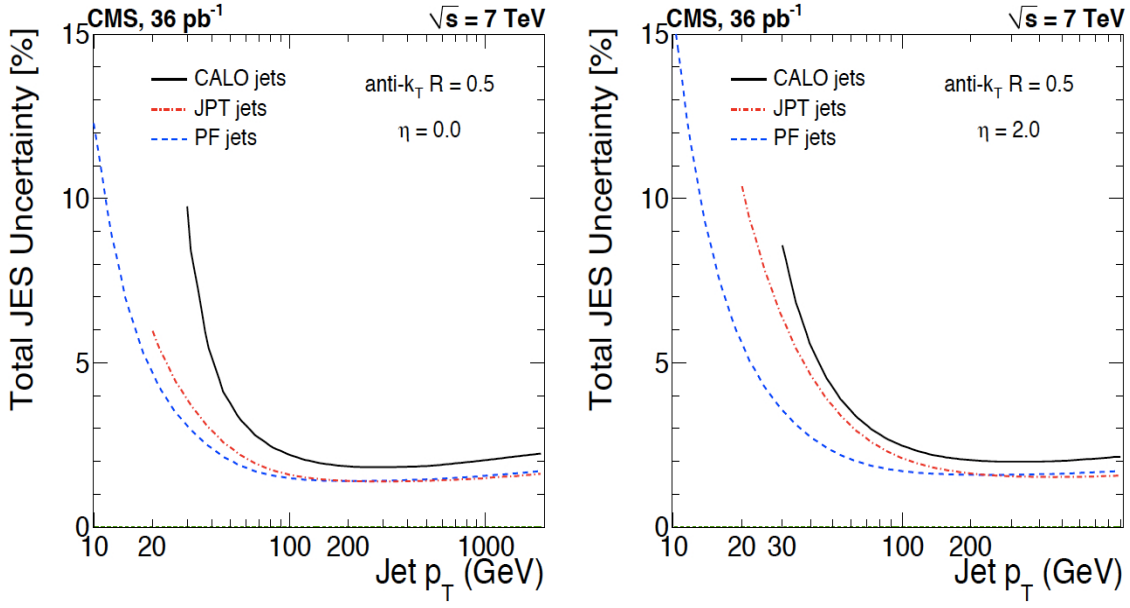


Figure 4.8: Total jet energy scale uncertainty as a function of jet p_T for various η values [66].

The rapidity of jets, calculated by $\frac{1}{2}\ln\left[(E+p_z)/(E-p_z)\right]$, does not change when multiplying jet's four-momentum vector with a scalar number. However, the number of Z+jet events may change due to the jet p_T threshold at 30 GeV/c. When we scale the jet energy up by its uncertainty, events with jet p_T less than threshold may be promoted to be Z+(n+1)jet events: Z+0jet events turn to be Z+1jet events. The opposite way can be happened with scaling the jet energy down by its uncertainty. The shape comparison results of p_T and rapidity are shown in Figure 4.9. Although there is high fluctuation at high p_T of jets due to low statistic, the JES uncertainty shows only small effect on rapidity. JES uncertainty contributes up to 1% relative value of $y_{diff} = |0.5(y_Z - y_{jet})|$ and $y_{sum} = |0.5(y_Z + y_{jet})|$.

[2]Payload GR_R_42_V22A_AK5PFchs

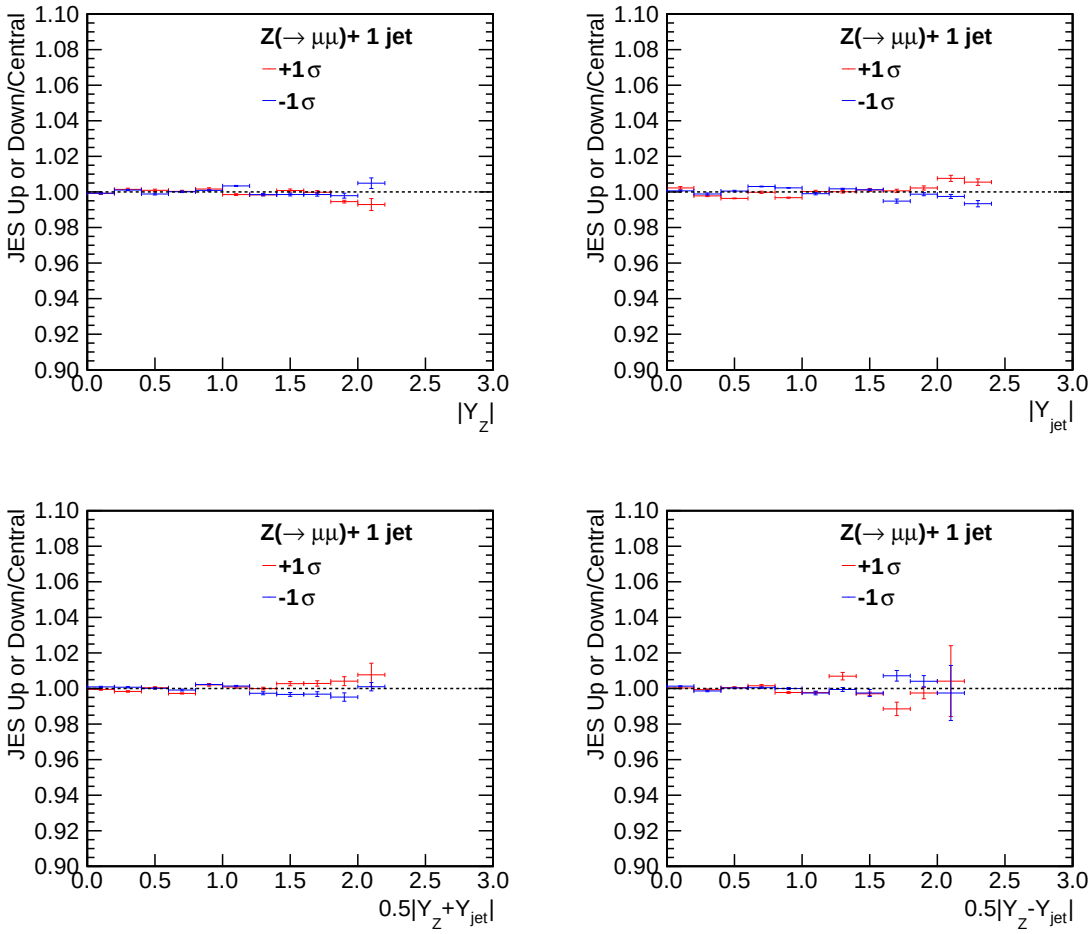


Figure 4.9: The ratio distributions between central value and smeared up and down 1-σ with JES uncertainty for $|y_Z|$, $|y_{jet}|$, y_{diff} =$|0.5(y_Z$-$y_{jet})|$, and y_{sum} = $|0.5(y_Z$+$y_{jet})|$.

4.7.2 Jet Energy Resolution

Finite jet energy resolution can be the threshold effects. The study of this uncertainty is based on simulated MADGRAPH by modified the reconstructed jet p_T with the p_T difference between matched reconstruction-level jets and generator-level jets by the following:

$$\begin{aligned} &\text{Upgraded} : p_T^{reco} = p_T^{gen} \\ &\text{Downgraded} : p_T^{reco} = p_T^{reco} + c \times (p_T^{reco} - p_T^{gen}) \end{aligned} \tag{4.1}$$

where c is a factor provided by Jet and Missing E_T group. Figure 4.10 shows the shape comparison of reconstruction-level jets before and after jet energy resolution modification. The contribution of jet energy resolution uncertainty mostly affects shapes of p_T but for rapidities is only a few percents of $|y_Z|$, $|y_{jet}|$, $y_{diff} = |0.5(y_Z - y_{jet})|$, and $y_{sum} = |0.5(y_Z + y_{jet})|$ except the last few bins due to low statistic.

Table 4.5: List of relative systematic uncertainties as in publication [1].

Source	Relative Value
Jet energy scale	$< 1\%$
Jet resolution	$< 2\%$
Pileup	$< 0.5\%$
Background	$< 0.2\%$
Theoretical (PDF variation)	$< 2\%$
Theoretical (QCD scale)	$< 2\%$ and $\approx 8\%$ for y_{diff}

4.8 Results and Theory Comparision

Theoretical prediction for massive vector boson Z are well established by several MC generations. The production of jets in association with Z boson generally relies on pQCD. Many MC generations have been developed to generate Z+jet events. Studies of such generations are the excellent testing ground for theoretical prediction. These measurements

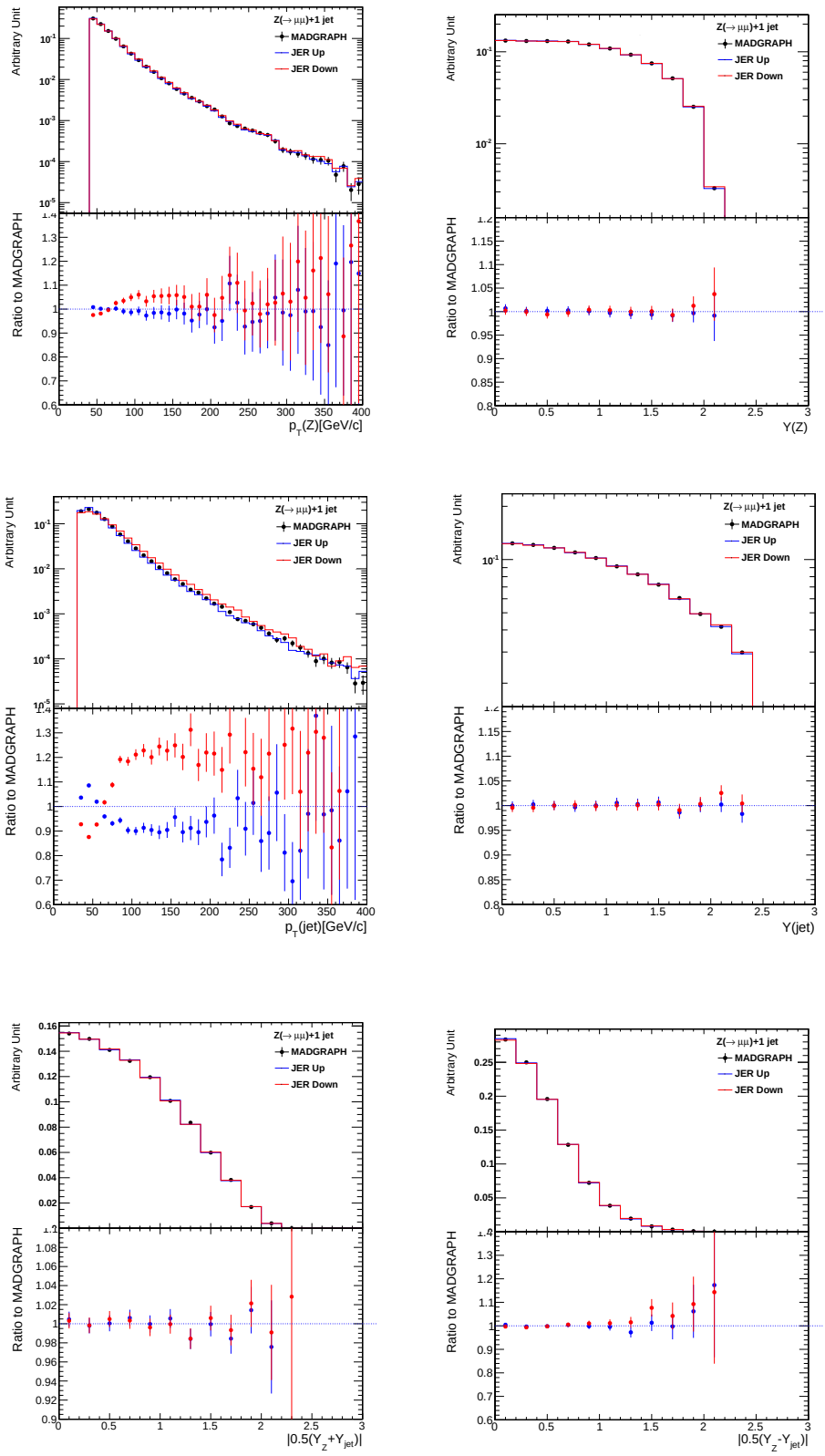


Figure 4.10: The distributions for $|y_Z|$, $|y_{jet}|$, $y_{diff} = |0.5(y_Z - y_{jet})|$, and $y_{sum} = |0.5(y_Z + y_{jet})|$ with jet energy resolution.

can be the input to improve MC simulations.

In this section, angular distributions of Z ($p_T > 40$ GeV) and jet ($p_T > 30$ GeV within $|\eta| < 2.4$) in Z+jet events will be studied. CMS data, MADGRAPH,and SHERPA are used to do shape comparisons with MCFM [81]. MADGRAPH is using the LO CTEQ 6L1 and MLM matching [82]. SHERPA uses tree-level matrix-element generator and also combine multi-parton matrix elements with the QCD parton cascades by CKKW [83]. PDF used to generate Z + 1 jet event in SHERPA is NLO CTEQ6.6M. MCFM is a parton-level MC giving NLO predictions and using CTQ6.1 PDF. Since MCFM does not have detector simulation, MADGRAPH and SHERPA generator level quantities are used and CMS data are corrected according to MADGRAPH generator information.

The angular distribution comparisons of the Z+jet events are shown in Figure 4.11. Rapidity comparisons to MCFM in lab frame for the Z and a jet show good agreement. The rapidity sum and difference between Z and jet are in good agreement for SHERPA and CMS data to MCFM but not for MADGRAPH. The results indicate that SHERPA can be better prediction of rapidity than MADGRAPH.

4.8.1 Conclusion

For the first time ever, the Z ($\rightarrow \mu\mu$) boson and a jet angular distributions in lab frame was measured at $\sqrt{s} = 7$ TeV. The data sample corresponds to an integrated luminosity of approximately 5.1 fb^{-1}. Events in which Z and a single jet for lab frame, with a jet p_T threshold of 30 GeV and jet $|\eta|$ less than 2.4, were selected for the analysis. Event reconstruction was performed using PF, an algorithm designed to optimizing particle identifications by exploiting information from all the subdetectors of the CMS. The measurement was compared to the MADGRAPH MC generator, which simulate the latest pQCD predictions. In lab frame, comparison of CMS data, MADGRAPH, and SHERPA to MCFM is preformed. The result shows that SHERPA is the best prediction in term of angular distributions. Our result is consistent with that of electron channel[84]. The normalized

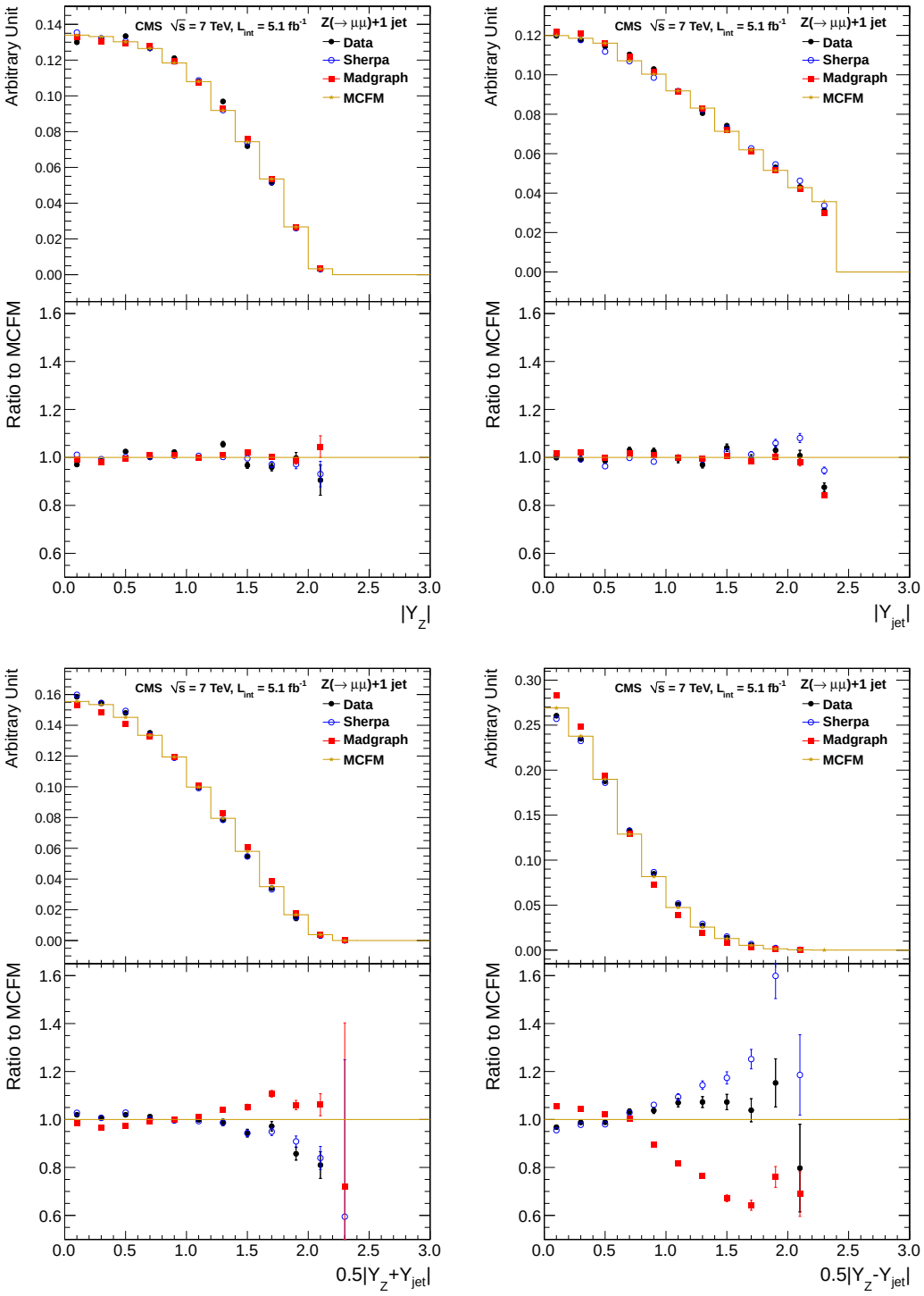


Figure 4.11: The theory comparison distributions for $|y_Z|$, $|y_{jet}|$, $y_{diff} = |0.5(y_Z - y_{jet})|$, and $y_{sum} = |0.5(y_Z + y_{jet})|$.

rapidity distributions of combined muon and electron channels for $|y_Z|$, $|y_{jet}|$, y_{diff}, and y_{sum} of the CMS data with theoretical predictions are shown in Fig. 4.12. The $|y_Z|$ and $|y_{jet}|$ distributions of data are well agreed within 5% with MADGRAPH, SHERPA, and MCFM. Since MADGRAPH and MCFM are using LO PDF, thus y_{sum} is less agreed with data at high range. The y_{diff} distribution is well agreed with MCFM up to y_{diff} = 1.0 but disagreed considerably for both MADGRAPH and SHERPA. Disagreement are from the same LO matrix elements and approaches to parton showering that these two program use. More discussions of combined results can be found at [1].

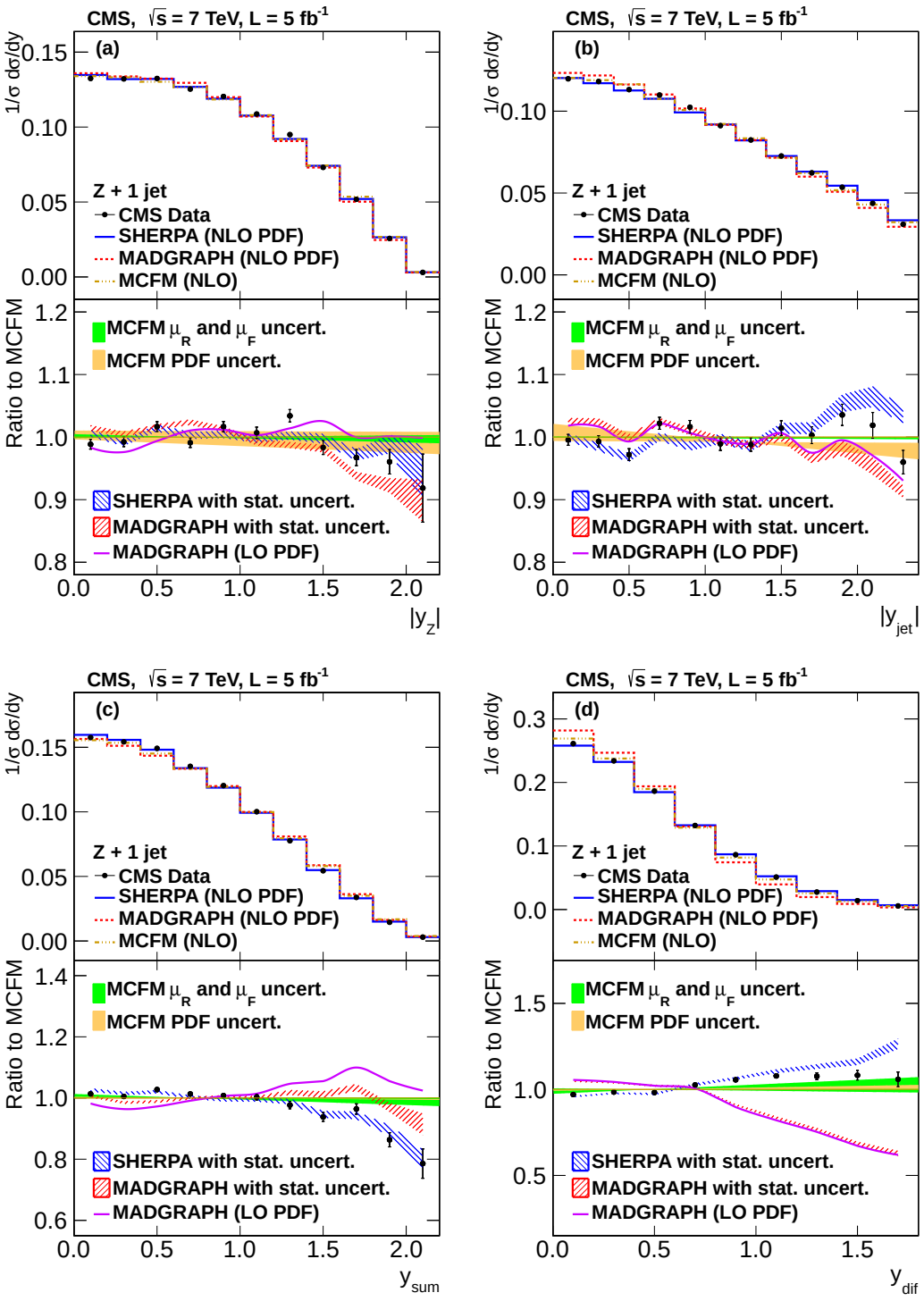


Figure 4.12: Distributions in absolute values of rapidities for (a) $|y_Z|$, (b) $|y_{jet}|$, (c) y_{sum}, and (d) y_{diff} after normalized to unity. The lower panel of each distribution is the ratio of data, MADGRAPH, and SHERPA to MCFM.

Chapter 5

Differential Cross Section of Jets Associated with the Z Boson in pp Collisions

5.1 Introduction

Jet production associated with the Z boson is an important aspect of the data analysis at the CMS. The understanding of such events is not only a test for pQCD, but also significant background for exploring new physics. The high energy pp collisions in the LHC enables us to measure cross sections of these jets with more precision than previous available.

The final state Z boson decay is the presence of a pair of particle and its anti-particle either hadronic or leptonic decay. The leptonic decay (a pair of electrons and muons) is much clearer signature, enabling to separate the Z decay from a large background form pure QCD events. The focus of this study is put on a Z boson decaying into an opposite sign pair of muons that is associated with jets[1]. Selected events are required to have a pair of muons with an invariant mass window of 40 GeV/c^2 around the known Z mass (91 GeV/c^2. PF jets are selected using anti-k_T clustering algorithm requiring $p_T > 30$ GeV/c and be in the tracker acceptance of $|\eta| < 2.4$. Corrections for efficiency of the trigger, selection, reconstruction and isolation are applied to all distributions. They are calculated by a tag and probe method that is provided by the CMS POG. The results will be shown after applying the unfolding procedure to deconvolve the distributions from detector effects. Results from the muon channel compared with the theoretical predictions are also shown.

[1] Z boson decay into an opposite sign pair of electrons is being studied by our colleagues at the CMS

These measurements are based on data collected by CMS during 2012 at $\sqrt{s} = 8$ TeV. The differential cross section is measured versus jet multiplicity; as well as the transverse momentum (p_T) and pseudo-rapidity (η) up to the second highest transverse momentum jets. Differential cross sections for all distributions are deconvoluted from detector effects, utilizing the unfolding procedure. The corresponding total integrated luminosity used is 19.8 fb^{-1}.

5.2 Data and Monte Carlo Samples

Data are reprocessed with CMSSW_5_3_11. The data sets used for this analysis and corresponding run periods are summarized in Table 5.1. Only "Good" run and luminosity sections, as certified by physics validation teams, are used for the analysis which are listed in the JSON file "Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt".

Table 5.1: 2012 Data sets, run ranges and recorded luminosity.

Dataset Name	Run range	Recorded Luminosity (fb^{-1})
/DoubleMu/Run2012A-22Jan2013-v1	190456-193621	0.89
/DoubleMu/Run2012B-22Jan2013-v1	193833-196531	4.42
/DoubleMu/Run2012C-22Jan2013-v1	198022-203742	7.098
/DoubleMu/Run2012D-22Jan2013-v1	203777-208686	7.289

The simulated signal and background samples used for this analysis are listed in Table 5.2. The signal is defined as Drell-Yan plus up to four jets. The Parton Shower technique is used to describe the QCD radiation which is a collinear approximation of parton splitting. Therefore, it is inadequate at describing multiple hard jets but can be improved by using information from matrix element calculations. Samples of events containing a Z boson that is accompanied by jets are generated by MADGRAPH [67], using the PDF CTEQ6L1 [86], and interfaced with PYTHIA [68], which simulates parton showering and hadronization. MADGRAPH generates tree-level events with up to four partons at Leading Order in the final state on the basis of a matrix-element calculation. The PYTHIA parameters for

underlying events have been set according to the Z2 tune [87].

Background processes include jets associated with double vector bosons (VV+jets) and $t\bar{t}$ pair processes. Background estimations are obtained from MC simulations generated by MADGRAPH interfaced with PYTHIA. The NNLO cross section is used to normalize the simulated Z+jets sample. For VV+jets samples, the NLO calculation for cross section is used. The next-to-next-to-leading-log (NNLL) cross section is used for $t\bar{t}$ sample [88]. The cross section values are taken from [31]. All MC samples are normalized to the integrated luminosity of the collision data. The detector simulation, trigger, and reconstruction sequence of these MC samples are carried out using the same CMSSW as used with the real data.

Table 5.2: Monte Carlo samples.

Dataset name	cross section (pb) $\times$ BR
/DYJetsToLL_M-50_TZ2Star_8TeV-mad-tarball/*	3503.75
/TTJets_MassiveBinDECAY_TZ2star_8TeV-mad-tauola/*	234
/ZZJetsTo2L2Q_TZ2star_8TeV-mad-tauola/*	17.654×0.14
/ZZJetsTo2L2Nu_TZ2star_8TeV-mad-tauola/*	17.654×0.04
ZZJetsTo4L_TZ2star_8TeV-mad-tauola/*	17.654×0.01
/WZJetsTo2L2Q_TZ2star_8TeV-mad-tauola/*	33.21×0.07
/WZJetsTo3LNu_TZ2_8TeV-mad-tauola*	33.21×0.03
/WWJetsTo2L2Nu_TZ2star_8TeV-mad-tauola/*	54.838×0.11

5.2.1 Pileup Re-weighting

In the MC production, an approximate number of pileup interactions are assumed. However, this assumption does not match the conditions for each data-taking period. Also, the final distribution for the number of reconstructed primary vertices is still sensitive to the details of the primary vertex reconstruction, differences in underlying events in the data versus MC samples and there is a bias caused by the offline event selection criteria and the trigger. In order to correct for these effects, the number of simulated pileup interactions in MC are re-weighted by the data pileup distribution using the entire data-taking period with

the minimum bias cross-section of 69.4 mb. After PU re-weighting, the number of vertices are in good agreement between data and MC for the Z events, as shown in Fig.5.1.

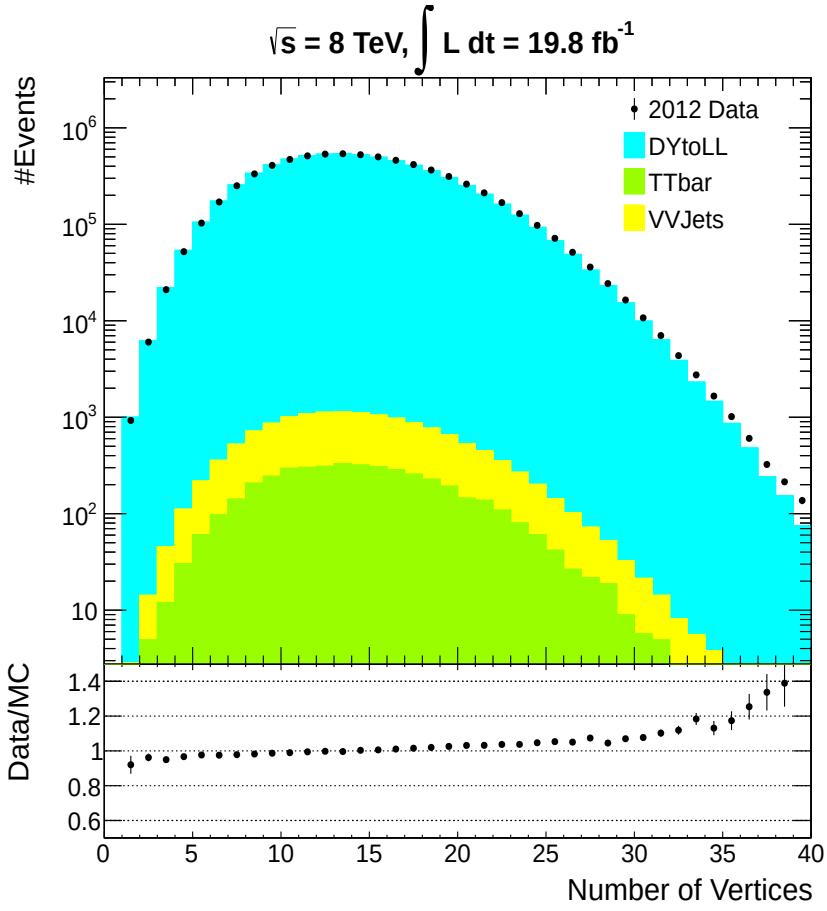


Figure 5.1: The number of primary vertices in data and MC in the di-muon channel.

5.3 Event Reconstruction and Selection

5.3.1 Trigger

The rates of various triggers were prescaled since the instantaneous luminosity increased throughout the 2012 run. Only non-prescaled triggers were used in this study. At the HLT stage, the double leptons are required to have a particular transverse energy p_T. The thresholds for double muons are $p_T > 17$ GeV/c for the leading muon and $p_T > 8$ GeV/c for the second muon. The HLT path is HLT_Mu17_Mu8_v* with

L1_DoubleMu3p5 seed.

5.3.2 Muons

The muon candidates are selected from the PF collection by matching them with appropriate trigger objects. The matching condition requires that a muon candidate and a trigger object to be separated by no more than $\Delta R \leq 0.3$. If there is more than one muon within the $\Delta R = 0.3$ range, then muons are sorted by minimum ΔR and the closest muon will be selected as our candidate.

Muon identification (as show in Fig. 5.2) is important in order to reject muons from cosmic rays, suppress hadronic punch-through, muons from decays in flight, and to guarantee a good p_T measurement. Tight identification criteria is used, as suggested by Muon POG. The PF based combined relative particle flow isolation (PFIsoCorr) is defined by

$$\text{PFIsoCorr} = \frac{[\sum^{Chhad} p_T + max(0., \sum^{Nhad} p_T + \sum^{EM} p_T - 0.5 \sum^{PU} p_T)]}{p_T^{\mu}} \leq 0.2, \qquad (5.1)$$

where the sum of charged hadrons (Ch had), neutral hadrons (N had), photons (EM), and charged particles from the pileup (PU), in cone of $\Delta R = 0.4$. The factor, 0.5, corresponds to a naive average of neutral to charged particles. This PFIsoCorr ≤ 0.2 is a loose cut that can be used to suppress the muon contamination within jets. Both muons must have opposite charge, $p_T > 20$ GeV and be located in the muon chamber region ($|\eta|$ <2.4).

The selections used in this analysis are listed in Table 5.3. The first and second muon candidates are obtained after all of the selection criteria are applied. Application of the full selection leave a set of events where most of the background, dominating by a few orders of magnitude, have been suppressed while retaining most of the signal. The p_T, η, and ϕ distributions of muon candidates are plotted in Fig. 5.3. Both ϕ and η distributions show good agreement between data and MC. The p_T plots show good agreement at the kinematic region up to 60 GeV, where we expect to find most of the muons coming from Z decays.

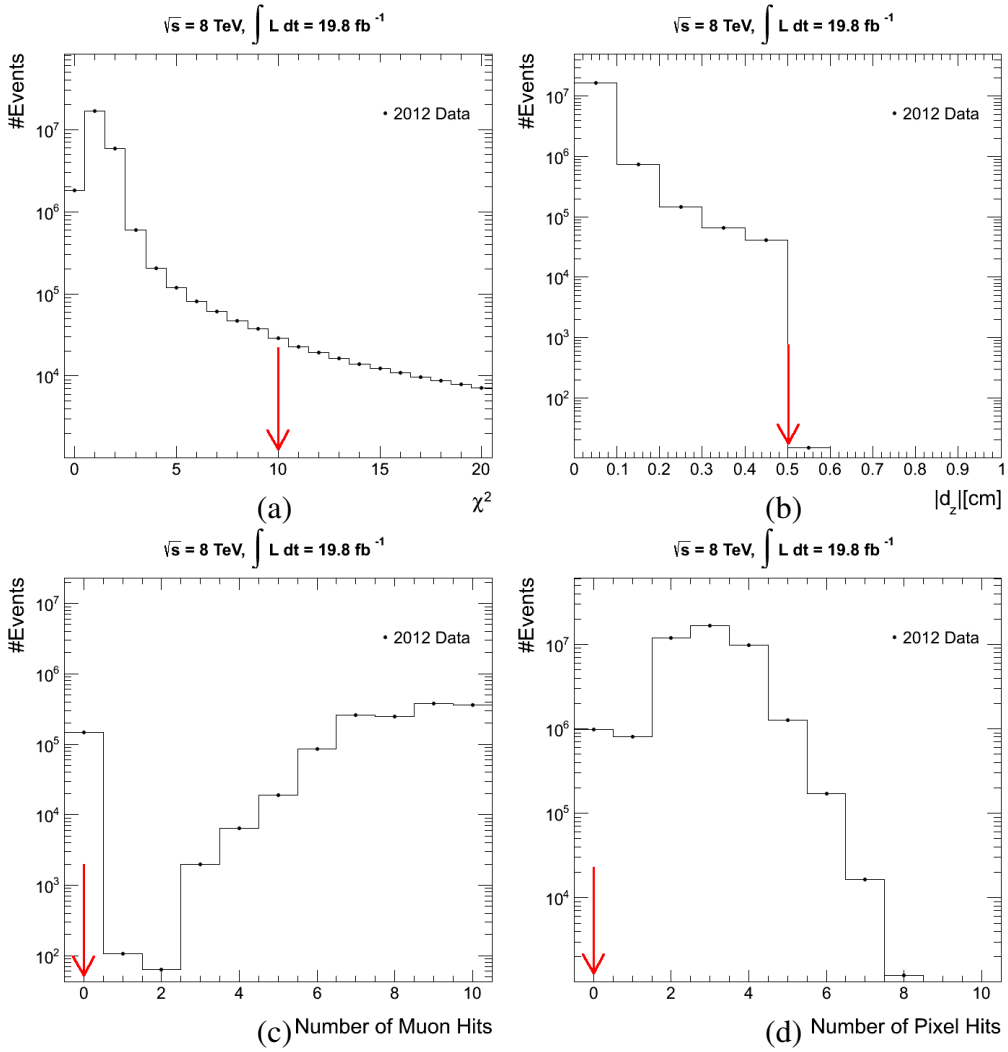


Figure 5.2: Properties of muon identifications before selection applied: (a) di-muon mass, (b) transverse momentum, (c) rapidity, and (d) azimuthal angle in data and MC.

Table 5.3: Muon selections used in this analysis.

Selection
$p_T^{\mu} > 20$ GeV
$\lvert\eta^{\mu}\rvert < 2.4$
recoMu.isGlobalMuon() & recoMu.isPFMuon()
Number of Pixel Hits > 0
Number of Muon Hits > 0
Number of Muon Stations > 1
Normalized $\chi^2 < 10$
$\lvert d_{xy}\rvert < 0.2$ cm
$\lvert d_z\rvert < 0.5$ cm
Number of tracker layer with hits > 5
PFIsoCorr ≤ 0.2

Efficiency Corrections

Since the efficiency is not one for both data and MC, the measured variables need to be corrected for efficiency. The efficiency can be factored into several contributions corresponding to the different steps of the selection processes. In order to extract particle efficiencies from the data without relying on the efficiency predications from MC, a data-driven approach to measure particle efficiencies called "Tag and Probe" method is utilised. Comparing data and MC efficiencies allows us to obtain necessary scale factors to correct the simulated samples in order for them to be consistent with data. In this analysis, the scale factors from Muon POG are used (more discussion in Appendix A).

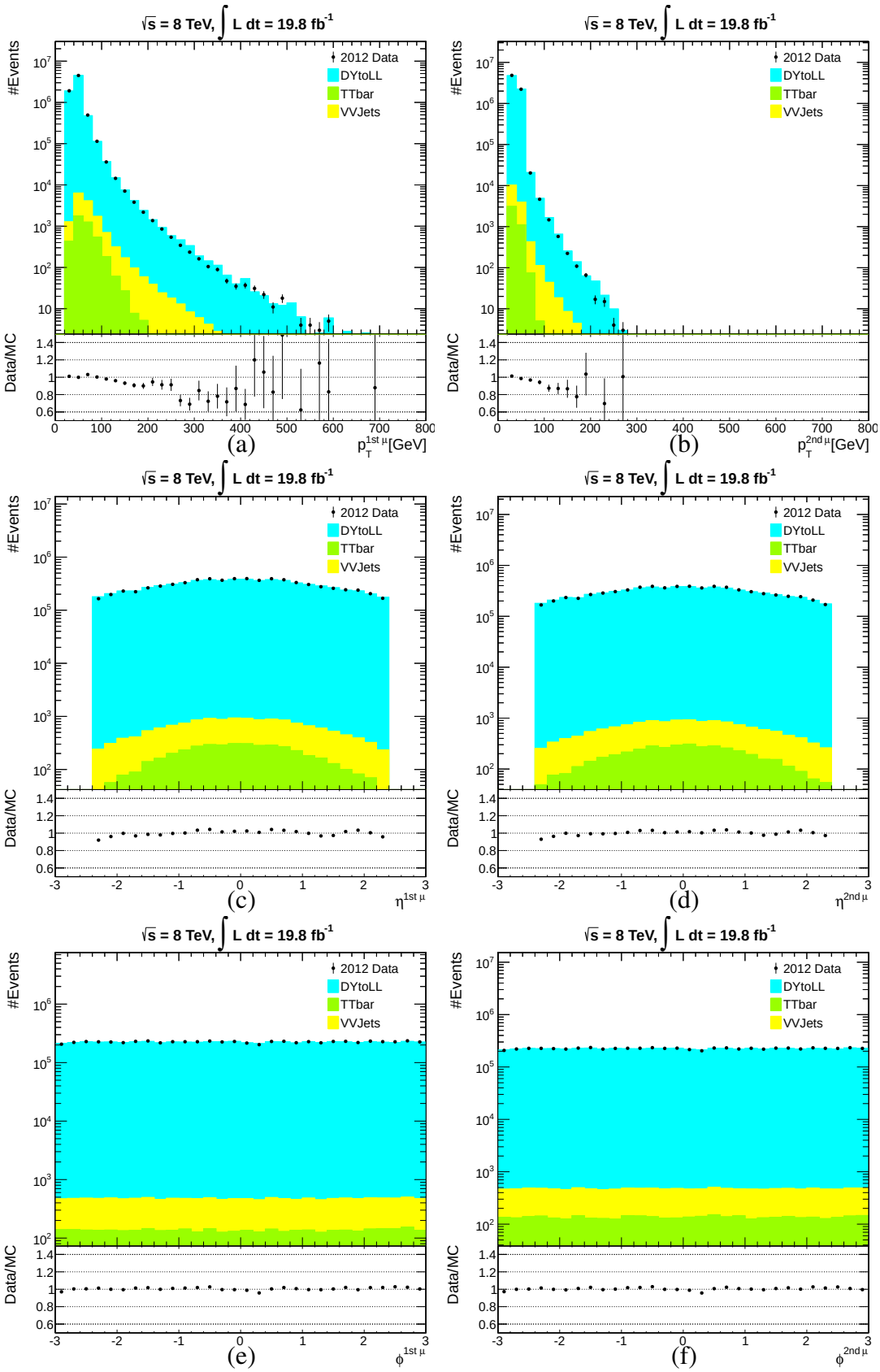


Figure 5.3: The first and second leading lepton p_T (a and b), η (c and d), and ϕ (e and f) in data and MC in the di-muon channel after all muon selections are applied.

5.3.3 Z Boson Reconstruction

Z bosons are reconstructed from the selected opposite charged di-muon events with the criteria explained in previous sections. The events lying in the Z mass window of $71 < M_Z < 111$ GeV/c^2 are used for the analysis. Figure. 5.4 shows the invariant mass, transverse momentum, rapidity, and azimuthal angle of the Z bosons respectively.

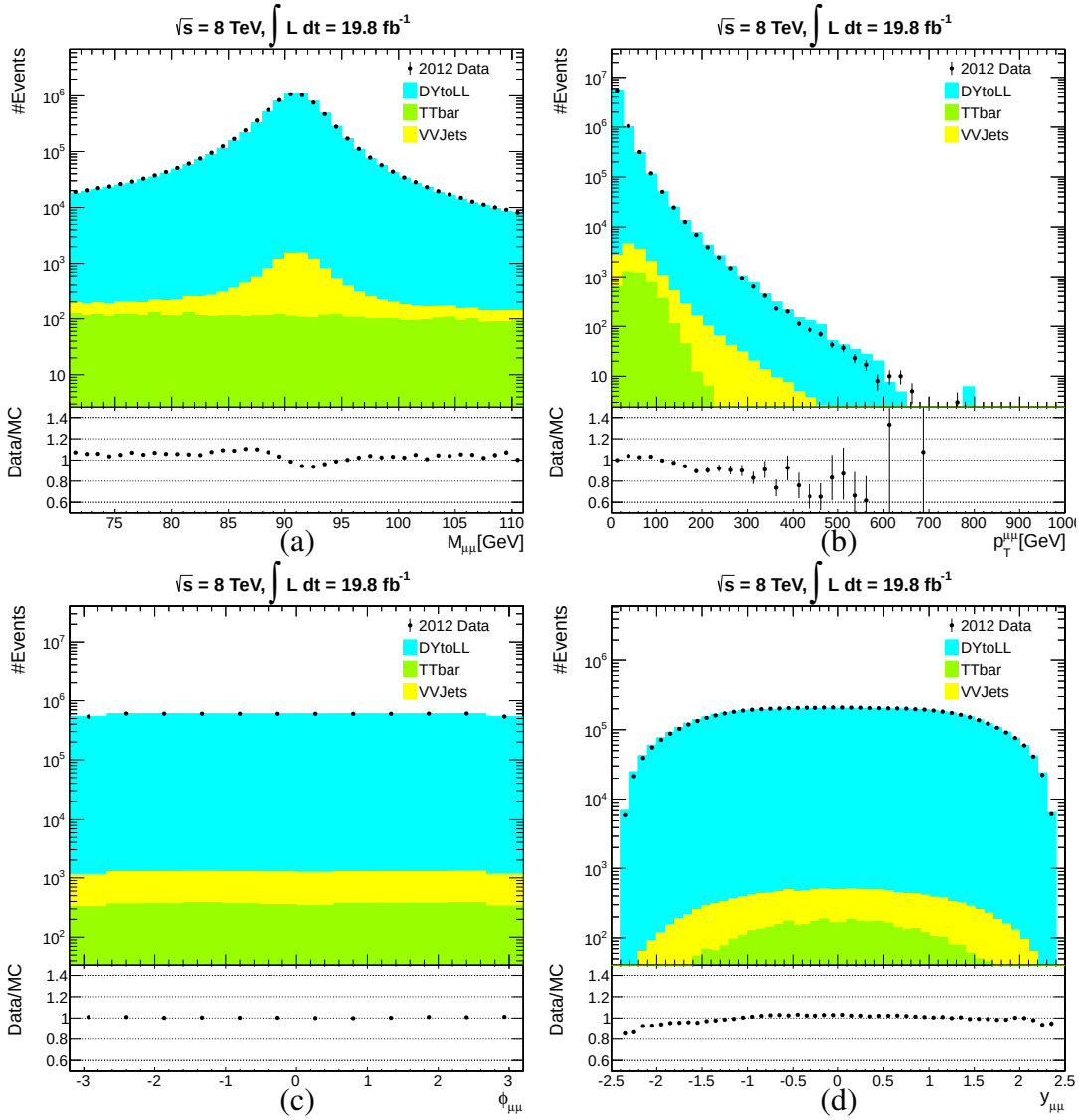


Figure 5.4: Reconstructed (a) di-muon mass, (b) transverse momentum, (c) rapidity, and (d) azimuthal angle in data and MC.

5.3.4 Jets

Jets are reconstructed using the PF algorithm. Jets are clustered using the anti-k_T algorithm with a size parameter of $\Delta R = \sqrt{\eta^2 + \phi^2} = 0.5$. Charged Hadron Subtraction is also used in order to ensure that jets are from the same primary vertex as the Z boson by requiring the z component of primary vertex to be $z < 24$ cm, the radius in $x - y$ plane to be $\rho < 2$ cm, and number of degrees of freedom of vertex fit to be $n.d.o.f. > 4$. Jet $p_T >$ 30 GeV threshold is used to reject low PU jets and increase jet energy resolution. Only jets within $|\eta| < 2.4$, within the silicon tracker acceptance, are selected to ensure the good quality of the tracking information.

During the event reconstruction for the final state partons, the decay leptons could mix with the jet collections. Leptons can intervene with jets in two possible ways: first, if the energy of a lepton is high enough, it might be reconstructed as a jet, which increases the jet multiplicity; second, the leptons can decay into a jet cone, adding to the total energy of the jets. These two categories of jets are vetoed if jets are in the vicinity of ΔR = 0.5 from lepton candidates. Jets are also applied JEC as in Section 4.2.2.

Jet and PU Identification

Loose identification criteria from JetMET POG are used to improve the jet quality by requiring that jet candidates must deposit energy in both ECAL and HCAL. The jet identification distributions are shown in Fig. 5.5.

Loose PU identification is applied to further reduce PU contamination. Due to the fact that PU jets primarily come from overlapping jets incurred during PU interactions, PU jets exhibit two characteristic features: they are both diffuse and, where charged particle identification is possible, some fraction of the charged particles will not point to the primary vertex. These characteristics allow for the identification of PU jets in regions where both charged particle tracking and jet shape identification are possible. Both vertex and shape information are combined through a multivariate analysis technique to give a single dis-

criminator targeting the identification of PU jets. This technique is known as the PU jet identification [89]. Table5.4 summarize the selection criteria used for jets.

Table 5.4: Jet selection criteria used in this analysis.

Selection
$p_T^{jet} > 30$ GeV
$\left\|\eta^{jet}\right\| < 2.4$
Neutral Hadron Fraction < 0.99
Neutral EM Fraction < 0.99
Number of Constituents > 1
Charged Hadron Fraction > 0
Charged Multiplicity > 0
Charged EM Fraction < 0.99
Loose PU ID

5.4 Measured Observables

The physical observables presented in this analysis are differential cross sections measured as a function of:

- Exclusive jet multiplicity (counting number of jets associated to the Z in each event).
- Transverse momentum and pseudorapidity of jets, up to the second highest p_T jet.

In this section, the comparisons between data and the predictions from MC simulation, including both signal and background, satisfying the selection criteria are shown. MC signal and backgrounds are reweighed for pileup effects and scale factor corrections are applied as previously described in Sections 5.2.1 and 5.3.2. The MC samples are normalised to the number of expected events:

$$N_{expected} = \frac{\sigma \times L}{N_{generated}} \tag{5.2}$$

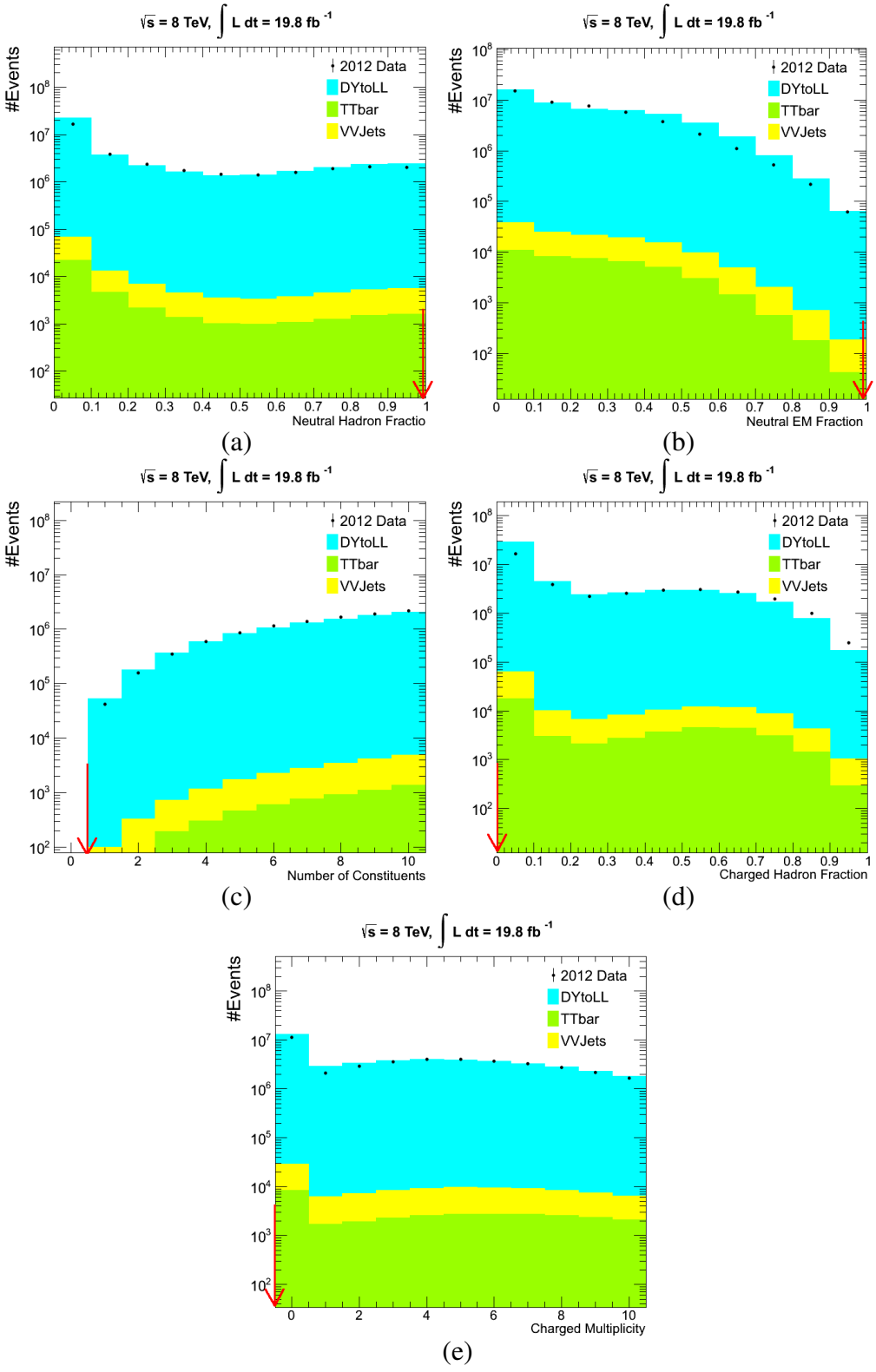


Figure 5.5: The jet identifications for (a) neutral hadron, (b) neutral EM fraction, (c) number of constituents, (d) charged hadron fraction, and (e) charged multiplicity before selections. The red arrow shows the selection values used.

where σ is the cross section of the process and L is the total integrated luminosity of the data.

The measurements of the exclusive and inclusive jet multiplicity distributions are shown in Fig. 5.6 (a) and (b). There is good agreement between data and MC up to 4 jets, where the theoretical calculation based on the LO matrix element up to 4 partons is available.

After the initial study on the statistic of jet p_T distributions, the results show the decreasing number of jets at high p_T. Thus, the variable bin ranges are used to compensate the jets with low statistics at high p_T. The p_T distributions of the first and second leading jets agree well at low p_T as shown in Fig. 5.6 (c) and (d). The η distributions of the first and second leading jets also agree well in the barrel region and show some discrepancy in the endcap region as shown in Fig.5.6 (e) and (f). The discrepancy of both p_T and η are mainly from the low statistic of MC.

5.5 Unfolding

The background contamination is subtracted bin by bin according to the MC predictions as in Table 5.2. The measured signal distributions are corrected for various experimental effects such as migration and acceptance by an unfolding procedure in order to obtain the cross sections at the particle level to be compared with theoretical predictions. In this analysis, the Iterative Baysian method is used for unfolding of the detector effects, as proposed by D'Agostini and implemented in the statistical analysis toolkit RooUnfold.

Response matrices (as shown in Fig. 5.7) are constructed from the MC DY+Jets reconstructed and generated quantities. The generated values refer to the stable leptons (status=1) from the decay of Z boson which are dressed with all the photons within the cone of radius 0.1 from the lepton to account for the final state radiation (FSR) and to generator-level hadrons. The phase space for the cross section measurement is $p_T > 20$ GeV/c and $|\eta| < 2.4$ for muons and Z mass cut as in the measured data between 71 and 111 GeV/c^2. Jets are at $p_T > 30$ GeV/c and $|\eta| < 2.4$.

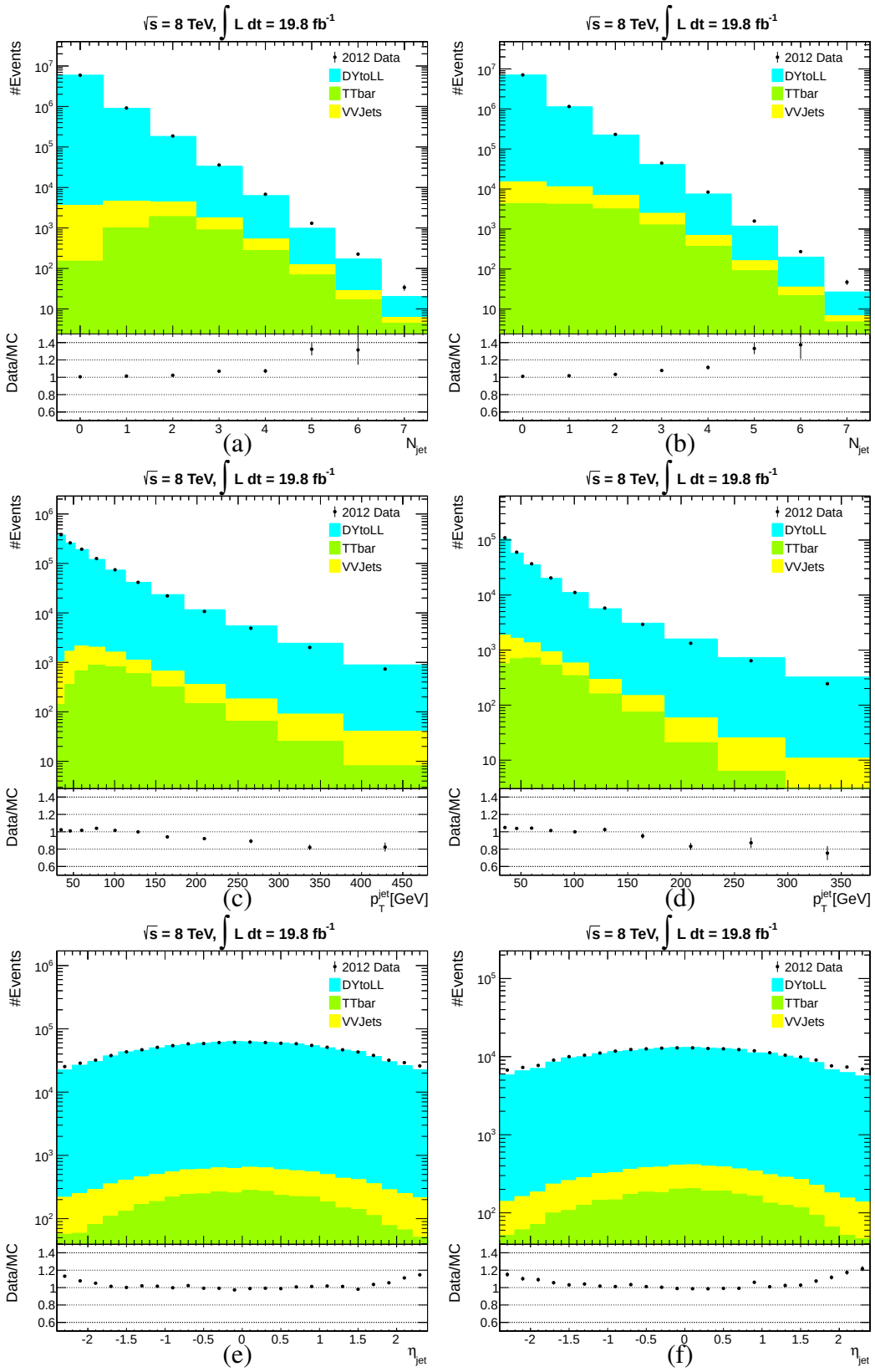


Figure 5.6: Plots show the measured variables of exclusive (a) and inclusive (b) jet multiplicity, the first leading jet p_T (c) and the second leading jet p_T (d), and the first leading jet rapidity (e) and the second leading jet η (f) in data and MC in the di-muon channel.

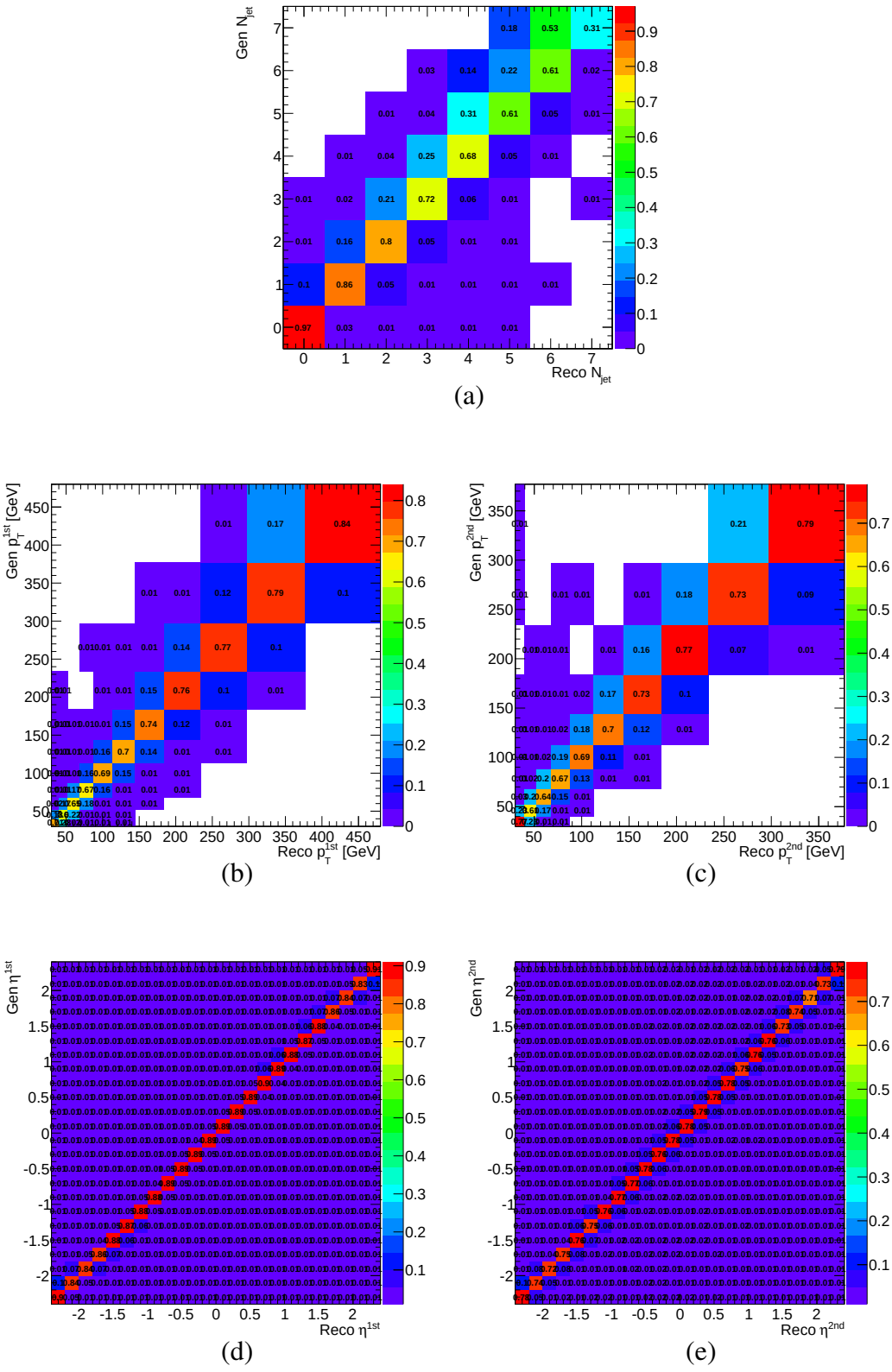


Figure 5.7: Unfolding response matrices of exclusive jet multiplicity (a), the first (b) and second (c) leading jet p_T , and the first (d) and second (e) leading jet η.

The unfolded differential cross section distributions as functions of exclusive jet multiplicity, the leading jet p_T, the second leading jet p_T, the leading jet η, and the second leading jet η are shown in Fig. 5.8. These plots show comparisons of the measured values before unfolding, iterative Baysian unfolding method (Bayes), singular value decomposition (SVD), and simple inversion of the response matrix without regularisation (Bin-by-Bin) with generator level values (Truth) for MADGRAPH interfaced with PYTHIA. The measured quantities are up to 30% less than the truth quantities. After measured quantities are unfolded to the generator level, the measured central values agree with the truth values and behave as seen in previous section. The Bayes method proves to be the best method of three because it give the less statistical error.

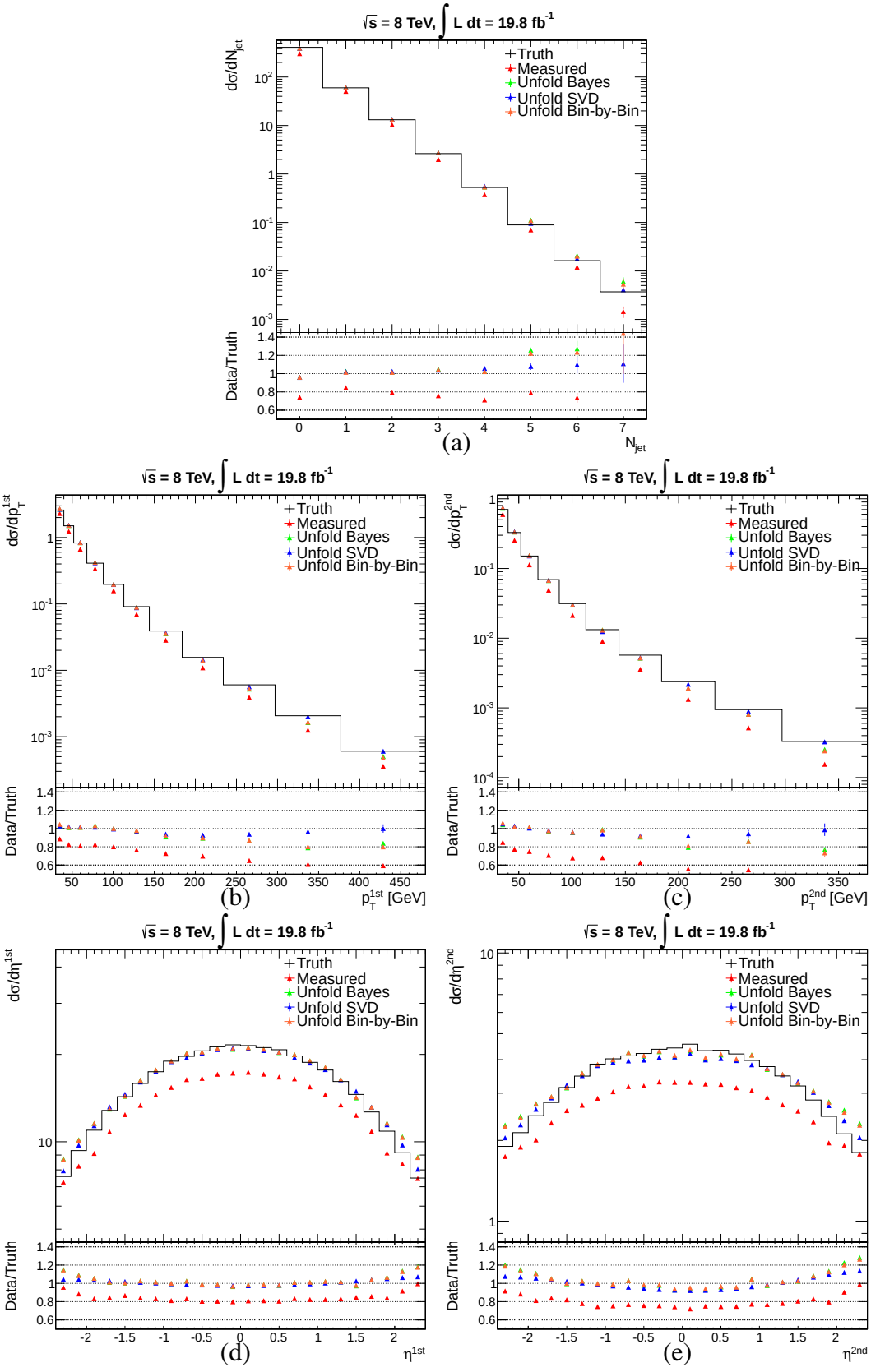


Figure 5.8: Unfolding results of exclusive jet multiplicity (a), the first (b) and second (c) leading jet p_T, and the first (d) and second (e) leading jet η.

5.6 Systematic Uncertainties

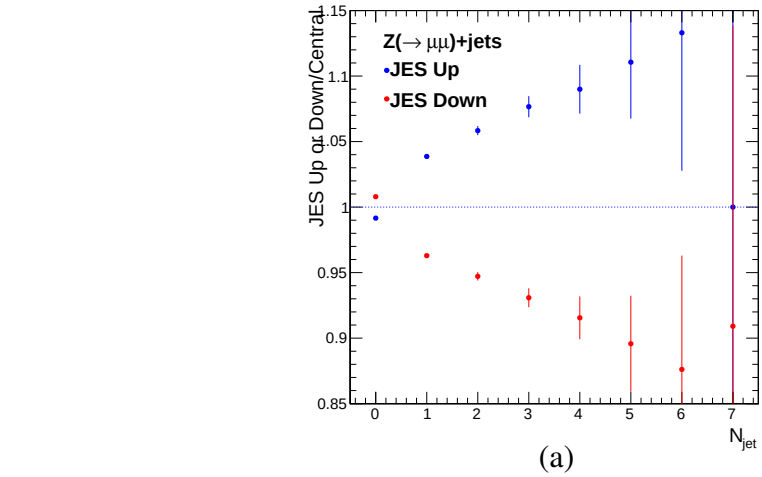
Systematic uncertainties in this analysis come from several sources and can be split between the experimental uncertainties and theoretical uncertainties. The experimental uncertainties include Jet Energy Scale (JES), Jet Energy Resolution (JER), background estimation, PU, unfolding, efficiency correction, and luminosity. The theoretical uncertainties include factorization and normalization scales of QCD and PDF.

In this section the two main sources of systematic uncertainties are described [2]. The largest systematic effects come from the uncertainty on the JES. This has been done by rescaling, up and down, the jet p_T spectrum in data. Plots illustrating the effect of jet energy scale variation on the data are shown in Fig. 5.9. The JES uncertainty affects the multiplicity dramatically due to several different factors such as non-uniformity and non-linearity of the detector response in pseudo-rapidity and momentum in the energy reconstruction or the PU subtraction. The contributions of p_T and η are estimated to be up to 5% for the phase space of interest. The other main contribution is the JER. The prescription is to scale a reconstructed (corrected) jet p_T based on the p_T difference between matched reconstructed and generated jets:

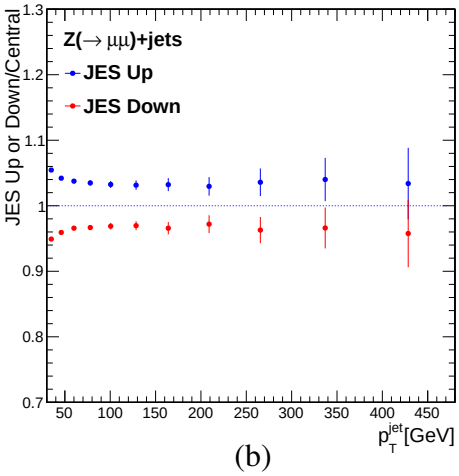
$$\text{scaled } p_T = \max[0,\ \text{gen } p_T + c \times (\text{reco } p_T - \text{gen } p_T)]. \tag{5.3}$$

Plots illustrating the effect on the data of the jet energy resolution variation are shown in Fig. 5.10. The JER uncertainties of p_T and η distributions are approximately 1-2%.

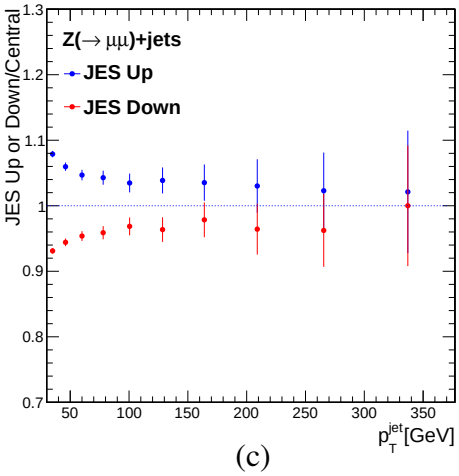
[2] Since this is the ongoing analysis, more uncertainties, such as PU and PDF, are in the study processes.



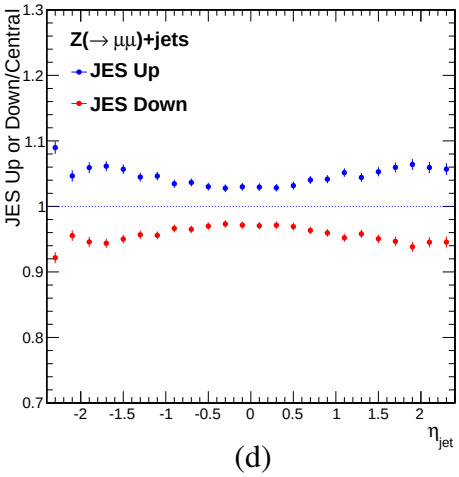
(a)



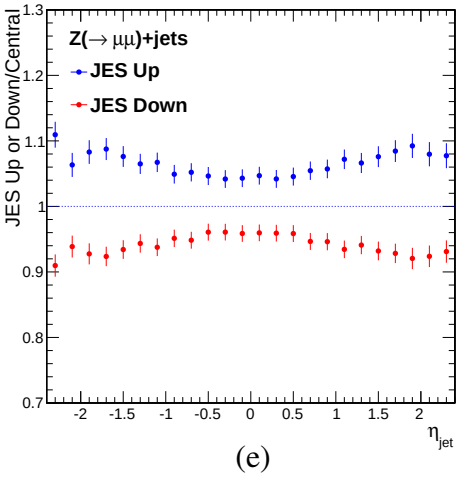
(b)



(c)

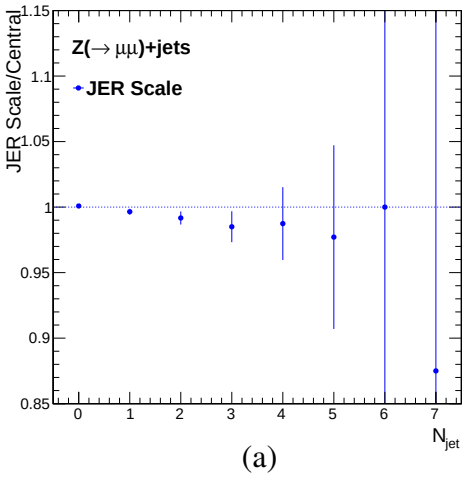


(d)

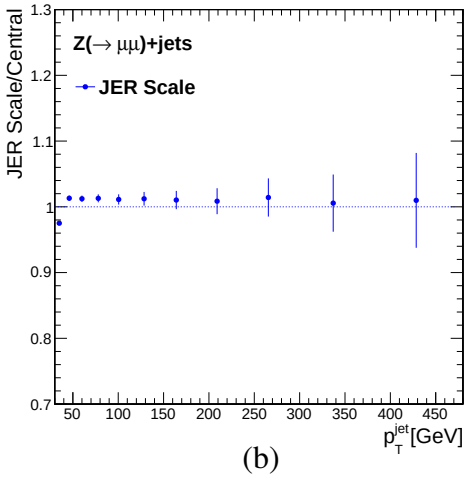


(e)

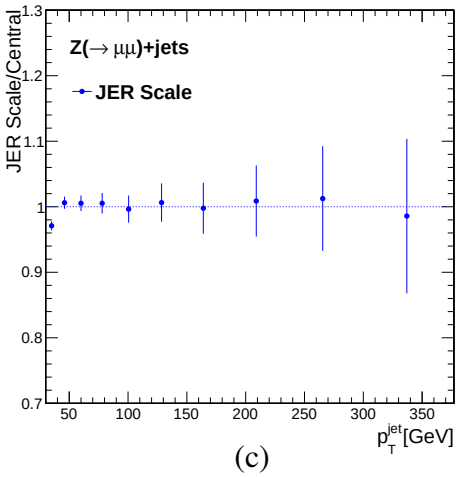
Figure 5.9: Systematic effects of jet energy scale of exclusive jet multiplicity (a), the first (b) and second (c) leading jet p_T, and the first (d) and second (e) leading jet η.



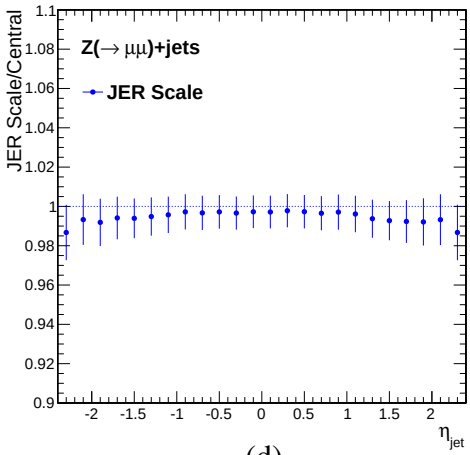
(a)



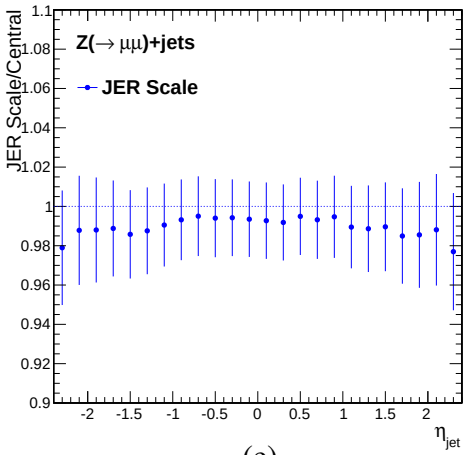
(b)



(c)



(d)



(e)

Figure 5.10: Systematic effects of jet energy resolution of exclusive jet multiplicity (a), the first (b) and second (c) leading jet p_T, and the first (d) and second (e) leading jet η.

5.7 Results

After applying the full selections presented in the previous chapters and the computation of uncertainties, differential cross sections as a function of exclusive jet multiplicities, the first and second leading jet p_T, and the first and second leading jet η of the Z ($\rightarrow \mu^+\mu^-$) + jets events are shown in Fig. 5.11. Table 5.5 - 5.9 have the final number of differential cross sections after efficiency corrections and unfolding. The given numbers are obtained from iterative Baysian Method unfolding method (in pb).

The differential cross section as a function of the exclusive jet multiplicity imply that the trend of the jet multiplicity represents the expectation of the pQCD prediction for an exponential decay with the number of jets. The jet differential cross sections $\frac{d\sigma}{dp_T}$ and $\frac{d\sigma}{d\eta}$ for the 2 highest p_T jets are consistent with the results in Sec. 5.4. MADGRAPH provides good prediction of jet multiplicity up to 4 jets, due to the recent pQCD calculation. The discrepancies at high p_T and high η jets are from the lack of statics of MC.

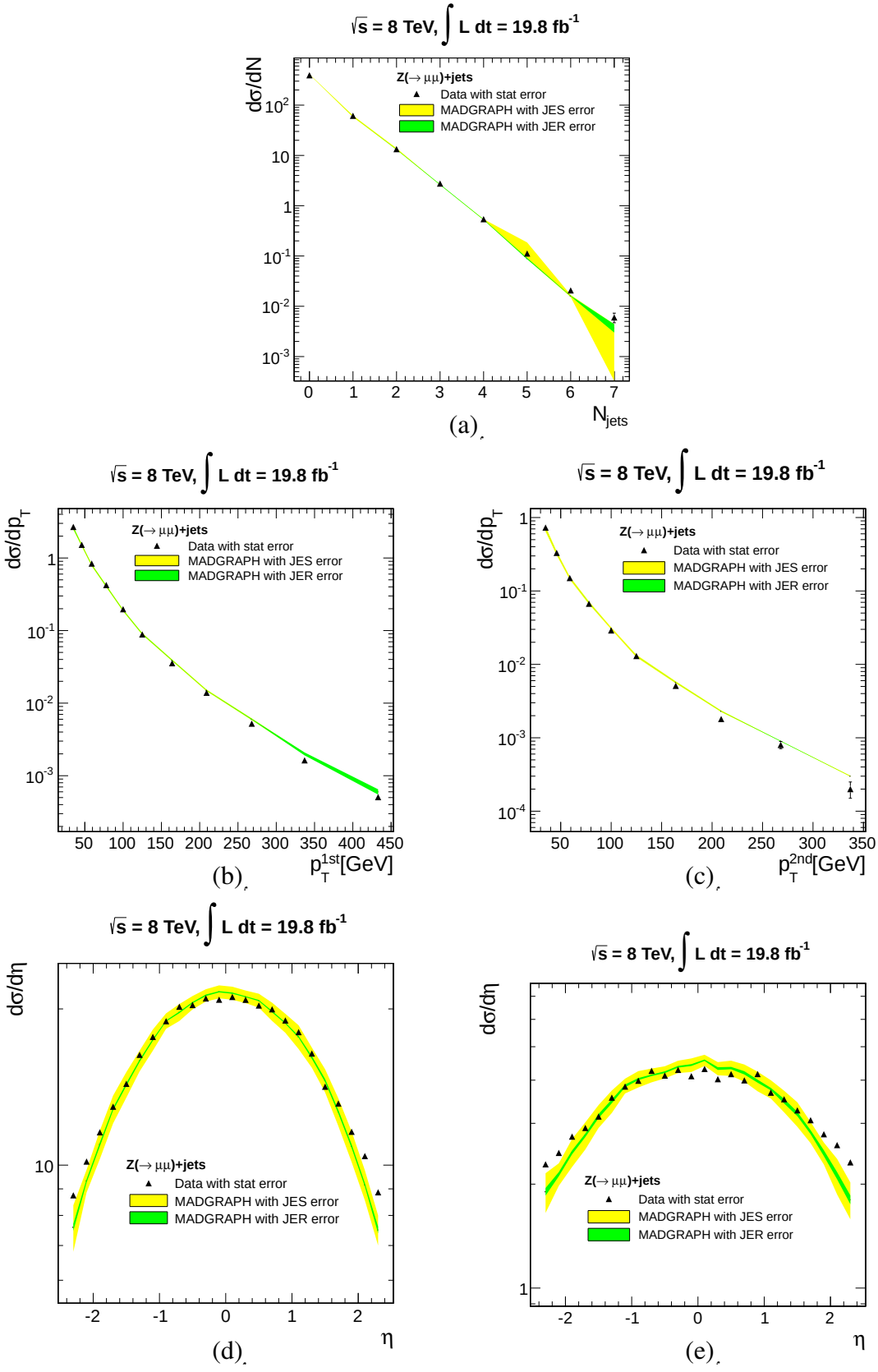


Figure 5.11: Differential cross section as a function of exclusive jet multiplicity (a), the first (b) and second (c) leading jet p_T, and the first (d) and second (e) leading jet η.

Table 5.5: Differential cross section as function of number of jets: muon channel

Njet	$d\sigma/dN_{jet}$[pb]	stat	JES	JER
0	393	0.17	3.3	0.33
1	61.36	0.077	2.27	0.21
2	13.35	0.038	0.801	0.11
3	2.76	0.018	0.104	0.04
4	0.54	0.0084	0.01	0.007
5	0.112	0.004	0.097	0.0026
6	0.0207	0.0019	0.00079	0.0002
7	0.00606	0.0013	0.0007	0.0007

Table 5.6: Differential cross section as function of leading jet p_T: muon channel

jet p_T [GeV/c]	$d\sigma/dp_T$[pb/(GeV/c)]	stat	JES	JER
30 - 40	2.68	0.0038	0.14	0.067
40 - 52	1.52	0.0031	0.064	0.02
52 - 68	0.838	0.0021	0.031	0.01
68 - 88	0.424	0.0014	0.015	0.0055
88 - 113	0.197	0.00086	0.0064	0.0022
113 - 144	0.0885	0.00052	0.0027	0.001
144 - 184	0.0355	0.00029	0.0011	0.00037
184 - 234	0.0139	0.00017	0.00041	0.0001
234 - 297	0.00522	0.00009	0.00018	7.4×10^{-5}
297 - 377	0.00163	0.00005	6.5×10^{-5}	9.2×10^{-5}
377 - 480	0.000507	0.00002	1.7×10^{-5}	5×10^{-5}

Table 5.7: Differential cross section as function of second leading jet p_T: muon channel

jet p_T [GeV/c]	$d\sigma/dp_T$[pb/(GeV/c)]	stat	JES	JER
30 - 40	0.732	0.0021	0.058	0.0044
40 - 52	0.334	0.0016	0.02	0.0018
52 - 68	0.153	0.00097	0.007	0.000345
68 - 88	0.0671	0.00059	0.003	0.00042
88 - 113	0.0299	0.00036	0.001	2.8×10^{-5}
113 - 144	0.0131	0.00022	0.0005	4.7×10^{-5}
144 - 184	0.00518	0.00012	0.00018	2.1×10^{-5}
184 - 234	0.00188	0.000067	5.7×10^{-5}	1×10^{-5}
234 - 297	0.000812	0.000038	5.4×10^{-6}	8.2×10^{-6}
297 - 377	0.000253	0.00002	5×10^{-6}	6.8×10^{-7}

Table 5.8: Differential cross section as function of leading jet η: muon channel

jet η	$d\sigma/d\eta$[pb]	stat	JES	JER
-2.4< η <-2.2	8.74	0.076	0.78	0.12
-2.2< η <-2.0	10.2	0.077	0.47	0.068
-2.0< η <-1.8	11.6	0.082	0.69	0.094
-1.8< η <-1.6	12.9	0.086	0.79	0.076
-1.6< η <-1.4	14.3	0.09	0.81	0.087
-1.4< η <-1.2	16.3	0.096	0.73	0.083
-1.2< η <-1.0	17.7	0.099	0.82	0.075
-1.0< η <-0.8	18.9	0.1	0.65	0.052
-0.8< η <-0.6	20.2	0.1	0.74	0.066
-0.6< η <-0.4	20.3	0.1	0.61	0.057
-0.4< η <-0.2	20.9	0.1	0.58	0.07
-0.2< η < 0.0	20.8	0.1	0.62	0.057
0.0< η < 0.2	21.09	0.1	0.62	0.06
0.2< η < 0.4	20.8	0.1	0.59	0.045
0.4< η < 0.6	20.3	0.1	0.64	0.054
0.6< η < 0.8	19.97	0.1	0.8	0.069
0.8< η < 1.0	19.001	0.1	0.79	0.055
1.0< η < 1.2	18.043	0.1	0.93	0.07
1.2< η < 1.4	16.41	0.096	0.72	0.1
1.4< η < 1.6	14.162	0.09	0.75	0.1
1.6< η < 1.8	13.145	0.087	0.78	0.1
1.8< η < 2.0	11.596	0.081	0.74	0.091
2.0< η < 2.2	10.41	0.077	0.61	0.07
2.2< η < 2.4	8.86	0.075	0.5	0.11

Table 5.9: Differential cross section as function of the second leading jet η: muon channel

jet η	$d\sigma/d\eta$[pb]	stat	JES	JER
-2.4< η <-2.2	2.28	0.042	0.25	0.048
-2.2< η <-2.0	2.46	0.041	0.16	0.03
-2.0< η <-1.8	2.74	0.043	0.23	0.032
-1.8< η <-1.6	2.91	0.045	0.25	0.032
-1.6< η <-1.4	3.14	0.046	0.24	0.044
-1.4< η <-1.2	3.56	0.049	0.26	0.044
-1.2< η <-1.0	3.833	0.052	0.19	0.036
-1.0< η <-0.8	3.98	0.052	0.22	0.027
-0.8< η <-0.6	4.26	0.054	0.19	0.021
-0.6< η <-0.4	4.12	0.053	0.17	0.025
-0.4< η <-0.2	4.28	0.054	0.18	0.025
-0.2< η < 0.0	4.11	0.053	0.2	0.027
0.0< η < 0.2	4.31	0.055	0.17	0.022
0.2< η < 0.4	4.03	0.052	0.19	0.039
0.4< η < 0.6	4.16	0.054	0.22	0.034
0.6< η < 0.8	3.99	0.052	0.24	0.046
0.8< η < 1.0	4.17	0.054	0.26	0.040
1.0< η < 1.2	3.68	0.05	0.23	0.031
1.2< η < 1.4	3.52	0.049	0.25	0.053
1.4< η < 1.6	3.27	0.047	0.26	0.045
1.6< η < 1.8	3.06	0.046	0.2	0.037
1.8< η < 2.0	2.79	0.044	0.18	0.048
2.0< η < 2.2	2.59	0.042	0.25	0.051
2.2< η < 2.4	2.31	0.042	0.22	0.053

Chapter 6

Conclusion

The study of jet productions in association with a weak vector boson in pp collision provides a good opportunity to test pQCD. Since theoretical predictions at next-to-next-to-leading order are only available for jet multiplicities up to 4 jets, the understanding of $Z/\gamma^* \rightarrow \mu^+\mu^-$ + jets final states from the CMS data is important as the input for theoretical physics community.In addition, new physics searches including the Higgs boson or SUSY would benefit from studying background processes. The final state decay of jets and the Z boson into high momentum leptons can be misidentified as a signature of new particles.

The angular distributions for the Z boson and a single jet have been analysed. Data of 5 fb^{-1} at $\sqrt{s}$ = 7 TeV are collected by the CMS experiment in 2011. The rapidity distributions of the Z boson ($|y_Z|$) and a jet ($|y_{jet}|$) are found to agree with predictions from SHERPA, MADGRAPH, and MCFM. The rapidity sum (y_{sum}) between the Z boson and a jet is described by all predictions up to 5% precision for $y_{sum} < 1.0$. At $y_{sum} > 1.0$, SHERPA is the best described due to the hybrid calculations that employ NLO PDF. The rapidity difference (y_{diff}) is best described with MCFM. Both SHERPA and MCFM are different by methods in which partons from matrix elements are matched to parton showers. MADGRAPH disagreed considerably because it uses LO PDF.

The jet productions associated with the Z boson, considering the muon final state decay, of the 19.8 fb^{-1} CMS data during 2012 at $\sqrt{s}$ = 8 TeV, is studied here. Differential cross sections as function of the jet multiplicity and of the transverse momentum and rapidity of the first and second leading jets have been measured. The measured differential

cross sections have been corrected for detector effects, back to the hadron level, using an unfolding method which also accounts for the efficiency of the selection criteria. The final distributions are compared to pQCD predictions obtained by MADGRAPH with LO PDF. In general, the LO calculations reproduce well the shape of the distributions. For future developments, NLO calculation will represent a considerable improvement and the analysis update with the 13 TeV data, available with a greater statistics, will be an interesting development putting more stringent constraints to the predictions.

Bibliography

[1] CMS Collaboration, "Rapidity distributions in exclusive Z+jet and γ+jet events in pp collisions at $\sqrt{s}$ = 7 TeV", Physical Review D 88, 112009 (2013) doi:10.1103/PhysRevD.88.112009

[2] F. Halzen and A. Martin, "Quarks and leptons: an introductory course in modern Particle Physics", Wiley, New York USA (1984).

[3] M.Y. Han and Y. Nambu, "Three Triplet Model with Double $SU(3)$ Symmetry", Phys.Rev., 139:B1006-B1010, (1965).

[4] S. L. Glashow, Nucl. Phys. 22, 579 (1961).

[5] S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967)

[6] A. Salam, "In Elementary Particle Theory", ed. N. Svartholm (Almquist and Wiksells, Stockholm (1969).

[7] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys. Lett., B716(1):30, arXiv:1207.7235. 2.1 (2012).

[8] R. P. Feynman, "Quantum Electrodynamics", Benjamin, New York (1962).

[9] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "Experimental test of parity convservation in beta decay", Phys. Rev., 105:1413-1414 (1957).

[10] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", Physical Review Letters 10 (12): 531-533 (1963) doi:10.1103/PhysRevLett.10.531.

[11] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", Progress of Theoretical Physics 49 (2): 652-657 (1973) doi:10.1143/PTP.49.652.

[12] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", Phys.Lett., 8:214-215 (1964).

[13] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture", Phys.Lett., B47:365-368 (1973).

[14] D.J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories", Phys.Rev., D8:3633-3652 (1973).

[15] S. Weinberg, "Non-Abelian Gauge Theories of the Strong Interactions", Phys.Rev.Lett., 31:494-497 (1973).

[16] CTEQ Collaboration, R. Brock et al., "Handbook of perturbative QCD: Version 1.0", Rev. Mod. Phys. 67 157-248 (1995).

[17] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "MRST2001: Partons and alpha(s) from precise deep inelastic scattering and Tevatron jet data", Eur. Phys. J. C23 73-87, arXiv:hep-ph/0110215 (2002).

[18] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", Nucl. Phys. B126 298 (1977).

[19] H. L. Lai et al., "Global QCD analysis and the CTEQ parton distributions", Phys. Rev. D51 4763-4782 arXiv:hep-ph/9410404 (1995).

[20] Particle Data Group Collaboration, K. Nakamura et al., "Review of particle physics", JP. G37, 075021 (2010).

[21] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and Collider Physics", volume 8 of 375 Cambridge Monographs on Part. Phys., Nucl. Phys. and Cosm. Cambridge (1996)

[22] CDF Collaboration, "W Boson + Jet Angular Distribution in p$\bar{p}$ Collsions at $\sqrt{s} =$ 1.8 TeV",Phys. Rev. Let.732296-2300 (1994). doi:10.1103/PhysRevLett.73.2296.

[23] P. M. Rubinov,"The Center-of-mass angular distribution of direct photons at $\sqrt{s} =$ 1.8 TeV observed with the D0 detector", FERMILAB-THESIS-1995-72 (1995). Ph.D. Thesis (Advisor: Guido Finocchario).

[24] CMS Collaboration,"Measurement of Dijet Angular Distributions and Search for Quark Compositeness in pp Collisions at $\sqrt{s} = 7$ TeV",Phys. Rev. Lett.106 201804 (May, 2011). doi:10.1103/PhysRevLett.106.201804

[25] ATLAS Collaboration, "High-pT dijet angular distributions in pp Collisions at $\sqrt{s} =$ 7 TeV measured with the ATLAS detector at the LHC",Technical Report ATLAS-CONF-2010-074, CERN, Geneva (Jul, 2010).

[26] T. Aaltonen et al., "Measurement of inclusive jet cross-sections in Z/$\gamma^*_{(e^+e^-)}$ + jets production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV", Phys.Rev.Lett., 100:102001 (2008).

[27] V.M. Abazov et al., "Measurements of differential cross sections of Z/γ^*+jets+X events in proton anti-proton collisions at $\sqrt{s}$ = 1.96 TeV", Phys.Lett., B678:45-54 (2009).

[28] D. D. Price, "W/Z properties and V +jets at the Tevatron", EPJ Web Conf. , 28:06006 (2012).

[29] G. Aad et al., "Measurement of the production cross section for Z/γ^* in association with jets in $p-p$ collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector", Phys.Rev., D85:032009 (2012).

[30] CMS Collaboration, "Jet Production Rates in Association with W and Z Bosons in $p-p$ Collisions at $sqrts$ = 7 TeV", JHEP, 1201:010 (2012).

[31] K. Melnikov and F. Petriello, "Electroweak Gauge Boson Production at Hadron Colliders Through O (α_s^2)", Phys. Rev. D 74, no. 11, 114017 (2006). doi:10.1103/PhysRevD.74.114017.

[32] H. Ita, Z. Bern, L.J. Dixon, F.F. Cordero, D.A. Kosower, et al., "Precise Predictions for Z + 4 Jets at Hadron Colliders", Phys.Rev., D85:031501 (2012).

[33] S. D. Drell and T. M. Yan, "Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies", Physical Review Letters 25 (5): 316-320 (1997) doi:10.1103/PhysRevLett.25.316. And erratum in S. D. Drell and T. M. Yan, Physical Review Letters 25 (13): 902 (1997) doi:10.1103/PhysRevLett.25.902.2.

[34] P. Nadolsky, J. Gao, M. Guzzi, J. Huston, H. Lai, et al., "Progress in CTEQ-TEA PDF analysis", arXiv:1206.3321 (2012).

[35] F. A. Berends, H. Kuijf, B. Tausk, and W.T. Giele, "On the production of a W and jets at hadron colliders", Nucl.Phys., B357:32-64 (1991).

[36] L. Cerrito, "Measurements of vector boson plus jets at the Tevatron", FERMILAB-CONF-10-311-E, arXiv:1008.2291 (2010).

[37] T. L. S. Group, "The Large Hadron Collider Conceptual Design", CERN-AC-95-05, arXiv:hep-ph/0601012 (1995).

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

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

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

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

[42] CMS Collaboration,"CMS Luminosity Public Results", Public Twiki (2012).

https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults

[43] CMS Collaboration,"The Magnet Project Technical Design Report", CERN/LHCC 1997-10 (1997).

[44] CMS Collaboration,"The CMS Tracker System Project Technical Design Report", CERN/LHCC 1997-10 (1997).

[45] CMS Collaboration, "CMS Tracker Detector Performance Results", Public Twiki (2012). https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK

[46] CMS Collaboration,"The CMS Electromagnetic Calorimeter Project Technical Design Report", CERN/LHCC 1997 (1997).

[47] G. Dissertori, "Electromagnetic Calorimetry and $e\gamma$ Performance in CMS", Tech. Rep. CMS-CR-2003-024, CERN, Geneva (2003)

[48] CMS Collaboration. "The CMS Hadron Calorimeter Project Technical Design Report", CERN/LHCC 1997-31 (1997).

[49] S. Abdullin et al, Eur. Phys. J. C, 53, 139 (2008).

[50] CMS Collaboration, "The CMS Muon Project Technical Design Report", CERN/LHCC 1997-32 (1997).

[51] CMS Collaboration, "Level-1 Trigger Technical Design Report", CERN/LHCC 2000-038 (2000).

[52] CMS Collaboration, “DAQ and High-Level Trigger Technical Design Report”, CERN/LHCC 2002-026 (2002).

[53] CMS Collaboration, “Measurement of CMS luminosity”, CMS-PAS-EWK-10-004 (2010).

[54] CMS Collaboration, “Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update”, CMS-PAS-SMP-12-008 (2012).

[55] CMS Collaboration, “CMS Physics Technical Design Report Volume I:Detector Performance and Software”, CERN/LHCC 2006-001 (2006).

[56] M.Lutz,“Programming Python”, Sebastopol, CA:O’Reilly (1996).

[57] R. Brun and F. Rademaker,“ ROOT: An Object oriented data analysis framework”, Nucl. Instrum. Meth. A389 81-86 (1997).

[58] M.Dobbs and J. B. Hansen,“ The HepMC C++ Monte Carlo Event Record for High Energy Physics”, Comput. Phys. Commun. 134 41-46 (2001).

[59] GEANT 4. Collaboration, “GEANT 4 - A Simulation Toolkit”, Nucl. Instrum. Meth. A506, no. 3, 250-303, (2003). doi:10.1016/S0168-9002(03)01368-8.

[60] W. Adam et al,“PAT: The CMS Physics Analysis Toolkit”, J. Phys.: Conf. Ser. 219 032017 (2010).

[61] H. Ita, Z. Bern, L. J. Dixon, F. Febres Cordero, D. A. Kosower, and D. MaÄśtre, “Precise predictions for Z-boson +4 jet production at hadron colliders”, Phys. Rev. D 85, 031501(R) (2012).

[62] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001 (2009).

[63] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002 (2010).

[64] CMS Collaboration, "Performance of Muon Identification in pp Collisions at s = 7 TeV", CMS Physics Analysis Summary CMS-PAS-MUO-10-002 (2010).

[65] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt Jet Clustering Algorithm", JHEP 4 63 (2008). doi:10.1088/1126-6708/2008/04/063.

[66] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", arXiv:1107.4277 (2011).

[67] F. Maltoni and T. Stelzer, "MadEvent: Automatic Event Generation with MadGraph", JHEP 2 027 (2003), doi:10.1088/1126-6708/2003/02/027.

[68] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 Physics and Manual", JHEP 5 026 (2006). doi:10.1088/1126-6708/2006/05/026.

[69] R. Field, "Early LHC Underlying Event Data - Findings and Surprises", arXiv:1010.3558v1 (2010).

[70] R. Kleiss and W. J. Stirling, "Top Quark Production at Hadron Colliders: Some Useful Formulae", Z. Phys. 40, no. 3, 419-423 (1988). doi:10.1007/BF01548856.

[71] GEANT 4 Collaboration, "GEANT 4 Developments and Applications", IEEE Trans. Nucl. Sci 53, no. 1, 270-278 (2006). doi:10.1109/TNS.2006.869826.

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

[73] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", CMS Physics Analysis Summary CMS-PAS-TRK-10-005 (2010).

[74] CMS Collaboration, "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s}$ = 7 TeV", arXiv:1107.4789 (2011).

[75] CMS Collaboration, "Rates of Jets Produced in Association with W and Z Bosons", CMS Physics Analysis Summary CMS-PAS-EWK-10-012 (2011).

[76] CMS Collaboration,"MuonPOG", Twiki (2012).

https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs

[77] T. Gleisberg and et al., "Event Generation with SHERPA 1.1", arXiv:0811.4622v1 (2008).

[78] S. Alekhin et al., "The PDF4LHC Working Group Interim Report",arXiv:1101.0536 (2011).

[79] H. Paukkunen and P. Zurita, "PDF reweighting in the Hessian matrix approach", arXiv:1402.6623 (2014).

[80] S. Linn and S. S. Yu, "Theoretical Predictions for the Measurement of Z+jet and γ+jet Angular Distribution in pp Collisions at $\sqrt{s}$ = 7 TeV", CMS Physics Analysis Note, AN-12-230 (2012)

[81] J. Campbell, K. Ellis, and C. Williams, "MCFM - Monte Carlo for FeMtobarn processes", Fermi Lab (2012). http://mcfm.fnal.gov/

[82] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-pair production in hadronic collisions ", JHEP01 013 (2007). doi:10.1088/1126-6708/2007/01/013

[83] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, "QCD Matrix Elements + Parton Showers", JHEP11 063 (2001). doi:10.1088/1126-6708/2001/11/063

[84] A. Singh, S. Yu, D. Mekterovic et al., "Measurement of Z^0/γ +Jet Angular Distributions in pp Collisions at $\sqrt{s}$ = 7 TeV with the Electron Decay Mode", CMS Physics Analysis Note, AN-12-135 (2012).

[85] G. D'Agostini. "A multidimensional unfolding method based on Bayes' theorem", Nuclear Instruments and Methods in Physics Research A, 362:487-498 (1995).

[86] P. Nadolsky et al., "Progress in CTEQ-TEA PDF Analysis", arXiv:1206.3321 (2012). doi:10.3204/DESY-PROC-2012-02/301

[87] CMS Collaboration,"Charged particle multiplicities in pp interactions at $\sqrt{s}$ = 0.9, 2.36, and 7 TeV", JHEP 1101 079 (2011). doi:10.1007/JHEP01(2011)079

[88] N. Kidonakis, "Differential and total cross sections for top pair and single top production", arXiv:1205.3453 (2012). doi:10.3204/DESY-PROC-2012-02/251

[89] CMS Collaboration, "Pileup Jet Identification", CMS Physics Analysis Summary CMS-PAS-JME-13-005 (2013).

[90] N. Adam, J. Berryhill, V. Halyo, A. Hunt, and K. Mishra, "Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data", CMS-AN-2009-111 (2009).

[91] CMS Collaboration,"Tag and Probe ", Twiki (2014).

https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe

[92] G. Cowens, "A Survey of Unfolding Methods for Particle Physics", in Proc. Advanced Statistical Techniques in Particle Physics. Durham (2002).

[93] T. Adye, "Unfolding algorithms and tests using RooUnfold", arXiv:1105.1160 (2011).

Appendix A

Tag and Probe

One of critical elements in high energy physics analysis is accurate efficiency measurements. Efficiency calculations from MC simulation produce large systematic errors from imperfection of particle and detector simulation. Consequently, efficiency measurements without reference from simulation are important to physics analysis. The method, called "Tag and Probe", is developed to fulfill this requirements by solely relying on real data. This method uses particles that ease to identify and are very low fake rate from the known di-object resonances such as Z or J/ψ. The "Tag and Probe" method requires one object to satisfy tight requirements designed to isolate the required particle type ("tag" object). The other object is required to satisfy the loose requirements ("probe" object). Invariant mass (in this case the Z from muons and electrons) will be calculated with this pair of objects. The efficiency is measured by:

$$\varepsilon = \frac{2N_{TT} + N_{TP}}{2N_{TT} + N_{TP} + N_{TF}}, \tag{A.1}$$

where T is a tag object, P is a "Pass" probe object, and F is a "Fail" probe object for signal yield (N). More details about this method are discussed at [90] and [91]. Muon ID, isolation, and trigger efficiencies of both Data and MC are calculated and provided by MuonPOG [76].

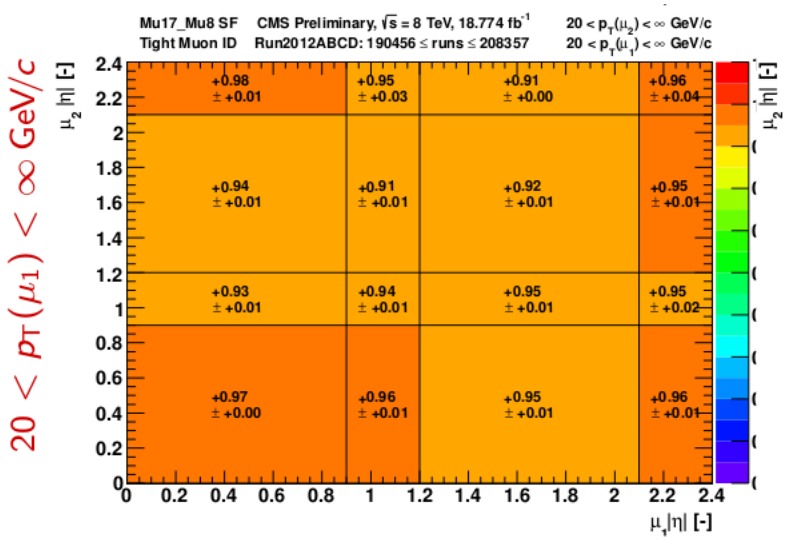


Figure A.1: 2D plot of trigger scale factor for 2012 (8 TeV) as provided by MuonPOG [76]

A.1 Efficiency Scale Factors

Both data and MC have their own efficiencies calculate from tag and probe method. In order to correct the differences and make the proper comparisons between the two, scale factors, the ratio between data and MC efficiencies, is applied on MC to compensate the measured differences. The trigger scale factors are given as a function of $|\eta|$ for both muons. The identification and isolation scale factors are given as a function of p_T and $|\eta|$. Trigger scale factor for 2011 (7 TeV) of double muons is considerably small, so it safely assumes to be 1. Scale factors of 2012 (8 TeV) are showed in Fig. A.1. The identification and isolation scale factors for 2011 (7 TeV) and 2012 (8 TeV) are shown in Table . A.1 - A.3.

Table A.1: Combined muon identification and isolation scale factors (Data/MC) for 2011 Tight ID.

Combined Scale Factors for Muons Tight ID

	$0 < \|\eta\| \leq 0.9$	$0.9 < \|\eta\| \leq 1.2$	$1.2 < \|\eta\| \leq 2.1$
$20 < p_T \leq 30$	0.973	0.994	0.980
$30 < p_T \leq 40$	0.973	0.995	0.980
$40 < p_T \leq 50$	0.972	0.996	0.978
$50 < p_T \leq 60$	0.970	0.994	0.981
$60 < p_T \leq 80$	0.973	0.996	0.980
$80 < p_T \leq 250$	0.972	0.990	0.981

Table A.2: Muon identification scale factors (Data/MC) for 2012 Tight ID.

Scale Factors for Muons Tight ID

	$0 < \|\eta\| \leq 0.9$	$0.9 < \|\eta\| \leq 1.2$	$1.2 < \|\eta\| \leq 2.1$	$2.1 < \|\eta\| \leq 2.4$
$20 < p_T \leq 25$	0.989	0.987	1.002	0.994
$25 < p_T \leq 30$	0.994	0.994	0.999	
$30 < p_T \leq 35$	0.994	0.991	0.998	
$35 < p_T \leq 40$	0.994	0.990	0.997	
$40 < p_T \leq 50$	0.992	0.990	0.997	
$50 < p_T \leq 60$	0.991	0.991	0.998	
$60 < p_T \leq 90$	0.990	0.986	0.994	
$90 < p_T \leq 140$	1.004	1.012	1.019	
$140 < p_T \leq 300$	1.028	0.956	1.016	
$300 < p_T \leq 500$	1.000	1.000	0.609	

Table A.3: Muon isolation scale factors (Data/MC) for 2012.

Scale Factors for Tight Muons CombRelIsodBeta < 0.2 (R=0.4)

	$0 < \|\eta\| \leq 0.9$	$0.9 < \|\eta\| \leq 1.2$	$1.2 < \|\eta\| \leq 2.1$	$2.1 < \|\eta\| \leq 2.4$
$20 < p_T \leq 25$	0.987	0.992	1.000	1.028
$25 < p_T \leq 30$	1.001	1.004	1.006	
$30 < p_T \leq 35$	0.999	1.002	1.005	
$35 < p_T \leq 40$	0.999	1.001	1.002	
$40 < p_T \leq 50$	0.998	1.000	1.000	
$50 < p_T \leq 60$	0.999	1.000	1.000	
$60 < p_T \leq 90$	1.001	1.000	1.001	
$90 < p_T \leq 140$	1.001	0.999	1.000	
$140 < p_T \leq 300$	1.001	1.000	1.002	
$300 < p_T \leq 500$	1.011	1.000	1.000	

Appendix B

Unfolding

The goal of a particle physics measurement is to compare it to a theoretical prediction. Complications arise when a measured value is subject to random fluctuations caused by detector effects. Each observation is then characterized by a true (and unknown) value t that is smeared by detector effects and becomes a measured value m. In general, one can simply smear the prediction to include the distortions of the detector. However, the measurement cannot be subsequently compared with the results of other experiments because the smearing is detector dependent. Therefore, the measurement should be "unfolded" of detector effects. For this analysis, the jet p_T and η resolution may have a large effect and thus the Z+jets measurement may need to be unfolded.

The continuous distributions for t and m are related by a convolution [92],

$$f_{\text{meas}}(m) = \int R(m|t)\, f_{\text{true}}(t) dt, \tag{B.1}$$

where R is called the response function and depends only on the measuring apparatus. For a measurement binned into a histogram, Eq. B.1 becomes

$$m_i = \sum_{j=1}^{N} R_{ij} t_j, \tag{B.2}$$

where N is the total number of bins. The response matrix then has the simple interpretation of a conditional probability: R_{ij} is the probability that an observed value in bin i

corresponds to a true value in bin j.

To obtain the true distribution, one may intuitively invert the response matrix such that

$$\mathbf{t} = R^{-1}\mathbf{m}, \tag{B.3}$$

where $\mathbf{t}$ and $\mathbf{m}$ are vectors of the true and measured values, respectively. However, unfolding techniques perform more complicated inversions (discussed in [92]) in order to minimize statistical fluctuations.

The unfolding procedure in the analysis was performed with the RooUnfold package[93]. This package provides a framework for different unfolding algorithms. There are two most widely used methods in CMS:

- *Iterative Bayes* is used Bayes' theorem [85] to invert the response matrix. The true estimated variables, called "cause", $C_i (i = 1, ..., n_C)$ are related to the measurement variables, called "effect", $E_j (j = 1, ..., n_E)$ by the probability $P(E_j|C_i)$, called the $n_E \times n_C$ migration matrix. The invert migration matrix, $P(C_i|E_j)$ called smearing matrix, can be calculated by using Bayes' theorem with a certain hypothesis that the effect value E_j is obtained starting from a true estimated value C_i with probability $P_0(C_i)$. Thus the smearing matrix can be written as

  $$P(C_i|E_j) = \frac{P(E_j|C_i)P_0(C_i)}{\sum_{k=1}^{n_C} P(E_j|C_k)P_0(C_k)} \tag{B.4}$$

  If we observed the $n(E_j)$ events of E_j, the expected number of events associated to each cause is $\hat{n}(C_i) = n(E_j)P(C_i|E_j)$. The expected true estimated value will be $\hat{P}(C_i) = \hat{n}(C_i)/\sum_{k=1}^{n_C} \hat{n}(C_k)$. This true estimated value $\hat{P}(C_i)$ will be substituted as the new hypothesis probability instead of $P_0(C_i)$ in iteration processes. The number of iterations can be specified as a regularization parameter.

- *Singular Value Decomposition (SVD)* is using a linear regularization algorithm through

a parameter k_{SVD} on a response matrix. The method defines the real response matrix (A) in factorized form as $A = U \cdot S \cdot V^T$ where U and V are orthogonal matrices and S is a diagonal matrix with non-negative diagonal elements. The measured value y can be represented in term of response matrix $\hat{A}$ and true value x as $y = \hat{A} \cdot x$. Therefore, the invert response matrix of SVD method is $\hat{A}^{-1} = V \cdot S^{-1} \cdot U^T$. This method requires the correct choice of k_{SVD} to smoothly cut-off the rapidly oscillating distributions. A too-small value will bias the unfolding result towards the MC truth input, a too-large value will give a result that is dominated by un-physically enhanced statistical fluctuations. Normally k_{SVD} will be chosen between 2 and the number of bins of particular distribution.